VACUUM TUBE AMPLIFIERS

VACUUM TUBE AMPLIFIERS

Edited by

GEORGE E. VALLEY, JR.

ASSISTANT PROFESSOR OF PHYSICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

HENRY WALLMAN

ASSOCIATE PROFESSOR OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

VACUUM TUBE AMPLIFIERS

COPYRIGHT, 1948, BY THE
MCGRAW-HILL BOOK COMPANY, INC.
PRINTED IN THE UNITED STATES OF AMERICA

x

THE MAPLE PRESS COMPANY, YORK, PA.
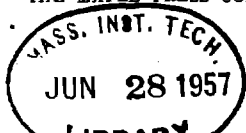
VACUUM TUBE AMPLIFIERS

EDITORIAL STAFF

GEORGE E. VALLEY, JR.
HENRY WALLMAN
HELEN WENETSKY

CONTRIBUTING AUTHORS

YARDLEY BEERS
ERIC DURAND
HAROLD FLEISCHER
JOHN W. GRAY
HARRY J. LIPKIN
DUNCAN MACRAE, JR.
E. JAY SCHREMP
RICHARD Q. TWISS
ROBERT M. WALKER
HENRY WALLMAN

vi

Foreword

THE tremendous research and development effort that went into the development of radar and related techniques during World War II resulted not only in hundreds of radar sets for military (and some for possible peacetime) use but also in a great body of information and new techniques in the electronics and high-frequency fields. Because this basic material may be of great value to science and engineering, it seemed most important to publish it as soon as security permitted.

The Radiation Laboratory of MIT, which operated under the supervision of the National Defense Research Committee, undertook the great task of preparing these volumes. The work described herein, however, is the collective result of work done at many laboratories, Army, Navy, university, and industrial, both in this country and in England, Canada, and other Dominions.

The Radiation Laboratory, once its proposals were approved and finances provided by the Office of Scientific Research and Development, chose Louis N. Ridenour as Editor-in-Chief to lead and direct the entire project. An editorial staff was then selected of those best qualified for this type of task. Finally the authors for the various volumes or chapters or sections were chosen from among those experts who were intimately familiar with the various fields, and who were able and willing to write the summaries of them. This entire staff agreed to remain at work at MIT for six months or more after the work of the Radiation Laboratory was complete. These volumes stand as a monument to this group.

These volumes serve as a memorial to the unnamed hundreds and thousands of other scientists, engineers, and others who actually carried on the research, development, and engineering work the results of which are herein described. There were so many involved in this work and they worked so closely together even though often in widely separated laboratories that it is impossible to name or even to know those who contributed to a particular idea or development. Only certain ones who wrote reports or articles have even been mentioned. But to all those who contributed in any way to this great cooperative development enterprise, both in this country and in England, these volumes are dedicated.

L. A. DuBridge.

Preface

Soon after Drs. I. I. Rabi and L. A. DuBridge decided that the technical knowledge of the Radiation Laboratory staff should be preserved, it was evident that at least one complete book would be required on lumped-parameter circuits. The early planning for that book was done during a series of conferences called by L. J. Hawworth, and attended by B. Chance and G. E. Valley, Jr.

It was difficult to arrange all the subject matter in a way that would be easy to read and economical of space. It would have been possible to describe the various electrical devices in order, but to describe each instrument completely would have involved an intolerable amount of repetition concerning basic circuits, such as multivibrators and amplifiers. It would also have required an intolerable amount of cross-indexing if the work were to be usable by those interested, not in the particular instruments described, but in the application of their design principles to completely different problems. It was apparent, too, that the work should not stress radar.

The material was therefore divided into two parts: the first part to include the basic principles of circuit design, the second to pertain to the assembly of basic circuits into functional instruments such as receivers and data display systems. These decisions were made in the interests of clarity and brevity. Even so, upon completion of the consequent outline, it was evident that several volumes would be required. Accordingly new outlines were prepared for each of these and were then revised separately for each volume by committees composed of the editors and authors concerned.

The first of these books, *Components Handbook*, discusses the physical embodiments of the lumped-parameters themselves: resistors, cables, motors, vacuum tubes, etc. Next, *Vacuum Tube Amplifiers* and *Waveforms* discuss the principles of circuit design, respectively, for circuits that are essentially linear (amplifiers) and for circuits that are essentially nonlinear (oscillators, electronic switches, and the like). The four following volumes concern themselves with the design of complex functional devices. They are *Electronic Time Measurements*, *Electronic Instruments*, *Cathode Ray Tube Displays*, and *Microwave Receivers*.

The amplifiers discussed in this volume are designed to have extreme

values in one of several of the pertinent characteristics: bandwidth, sensitivity, linearity, constancy of gain over long periods of time, etc. In most cases the design of such amplifiers, in which the ultimate performance is obtained from available types of components, cannot be carried out by simple rules of thumb.

The volume therefore begins with a chapter on "Linear Analysis and Transient Response" which lays the theoretical basis for the high-fidelity reproduction of transient signals, such as rectangular pulses. Although the chapter is rather theoretical, a summary is contained in Sec. 1·10 of the precise steps needed to determine the transient response of a given network. The practical application of these principles is examined in the next chapter, "High-fidelity Pulse Amplifiers," for direct, or "video," pulses. The resemblance of this material to that contained in Chap. 3 is only superficial; "Pulse Amplifiers of Large Dynamic Range" is about the design of amplifiers intended to deal with pulses of widely varying magnitude, all other characteristics being secondary. Chapters 4 through 7 deal with the theoretical and practical aspects of several methods of amplifying, with varying degrees of fidelity, pulse-modulated carrier frequencies as high as 200 Mc/sec. Although the design principles are examined in these chapters chiefly from the standpoint of relatively high frequencies, they are perfectly general in their application. That this is true is exemplified by Chap. 10, "Low-frequency Feedback Amplifiers," wherein some of the results of Chap. 4 are applied to filter amplifiers operating at frequencies as low as 50 cps.

Chapter 8 deals with the examination and adjustment of the amplifiers previously described, especially when they are employed as intermediate frequency amplifiers in superheterodyne receivers. Chapter 9 discusses some of the innumerable ways in which inverse feedback can be employed to stabilize the gain of an amplifier. The well-known principles of Nyquist, Bode, and others are applied particularly to circuits in which inductances do not appear, and use is made of this fact to simplify the analysis; in addition the chapter describes the successively less approximate phases through which the design of such an amplifier can proceed. Chapter 11 recounts the experience at the Radiation Laboratory concerning the design of rugged and reliable direct-coupled amplifiers, no particular emphasis being placed upon extreme sensitivity.

Chapter 12, "Amplifier Sensitivity," examines the subject of noise in a rigorous and very theoretical manner. The design of amplifiers for best signal-to-noise ratio is discussed in Chap. 13, "Minimal-noise Input Circuits," and in Chap. 14 the experimental measurement of amplifier sensitivity is explained.

Appendix A contains an existence theorem on the physical realizability of filter amplitude characteristics.

In addition to the material contained in this volume, information concerning the application of amplifiers to specific purposes will be found in other volumes. In particular the use of amplifiers in computers and servomechanisms is discussed in *Electronic Instruments*. In *Cathode Ray Tube Displays* is included a chapter devoted to amplifiers specifically designed to drive inductive loads (i.e., cathode-ray tube deflection coils). *Microwave Receivers* contains a good deal of information on the use in microwave receivers of the types of amplifier described in Chaps. 3 through 7. It also contains a discussion of the noise problem as it affects superheterodyne receiving systems.

The editors wish to acknowledge the inspiration and guidance of the Editor-in-Chief, L. N. Ridenour, and of his editorial board. This book is the product of a large organization, much of the credit for whose successful operation goes to Charles Newton and his able assistants Dr. V. Josephson, M. Dolbeare, and M. Phillips. Whatever uniformity of style and format the book may present is largely due to the Technical Coordination Group operating under the direction of Drs. L. B. Linford and A. M. Stone. To the authors, the editors extend their thanks for a task conscientiously performed and their congratulations upon its completion. The assistance of Mr. J. H. Irving in furnishing important background material for Chap. 1 is gratefully acknowledged. It is due to the generosity of the British Air Commission that Mr. R. Q. Twiss was able to work on the several important chapters that bear his name.

The preparation of the illustrations for the volume was ably supervised by Martha Murrell. The timely assistance of Margot Cheney and Beka Hepner resulted in the volume's being prepared within the allotted time. It was the task of Doris Williams to type over the most illegible of the original manuscript.

THE EDITORS.

CAMBRIDGE, MASS.,
July, 1946.

·

Contents

xviii

CHAPTER 1

LINEAR-CIRCUIT ANALYSIS AND TRANSIENT RESPONSE

By Richard Q. Twiss

1·1. Introduction.—The purpose of this chapter is to provide a systematic procedure for finding the response of a linear network to an applied signal under arbitrary initial conditions. The mathematical machinery used in deriving this procedure is based upon the Laplace transform.[1]

The Laplace transform analysis is not the only method that has been used to solve linear network problems; and because some of its present-day aspects have been influenced by earlier work, a short historical discussion will be given.

Long before the first statement of Kirchhoff's laws or the development of linear-network analysis, Laplace had used transform theory to solve differential equations while Cauchy had employed the Fourier transform as the theoretical basis for an operational calculus. However, in the early treatment of the linear network by Clerk Maxwell, the classic theory of linear-differential equations was used. This method provides a solution in the form of the sum of a "particular integral" and the "complementary function," the latter containing a number of arbitrary constants that have to be determined, from the initial conditions, by an auxiliary set of equations.

This treatment can be made entirely rigorous if the scope of application is sufficiently restricted, and it is still used in elementary textbooks. Unfortunately it is cumbersome in application, particularly when there are a large number of arbitrary constants to be determined. It was largely the search for a compact and simple solution that led Oliver Heaviside to develop the attack usually referred to as the Heaviside operational calculus. Heaviside himself was either ignorant of or indifferent to the work of his predecessors, and his system was presented as a set of disconnected and arbitrary rules which did, in fact, solve a wide range of problems with a minimum of computation but which totally lacked a valid theoretical basis.

The wide application of linear-differential equations in physics and engineering, the fact that Heaviside's calculus (as it stands) is applicable

[1] For an excellent general reference see M. F. Gardner and J. L. Barnes, *Transients in Linear Systems*, Wiley, New York, 1942.

1

only to the case where the initial electrostatic and magnetic energy in the circuit is zero, and the desire to furnish a rigorous mathematical foundation led a whole army of workers, most of whom appeared equally ignorant of Cauchy's original work, into this field; it was not until a vast contribution to the theory of the operational calculus had been made that it was realized that this calculus is intimately related to the theory of the Laplace transform and that a proof of all Heaviside's results can be provided on the functional transform basis.

This result naturally suggested the question as to whether or not the direct application of the Laplace transform theory could provide a solution as simple and compact as that given by the operational calculus, but wider in application and completely rigorous and systematic in method. This question has now been answered in the affirmative; hence it can be expected that the Heaviside calculus will drop into desuetude and will be replaced by the transform analysis.

At the same time as the operational calculus was being investigated and extended, other writers were using the Fourier series and Fourier integral to derive the response of a network to an applied signal from its steady-state response. This line of approach, at least in its elementary form, is less powerful than the Heaviside calculus and is also directly applicable only to the case in which the initial electrostatic and magnetic energy in the network is zero. When the method is extended to overcome these disadvantages, it becomes virtually equivalent to the Laplace transform method. It is shown in Sec. 1·9 that both these methods can be regarded as special cases of the general transform theory in the complex plane.

It must be emphasized that in this chapter the mathematical machinery is regarded merely as scaffolding, as much as possible of which is to be dispensed with as soon as the fundamental form of the general solution has been found. Accordingly, only the most important results of the theory are proved, the auxiliary theorems being stated without justification, and no attempt is made to discuss those aspects of the Laplace transform theory which do not bear directly upon the linear-network problem.

Instead, emphasis is placed upon the methods for setting up the network equations, illustrated by practical examples. In Sec. 1·2 a condensed account is given of the basic properties of linear networks. In Sec. 1·3 the integro-differential equations for several practical networks are derived, the results extended to the general case, and the mesh and nodal methods of setting up those equations are compared. In Sec. 1·4, which contains all the fundamental mathematical theory, the concept of the Laplace transform is introduced, its principal properties derived, and its ability to transform a set of differential equations into a

set of algebraic equations is illustrated for the case of the nth-order equation in a single independent variable.   The inverse Laplace transform is then derived and used to complete the solution of the nth-order equation.   The section terminates with a discussion of the convolution theorem and its application to network analysis.   The theory is then applied to the equations of Sec. 1·3; the general solution is derived; and its form discussed.

It is at this stage that the concept of the "transform network" is introduced.   This concept can often be of great help in solving practical problems; and because its use is not widespread, it is treated very fully. It might be regarded as a mistake in emphasis to attach such prominence to what is, in essence, merely a mechanism for shortening a part of the analysis, but this view can be countered by two powerful arguments. In the first place the most difficult and least standardized part of network analysis is the setting up of the transform equations.   Once these equations have been derived, a condensed procedure can be followed that is always the same whatever the particular problem to be solved.   Secondly, the transform equations can be derived from the transform network by the conventional methods of the steady-state analysis without introducing the integro-differential equations at all.   Thus the use of the transform network not only shortens the analysis but greatly simplifies it as well.

The standard procedure for solving network problems is summarized in Sec. 1·11; and in Sec. 1·12 this procedure is applied to derive the complete solution of two practical examples.

The chapter also contains a section on the Fourier transform theory, the merits of which are compared with those of the Laplace transform.

1·2. The Basic Properties of Linear Networks.—Most of the concepts introduced in this section will already be familiar to the reader acquainted with the conventional steady-state network analysis.   Reference can be made to any of the standard textbooks and in particular to Gardner and Barnes *Transients in Linear Systems* for a detailed discussion.[1]

The Class of Networks Considered.—The networks considered in this chapter are composed partly of lumped passive elements (inductors, capacitors, and resistors) and partly of active elements (voltage and current generators).   These active elements are not physical entities but mathematical idealizations; in practice active networks will contain not pure generators but vacuum tubes.   However, it is assumed that all the vacuum tubes treated in this analysis can be completely represented by a three-terminal network consisting of lumped-passive elements together with voltage and current generators.

---

[1] The whole of the treatment of this section follows along lines similar to those adopted by M. F. Gardner and J. L. Barnes, *op. cit.*

All the elements, both passive and active, are assumed to be linear (thus ensuring that the superposition principle can be applied) and invariant with time.

The Network Elements.—The integro-differential equations relating the voltage across a passive element to the current through it are set out in Table 1·1.

<div align="center">TABLE 1·1.—PASSIVE-ELEMENT REPRESENTATION</div>

Element	Symbol	Impedance basis	Admittance basis
Inductor		$v_L(t) = L \dfrac{di_L(t)}{dt}$ where L is the self inductance of the element, $L = \Gamma^{-1}$	$i_\Gamma(t) = i_\Gamma(0+) + \Gamma \displaystyle\int_0^t v_\Gamma(t)\,dt,$ where Γ is the inverse self inductance of the element, $\Gamma = L^{-1}$
Capacitor		$v_S(t) = v_S(0+) + S \displaystyle\int_0^t i_S(t)\,dt,$ where S is the elastance of the element, $S = C^{-1}$	$i_C(t) = C \dfrac{dv_C(t)}{dt},$ where C is the capacitance of the element, $C = S^{-1}$
Resistor		$v_R(t) = Ri_R(t),$ where R is the resistance of the element, $R = G^{-1}$	$i_G(t) = Gv_G(t),$ where G is the conductance of the element, $G = R^{-1}$

NOTE 1. The values $i_\Gamma(0+)$ and $v_S(0+)$ are the limiting values of $i_\Gamma(t)$ and $v_S(t)$ respectively as t decreases toward zero.

NOTE 2. The parameters L, S, R, Γ, C, G are descriptive of the element only and are independent of the time or of the magnitude of the voltage across the element.

It will be shown later in this section that there are two fundamental methods of deriving the basic equations: One uses the concepts voltage generator, impedance, and mesh; the other uses the concepts current generator, admittances, and node; and it is for this reason that in one column of Table 1·1 the voltage-current relation has been set out on an impedance basis and in the neighboring column on an admittance basis.
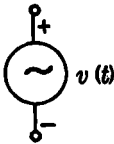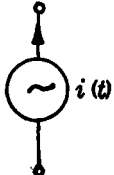
In the special case in which $v(t)$ and $i(t)$ are exponential functions of the time, $V(p) \exp pt$ and $I(p) \exp pt$ respectively,

$$\frac{v_L(t)}{i_L(t)} = Lp, \qquad i_\Gamma(t) = \frac{\Gamma}{p}\,[i_\Gamma(0+) = 0],$$

$$\frac{v_S(t)}{i_S(t)} = \frac{S}{p}, \qquad \frac{i_C(t)}{v_C(t)} = Cp[v_C(0+) = 0],$$

$$\frac{v_R(t)}{i_R(t)} = R, \qquad \frac{i_R(t)}{v_R(t)} = G,$$

and the ratios Lp, S/p, R, are called p impedances; Γ/p, Cp, G, are called p-admittances. In the particular case where $p = j\omega$ is purely imaginary, the ratios are simply called impedances and admittances and are the familiar steady-state concepts.

The main properties of the active elements are set out in Table 1·2.

TABLE 1·2.—ACTIVE-ELEMENT REPRESENTATION

Element	Symbol	Internal impedance	Internal admittance
Voltage source		Zero	Infinite
Current source		Infinite	Zero
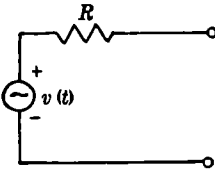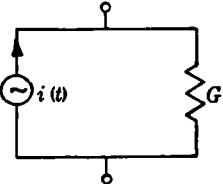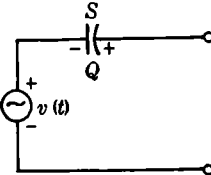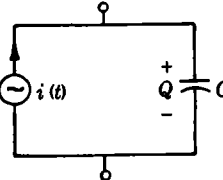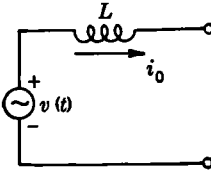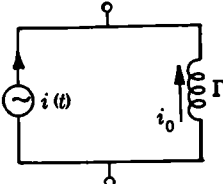
Voltage-current Source Transformation.—In the mesh analysis it is desirable that all the sources be voltage generators, but in the nodal analysis it is desirable that all the sources be current generators. In many cases it will be possible to achieve this state of affairs with the aid of one or the other of the transformations set out in Table 1·3.[1]

In the steady-state case when $v(t)$ and $i(t)$ are exponential functions of the time $V(p) \exp pt$ and $I(p) \exp pt$ respectively, these transformations are all special cases of the general transformations of Fig. 1·1 where $Z(p)$ and $Y(p)$ are p-impedances and p-admittances respectively related by the equations

$$Z(p) = \frac{1}{Y(p)},$$

$$V(p) = Z(p)I(p),$$

[1] For a proof of these transformations see Gardner and Barnes, *op. cit.* p. 23.

TABLE 1·3.—Voltage-current Equivalences

Voltage generator	Current generator

$v(t) = Ri(t),$
where $R = G^{-1}$

$i(t) = Gv(t),$
where $G = R^{-1}$

$v(t) = S \int_0^t i(t)\, dt,$

where $S = C^{-1}$ and Q is the initial charge on the condenser

$I(t) = C \dfrac{dv(t)}{dt},$

where $C = S^{-1}$ and Q is the initial charge on the condenser

$v(t) = L \dfrac{di(t)}{dt},$

where $L = \Gamma^{-1}$ and i_0 is the initial current through the inductor

$i(t) = \int_0^t v(t)\, dt,$

where $\Gamma = L^{-1}$ and i_0 is the initial current through the inductor

where $V(p)$ and $I(p)$ are constant-voltage and constant-current p-generators respectively.

This general transformation can be used when deriving the transform equations from the transform network, but it cannot be used directly when setting up the integro-differential equations.
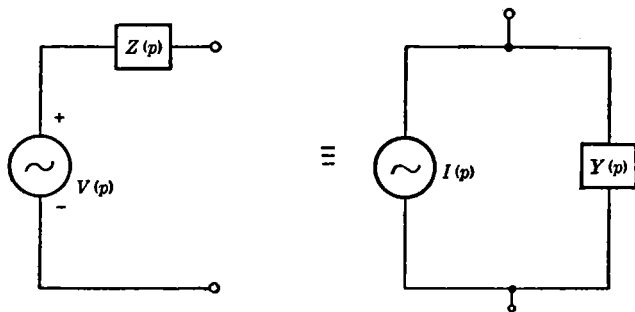
FIG. 1·1.—Voltage-current source equivalence.

Mutual Inductance.—Mutual inductance is not an element of a network in the sense discussed above; but, since the presence of magnetic coupling between the various inductances of the network affects the form of the network equations, this concept must be introduced.

A full discussion of mutual inductance will be found in all the standard textbooks on electromagnetism.[1]

Consider the simple transformer of Fig. 1·2, where the two inductances L_1 and L_2 are coupled together by mutual inductance M.

If the inductor L_2 is open circuited so that $i_2 = 0$, the voltage across 2-2 is $\pm M\, di_1/dt$. Similarly if L_1 is open circuited, the voltage across the terminals 1-1 is $\pm M\, di_2/dt$. The ambiguity of sign is resolved as follows: If a positive rate-of-change of i_2 induces in Circuit 1 a voltage

FIG. 1·2.—Mutual-inductance coupling.

drop in the arrow direction, the sign is positive; if a voltage rise is induced, the sign is negative.

More generally if a coil L_j is coupled by mutual inductances M_{jk} to a system of coils $L_k(j \neq k, k = 1, \cdots, n)$, then the open-circuit voltage across the coil L_j is
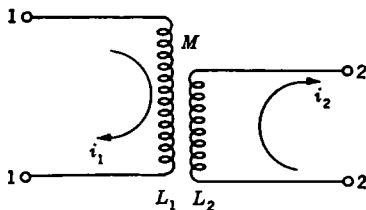
[1] M. Abraham and R. Becker, *Classical Electricity and Magnetism*, Blackie, Glasgow, 1932; Sir James Jeans, *Mathematical Theory of Electricity and Magnetism*, 4th ed., Cambridge, London, 1923.

$$\pm M_{j1}\frac{di_1}{dt} \pm M_{j2}\frac{di_2}{dt} \pm \cdots \pm M_{jn}\frac{di_n}{dt},$$

where the signs of the various terms are determined by the rule above.

The Network Structure.—In setting up the basic equations such concepts as node, branch, mesh, etc., will be used, and in comparing the merits of the mesh and nodal analysis, it is necessary to enumerate the number of independent node pairs and meshes. Because different writers have given these terms different meanings, it seems desirable to remove the possibility of ambiguity by defining them here. The notation used is similar to that of Gardner and Barnes (*loc. cit.*), but a few minor changes, intended to bring out the difference between the steady-state and the transient analysis, have been made.

Fig. 1·3.—Universal element symbol.

Such terms as "node" and "mesh" are primarily geometrical and do not depend upon the detailed nature of the elements making up the network. Accordingly the symbol of Fig. 1·3 will be used to represent any of the five basic elements (three passive and two active) whenever this appears desirable in the interests of clarity.

The various terms descriptive of the network structure will now be defined.

1. *Terminal.* The end points of an element are called its terminals.
2. *Node.* The junction point of two or more terminals is called a *node*. The junction point of no more than two terminals is called a *simple*
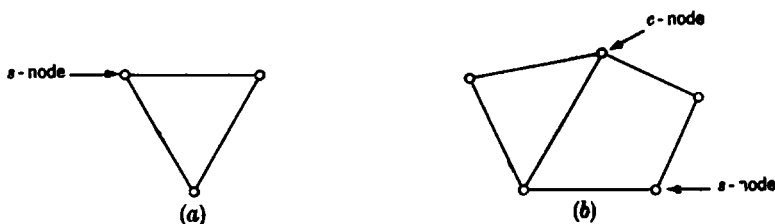
FIG. 1·4.—Network geometry I.

node or *s-node*. The junction point of more than two terminals is called a *complex node* or *c*-node. The network in Fig. 1·4*a* has three *s*-nodes; the network in Fig. 1·4*b* has three *s*-nodes and two *c*-nodes.

3. *Branch.* The series connection of elements, none of whose internal nodes is a *c*-node, is called a *closed branch* if the series connection is closed in an *s*-node or *c*-node and an *open branch* if the end points of the branch are both *c*-nodes. Examples of both closed and open branches are shown in Fig. 1·5.

4. *Mesh.* Any closed path via one or more branches in series forms a mesh. A closed branch is a special case of a mesh.

5. *Separable part.* A part of the network having only mutual inductive coupling with the rest of the network is called a separable part. Thus the network of Fig. 1·6, which has four c-nodes, two s-nodes, six open branches, and one closed branch, has three separable parts.

FIG. 1·5.—Network geometry II.

6. *Reference node.* One node of each separable part of a network is called the reference node of that part. The choice of reference node is arbitrary.

7. *Independent node pair.* The combination of the reference node with any other node of the same separable part is called an independent node pair.
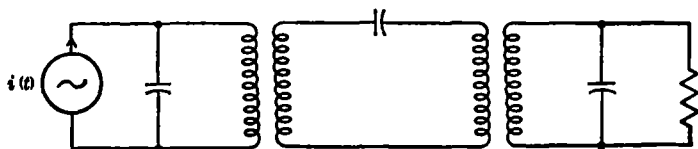
FIG. 1·6.—Network with three separable parts.

It is clear from the above definitions that if N is the total number of nodes in the network, P the number of separable parts of the network, and N_P the number of independent node-pairs, then

$$N_P = N - P.$$

Another most important result, which can be proved by topological methods,[1] is that M, the number of independent meshes, is given by

$$M = E - N_P,$$

where E is the number of elements in the network. A heuristic proof of this result is given in Sec. 1·3.

It remains only to discuss the reason for dividing the nodes into s- and c-nodes. In the elementary steady-state analysis s-nodes are not regarded as nodes at all. The condition that no charge accumulate at any point of the network will be satisfied automatically at an s-node if it is assumed that the instantaneous current flowing in any element of a branch is equal to that flowing in any other element of the same branch. The impedance

[1] See M. F. Gardner and J. L. Barnes, *op. cit.*, for a bibliography of papers on this subject.

or admittance of a branch can then be written down by inspection.   This cannot be done, however, when the integro-differential equations for the network are being set up.

The point will become clearer after the concept of the transform network has been introduced, when it will appear that an *s*-node can be neglected only if the initial potential of the node is zero and if the initial currents flowing into it from inductors are zero.[1]

1·3. The Integro-differential Equations of the Linear Network.—It was pointed out in the introduction that the solution of practical problems is effected by deriving the trans orm equations from the transform network[2] so that the integro-differential equations are not formulated directly.   Nevertheless the methods of setting up these integro-differential equations are discussed in some detail and illustrated with several examples in this section.   This course of action is followed for two reasons. In the first place the transform network can be employed with confidence only when the theory underlying it is fully understood, and this theory can be developed only from the integro-differential equations themselves. Second, the method of setting up the transform equations from the transform network is very similar in form, though more powerful in scope, to the methods of setting up the integro-differential equations from the original network.

The examples given in this section are not necessarily practical networks; they have rather been chosen to demonstrate some particular point of the theory.   This is not true, however, of the examples given at the end of this chapter, which are selected to illustrate the power of the method in solving practical problems.

Kirchhoff's Laws and the Mesh-nodal Analysis.—Kirchhoff's laws, which are the fundamental basis for the whole of network theory, can be stated as follows:

LAW 1.   The total voltage drop around any mesh of the circuit is zero.
LAW 2.   The total instantaneous current flowing into any node of the network is zero.

The basic equations can be obtained by applying these laws to every branch and node of the network, but this is usually an excessively clumsy procedure.   A much shorter analysis can be obtained from the so-called "mesh-nodal analysis" which will now be described.

It may be emphasized here that Kirchhoff's laws apply to both linear and nonlinear circuits, but the mesh-nodal analysis can be applied only to the former.

---

[1] It will be shown that only under these circumstances will an *s*-node in the original network transform into an *s*-node in the transform network.

[2] See Sec. 1·7.

In the mesh method all current sources are replaced by voltage sources,[1] and the variables are the M currents flowing in the M independent meshes.   This system automatically satisfies Kirchhoff's second law at every node, because as much current flows into any node as flows out of it. M independent equations are obtained by applying Kirchhoff's first law to the M independent meshes in turn, thus determining the system.

In the nodal analysis all voltage sources are replaced by current sources.   If the network has N nodes and P separable parts, the variables are the $N - P$ voltage differences across the $N - P$ independent node pairs. $N - P$ independent equations are obtained[2] by applying Kirchhoff's second law to every node of the network in turn, thus determining the system.

Before the mesh-nodal analysis is applied to a few typical network structures, a heuristic proof will be given of the formula stated in Sec. 1·2 relating the number of independent meshes to the number of elements and nodes in the network.

Let E be the number of elements, P the number of separable parts, and N the number of nodes.   Then there will be E equations relating the currents flowing in the elements to the potentials across them, and $N - P$ equations obtained by applying Kirchhoff's second law to every independent node pair in turn.   These $N - P$ equations determine $N - P$ of the currents in terms of the others; hence if the analysis is set up on the mesh basis, $E - (N - P)$ equations will be needed to determine the unknowns, and this is therefore the number of independent meshes.

It may be observed that the mesh analysis will require fewer equations than the nodal if $E < 2(N - P)$.

The Mesh Analysis and Examples.—In general the mesh equations may be set up in a large number of ways, subject only to the two provisos:

There are $E - N + P$ mesh equations.

Every element of the network is included in at least one mesh.

The actual choice will depend upon which voltage or current is to be calculated.   Usually it is required to find the voltage across a single element or branch of the network when an arbitrary signal is applied across some part of the network.   If this is the case, it will be best to set up the mesh analysis in such a way that this element or branch is contained in only one mesh.   In the general theoretical case, of course, where the solution for every current and voltage in the network is required, one choice of the independent mesh system is as good as another.

[1] By one of the transformations of Table 1·3.

[2] Not N equations.   If Kirchhoff's second law is satisfied at all but one of the nodes of a separable part, it will automatically, be satisfied at the remaining node.

Two-mesh Network.—As a first example, consider the network of Fig. 1·7. Initially it will be supposed that the condensers have charges Q_1, Q_{12}, Q_2, respectively, and that currents ρ_1, ρ_2 are flowing in coils L_1, L_2 in the directions indicated. It will be assumed that the inductances are coupled by mutual inductance M and are wound in phase. This network has one separable part, eight elements, and seven nodes; hence there are two independent meshes. The mesh system of Fig. 1·7 is the natural one to take if the voltage across R_2 is to be determined. Had the voltage across S_{12} been of primary interest it would have been better to have taken the mesh system of Fig. 1·8.
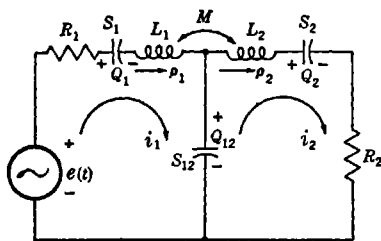
FIG. 1·7.—Two-mesh network.
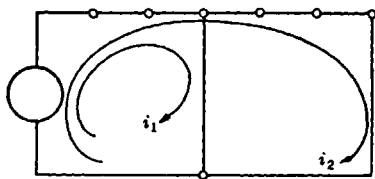
FIG. 1·8.—Alternative two-mesh network.

If Kirchhoff's first law is applied to the first mesh of Fig. 1·1 and the results given in Table 1·1 are utilized, one gets

$$e(t) = R_1 i_1 + Q_1 S_1 + S_1 \int_0^t i_1 \, d\tau + L_1 \frac{di_1}{dt} - M \frac{di_2}{dt} + Q_{12} S_{12}$$
$$+ S_{12} \int_0^t (i_1 - i_2) \, d\tau;$$

applying the law to the second mesh gives

$$0 = -Q_{12} S_{12} - S_{12} \int_0^t (i_1 - i_2) \, d\tau + L_2 \frac{di_2}{dt}$$
$$- M \frac{di_1}{dt} + Q_2 S_2 + S_2 \int_0^t i_2 \, d\tau + R_2 i_2.$$

Collecting similar terms and transposing the equations gives

$$e(t) - Q_1 S_1 - Q_{12} S_{12} = \left[R_1 + L_1 \frac{d}{dt} \right.$$
$$\left. + (S_1 + S_{12}) \int_0^t d\tau \right] i_1 - \left[M \frac{d}{dt} + S_{12} \int_0^t d\tau \right] i_2,$$

$$- Q_{12} S_{12} - Q_2 S_2 = - \left(M \frac{d}{dt} + S_{12} \int_0^t d\tau \right) i_1$$
$$+ \left[R_2 + L_2 \frac{d}{dt} + (S_2 + S_{12}) \int_0^t d\tau \right] i_2.$$

Two of the initial conditions (the total voltage drop around the condensers in the two meshes) have already been included in the equations; the two remaining conditions are

$$i_1(0) = \rho_1, \qquad i_2(0) = \rho_2.$$

If the following integro-differential operators are defined,

$$\left.\begin{aligned}
z_{11} &= R_1 + (S_1 + S_{12}) \int_0^t d\tau + L_1 \frac{d}{dt}, \\
z_{12} &= -M \frac{d}{dt} - S_{12} \int_0^t d\tau, \\
z_{21} &= -M \frac{d}{dt} - S_{12} \int_0^t d\tau, \\
z_{22} &= R_2 + (S_2 + S_{12}) \int_0^t d\tau + L_2 \frac{d}{dt},
\end{aligned}\right\} \tag{1}$$

and if $u_1(0) = Q_1 S_1 + Q_{12} S_{12}$ is the initial voltage drop around the first mesh in a clockwise direction and if $u_2(0) = -Q_{12} S_{12} + Q_2 S_2$ is the initial voltage drop around the second mesh in a clockwise direction, then the equations can be put in the condensed form

$$\left.\begin{aligned}
e(t) - u_1(0) &= z_{11} i_1(t) + z_{12} i_2(t), \\
-u_2(0) &= z_{21} i_1(t) + z_{22} i_2(t),
\end{aligned}\right\} \tag{2}$$

where

$$i_1(0) = \rho_1 \qquad \text{and} \qquad i_2(0) = \rho_2. \tag{3}$$

It will be noted that $z_{12} = z_{21}$. This is a characteristic of all passive networks.

In normal network parlance $e_1(t)$ is called the driving function, and $i_1(t)$ and $i_2(t)$ are called the response functions.

The General Mesh Equations for the Passive Network.—The discussion in the first example was made as general as possible so that the development of the general case would not be too abstract when the condensed notation, defined below, is employed.

Let L_{jj}, R_{jj}, S_{jj} be the total self-inductance, resistance, and elastance, respectively, in mesh j.

Let L_{jk}, R_{jk}, S_{jk} be the total self- and mutual inductance, resistance, and elastance, respectively, common to meshes j and k.

Let $e_j(t)$ be the driving voltage in mesh j.

Let $u_j(0)$ be the total initial voltage drop around the elastances in mesh j, in a clockwise direction.

Let $i_j(t)$ be the current flowing in the jth mesh in a clockwise direction.

Let

$$z_{jj} = L_{jj}\frac{d}{dt} + R_{jj} + S_{jj}\int_0^t d\tau.$$

Let

$$z_{jk} = -L_{jk}\frac{d}{dt} - R_{jk} - S_{jk}\int_0^t d\tau.$$

Then z_{jj} is the total impedance in the jth mesh and $-z_{jk}$ is the impedance common to the jth and kth meshes. If Kirchhoff's first law
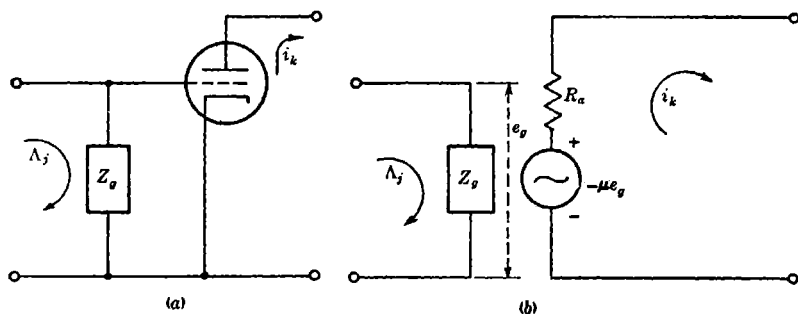
Fig. 1·9.—Equivalent circuit for a tube.

is applied to each of the n independent meshes in turn, the following system of equations results.

$$e_j(t) - u_j(0) = \sum_{k=1}^n z_{jk}i_k(t) \qquad (j = 1, \cdots, n), \qquad (4)$$

where $z_{jk} = z_{jk}$. Of the initial conditions n are already included in the equations. The remaining n are given by the initial values of $i_j(t)$ $(j = 1, \cdots, n)$.

It will be seen that Eqs. (2) are a special case of Eqs. (4) when $n = 2$ and $e_2(t) = 0$.

The General Mesh Equations for the Active Network.—In this chapter attention will be confined to the case where the three-terminal vacuum-tube network of Fig. 1·9a can be replaced by the equivalent network of Fig. 1·9b.

To demonstrate the modification produced in the network equations when vacuum tubes are present, let it be supposed that the grid-cathode circuit impedance operator

$$Z_g \equiv L_g\frac{d}{dt} + R_g + S_g\int_0^t d\tau$$

pertains to the jth circuit and that the anode-cathode path is in the kth circuit.

Then application of Kirchhoff's first law to the kth circuit gives

$$z_{k1}i_1 + z_{k2}i_2 + \cdots + z_{kj}i_j + \cdots + z_{kn}i_n = e_k(t) - \mu e_g;$$

and since $e_g = z_g i_j$, this equation can be written

$$z_{k1}i_1 + z_{k2}i_2 + \cdots + (z_{kj} + \mu z_g)i_j + \cdots + z_{kn}i_n = e_k(t);$$

thus the general form of the active network equations are the same as those of the passive network save that it is no longer possible to assume

$$z_{jk} = z_{kj}.$$

The Nodal Analysis and Examples.—The nodal analysis is less widely used than the mesh analysis. However, in many cases, particularly for
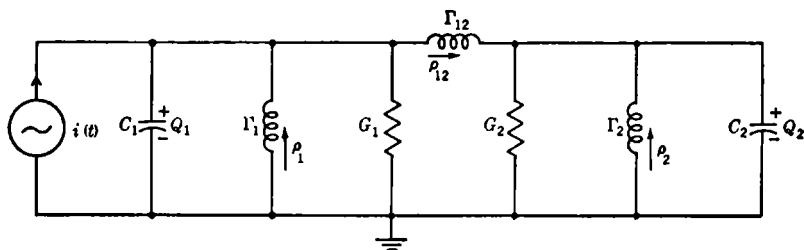
Fig. 1·10.—Two-node network.

active networks, it is much more convenient, as the succeeding examples show.

Two-node Network.—Consider the network of Fig. 1·10, where the condensers C_1, C_2 initially carry charges Q_1, Q_2 and where initial currents ρ_1, ρ_{12}, ρ_2 flow in the inverse inductances Γ_1, Γ_{12}, Γ_2 in the directions shown in the figure. This network has one separable part, three nodes, and nine branches; hence in the mesh analysis seven equations would be needed to determine the system. But th s network has only two independent node pairs, whose voltages may be taken as $e_1(t)$, $e_2(t)$, the grounded node being taken as the reference node. If Kirchhoff's second law is applied to these two nodes in succession and if the results given in Table 1·1 are utilized, one gets

$$i(t) = C_1 \frac{d}{dt} e_1 + \Gamma_1 \int_0^t e_1 \, d\tau - \rho_1 + G_1 e_1 + \Gamma_{12} \int_0^t (e_1 - e_2) \, d\tau + \rho_{12},$$

$$0 = C_2 \frac{d}{dt} e_2 + \Gamma_2 \int_0^t e_2 \, d\tau - \rho_2 + G_2 e_2 + \Gamma_{12} \int_0^t (e_2 - e_1) \, d\tau - \rho_{12};$$

collecting similar terms and transposing the equations one gets

$$i(t) + \rho_1 - \rho_{12} = \left[C_1 \frac{d}{dt} + G_1 + (\Gamma_1 + \Gamma_{12}) \int_0^t d\tau \right] e_1 - \left(\Gamma_{12} \int_0^t d\tau \right) e_2,$$

$$\rho_2 + \rho_{12} = - \left(\Gamma_{12} \int_0^t d\tau \right) e_1 + \left[\Gamma_2 \frac{d}{dt} + G_2 + (\Gamma_2 + \Gamma_{12}) \int_0^t d\tau \right] e_2.$$

If the following integro-differential operators be defined

$$y_{11} = C_1 \frac{d}{dt} + G_1 + (\Gamma_1 + \Gamma_{12}) \int_0^t d\tau, \qquad y_{12} = -\Gamma_{12} \int_0^t d\tau,$$

$$y_{21} = -\Gamma_{12} \int_0^t d\tau, \qquad y_{22} = C_2 \frac{d}{dt} + G_2 + (\Gamma_2 + \Gamma_{12}) \int_0^t d\tau,$$

and if $j_1(0) = +\rho_1 - \rho_{12}$ is the initial current in the inductances flowing into the first node and if $j_2(0) = +(\rho_2 + \rho_{12})$ is the initial current in the inductances flowing into the second node, then the equations can be put in the condensed form

$$\left. \begin{array}{r} i_1(t) + j_1(0) = y_{11}e_1(t) + y_{12}e_2(t), \\ +j_2(0) = y_{21}e_1(t) + y_{22}e_2(t), \end{array} \right\} \tag{5}$$

where

$$e_1(0) = \frac{Q_1}{C_1}, \qquad e_2(0) = \frac{Q_2}{C_2}.$$

In this case $i_1(t)$ is called the driving function, and $e_1(t)$, $e_2(t)$ the response functions.
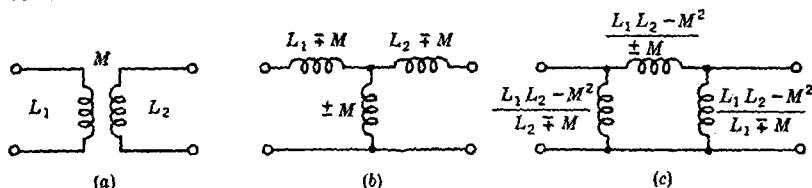
Fig. 1·11.—Transformer equivalent networks.

Mutual Inductance in the Nodal Analysis.—In general where mutual inductance is present it is better to employ the mesh analysis. In the case where the d-c levels of the various parts of the network are not of interest it is possible, however, to replace a transformer by its equivalent T- or Π-network. Thus except for their d-c characteristics the three networks of Fig. 1·11a, b, c are equivalent. The number of independent node pairs is not increased by this transformation because, provided the d-c characteristics are not important, the reference nodes for the various separable parts can be taken as coincident. Thus, although the number of separable parts in the nodal analysis is always reduced from P to 1 by

this transformation, the number of nodes is reduced from N to $N - P + 1$, and hence there are still $N - P$ independent node pairs.

To illustrate the above discussion by a practical example, consider the network of Fig. 1·12.

Since this network as it stands has two separable parts, seven elements, and five nodes, there are no less than four independent meshes, and four equations would be needed to describe the network on the mesh basis. If, however, the voltage generator is replaced by the equivalent current generator and the transformer is replaced by its equivalent Π-network, the network assumes the general form of Fig. 1·10, which, as has already been shown, can be described on the nodal basis by only two equations.

Fig. 1·12.—Mutual-inductance-coupled network.

Two-node Active Network.—The chief importance of the nodal analysis lies in the fact that it is especially suitable for application to active net-
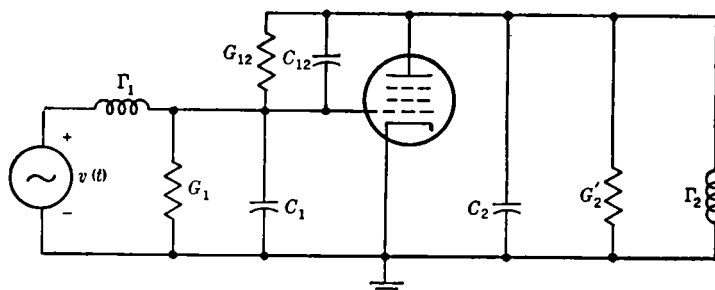
Fig. 1·13.—Two-node active network.
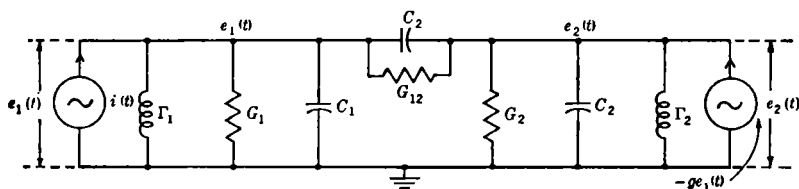
Fig. 1·14.—Equivalent two-node active network.

works in which the vacuum tube acts not merely as a buffer between successive stages but is embedded in the network. The equations for the active network of Fig. 1·13 will illustrate this point. This network is the basic circuit of a negative feedback pair when stray anode-grid capacity is taken into account.

The equivalent circuit for this active network is drawn in Fig. 1·14, where all voltage generators have been replaced by current generators. It is assumed that all the initial currents in the coils and the charges on the condensers are zero; g is the transconductance of the tube; R_a is the internal resistance of the tube; G_2 equals $G_2' + (1/R_a)$; and

$$i(t) = \Gamma_1 \int_0^t v(\tau) \, d\tau.$$

This network has two independent node pairs. If ground is taken as the reference node and $e_1(t)$, $e_2(t)$ as the potentials of the independent node pairs, then application of Kirchhoff's second law to the two independent nodes in turn gives

$$i = \Gamma_1 \int_0^t v \, d\tau = \Gamma_1 \int_0^t e_1 \, d\tau + C_1 \frac{de_1}{dt} + G_1 e_1 + C_{12} \frac{d}{dt}(e_1 - e_2) \\ + G_{12}(e_1 - e_2),$$

$$-ge_1 = \Gamma_2 \int_0^t e_2 \, d\tau + C_2 \frac{de_2}{dt} + G_2 e_2 + C_{12} \frac{d}{dt}(e_2 - e_1) + G_{12}(e_2 - e_1).$$

After rearrangement the equations can be put in the form

$$\Gamma_1 \int_0^t v \, d\tau = \left[(C_1 + C_{12}) \frac{d}{dt} + G_1 + G_{12} + \Gamma_1 \int_0^t d\tau \right] e_1 \\ - \left(G_{12} + C_{12} \frac{d}{dt} \right) e_2,$$

$$0 = \left(g - G_{12} - C_{12} \frac{d}{dt} \right) e_1(t) \\ + \left[(C_2 + C_{12}) \frac{d}{dt} + G_2 + G_{12} + \Gamma_2 \int_0^t d\tau \right] e_2.$$

It may be seen that these equations are of the same general form as Eq. (5) save that $y_{12} \neq y_{21}$.

Because the network of Fig. 1·14 has no less than 10 branches, it is obvious that the mesh analysis would prove formidable indeed.

The General Nodal Equations.—When setting up the general nodal equations for the passive network, it will be assumed that all voltage sources have been replaced by current sources and all mutual-inductance-coupled circuits replaced by the equivalent Π- or T-networks.[1] The following notation will be used.

---

[1] Thus the network has only one reference node, the reference nodes of the various separable parts being taken as coincident.

Let

Γ_j, G_j, C_j be the inverse inductance, conductance, and capacitance linking the jth node to the reference node.

Γ_{jk}, G_{jk}, C_{jk} be the inverse inductance, conductance, and capacitance linking the jth node to the kth node.

$i_j(t)$ be the driving current in the jth node.

ρ_j be the initial current flowing into the jth node from the inductances.

$e_j(t)$ be the voltage of the jth node with respect to the reference node.

$$y_j = \Gamma_j \int_0^t dt + G_j + C_j \frac{d}{dt},$$

$$y_{jk} = -\left(\Gamma_{jk} \int_0^t dt + G_{jk} + C_{jk} \frac{d}{dt} \right),$$

$$y_{jj} = y_j - \sum_{\substack{k \neq j = 1}}^{n} y_{jk},$$

where n is the number of independent node pairs. Then y_{jj} is the admittance linking the jth node to all other nodes; y_{jk} is the admittance linking the jth node to the kth node. Application of Kirchhoff's second law to each independent node pair of the network in succession yields the system of equations

$$i_j(t) + \rho_j = \sum_{k=1}^{n} y_{jk} e_k(t) \qquad (j = 1, \cdots, n); \tag{6}$$

n of the initial conditions are contained in this system of equations. The remaining n conditions are given by the initial values of $e_k(t)$ ($k = 1$, . . . , n).

In the active network the equations assume the same general form as those of Eq. (6) save that it is no longer possible to assume $y_{jk} = y_{kj}$.

Comparison of the Mesh and Nodal Analyses.—As developed in this section, there is a formal symmetry between the mesh-nodal methods of setting up the integro-differential equations which is summarized in Table 1·4.

In any particular case the mesh analysis requires more equations if $E > 2(N - P)$ and fewer equations if $E < 2(N - P)$. The mesh analysis is usually to be preferred where there are a considerable number of mutual inductances, and the nodal analysis is usually better when there are vacuum tubes embedded in the network.

In general, the above comparison of the mesh and nodal analyses is valid for the steady-state equations of the transform network. It is no longer possible to choose the shorter solution simply by comparing E

TABLE 1·4.—COMPARISON OF MESH-NODAL ANALYSES

Comparative basis	Mesh analysis	Nodal analysis
Passive elements...	Resistance, elastance, inductance; more generally impedance	Conductance, capacitance, inverse inductance; more generally admittance
Active elements....	Voltage sources	Current sources
Connections.......	Elements in series	Elements in parallel
Basic law.........	Kirchhoff's first law: The voltage drop around a closed mesh is zero	Kirchhoff's second law: The current flowing into any node is zero
Number of independent equations.	$M = E - N + P$, where M is the number of independent meshes	$N_p = N - P$, where N_p is the number of independent node pairs

with $2(N - P)$, however, because more condensed methods of setting up the transform equations from the transform network may be available, giving in the mesh case fewer than $E - N + P$ equations and in the nodal case fewer than $N - P$ equations.

The Principle of Superposition.—It is one of the most important results of the theory of linear integro-differential equations that the solutions satisfy the principle of superposition. This principle may be stated in its simplest form as follows:

Any linear combination of two or more solutions of a set of linear integro-differential equations is itself a solution.

In network theory the principle takes a form that will now be illustrated for the general mesh case.[1]

Let $i_{k1}(t)$ $(k = 1, \ldots, n)$ be a set of solutions of the general mesh equations

$$\sum_{k=1}^{n} Z_{jk} i_k(t) = e_j(t) \qquad (j = 1, \cdots, n),$$

and let $i_{k2}(t)$ $(k = 1, \ldots, n)$ be a set of solutions of the general mesh equations

$$\sum_{k=1}^{n} Z_{jk} i_k(t) = e_j'(t) \qquad (j = 1, \cdots, n);$$

then $i_{k1}(t) + i_{k2}(t)$ is a solution of the equations

$$\sum_{k=1}^{n} Z_{jk} i_k(t) = e_j(t) + e_j'(t) \qquad (j = 1, \cdots, n).$$

[1] The analysis for the nodal case follows exactly the same course.

This result follows from the fact that

$$Z_{jk}i_{k1}(t) + Z_{jk}i_{k2}(t) = Z_{jk}[i_{k1}(t) + i_{k2}(t)].$$

The principle of superposition may be extended by continuity to state that the response of a network to the convergent sum of an infinite set of driving functions is equal to the sum of the responses of the network to each driving function taken separately, and it will be so used in Sec. 1·8 when considering the response of a network to a periodic driving function.

A particular use of the superposition principle occurs in regarding the system of driving functions

$$e_1(t) = f_1(t), \qquad e_2(t) = f_2(t), \qquad \cdots, \qquad e_n(t) = f_n(t)$$

as a superposition of n systems, the first consisting of driving function

$$e_1(t) = f_1(t), \qquad e_2(t) = 0, \qquad \cdots, \qquad e_n(t) = 0,$$

the second of

$$e_1(t) = 0, \qquad e_2(t) = f_2(t), \qquad \cdots, \qquad e_n(t) = 0,$$

and so forth, thus reducing the general problem of finding the network response to an arbitrary system of generators to one of finding the response to a single generator.

The superposition principle is not given much prominence in this chapter and is introduced here chiefly because it will be needed in the analysis of the Fourier transform. It may be remarked, however, that it can be utilized to derive, heuristically at least, an expression for the transient response from the steady-state response.

1·4. The Theory of the Laplace Transform.   *Introduction.*—In this section the theory of the Laplace transform is developed as far as is necessary for the solution of the linear-network equations.

There are two main alternative methods for deriving this theory. One, which employs the principle of superposition to synthesize the response to component generators, is similar to that often used in heuristic discussions of the Fourier transform theory, the response to an aperiodic function being regarded as the limiting case of the response to periodic functions.   Unfortunately a considerable amount of discussion is required if the full power and rigor of the Laplace transform are to be made apparent on this basis.   The second attack, which is followed exclusively in this chapter, introduces the Laplace transform as a mathematical concept with properties that fit it for use in the solution of linear integro-differential equations.   The latter method is much more compact than the former and has the additional advantage of starting from fundamentals and of requiring no special circuit theory.   A possible drawback,

however, is that the abstract and formal nature of the discussion may obscure both the reasons for the various steps in the development and the physical interpretation of the results obtained. In order to mitigate this possibility the purpose of the various theorems will be pointed out when they are stated. In addition, the theory will be illustrated by applying it to a particular problem, the various steps of whose solution will be made as the appropriate stages are reached.

The Laplace transform sets up a one-to-one correspondence between a function $f(t)$ defined along the real axis and a function $F(p)$ defined in the right half of the complex plane. Since it often happens that the transform of a function is a much simpler function than the original, a complicated relation involving the original function becomes a simple relation involving the transform function. In particular, linear integro-differential equations in a real variable transform into linear algebraic equations in the complex plane. It is this last property that makes the Laplace transform such a powerful tool in the solution of integro-differential equations.

The procedure for solving problems with the aid of the Laplace transform follows essentially the same course in all cases.[1] A mathematical process is found that transforms the set of equations in the original variables into a set of equations in the Laplace transforms of these variables. These equations are then solved to give explicit relations for the Laplace transforms of these variables. The solution for the original variables is then obtained with the aid of the inverse Laplace transform.

The development given in this section is along lines parallel to the procedure outlined in the previous paragraph. The Laplace transform is first defined; the transforms of some typical driving functions are derived; and the convergence of the defining integrals is discussed. The mathematical process for obtaining the transform equations is then outlined; expressions are obtained for the Laplace transform of the integral and derivative of a function; and the solution of the transform equation is obtained in a particular case. The basic theory is completed with the introduction of the inverse Laplace transform and the derivation of its relation to the direct Laplace transform. The section terminates with a proof of the so-called "convolution theorem" and a short discussion of its application to network theory.

---

[1] The Laplace transform is used to solve not only linear integro-differential equations in a single independent variable but also linear equations with an arbitrary number of independent variables, as well as partial differential, integral, and finite difference equations. The Laplace transform may also be applied to evaluate definite integrals, to sum series, and to develop the theory of functions both of the real and the complex variable.

On the first reading it may be desirable to turn, at this stage, to Sec. 1·10 where a summary of the $\mathcal{L}$-transform theory is given. The reader untrained in mathematics can utilize this summary as a guide to the important parts of this section and can ignore the remainder, which is not essential for the practical application of the method, although it cannot be omitted if the underlying theory is to be understood clearly and its limitations, as developed in this chapter, made apparent.

The Laplace Transform.—As stated in Sec. 1·3 no attempts are made to provide a theory more general than will be needed in circuit analysis. Accordingly the class of functions discussed will be restricted to those normally encountered in practice. These functions, which are all of exponential type, will now be defined.

DEFINITION 1. THE E-FUNCTIONS. Let $f(t)$ be a function of t defined at least for all positive values of t. Then if real positive finite numbers A and k exist such that

$$|f(t)| < Ae^{kt}$$

for all positive t, $f(t)$ is said to be an E-function, or a function of exponential type. If values of t exist such that $|f(t)| > Be^{k't}$ where $k' < k$ and B is any real positive finite number, then k is called the normal exponent.

The E-functions have a number of interesting properties of which only one will be stated here.[1] It will be needed when the Laplace transform of the integral[2] of a function is being considered.

THEOREM 1. Let $f(t)$ be an E-function; then $\int_0^t f(\tau)\,d\tau$ is also an E-function of normal exponent k.

It might be thought that by confining attention to the E-functions an appreciable loss would be suffered in the power and flexibility of the method. Thus all functions possessing poles on the real axis are barred from discussion and in particular the Kirchhoff-Dirac δ-function defined by

$$\left.\begin{array}{l}\delta(t) = 0, \qquad t \neq 0, \\ \int_{-\infty}^{\infty} \delta(t)\,dt = 1.\end{array}\right\} \tag{7}$$

[1] See Gustave Doetsch, *Theorie und Anwendung der Laplace-transformation*, Dover Publications, New York, 1943, for a discussion of those properties relevant to the Laplace transform theory.

[2] All the integrals considered here are to be taken in the sense of Lebesgue. The use of a more general integral such as the Young-Stieltjes integral would have considerable advantages, and there would certainly be a good case for deriving the $\mathcal{L}$-transform theory on this basis in a mathematical monograph, but the resulting analysis is too lengthy for a book of this kind.

Fortunately the difficulty can be overcome simply by defining these functions as the limit of E-functions. Thus $\delta(t)$ can be defined, among many possible alternatives, by the equations

$$\left.\begin{array}{l} \delta(t) = 0 : t < 0, \\ \delta(t) = \lim_{\alpha \to \infty} \alpha e^{-\alpha t} : t \geqq 0, \end{array}\right\} \tag{7a}$$

where it is to be understood that the limiting process will be carried out only at a stage in the analysis such that all subsequent operations are valid in the present theory.[1]

The arguments employed to justify the use of an analysis restricted to the class of E-functions naturally provoke the query as to whether or not further restrictions are advisable. Would anything be lost if attention were confined to those functions where the integral $\int_{\infty}^{\infty} |f(t)| \, dt$ is convergent? After all, every function realizable in practice is of this type. It is not possible to obtain infinite voltages and currents or even infinite rates of change of these quantities, because none of the applied signals can ever be infinite in duration. Nevertheless this further restriction will not be made in this chapter, because a considerable simplification of the mathematics is often achieved by employing mathematically idealized driving functions. If this additional restriction was imposed, it would not be possible to deal directly with the simple Heaviside step function

$$\begin{array}{ll} f(t) = 0, & t < 0, \\ f(t) = 1, & t \geqq 0; \end{array}$$

a limiting process, as used for the δ-functions, must be employed. This process gets increasingly clumsy when functions such as

$$f(t) = t, \qquad t > 0,$$

or

$$f(t) = e^{ct}, \qquad t > 0$$

are discussed. It is true that such functions cannot in practice be obtained, but when considering the response of a network to a sawtooth waveform or to the built-up waveform of an oscillator it is much simpler to regard those processes as continuing indefinitely rather than ending after an arbitrary time T. It may be further noted that in the theory of linear networks, unstable networks or even networks with zero damping cannot be discussed with the restricted theory.

[1] It may be noted that nothing would be gained, even if the restriction that all functions be E-functions, were removed. The integral in Eq. (7) is not a Lebesgue integral, but a Young-Stieltjes integral; such a function could not be introduced into the theory unless the latter were based throughout on the more general integral, a course of action already rejected because of its associated complexity of proofs.

The functions to be dealt with have now been defined   The concept of the Laplace transform may now be introduced.

DEFINITION 2. THE LAPLACE TRANSFORM. The Laplace transform $F(p)$ of a function $f(t)$, which is defined "almost everywhere" for nonnegative values of t, is defined by

$$F(p) \equiv \int_0^\infty e^{-pt} f(t) \, dt$$

for these values of p for which the integral converges.[1]

$F(p)$ can be written symbolically as

$$F(p) \equiv \mathfrak{L}[f(t)],$$

where $\mathfrak{L}$ is called the direct Laplace operator and is equal to $\int_0^\infty e^{-pt} \, dt$. In the remainder of this book the Laplace transform, operator, integral, and so forth, will be written $\mathfrak{L}$-transform, $\mathfrak{L}$-operator, $\mathfrak{L}$-integral, respectively.

In those regions of the complex plane where the $\mathfrak{L}$-integral is not convergent, $F(p)$ is defined by analytic continuation. As an example consider the case when $f(t) = 1$. The $\mathfrak{L}$-integral of $f(t)$ is convergent for all $R(p) > 0$ and is equal to $1/p$. But $1/p$ is a function regular all over the complex plane except for a simple pole at the origin. Hence $F(p)$ may be defined as equal to $1/p$ in the left half of the complex plane, where the $\mathfrak{L}$-integral of $f(t)$ diverges.

It may be emphasized at this point that no restrictions whatever are placed upon $f(t)$ for negative values of t. Thus the three functions

(a)	$f(t) = 0,$	$t < 0;$	$f(t) = 1,$	$t \geqq 0;$
(b)	$f(t) = e^{-kt},$	$t < 0;$	$f(t) = 1,$	$t \geqq 0;$
(c)	$f(t) = \Gamma(t),$	$t < 0;$	$f(t) = 1,$	$t \geqq 0;$

all have the same $\mathfrak{L}$-transform, namely, $1/p$. As will appear later, some advantages are to be gained by assuming that $f(t) = 0$ for $t < 0$, but there is no necessity that this be so.

[1] This definition is the most general one possible if the integral, regarded as the improper integra

$$\lim_{\substack{\Sigma \to 0 \\ \omega \to \infty}} \int_\Sigma^\omega e^{-pt} f(t) \, dt,$$

is defined in some sense.

A function is defined "almost everywhere" if it is defined everywhere save in a set of "zero measure." In most practical cases the words "almost everywhere" can be omitted from the above definition, the term being introduced for reasons connected with the proof of the theory, not its application.

The $\mathcal{L}$-transform of a few simple functions common in network theory will now be derived.

1. $f(t) = t^n$:

$$F(p) = \int_0^\infty e^{-pt}t^n \, dt \text{ is convergent for } R(p) > 0.$$

Let $pt = \tau$, then

$$F(p) = \int_0^\infty p^{-(n+1)}\tau^n e^{-\tau} \, d\tau = \frac{\Gamma(n+1)}{p^{n+1}}$$
$$= \frac{n!}{p^{n+1}}$$

if n is an integer.

2. $f(t) = e^{\alpha t}$, α *complex:*

$$F(p) = \int_0^\infty e^{-(p-\alpha)t} \, dt \text{ is convergent for } R(p) > R(\alpha) \text{ and } = 1/(p - \alpha).$$

From the transform for $e^{\alpha t}$ it is possible to deduce the $\mathcal{L}$-transform of $\cos \beta t$, $\sin \beta t$.

Thus

$$\mathcal{L}(\cos \beta t) = \mathcal{L}\frac{e^{i\beta t} + e^{-i\beta t}}{2} = \frac{1}{2}\left(\frac{1}{p - i\beta} + \frac{1}{p + i\beta}\right) = \frac{p}{p^2 + \beta^2};$$

similarly

$$\mathcal{L}(\sin \beta t) = \frac{\beta}{p^2 + \beta^2}.$$

The $\mathcal{L}$-transform of the hyperbolic sine and cosine can be found similarly.

3. $f(t) = \delta(t)$: From the transform of $e^{-\alpha t}$ it follows that

$$F(p) = \lim_{\alpha \to \infty} \frac{\alpha}{p + \alpha}.$$

$F(p)$ is commonly taken equal to 1 (that is, the limiting process is carried out immediately), but this will not be done here until it has been shown that this course of action is legitimate in network analysis.

A fuller list of $\mathcal{L}$-transforms will be found in Table 1·5 at the end of this chapter. A number of basic theorems needed later in this section will now be stated.

THEOREM 2. If $f(t)$ is an E-function with normal exponent k, then $F(p)$ is convergent for all $R(p) > k$, so that $F(p)$ converges over a right half of the complex plane.

THEOREM 3. If $f(t)$ is an E-function with normal exponent k, then $pF(p)$ is bounded for all $R(p) > k$ and $F(p)$ tends uniformly to zero as $p \to \infty$. This result is needed to establish the inverse $\mathcal{L}$-transform.

THEOREM 4.   If $f(t)$ is an E-function of normal exponent k, then $F(p)$ is regular in the half plane $R(p) > k$ and

$$\frac{d^n F(p)}{dp^n} = (-1)^n \int_0^\infty e^{-pt} t^n f(t) \, dt.$$

The next theorem is concerned with the important questions of the uniqueness of the Laplace transform.[1]

THEOREM 5.   If $f_1(t)$, $f_2(t)$ are E-functions with $\mathcal{L}$-transforms $F_1(p)$, $F_2(p)$ such that $F_1(p) = F_2(p)$ wherever the $\mathcal{L}$-integrals of $f_1(t)$ and $f_2(t)$ converge, then $f_1(t) = f_2(t)$ almost everywhere for nonnegative values of t.

As stated above, the concept of "equals almost everywhere" is not one that plays an important part in network theory; and if we neglect this refinement, the theorem can be loosely stated: "If two functions have the same $\mathcal{L}$-transform, then they are equal for nonnegative real values of t."

The $\mathcal{L}$-Transform of Integro-differential Equations.—In the previous section the $\mathcal{L}$-transform concept has been introduced, and some of its fundamental properties stated.   In this section it is proposed to show how it can be applied to transform integro-differential equations.

To keep the discussion concrete, attention will be focused on the linear equation with one dependent variable,

$$a_n \frac{d^n i(t)}{dt^n} + a_{n-1} \frac{d^{n-1} i(t)}{dt^{n-1}} + \cdots + a_1 \frac{di(t)}{dt} + a_0 i(t) + a_{-1} \int_0^t i(\tau) \, d\tau$$
$$+ \cdots + a_{-m} \left(\int_0^t d\tau \right)^m i(\tau) = e(t). \quad (8)$$

The solution of this equation will illustrate all the fundamental aspects of the general network solution, and in fact the network equations can often be reduced to the form of Eq. (8) by a judicious elimination process.

In the particular case of the solution of linear equations with one independent variable the mathematical process, mentioned above, for obtaining the transform equations is almost trivial.[2]   It consists merely in taking the $\mathcal{L}$-transform of both sides of the equation.   In order to effect this, three auxiliary theorems will be needed to give (1) the transform of a sum of functions, (2) the transform of the integral of a function, and (3) the transform of the derivative of a function.   As these theorems are fundamental to this chapter, the proofs are given in full.

THEOREM 6.   ADDITIVITY.   Let $f_1(t)$, $f_2(t)$ be two E-functions with normal exponent k and $\mathcal{L}$-transform $F_1(p)$, $F_2(p)$, respectively.   Then

---

[1] See Gustave Doetsch, *op. cit.*, Chap. 3, p. 7; M. Lerch, *Acta Math.*, **27**, 339–351, (1903).

[2] In other applications of the $\mathcal{L}$-transform this is far from being the case.

$a_1f_1(t) + a_2f_2(t)$ is an E-function with normal exponent k and $\mathcal{L}$-transform $a_1F_1(p) + a_2F_2(p)$, where a_1 and a_2 are arbitrary constants.

Proof: The first part of the theorem is trivial.   To prove the second part one has by definition,

$$\mathcal{L}[a_1f_1(t) + a_2f_2(t)] = \int_0^\infty e^{-pt}[a_1f_1(t) + a_2f_2(t)]\, dt = a_1 \int_0^\infty e^{-pt}f_1(t)\, dt$$
$$+ a_2 \int_0^\infty e^{-pt}f_2(t)\, dt = a_1F_1(p) + a_2F_2(p),$$

as required.

This result may be generalized by induction to give

$$\mathcal{L}\left[\sum_{r=1}^N a_r f_r(t)\right] = \sum_{r=1}^N a_r F_r(p). \tag{9}$$

THEOREM 7.   TRANSFORM OF THE INTEGRAL OF A FUNCTION.   If $f(t)$ is an E-function with normal exponent k, then $\int_0^t f(\tau)\, d\tau$ has an $\mathcal{L}$-transform $F(p)/p$, where $F(p)$ is the $\mathcal{L}$-transform of $f(t)$.

Proof: By definition,

$$\mathcal{L}\left[\int_0^t f(\tau)\, d\tau\right] = \lim_{x \to \infty} \int_0^x e^{-pt}\left[\int_0^t f(\tau)\, d\tau\right] dt;$$

now since $f(t)$ is an E-function, $\int_0^t f(\tau)\, d\tau$ is differentiable with respect to t for all finite t.   Hence integrating by parts one gets

$$\mathcal{L}\left[\int_0^t f(\tau)\, d\tau\right] = \lim_{x \to \infty}\left[-\frac{e^{-pt}}{p}\int_0^t f(\tau)\, d\tau\right]_0^x + \lim_{x \to \infty}\int_0^x \frac{e^{-pt}}{p}f(t)\, dt.$$

But, since by Theorem 1 $f(t)$ has normal exponent k, there exists A such that

$$\int_0^t f(\tau)\, d\tau < Ae^{kt};$$

hence for $R(p) > k$,

$$\lim_{x \to \infty} -\left[\frac{e^{-pt}}{p}\int_0^t f(\tau)\, d\tau\right]_0^x = 0.$$

Accordingly,

$$\mathcal{L}\left[\int_0^t f(\tau)\, d\tau\right] = \lim_{x \to \infty}\frac{1}{p}\int_0^x e^{-pt}f(t)\, dt = \frac{F(p)}{p}.$$

This theorem may be generalized by induction to give

THEOREM 7a.   If $f(t)$ is an E-function of normal exponent k and $\mathcal{L}$-transform $F(p)$, then the n-fold iterated integral $\left(\int_0^t d\tau\right)^{(n)} f(\tau)$ (n a

positive integer) is also an E-function of normal exponent k, $\mathcal{L}$-transformable at least for $R(p) > k$, such that

$$\mathcal{L}\left\{\left(\int_0^t d\tau\right)^n f(\tau)\right\} = \frac{F(p)}{p^n}. \tag{10}$$

THEOREM 8.   If $f(t)$ is an E-function with $\mathcal{L}$-transform $F(p)$ whose derivative $f'(t)$ is also an E-function with normal exponent k, and if $F(p)$ is the $\mathcal{L}$-transform of $f(t)$,

$$\mathcal{L}[f'(t)] = pF(p) - f(0).$$

Proof: Since $f'(t)$ is an E-function with normal exponent k, then, by Theorem 1, $\int_0^t f(\tau)\,d\tau + f(0)$ is an E-function with normal exponent k. But

$$f(t) = \int_0^t f'(\tau)\,d\tau + f(0+)$$

and hence

$$\mathcal{L}[f(t)] = \mathcal{L}\left[\int_0^t f'(\tau)\,d\tau + f(0+)\right]$$
$$= \mathcal{L}\left[\int_0^t f'(\tau)\,d\tau\right] + \mathcal{L}[f(0+)]$$

by Theorem 6.

But, by Theorem 7,

$$\mathcal{L}\left[\int_0^t f'(\tau)\,d\tau\right] = \frac{\mathcal{L}[f(t')]}{p};$$

and, since $\mathcal{L}[f(0+)] = f(0+)/p$,

$$\mathcal{L}[f'(t)] = p\mathcal{L}[f(t)] - f(0+) = pF(p) - f(0+).$$

This result may be generalized by induction to give

THEOREM 8a.   If $f^{(n)}(t)$, the nth derivative of $f(t)$, is an E-function with normal exponent k, if

$$\lim_{t\to +0} f(t) = f(0+), \qquad \lim_{t\to +0} f'(t) = f(0+), \cdots,$$
$$\lim_{t\to +0} f^{(n-1)}(t) = f^{(n-1)}(0+),$$

and if the $\mathcal{L}$-transform of $f(t)$ is $F(p)$, then

$$\mathcal{L}[f^{(n)}(t)] = p^n F(p) - [p^{n-1}f(0+) + p^{n-2}f'(0+) + \cdots + pf^{(n-2)}(0+) + f^{(n-1)}(0+)]. \tag{11}$$

It is now possible to obtain the transform of Eq. (8).   Taking the $\mathcal{L}$-transform of both sides and using the results of Theorems 6, 7, and 8 give

$$a_n p^n I(p) + a_{n-1} p^{n-1} I(p) + \cdots + a_0 I(p) + \frac{a_{-1} I(p)}{p}$$
$$+ \cdots + \frac{a_{-m} I(p)}{p^m}$$
$$= E(p) + a_n[i(0)p^{n-1} + i'(0)p^{n-2} + \cdots + i^{n-1}(0)]$$
$$+ a_{n-1}[i(0)p^{n-2} + i'(0)p^{n-3} + \cdots + i^{n-2}(0)]$$
$$\cdots$$
$$+ a_1 i(0),$$
$$= V(p), \tag{12}$$

where $I(p)$, $E(p)$ are the $\mathcal{L}$-transforms of $i(t)$ and $e(t)$, respectively. $V(p)$ is called the excitation transform function of the network. It contains the driving-function transform and the initial conditions.

It is seen that the use of the Laplace transform has in fact reduced the original integro-differential equation to an algebraic equation which may be solved for $I(p)$ to give

$$I(p) = \frac{1}{a_n p^n + a_{n-1} p^{n-1} + \cdots + a_0 + \dfrac{a_{-1}}{p} + \cdots + \dfrac{a_{-m}}{p^m}} V(p); \tag{13}$$

the denominator of Eq. (13) is called the "characteristic function," and the equation obtained by setting the characteristic function equal to zero is called the "characteristic equation."

Equation (13) is of the form

Response transform = system transform $\times$ excitation transform (14)

and this is typical of all network solutions.

The system function, which is usually a more complicated expression than the reciprocal of the characteristic function, is the transform of the response of the system to the driving function whose transform is unity. It has been proved earlier in this section that the δ-function has, in the limit, the $\mathcal{L}$-transform unity, and this is one of the reasons for the importance of the δ-function in network theory. The conditions under which it is legitimate to take the $\mathcal{L}$-transform of the δ-function equal to unity are discussed in the next section, where it is shown that this operation is valid for all but an unimportant set of networks.

To complete the solution of Eq. (8), a means must be found for deriving $i(t)$ from the expression for $I(p)$ given in Eq. (13). There are two main alternative methods for effecting this. One is based upon the concept of the inverse $\mathcal{L}$-transform. In the other, which is used by a number of writers, the transform function is broken down into a sum of simpler functions each of which may be recognized as one of the trans-

forms tabulated in Table 1·5 at the end of the chapter or in more complete tables given in other sources.[1]

The latter method has the advantage of mathematical simplicity in that no fresh ideas are introduced and no knowledge of contour integration is required, but the former will be followed here. For one thing it is quite easy to find network problems where the elementary method would fail or prove very cumbersome unless a formidable table of transforms was compiled. Furthermore, it will be shown that the standard procedure based upon the inverse Laplace transform is at least as short and compact as any alternative.

The Inverse £-transform.—A number of the results needed in this section have already been stated in the discussion of the direct Laplace transform. One or two auxiliary theorems are required, however, to complete the introductory material. The relevance of these auxiliary theorems may become clearer if the explicit expression for the inverse £-transform, hereafter called the $\mathcal{L}^{-1}$-transform, is stated, the proof being deferred until later.

If $F(p)$ is the £-transform of an E-function $f(t)$ with normal exponent k, then $f(t) = \mathcal{L}^{-1}[F(p)]$ is given by[2]

$$f(t) = \frac{1}{2\pi j} \lim_{\omega \to \infty} \int_{c-j\omega}^{c+j\omega} F(p)e^{pt}\,dp, \qquad t \geqq 0,$$

where c is a real positive number greater than k.

THEOREM 9. If $F(p)$ is the £-transform of an E-function $f(t)$ with normal exponent k, then

$$\lim_{\omega \to \infty} \int_{c-j\omega}^{c+j\omega} e^{pt}F(p)\,dp$$

is uniformly convergent for finite t if $c > k$. This result follows simply from the theorem that states that $pF(p)$ is bounded in the right half of the complex plane.

THEOREM 10. Let $F(p)$ be the £-transform of an E-function $f(t)$ with normal exponent k. Then

$$g(t) = \frac{1}{2\pi j} \lim_{\omega \to \infty} \int_{c-j\omega}^{c+j\omega} F(p)e^{pt}\,dp$$

is also an E-function with normal exponent k.

---

[1] See M. F. Gardner and J. L. Barnes *Transients in Linear Systems*, Vol. **1**, Wiley, New York, 1942.

[2] Here and throughout this section $f(t)$ is taken equal to $\frac{1}{2}[f(t+0) + f(t-0)]$ at a point of discontinuity.

This result is proved by showing that $e^{-ct}g(t)$ is bounded for all t; and since c is any number $> k$, $g(t)$ is an E-function with normal exponent k.

The main result of this section will now be proved. Once it is obtained, the derivation of the explicit relation for the $\mathcal{L}^{-1}$-transform can be achieved in a few lines.

THEOREM 11. Let $F(p)$ be the $\mathcal{L}$-transform of an E-function $f(t)$ with normal exponent k. If

$$g(t) = \frac{1}{2\pi j} \lim_{\omega \to \infty} \int_{c-j\omega}^{c+j\omega} F(p)e^{pt}\, dp \qquad \text{where } t > 0,\, c > k,$$

then

$$F(p) = \int_{0}^{\infty} e^{-pt} g(t)\, dt.$$

Proof: Let

$$G(p,\tau) = \int_{0}^{\tau} e^{-pt} g(t)\, dt = \int_{0}^{\tau} e^{-pt}\, dt\, \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(\sigma)e^{\sigma\tau}\, d\sigma$$

when $R(\sigma) > c$ on setting ω equal to its limiting value, a step that is justified by the convergence of the integral. Now the order of integration may be interchanged, since, by Theorem 9, the integrals involved are uniformly convergent. Hence

$$G(p,\tau) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(\sigma)\, d\sigma \int_{0}^{\tau} e^{-(p-\sigma)t}\, dt$$

$$= \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(\sigma)\, \frac{1 - e^{-(p-\sigma)\tau}}{p - \sigma}\, d\sigma.$$

But $\int \dfrac{F(\sigma)\, d\sigma}{p - \sigma}$ is absolutely convergent along the line $R(\sigma) = c$ by Theorem 3; and since $R(p) > c$, this implies that $\displaystyle\int_{c-j\infty}^{c+j\infty} \frac{F(\sigma)e^{-(p-\sigma)\tau}}{p - \sigma}\, d\sigma$ can be made arbitrarily small by choosing τ great enough. Accordingly if $G(p) = \lim\limits_{\tau \to \infty} G(p,\tau)$,

$$G(p) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{F(\sigma)}{p - \sigma}\, d\sigma.$$

Let $\oint d\sigma$ be the symbol representing integration in a clockwise direction around the contour composed of the straight line $R(\sigma) = c$ and the infinite half circle in the right half of the complex plane (Fig. 1·15). Let $\int_c ds$ be the integral around the half circle alone. Now $\sigma F(\sigma)$ is bounded all over the half plane $R(\sigma) > c$; hence $\displaystyle\int_c \frac{F(\sigma)}{p - \sigma}$ is zero for all $R(p) > c$ and

accordingly

$$G(p) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{F(\sigma)}{p - \sigma}\, d\sigma = \frac{1}{2\pi j} \oint \frac{F(\sigma)}{p - \sigma}\, d\sigma.$$

But $F(\sigma)$ is regular in and on the contour $\oint$; hence by the theory of residues,

$$\frac{1}{2\pi j} \oint \frac{F(\sigma)}{p - \sigma}\, d\sigma = F(p).$$

Hence

$$F(p) = \int_0^\infty e^{-pt} g(t)\, dt.$$
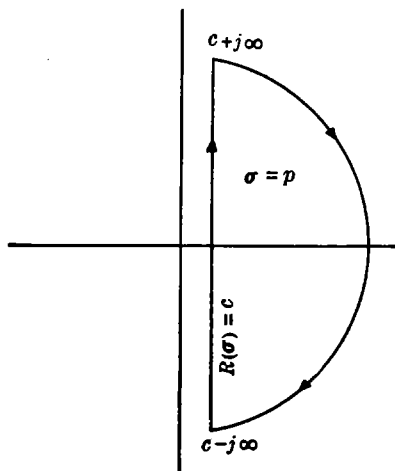
FIG. 1·15.—The infinite right-half circle.

THEOREM 12.   Let $f(t)$ be an E-function with normal exponent k and $\mathcal{L}$-transform $F(p)$; then

$$\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} e^{pt} F(p)\, dp = f(t) \qquad \text{almost everywhere for } t > 0,$$
$$= 0 \qquad \text{for } t < 0.$$

The proof of the first part of this theorem follows from Theorems 5 and 11.   By Theorem 11, if

$$g(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} e^{pt} G(p)\, dp,$$
$$G(p) = \int_0^\infty e^{-pt} g(t)\, dt,$$

for $R(p) > k$.

But by definition

$$F(p) = \int_0^\infty e^{-pt}f(t)\,dt$$

for $R(p) > k$.

Hence by the uniqueness Theorem 5,

$$f(t) = g(t)$$

almost everywhere for $t > 0$.

To prove the second part of the theorem it may be remembered that $pF(p)$ is bounded for all $R(p) > k$. Hence if $t < 0$,

$$\int_{c-j\infty}^{c+j\infty} e^{pt}F(p)\,dp = \oint e^{pt}F(p)\,dp,$$

where $\mathscr{S}$ has the same meaning as in Theorem 11, by Jordan's lemma.[1] Now $e^{pt}F(p)$ is regular (and bounded) for all $R(p) > k$. Hence by Cauchy's theorem

$$\oint e^{pt}F(p)\,dp = 0$$

for $t < 0$.

The second part of Theorem 12 provides the main reason why it is sometimes desirable to assume $f(t) = 0$ for $t < 0$, since in this case

$$f(t) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} e^{pt}F(p)\,dp$$

for almost all t.

The $\mathcal{L}^{-1}$-transform is an operator that may be written symbolically

$$\mathcal{L}^{-1} \equiv \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} e^{pt}\,dp. \tag{15}$$

Returning now to Eq. (8) it may be seen that the explicit expression for the response function is

$$i(t) =$$

$$\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} \frac{e^{pt}V(p)\,dp}{a_n p^n + a_{n-1}p^{n-1} + \cdots + a_1 p + a_0 + \dfrac{a_{-1}}{p} + \cdots + \dfrac{a_{-m}}{p^m}}. \tag{16}$$

This integral can most easily be evaluated by the theory of residues[2] which will be discussed briefly below.

---

[1] See E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th ed., Cambridge, London, 1927, p. 115.

[2] For a fuller treatment see E. T. Whittaker and G. N. Watson, *op. cit.*, pp. 164–189.

The Theory of Residues.—The evaluation of integrals in the complex plane is often best accomplished by transforming them into integrals around a closed contour. In the £-transform theory the integrals to be evaluated are of the form

$$\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(p)e^{pt}\, dp,$$

and these can be transformed into integrals around a closed contour if the integral of $F(p)e^{pt}$ around the infinite half circle in either the right or left half of the complex plane is zero.

A sufficient condition that this is so can be deduced from Jordan's lemma which states that

$$\int e^{pt}F(p)\, dp$$

is zero for all positive t when the path of integration is the infinite left half circle, provided that $F(p) = O(1/p)$ for all $R(p) < 0$.[1]

Two corollaries of this result may be stated:

1. $\int e^{pt}F(p)\, dp$ integrated over the infinite *left* half circle is zero for all $t > T$ if $e^{pT}F(p) = O(1/p)$ for all $R(p) < 0$.
2. $\int e^{pt}F(p)\, dp$ integrated over the infinite *right* half circle is zero for all $t < T$ if $e^{pT}F(p) = O(1/p)$ for all $R(p) > 0$. In particular the integral is zero for all $t < 0$ if $F(p) = O(1/p)$ for all $R(p) > 0$.

It has already been noted in the previous section that the £-transform of an E-function is $O(1/p)$ for $R(p) > c$, and the majority of the £-transforms occurring in network theory are, in fact, $O(1/p)$ for all $R(p)$. The chief exceptions to this rule are functions of the form $e^{-pT}G(p)$ where $G(p)$ is $O(1/p)$.

For the time being let it be assumed that $\dfrac{1}{2\pi j} \displaystyle\int_{c-j\infty}^{c+j\infty} F(p)e^{pt}\, dp$ is equal to either

$$\frac{1}{2\pi j} \int_{\Gamma_1} F(p)e^{pt}\, dp$$

or

$$\frac{1}{2\pi j} \int_{\Gamma_2} F(p)e^{pt}\, dp,$$

where Γ_1 is the contour formed by the straight line $R(p) = c$ and the infinite *right* half circle and Γ_2 is the contour formed by the straight line

---

[1] $F(p)$ is said to be $O(1/p)$ if $pF(p)$ is bounded for all sufficiently large p.

$R(p) = c$, the infinite *left* half circle and the two straight-line segments[1] jointing the points $p = c \pm j\infty$ and $p = \pm j\infty$.

It is a well-known result of the theory of functions that the integral of a meromorphic function taken in an anticlockwise direction around a closed contour containing a number of simple poles is equal to $2\pi j$ times the sum of the residues of these poles.[2]   The residue of a pole is defined as follows:

Let $H(p)$ be a function regular and nonzero in the neighborhood of the point $p = p_0$.   Then the residue of the function $H(p)/(p - p_0)^n$ at its nth order pole $p = p_0$ is

$$\frac{1}{(n - 1)!}\left[\frac{d^{n-1}H(p)}{dp^{n-1}}\right]_{p=p_0}.$$

It follows that if $F(p)$ is a function $O(1/p)$ with poles at p_1, p_2, p_3, . . . , p_n of order r_1, r_2, r_3, . . . , r_n respectively, so that

$$F(p) = \frac{H(p)}{\displaystyle\prod_{s=1}^{n} (p - p_s)^{r_s}}, \tag{17}$$

where $H(p)$ is a function regular all over the finite part of the complex p-plane, then

$$\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(p)e^{pt}\, dp = \sum_{t=1}^{n}\frac{1}{(r_t - 1)!}\left\{\frac{d^{r_t-1}}{dp^{r_t-1}}[F_t(p)e^{pt}]\right\}_{p=p_t}, \tag{18}$$

where

$$F_t(p) = (p - p_t)^{r_t}F(p). \tag{19}$$

All the system transforms of the networks considered in this chapter are similar in form to Eq. (17); since the common excitation transforms are of this form also, the response of any network to one of the common driving functions can be evaluated from the formula of Eq. (18).   It is possible to find excitation functions whose transforms are not of the type of Eq. (17), but a discussion of such cases is outside the scope of this book and will not be attempted here.[3]

To illustrate the above theory the $\mathcal{L}^{-1}$-transforms of several functions will be derived.

---

[1] $\int F(p)e^{pt}\, dp$ over the straight line $p = c + j\infty$ to $p = j\infty$ is zero if $\displaystyle\int_{c-j\infty}^{c+j\infty} F(p)e^{pt}\, dp$ converges.

[2] The function must be regular upon the contour.

[3] See Gustave Doetsch, *op. cit.*

1. $F(p) = \dfrac{1}{p + a}$:

$$f(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{e^{pt}}{p + a}\, dp$$
$$= \frac{1}{2\pi j} \oint_{\Gamma_2} \frac{e^{pt}}{p + a}\, dp,$$

since $F(p) = O(1/p)$; $R(p) < 0$.

But $e^{pt}/(p + a)$ has a simple pole at $p + a = 0$ with residue e^{-at}. Hence $f(t) = e^{-at}$.

2. $F(p) = \dfrac{p + b}{(p + a)^2}$:

$$f(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{e^{pt}(p + b)}{(p + a)^2}\, dp$$
$$= \frac{1}{2\pi j} \oint_{\Gamma_2} \frac{e^{pt}(p + b)}{(p + a)^2}\, dp,$$

since $F(p) = O(1/p)$, $R(p) < 0$.

But $e^{pt}(p + b)/(p + a)^2$ has a double pole at $p + a = 0$, with residue

$$\frac{d}{dp}\,[(p + b)e^{pt}]_{p=-a} = e^{-at} + t(b - a)e^{-at}.$$

Hence $f(t) = [1 + (b - a)t]e^{-at}$.

3. $F(p) = \dfrac{p}{(p + a)(p + b)(p + c)}$ where $a \neq b \neq c \neq 0$:

$$f(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{pe^{pt}}{(p + a)(p + b)(p + c)}\, dp$$
$$= \frac{1}{2\pi j} \oint_{\Gamma_2} \frac{pe^{pt}}{(p + a)(p + b)(p + c)}\, dp,$$

since $F(p) = O(1/p)$, $R(p) < 0$.

But $\dfrac{pe^{pt}}{(p + a)(p + b)(p + c)}$ has three poles, at $p + a = 0$, $p + b = 0$, $p + c = 0$, with residues

$$\frac{ae^{-at}}{(a - b)(c - a)}, \qquad \frac{be^{-bt}}{(a - b)(b - c)}, \qquad \frac{ce^{-ct}}{(c - a)(b - c)}$$

respectively.

Hence

$$f(t) = \frac{ae^{-at}}{(a - b)(c - a)} + \frac{be^{-bt}}{(a - b)(b - c)} + \frac{ce^{-ct}}{(c - a)(b - c)}.$$

4. $F(p) = \dfrac{e^{-pr}}{p}$:

Here

$$f(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{e^{-pr}\, e^{pt}\, dp}{p}$$
$$= \frac{1}{2\pi j} \oint_{\Gamma_1} \frac{e^{-p(r-t)}\, dp}{p}$$

for $t < \tau$, since $e^{pr}F(p)$ is $O(1/p)$ for $R(p) > 0$ and $t < \tau$, and

$$f(t) = \frac{1}{2\pi j} \oint_{\Gamma_2} \frac{e^{+p(t-r)}\, dp}{p}$$

for $t > \tau$, since $e^{pt}F(p)$ is $O(1/p)$ for $R(p) < 0$ and $t > \tau$. Now since $e^{-p(r-t)}/p$ has no poles inside contour Γ_1,

$$f(t) = 0$$

for $t < \tau$.

The pole of $e^{-p(r-t)}/p$ is at $p = 0$, which lies inside Γ_2, and has residue 1. Hence

$$f(t) = 1$$

for $t > \tau$.

The integral defining $f(t)$ does not converge at the point of discontinuity $t = \tau$, but following the course of action adopted throughout this chapter, $f(\tau)$ is taken as

$$f(\tau) = \frac{f(\tau - 0) + f(\tau + 0)}{2} = \frac{1}{2}.$$

To conclude this section the conditions under which it is legitimate to take the transform of the δ-function as unity will be considered.

As a basis for discussion consider Eq. (16) where $V(p) = \alpha/(p + \alpha)$. Then if $n \geqq 0$,

$$\frac{\alpha p^m}{(p + \alpha)(a_n p^{m+n} + a_{n-1} p^{m+n-1} \cdots a_0 p^m + a_{-1} p^{m-1} + \cdots + a_{-m})}$$

is $O(1/p)$ for $R(p) < 0$.

Hence

$$i(t)$$
$$= \frac{1}{2\pi j} \oint_{\Gamma_2} \frac{\alpha p^m e^{pt}\, dp}{(p + \alpha)(a_n p^{m+n} + a_{n-1} p^{m+n-1} + \cdots + a_0 p^m + \cdots + a_{-m})}; \tag{20}$$

this integrand has poles at

$$p + \alpha = 0$$

and

$$a_n p^{m+n} + a_{n-1} p^{m+n-1} + \cdots + a_0 p^m + a_{-1} p^{m-1}$$
$$+ \cdots + a_{-m} = 0. \quad (21)$$

Two conditions have to be satisfied if $\mathcal{L}[\delta(t)]$ is to be taken equal to unity.

1. The residues of the poles of Eq. (21) must be the same whether $\mathcal{L}[\delta(t)] = 1$ or $\mathcal{L}[\delta(t)] = \lim_{\alpha \to \infty} \alpha e^{-\alpha t}$.

2. The residue of the pole at $p + \alpha = 0$ must $\to 0$ almost everywhere when $\alpha \to \infty$. The first condition is obviously satisfied, because $\frac{\alpha}{p_t + \alpha} \to 1$ and $\frac{\alpha}{(p_t + \alpha)^m} (m > 1) \to 0$ for all finite p_t when $\alpha \to \infty$. The second condition is satisfied only if $n > 1$, for the residue of $p + \alpha = 0$ is

$$\frac{(-\alpha)^{m+1} e^{-\alpha t}}{a_n(-\alpha)^{m+n} + a_{n-1}(-\alpha)^{m+n-1} + \cdots + a_{-m}},$$

which, as may be seen by inspection, $\to 0$ as $\alpha \to \infty$ for $t \geqq 0$ if $n \geqq 2$, $\to 0$ as $\alpha \to \infty$ for $t > 0$ and $\to 1/(a_n)$ as $\alpha \to \infty$ for $t = 0$ if $n = 1$ and diverges as $\alpha \to \infty$ if $n = 0$.

It may be concluded that the $\mathcal{L}$-transform of the δ-function may be taken equal to unity when the function is applied to a system whose system function is $O(1/p)$.

The Convolution Theorem.—One more theorem will be proved in this section: the so-called convolution or Faltung theorem. Although this does not play an essential role in the $\mathcal{L}$-transform theory, it is sometimes useful in solving practical problems and is certainly valuable when discussing the general solution of the network.

THEOREM 13.   THE CONVOLUTION THEOREM.   Let $f(t)$ and $g(t)$ be two E-functions with normal exponents k_1, k_2 and $\mathcal{L}$-transforms $F(p)$, $G(p)$ respectively. Then if $h(t)$ is a function whose $\mathcal{L}$-transform is $F(p)G(p)$,

$$h(t) = \int_0^t f(\tau) g(t - \tau) \, d\tau \quad (22)$$

almost everywhere.

Proof: Let c be a number $> k_1, k_2$.   Then

$$h(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(p)G(p) e^{pt} \, dp$$

almost everywhere, since if $F(p)$ and $G(p)$ are both $\mathcal{L}^{-1}$-transformable, $F(p)G(p)$ are so *a fortiori*.
But

$$F(p) = \int_0^\infty f(\tau) e^{-p\tau} \, d\tau;$$

hence

$$h(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} G(p) \left[\int_0^\infty f(\tau)e^{-p\tau}\,d\tau \right] e^{+pt}\,dt$$
$$= \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \left[\int_0^\infty G(p)f(\tau)e^{p(t-\tau)}\,d\tau \right] dp.$$

Since the two integrals are uniformly convergent in the region under consideration, the order of integration can be changed to give

$$h(t) = \int_0^t f(\tau) \left[\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} G(p)e^{p(t-\tau)}\,dp \right] d\tau$$
$$+ \int_t^\infty f(\tau) \left[\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} G(p)e^{p(t-\tau)}\,dp \right] d\tau.$$

Now by Theorem 12

$$\frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} G(p)e^{p(t-\tau)}\,dp \begin{array}{l} = g(t-\tau) \quad \text{if } t \geqq \tau, \\ = 0 \qquad\qquad \text{if } t < \tau. \end{array}$$

Hence

$$h(t) = \int_0^t f(\tau)g(t-\tau)\,d\tau,$$

where $h(t)$ is called the convolution of $f(t)$ and $g(t)$ and is written symbolically

$$h(t) = f(t) * g(t). \tag{23}$$

This theorem will be applied in the next section to find the solution of Eq. (8).

Conditions for Stability, Steady-state Response, etc., of Integro-differential Equations in a Single Variable.—In this section some of the general features of the solution of Eq. (15) will be discussed.   Most of the points made will be applicable *mutatis mutandis* to the solutions of the general network case.

It has been shown above that the solution of Eq. (15) is

$$i(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{p^m V(p)e^{pt}}{(a_n p^{m+n} + a_{n-1}p^{m+n-1} + \cdots + a_{-m})}\,dp; \tag{24}$$

and provided that e^{-pt} times the integrand is $O(1/p)$, this may be written

$$i(t) = \frac{1}{2\pi j} \oint_{\Gamma_2} \frac{p^m V(p)e^{pt}}{(a_n p^{m+n} + a_{n-1}p^{m+n-1} + \cdots + a_{-m})}\,dp. \tag{25}$$

It is always possible to factor the polynomial denominator of Eq. (25) into the form

$$\prod_{l=1}^{s} a_n(p - p_l)^{r_l},$$

where

$$\sum_{l=1}^{s} r_l = m + n.$$

Now the residue of the pole at $p = p_l$ is, in general, a function of the form

$$e^{p_l t} B_l(t),$$

where $B_l(t)$ is a polynomial in t of degree $r_l - 1$.

If $a_n, a_{n-1}, \ldots, a_{-m}$ are all real, then to every root $p_l = \alpha_l + j\beta_l$ there corresponds a conjugate root $\bar{p}_l = \alpha_l - j\beta_l$ so that $i(t)$ is a sum of terms of the kind

$$e^{\alpha_l t} \cos \beta_l t B_{lc}(t) + e^{\alpha_l t} \sin \beta_l t B_{ls}(t), \tag{26}$$

where $B_{lc}(t)$ and $B_{ls}(t)$ are both polynomials of degree $r_l - 1$ and, provided that $V(p)$ is also a rational or regular function of p, $i(t)$ consists solely of terms of the form of Eq. (26).

If $\alpha_l > 0$, the corresponding term in $i(t)$ tends to infinity with t. Hence a necessary and, in fact, sufficient condition that the system be stable is that all the poles of the system transform lie on or to the left of the imaginary axis in the complex p-plane. If all the poles lie to the left of the imaginary axis, the system is called absolutely stable. One of the most important cases occurs when the driving function is of the form

$$e(t) = e^{i\omega t}.$$

Then, if all the initial conditions are zero,

$$V(p) = \frac{1}{p + j\omega}.$$

If the system is absolutely stable, the residues of the poles of the system transform all tend to zero as t tends to infinity and the response tends in the limit to the residue of the pole at $p + j\omega \backsim 0$. This residue is of the form

$$i(t) = Y(j\omega)e^{i\omega t},$$

where $Y(j\omega)$ is a rational function of $j\omega$ independent of the time. This response is the familiar steady-state response to a sinusoidal generator of angular frequency ω. A further discussion of the steady-state response is given below in Sec. 1·8.

To conclude the discussion of Eq. (8) consider the application of the convolution theorem to its solution. Let $v(t)$ be the excitation function

of the system,[1] and let the system function, that is, the response of the system to a δ-function, be $Q(t)$.   Then

$$i(t) = v(t) * Q(t) = \int_0^\tau Q(\tau)v(t - \tau) \, d\tau = \int_0^\tau Q(t - \tau)v(\tau) \, d\tau. \quad (27)$$

In practice it is usually simpler to evaluate $i(t)$ directly from Eq. (16) with the aid of the theory of residues than to use Eq. (27).   However, Eq. (27) is a compact form in which to present the solution and may be preferable should it be desired to evaluate the response of the system to a whole series of driving functions.

1·5. The Use of the £-transform in the Solution of Network Problems. In this section the use of the £-transform is illustrated by applying it to
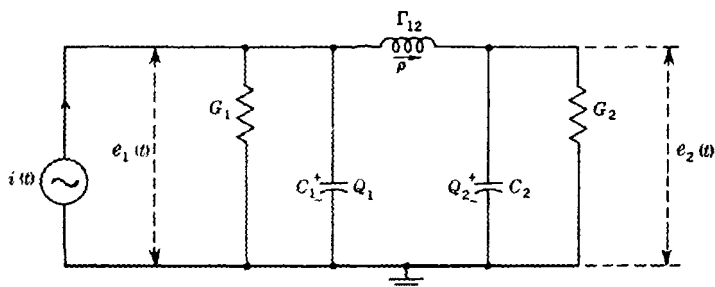
FIG. 1·16.—Two-node original network.

solve a special problem.   As was the case in deriving the network equations, the discussion is made more detailed than is necessary in order to bring out the general aspects of the solution.   The extension to the general case can then be made with a minimum of explanation.

In the course of the treatment the concept of the transform network will be introduced in such a way as to make its practical value clear.

The network chosen as an example is shown in Fig. 1·16.   It represents a terminated low-pass filter section driven from a pentode source. The response of this network to a δ-function will now be found.

To keep the discussion as general as possible it is assumed that there is an initial current ρ flowing in the inductor Γ_{12} in the direction shown in Fig. 1·16, and that the condensers have initial charges Q_1, Q_2, so that

$$\lim_{t \to 0} e_1(t) = e_1(0) = \frac{Q_1}{C_1},$$

$$\lim_{t \to 0} e_2(t) = e_2(0) = \frac{Q_2}{C_2},$$

---

[1] Assuming that this is an E-function, which is not the case if the initial conditions are not all zero.

where $e_1(t)$, $e_2(t)$ are the voltages across the two independent node pairs, the grounded node being taken as reference node.

This network is a special case of the network of Fig. 1·10, and hence from Eqs. (5) the equations of state can be written

$$\left.\begin{array}{l} i_1(t) \, - \, \rho \, = \, y_{11}e_1(t) \, + \, y_{12}e_2(t), \\ \qquad \quad +\rho \, = \, y_{21}e_1(t) \, + \, y_{22}e_2(t), \end{array}\right\} \tag{28}$$

where y_{11}, y_{22}, y_{12}, and y_{21} are operators given by

$$\left.\begin{array}{ll} y_{11} = G_1 + C_1 \dfrac{d}{dt} + \Gamma_{12} \displaystyle\int_0^t d\tau, & y_{12} = -\Gamma_{12} \displaystyle\int_0^t d\tau, \\[3mm] y_{21} = -\Gamma_{12} \displaystyle\int_0^t d\tau, & y_{22} = G_2 + C_2 \dfrac{d}{dt} + \Gamma_{12} \displaystyle\int_0^t d\tau \end{array}\right\} \tag{29}$$

If

    $I(p)$ is the £-transform of $i(t)$,

    $E_1(p)$ is the £-transform of $e_1(t)$,

    $E_2(p)$ is the £-transform of $e_2(t)$,

then taking the £-transform of both sides of the Eqs. (28) yields

$$\left.\begin{array}{l} I(p) \, - \, \dfrac{\rho}{p} \, = \, \left(C_1 p + G_1 + \dfrac{\Gamma_{12}}{p}\right) E_1(p) \, - \, \dfrac{\Gamma_{12}}{p} E_2(p) \, - \, Q_1, \\[3mm] \qquad \dfrac{\rho}{p} \, = \, - \dfrac{\Gamma_{12}}{p} E_1(p) \, + \, \left(C_2 p + G_2 + \dfrac{\Gamma_{12}}{p}\right) E_2 p \, - \, Q_2. \end{array}\right\} \tag{30}$$

If

$$\left.\begin{array}{l} J_1(p) \, = \, I(p) \, - \, \dfrac{\rho}{p} \, + \, Q_1, \\[3mm] J_2(p) \, = \, \dfrac{\rho}{p} \, + \, Q_2, \end{array}\right\} \tag{31}$$

then Eq. (30) assumes the general form

$$\left.\begin{array}{l} J_1(p) \, = \, Y_{11}(p)E_1(p) \, + \, Y_{12}E_2(p), \\ J_2(p) \, = \, Y_{21}(p)E_1(p) \, + \, Y_{22}E_2(p), \end{array}\right\} \tag{32}$$

where

$$\left.\begin{array}{ll} Y_{11}(p) = C_1 p + G_1 + \dfrac{\Gamma_{12}}{p}, & Y_{12}(p) = - \dfrac{\Gamma_{12}}{p}, \\[3mm] Y_{21}(p) = - \dfrac{\Gamma_{12}}{p}, & Y_{22}(p) = C_2 p + G_2 + \dfrac{\Gamma_{12}}{p}. \end{array}\right\} \tag{33}$$

$J_1(p)$, $J_2(p)$ are the excitation transforms for node pairs 1 and 2 respectively.   They are equal to the sum of the driving transform across the node pair, the initial charges on the condensers, and p^{-1} times the initial inductor currents flowing into the nodes.

Equations (32), the transform equations, are very similar to the normal steady-state equations for a network, a fact that enables one to derive them directly from the so-called transform network, which will now be introduced. Consider the diagram of Fig. 1·17, where the normal circuit elements have been replaced by the corresponding p-admittances. If Kirchhoff's second law is applied to the two node pairs of the network on the assumption that the current through an element is equal to the product of the voltage across it and the p-admittance, one gets the system of Eqs. (32). But since the transform diagram can be constructed directly from the network, the transform equations can be written down immediately without having first to set up the integro-differential equations and then to transform them.
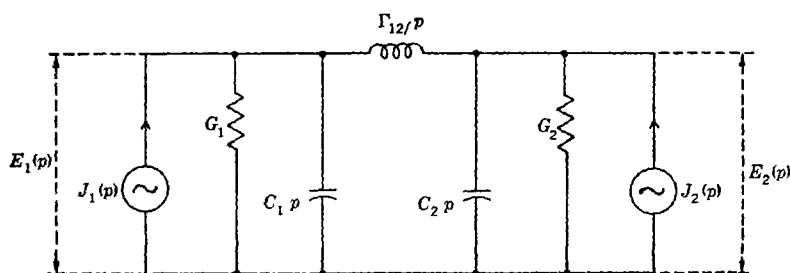
Fig. 1·17.—Two-node transform network.

The simplification of the analysis that results from using the transform network is often considerable, especially if there are a number of nonzero constants of integration. Accordingly a special section is devoted to deriving the procedure for setting it up in the general case, both for the mesh and nodal analysis, and to developing in greater detail the inherent advantages thus gained.

This subject is put aside for the moment, however, in order to return to Eqs. (32). These can be solved explicitly for $E_1(p)$, $E_2(p)$ to give

$$E_1(p) = \frac{J_1(p)Y_{22}(p) - J_2(p)Y_{12}(p)}{Y_{11}(p)Y_{22}(p) - Y_{12}(p)Y_{21}(p)},$$
$$E_2(p) = \frac{J_2(p)Y_{11}(p) - J_1(p)Y_{21}(p)}{Y_{11}(p)Y_{22}(p) - Y_{12}(p)Y_{21}(p)}.$$

The connection of the special with the general case is made more evident if this solution is itself put in more general terms.

Let

$$\Delta(p) = \begin{vmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{vmatrix}, \tag{34}$$

and let $\Delta_j(p)$ be the determinant formed by replacing the jth column of $\Delta(p)$ by $\begin{pmatrix} J_1 \\ J_2 \end{pmatrix}$. Then if $\Delta_{jk}(p)$ is the cofactor of y_{jk} in $\Delta(p)$, so that

$$\sum_{k=1}^{2} Y_{jk}\,\Delta_{jk} = \Delta(p)$$

and

$$\sum_{k=1}^{2} J_k\,\Delta_{kj} = \Delta_j(p),$$

the solution can be written

$$E_1(p) = \frac{\Delta_1(p)}{\Delta(p)} = \frac{J_1(p)\,\Delta_{11}(p) + J_2(p)\,\Delta_{12}(p)}{\Delta(p)}$$
$$= J_1(p)\,\Omega_{11}(p) + J_2(p)\,\Omega_{12}(p),$$
$$E_2(p) = \frac{\Delta_2(p)}{\Delta(p)} = \frac{J_1(p)\,\Delta_{21}(p) + J_2(p)\,\Delta_{22}(p)}{\Delta(p)}$$
$$= J_1(p)\,\Omega_{21}(p) + J_2(p)\,\Omega_{22}(p),$$

or finally

$$E_j(p) = \sum_{i=1}^{2} J_i(p)\,\Omega_{ji}(p) \qquad (j = 1, 2), \tag{35}$$

where

$$\Omega_{ji}(p) = \frac{\Delta_{ji}(p)}{\Delta(p)}.$$

The elements $\Omega_{ji}(p)$ are system transforms in the sense defined above in Sec. 1·4; thus it may be seen that the transforms of the response functions can be expressed as the sum of a number of simple products of excitation and system transforms.

In the important case when all the initial constants are zero,

$$\left.\begin{aligned} E_1(p) &= \frac{I(p)\,\Delta_{11}(p)}{\Delta(p)}, \\ E_2(p) &= \frac{I(p)\,\Delta_{21}(p)}{\Delta(p)}. \end{aligned}\right\} \tag{36}$$

To complete the solution it is necessary to obtain the explicit expressions for $e_1(t)$, $e_2(t)$. Now the $\Omega_{jk}(p)$ are rational functions of p; hence the $J_k(p)\,\Omega_{jk}(p)$ terms are of the same form as the response transform of Eq. (13) and response functions may be found by the same means. As in the case of Eq. (8) the direct extraction of the residues is usually the quickest

and most convenient method of obtaining the solution, and this procedure is illustrated numerically below. The convolution theorem, however, can be used to express the response function as a sum of convolutions of an excitation and a system function. Thus if

$$\omega_{jk}(t) = \mathcal{L}^{-1}[\Omega_{jk}(p)]$$

and

$$j_k(t) = \mathcal{L}^{-1}[J_k(p)],$$

then

$$e_i(t) = \sum_{k=1}^{2} \omega_{jk}(t) * j_k(t).$$

In the general case $\omega_{jk}(t)$ is the voltage (current) response of the kth node (mesh) to a δ-function impulse applied across the jth node (mesh).

The poles of $\Omega_{jk}(p)$ occur at the roots of the characteristic equation of the network, that is, at those values of p for which $\Delta(p) = 0$. A necessary condition that the network be stable is that all the roots of $\Delta(p)$ lie in the left half of the complex p-plane.

The solution of the chosen example will now be considered for the case when $i(t)$ is a δ-function. In order to simplify the labor of computation it is assumed that

$$Q_1 = Q_2 = 0, \qquad \rho = 0, \qquad G_1 = G_2 = G, \qquad C_1 = C_2 = C.$$

The fact that the system transforms are all $O(1/p)$ justifies taking $\mathcal{L}[\delta(t)] = 1$. Substituting these values for the driving function and initial conditions in Eq. (35) gives

$$E_2(p) = \frac{Cp + G + \dfrac{\Gamma_{12}}{p}}{\left(Cp + G + \dfrac{\Gamma_{12}}{p}\right)^2 - \dfrac{\Gamma_{12}^2}{p^2}} \cdot 1 = \frac{Cp + G + \dfrac{\Gamma_{12}}{p}}{(Cp + G)\left(Cp + G + \dfrac{2\Gamma_{12}}{p}\right)}.$$

If $8C\Gamma_{12}/G^2 > 1$ the roots of $Cp + G + 2\Gamma_{12}/p = 0$ are complex and are of the form $-p_1 \pm jp_2$, where

$$p_1 = \frac{G}{2C}, \qquad p_2 = \frac{G}{2C}\sqrt{\frac{8\Gamma_{12}C}{G^2} - 1}.$$

$E_2(p)$ has poles at $Cp + G = 0$ and $p = -p_1 \pm jp_2$, and the residues of $e^{pt}E_2(p)$ at these poles are as follows.

Residue at $Cp + G = 0$ is $\dfrac{1}{2C}\, e^{-\left(\frac{Gt}{C}\right)}$.

Residue at $p = -p_1 + jp_2$ is $\dfrac{e^{-p_1 t}e^{jp_2 t}}{2jp_2} \cdot \dfrac{-p_1 + jp_2}{2C}$.

Residue at $p = -p_1 - jp_2$ is $\dfrac{e^{-p_1 t}e^{-jp_2 t}}{-2jp_2} \cdot \dfrac{-p_1 - jp_2}{2C}$.

But $e_2(t)$ is the sum of the residues of $E_2(p)e^{pt}$; hence

$$e_2(t) = \frac{1}{2C} e^{-\frac{Gt}{C}} - \frac{p_1 e^{-p_1 t} \sin p_2 t}{2C p_2} + \frac{e^{-p_1 t} \cos p_2 t}{2C},$$

where

$$p_1 = \frac{G}{2C} \quad \text{and} \quad p_2 = \frac{G}{2C} \sqrt{\frac{8C\Gamma_{12}}{G^2} - 1}.$$

1·6. General Solution of the Network Equations. *The Nodal Case.--* The procedure in the general case is a simple extension of that already illustrated for the two-node network of Sec. 1·5; thus a detailed derivation of the results is not necessary.

It was shown in Sec. 1·3 that the general equation in the nodal case may be written in the form

$$i_j(t) + \rho_j = \sum_{k=1}^{n} y_{jk} e_k(t) \qquad (j = 1, \cdots, n), \qquad (37)$$

where

$$y_j = C_j \frac{d}{dt} + G_j + \Gamma_j \int_0^t dt,$$

$$-y_{jk} = C_{jk} \frac{d}{dt} + G_{jk} + \Gamma_{jk} \int_0^t dt,$$

$$y_{jj} = y_j - \sum_{k \neq j = 1}^{n} y_{jk},$$

and ρ_j is taken positive if the initial inductor current is flowing into the jth node.   Here n of the initial conditions are included in Eq. (37), and n more are given by the initial values of the node voltages

$$\lim_{t\to 0} e_1(t) = e_1(0), \quad \lim_{t\to 0} e_2(t) = e_2(0), \quad \cdots,$$
$$\lim_{t\to 0} e_n(t) = e_n(0).$$

Taking the $\mathcal{L}$-transform of both sides of each of Eqs. (37) gives[1]

$$I_j(p) + \frac{\rho_j}{p} + \sum_{k \neq j = 1}^{n} C_{jk}[e_j(0) - e_k(0)] + C_j e_j(0)$$
$$= \sum_{k=1}^{n} Y_{jk}(p) E_k(p) \qquad (j = 1, \cdots, n), \quad (38)$$

[1] $\mathcal{L}\left[C_{jk} \dfrac{d}{dt} e_k(t) \right] = pC_{jk}E_k(p) - C_{jk}e_k(0)$, etc.

where $I_i(p)$, $E_k(p)$ are the $\mathcal{L}$-transforms of $i_i(t)$ and $e_k(t)$ respectively and

$$-Y_{jk}(p) = C_{jk}(p) + G_{jk} + \frac{\Gamma_{jk}}{p}, \; i \neq k;$$

$$Y_j = C_j p + G_j + \frac{\Gamma_j}{p}, \qquad Y_{jj} = Y_j - \sum_{k=1}^{n} Y_{jk}. \tag{38a}$$

Let

$$\gamma_j = \sum_{k \neq j, k=1}^{n} C_{jk}[e_j(0) - e_k(0)] + C_j e_j(0).$$

Then γ_j is the sum of the initial charges in the condensers linking the jth node with all the other nodes of the network, the charge being reckoned as positive if the plate of the condenser connected to the jth node is positive.

Then, if the excitation transforms

$$J_j(p) = I_j(p) + \frac{i_j(0)}{p} + \gamma_j \qquad (j = 1, \cdots, n) \tag{39}$$

are introduced, the equations assume the general form

$$J_j(p) = \sum_{k=1}^{n} Y_{jk} E_k(p). \tag{40}$$

The method of arriving at Eq. (40) directly from the transform network is an immediate generalization of that given in Sec. 1·5, but a detailed discussion is deferred until Sec. 1·7 in order to avoid duplication.

The close similarity between Eqs. (40) and (32) permits carrying out the remaining steps of the solution with a minimum of comment. Equation (40) may be solved by Cramer's rule to give

$$E_j(p) = \sum_{k=1}^{n} \frac{J_k(p) \, \Delta_{jk}(p)}{\Delta(p)} = \sum_{k=1}^{n} J_k(p) \, \Omega_{jk}(p), \tag{41}$$

where $\Delta(p)$ is the principal determinant of the Y_{jk}'s, $\Delta_{jk}(p)$ is the cofactor of Y_{jk} in $\Delta(p)$, and

$$\Omega_{jk}(p) = \frac{\Delta_{jk}(p)}{\Delta(p)}.$$

If

$$\mathcal{L}^{-1}[J_k(p)] = j_k(t)$$

and

$$\omega_{jk}(t) = \mathcal{L}^{-1}[\Omega_{jk}(p)],$$

then the convolution theorem may be invoked to give the solution in the form

$$e_j(t) = \sum_{k=1}^{n} j_k(t) * \omega_{jk}(t).$$

The Mesh Case.—In Sec. 1·3 it was shown that the general equations in the mesh case may be put in the form

$$e_j(t) - u_j(0) = \sum_{k=1}^{n} z_{jk} i_k(t) \qquad (j = 1, \cdots, n), \tag{42}$$

where $u_j(0)$ is the initial voltage drop around the elastances in mesh j in a clockwise direction and

$$-z_{jk} = L_{jk}\frac{d}{dt} + R_{jk} + S_{jk}\int_0^t dt, \qquad z_{jj} = L_{jj}\frac{d}{dt} + R_{jj} + S_{jj}\int_0^t dt.$$

Here n of the initial conditions are included in Eq. (42), and n more are given by the initial values of the mesh currents:

$$\lim_{t\to 0} i_1(t) = i_1(0), \qquad \lim_{t\to 0} i_2(t) = i_2(0), \qquad \cdots,$$
$$\lim_{t\to 0} i_n(t) = i_n(0).$$

If $E_j(p)$, $I_k(p)$ are the $\mathcal{L}$-transforms of $e_j(t)$, $i_k(t)$ respectively, then taking the $\mathcal{L}$-transform of both sides of Eq. (42) gives

$$E_j(p) - \frac{u_j(0)}{p} + L_{jj}i_j(0) - \sum_{k\neq j=1}^{n} L_{jk}i_k(0) = \sum_{k=1}^{n} Z_{jk}(p)I_k(p).$$

Let

$$\phi_j(0) = L_{jj}i_j(0) - \sum_{k\neq j=1}^{n} L_{jk}i_k(0); \tag{43}$$

$\phi_j(0)$ is, therefore, the total initial flux through mesh j. Then, if

$$V_j(p) = E_j(p) - \frac{u_j(0)}{p} + \phi_j(0), \tag{44}$$

the equations assume the simple form

$$V_j(p) = \sum_{k=1}^{n} Z_{jk}(p)I_k(p). \tag{45}$$

From here the solution is carried out exactly as for the nodal case.

Unfortunately, $J_k(p)\,\Omega_{jk}(p)$ in Eq. (41) is not always the $\mathcal{L}$-transform of an E-function, although this is the only case with which the analysis developed in the previous section is competent to deal; it is, therefore, impossible to say that a solution always exists. For example, consider the case where pentodes with infinite internal impedances are connected by two-terminal interstage-coupling networks composed of a pure inductance, as in Fig. 1·18. Then if this system is driven by a current source $i(t)$ and if the initial currents are zero, the expression for the output transform voltage is

$$E(p) = I(p)p^n,$$

where n is the number of stages. This function has no $\mathcal{L}^{-1}$-transform.

All that can be said is that Eq. (41) is the general solution for the nodal network if the response transforms are $\mathcal{L}^{-1}$-transformable. In any
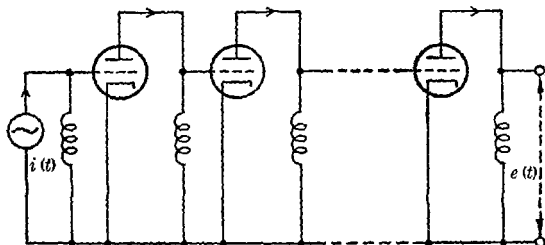
FIG. 1·18.—The n-stage inductance amplifier.

achievable physical network this is, of course, always the case; it is also the case for most of the idealized networks with which one wishes to deal.

1·7. The Transform Network.—The concept of the transform network was first introduced in Sec. 1·5. In this section it will be shown how to set up the general transform network both in the mesh and the nodal case and how to apply it to derive the transform equations. A discussion is given of the value of the concept in practical cases.

The General Nodal Case.—The general equations for the nodal case are given by Eq. (40). These equations can be derived without setting up the general integro-differential equations, as follows. Replace the various elements of the network by their appropriate p-admittances.[1] Place across the kth independent node pair a constant-current p-generator $J_k(p)$ of magnitude equal to the sum of three quantities:

1. $I_k(p)$, the $\mathcal{L}$-transform of $i_k(t)$, the driving function current generator in the kth independent node pair.

---

[1] See Sec. 1·2 for the definitions of these quantities. It is assumed that all transformers are replaced by their equivalent II- or T-networks.

2. ρ_k/p, where ρ_k is the total initial current flowing *into* the nth node from the inductors linking the nth node with the other nodes of the network.

3. γ_k, where γ_k is the sum of the initial charges on the condensers linking the kth node with all the other nodes of the network, the charge being reckoned as positive if the plate of the condenser connected to the kth node is positive.

Thus

$$J_k(p) = I_k(p) + \frac{\rho_k}{p} + \gamma_k.$$

The general nodal equations are now set up by applying Kirchhoff's second law to the independent node pairs of the transform network in turn, remembering that the current flowing into a p-admittance $Y(p)$ is equal to $Y(p)$ times the voltage across it.

The procedure has been illustrated for a two-node network in Sec. 1·5.

The General Mesh Case.—Equations (45), the transform equations, of the general mesh case can be obtained in a similar fashion as follows. Replace the various elements of the network by their appropriate p-impedances.[1] The method of taking account of mutual inductance will be discussed below. Place in the jth mesh a constant voltage generator $V_j(p)$ equal in magnitude to the sum of three quantities:

1. $E_j(p)$, the $\mathcal{L}$-transform of the driving function voltage generator in the jth mesh.
2. $-u_j(0)/p$, where $u_j(0)$ is the initial voltage drop round the capacitors in mesh j, in a clockwise direction.
3. $\phi_j(0)$, where $\phi_j(0)$ is the initial flux through the jth mesh, given by Eq. (46).

Thus

$$V_j(p) = E_j(p) - \frac{u_j(0)}{p} + \phi_j(0).$$

The general mesh equations are now set up by applying Kirchhoff's first law to each successive mesh of the transform network in turn, remembering that the voltage across a p-impedance $Z(p)$ is equal to $Z(p)$ times the current through the impedance.

Mutual inductance terms are allowed for, as in the conventional steady-state analysis, by assuming that the open-circuit voltage across coil L_j coupled by mutual inductances M_{jk} to a system of coils $L_k(k \neq j,$ $j = 1, \cdots, n)$ is simply

[1] See Sec. 1·2.

$$\sum_{\substack{k \neq j = 1}}^{n} \pm\, pM_{jk},$$

where the ambiguity of sign is resolved as in Sec. 1·2.

If this procedure is applied to the network of Fig. 1·9, then the transform network will be obtained in the form of Fig. 1·19,
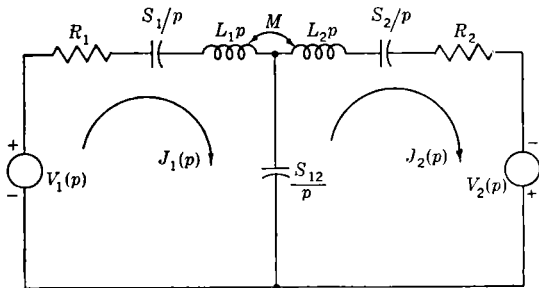
FIG. 1·19.—Two-mesh transform network.

where

$$\left.\begin{aligned}
V_1(p) &= E_1(p) - Q_1 S_1 - Q_{12} S_{12} + L_1 \rho_1 - M \rho_2, \\
V_2(p) &= +Q_{12} S_{12} - Q_2 S_2 + L_2 \rho_2 - M \rho_1.
\end{aligned}\right\} \tag{46}$$

Applying Kirchhoff's first law to the first and second meshes consecutively gives

$$\left.\begin{aligned}
V_1(p) &= I_1(p)\left(R_1 + \frac{S_1}{p} + L_1 p + \frac{S_{12}}{p}\right) - I_2(p)\left(Mp + \frac{S_{12}}{p}\right), \\
V_2(p) &= -I_1(p)\left(Mp + \frac{S_{12}}{p}\right) + I_2(p)\left(R_2 + \frac{S_2}{p} + \frac{S_{12}}{p} + L_2 p\right).
\end{aligned}\right\} \tag{47}$$

These equations may be compared with the original Eq. (2) and it is easily verified that Eqs. (47) are, indeed, the transforms of Eqs. (2).

Advantages of the Transform Network.—The value of the transform network is chiefly evident when the initial electrostatic and magnetic energy in the circuit is nonzero, this being the case in which use of the original Heaviside calculus was most cumbersome. It is also the case where the possibility of getting a sign wrong or of miswriting a term is most serious. The transform network not only is valuable as a reliable short cut to writing down the transform equations but also can be used to derive these equations in a simpler form.

In the first place it is now possible to use Thévenin's theorem in the general form given in Fig. 1·1, and this may decrease the number of equations required.

Second, consider the transform network of Fig. 1·20.   The transfer
voltage of this network can be written down by inspection to give

$$V_2(p) = \frac{J_1(p)}{Y_1(p) + Y_2(p) + Y_1(p)Y_2(p)Z_{12}(p)};$$

and if Y_1, Y_2, Z_{12} are complicated networks, this may represent a con-
siderable saving in labor.

Examples of this kind could be
multiplied indefinitely, but enough
has now been written to demon-
strate the statement that the
£-transform method, using the
concept of the transform network,
is as simple and compact as the
Heaviside calculus but is more
rigorous and also more powerful

Fig. 1·20.—Transform Π-network.

because of the ease with which it can be applied to the case of nonzero
initial conditions.
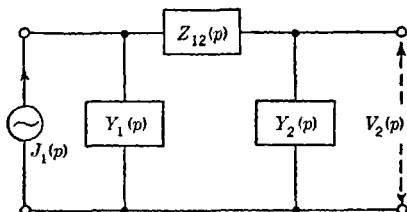
1·8. The Steady-state Response of the General Linear Network.—Up
to now attention has been focused solely on the response of a network to
a nonperiodic driving function.   There has been nothing in the analysis
making it inapplicable to the periodic case, but the necessity of knowing
the initial conditions prevents direct use of the £-transform theory.
There are two ways out of this difficulty, both of which will be discussed
in this section.   In the second and more valuable method a modified
£-transform, the so-called S-transform, is employed.   In the former, the
periodic driving function is expressed as a series of sinusoidal driving
functions.   Because the response to a sinusoidal driving function can be
written down from the transform network, the response function can be
represented in the form of a series that may converge rapidly enough to
justify the neglect of all but a few terms.

The first step in outlining the application of this method is to con-
sider the response of the network to a signal $Ie^{j\omega t}$ applied at time $t = 0$.

Now the £-transform of $Ie^{j\omega t}$ is

$$\frac{I}{p - j\omega};$$

thus, if the signal is applied across the jth node[1] (mesh), the transform
response across the kth node (mesh) is

[1] From the superposition principle of Sec. 1·3 there is no loss of generality involved
in confining attention to the case where the driving function is applied across a single
node (mesh).

$$E_k(p) = \frac{I}{p - j\omega} \Omega_{jk}(p) + \sum_{i=1}^{n} \Omega_{ik}(p)V_i(p),$$

where $\Omega_{jk}(p)$ are system functions of the type discussed in Sec. 1·6 and $V_i(p)$ are excitation functions due to the initial conditions and are all of the form $a_i + (b_i/p)$.

Let us assume that $\Omega_{ik}(p)V_i(p)$ are $O(1/p)$ and that all the poles of $\Omega_{ik}(p)V_i(p)$ lie to the left of the imaginary axis[1] in the complex p-plane. Then the residues of all these poles decay exponentially with time, and after a sufficiently long time their sum is vanishingly small. The residue of the pole at $p - j\omega = 0$, however, is $Ie^{j\omega t}\Omega_{jk}(j\omega)$, which is oscillatory for real values of ω.

It is clear from the above that after a sufficiently long time the response of the network differs by an arbitrarily small amount from

$$e(t) = Ie^{j\omega t}\Omega_{jk}(j\omega). \tag{48}$$

More loosely, Eq. (48) is the response across the kth node (mesh) to a signal $Ie^{j\omega t}$ applied at time $t = -\infty$ across the jth node (mesh). It is, of course, the familiar steady-state response and could have been obtained directly from the transform network by assuming p to have the special value $j\omega$.

The steady-state response to an arbitrary periodic waveform can be obtained from the above with the aid of the principle of superposition as follows:

If $f(t)$ is a periodic function of t with period T, that is,

$$f(t + T) = f(t)$$

for all t, then $f(t)$ may be expressed as a Fourier series of the form

where

$$\left. \begin{array}{l} f(t) = \displaystyle\sum_{-\infty}^{\infty} a_n e^{\frac{2\pi j n t}{T}}, \\[3ex] a_n = \dfrac{1}{T}\displaystyle\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)e^{-\frac{2\pi j n t}{T}}\, dt, \end{array} \right\} \tag{49}$$

provided that $f(t)$ satisfies suitable conditions[2] which are always satisfied for practical driving functions.

[1] Both these assumptions are essential if the network is to have a steady-state response at all, at least in the sense employed here.

[2] Sufficient conditions, general enough for the present purpose, were given by Dirichlet. These are that

1. $f(t + T) = f(t)$.

2. $f(t)$ defined and bounded in the range $-T/2$ to $T/2$.

3. $f(t)$ has only a finite number of maxima and minima in the range $-T/2$ to $T/2$.

4. $f(t)$ has only a finite number of discontinuities in the range $-T/2$ to $T/2$.

If $\omega = 2\pi/T$ then Eq. (49) may be written as

$$f(t) = \sum_{-\infty}^{\infty} a_n e^{jn\omega t},$$

where

$$a_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) e^{-jn\omega t} \, dt.$$

But the response of the system to a signal $a_n e^{jn\omega t}$ applied at time $t = -\infty$ is, by Eq. (48),

$$a_n e^{jn\omega t} \Omega_{jk}(jn\omega);$$

hence, by the principle of superposition, the output response to the periodic driving function $f_j(t)$ is

$$e_k(t) = \sum_{n=-\infty}^{\infty} e^{jn\omega t} \Omega_{jk}(jn\omega) \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f_j(t) e^{-jn\omega t} \, dt. \tag{50}$$

It may happen that only a few of the terms of this series are significant, but often the series must be summed in closed form if the expression for the response function is to be useful.

There is another method of tackling the steady-state analysis, based upon a generalized form of the £-transform theory. This will now be discussed and illustrated by a practical problem.

Let $f(t)$ be a periodic function with period T. The "steady-state Laplace transform" of $f(t)$ which, following Waidelich,[1] will be called the $\mathcal{S}$-transform and written symbolically

$$\mathcal{S}[f(t)] = F_s(p), \tag{51}$$

is defined by the equation

$$\mathcal{S}[f(t)] = F_s(p) = \int_0^T e^{-pt} f(t) \, dt. \tag{52}$$

The "inverse $\mathcal{S}$-transform," or "$\mathcal{S}^{-1}$-transform," is given by

$$f(t) = \mathcal{S}^{-1}[F_s(p)] = \frac{1}{2\pi j} \int_W \frac{e^{pt} F_s(p)}{1 - e^{-pT}} \, dp. \tag{53}$$

For the inverse transform, the path of integration W in the complex plane is the closed contour of Fig. 1·21 composed of W_1, W_2 and two

[1] D. L. Waidelich "The Steady-state Operational Calculus," *Proc. I.R.E.*, **34**, 38–83, February 1946. The treatment given in this section is based directly upon this paper, and reference should be made to it for more complete references and for an outline of the proof of the results quoted.

rectangular lines parallel to the real axis joining the ends of W_1 and W_2. In the limit in which these end points tend to infinity, the integral of $F_s(p)e^{pT}/(1 - e^{-pT})$ over those short rectilinear arcs is zero. The path W_3, to be used later, is the same as W_2 except that the direction is reversed. All the poles of $F_s(p)/(1 - e^{-pt})$ must lie to the left of W_3 except for the points $p = (jn2\pi)/T$, where n is an integer, which must lie between W_1 and W_3. If $F_s(p)$ has any poles on the imaginary axis or in the right half of the complex plane, W_3 must be indented to ensure that these poles lie to the left of W_3.
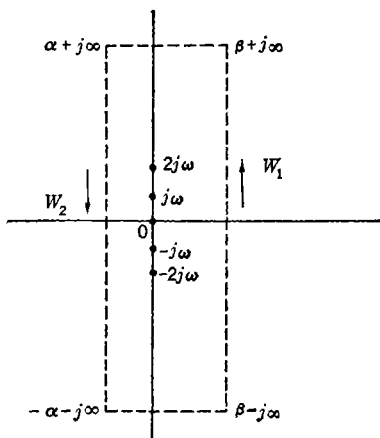
The method of applying the $\S$-transform theory to the steady-state solution of linear integro-differential equations follows essentially the same course as in the $\mathcal{L}$-transform theory. One takes the $\S$-transform of both sides of the equations, thus transforming them into a set of algebraic equations that can be solved for the $\S$-transform by the conventional process. The $\S^{-1}$-transform is then used to derive the final expression for the steady-state response.

Fig. 1·21.—The $\S$-transform contour.

To carry out this procedure three theorems similar to Theorems 6, 7, and 8 of Sec. 1·4 will be needed.

Theorem 6b. **Additivity.** If $F_s(p) = \S[f(t)]$ and $G_s(p) = \S[g(t)]$ then

$$\S[f(t) + g(t)] = F_s(p) + G_s(p). \tag{54}$$

Theorem 7b. $\S$-**transform of an Integral of a Function.** Let $f(t)$ be an E-function of period T and $\S$-transform $F_s(p)$; if

$$g(t) = g(0) + \int_0^t f(t)\, dt,$$

then

$$\S[g(t)] = \frac{1 - e^{-pT}}{p}\, g(0) + \frac{F_s(p)}{p}. \tag{55}$$

Theorem 8b. $\S$-**transform of the Derivative of a Function.** If $f'(t)$ is an E-function with period T and if $f(t)$ has $\S$-transform $F_s(p)$, then

$$\S[f'(t)] = pF_s(p) - (1 - e^{-pT})f(0). \tag{56}$$

The above theorems can be generalized by induction as were Theorems 6, 7, and 8.

At first sight it may be perplexing to see the occurrence of values of $f(t)$ at time zero.   In a steady-state analysis one would not expect such terms to be present.   The difficulty is more apparent than real, however, in that these terms do not, in general, contribute anything to the inverse $\mathcal{S}$-transform.

The above theory is best illustrated by a practical example, which is taken directly from Waidelich.[1]

Fig. 1·22.—Two-node network.

Let a square wave current $i(t)$, where

$$i(t) = A, \qquad 0 < t < \frac{T}{2},$$

$$i(t) = -A, \qquad \frac{T}{2} < t < T,$$
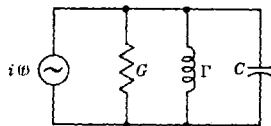
be applied to the network of Fig. 1·22.   Then the voltage $e(t)$ across the network is given implicitly by

$$i(t) = \frac{C\,de(t)}{dt} + Ge(t) + \Gamma \int_0^t e(t)\,dt - \rho,$$

where ρ is the current flowing in the coil at time $t = 0$.

Taking the $\mathcal{S}$-transform of both sides of the equation, and using the results of Theorems 6b, 7b, and 8b give

$$I_s(p) = \left(Cp + G + \frac{\Gamma}{p}\right) E_s(p) - \left[Ce(0) + \frac{\rho}{p}\right](1 - e^{-pT}),$$

where $e(0)$ is the voltage across the circuit at time $t = 0$ and $I_s(p)$, $E_s(p)$ are the $\mathcal{S}$-transforms of $i(t)$, $e(t)$ respectively.   Hence

$$E_s(p) = \frac{I_s(p) + \left[Ce(0) + \dfrac{\rho}{p}\right](1 - e^{-pT})}{Cp + G + \dfrac{\Gamma}{p}};$$

and

$$e(t) = \mathcal{S}^{-1}[E_s(p)] = \frac{1}{2\pi j}\int_W \frac{pI_s(p)e^{pt}}{(\Gamma + Gp + Cp^2)(1 - e^{-pT})}\,dp$$
$$+ \frac{1}{2\pi j}\int_W \frac{\rho + pCe(0)}{\Gamma + Gp + Cp^2}\,e^{pt}\,dp.$$

[1] *Ibid.*

The second integral has no poles inside the contour W; hence by Cauchy's theorem it is zero.   Accordingly the initial conditions vanish from the steady-state solution, and we get simply

$$e(t) = \frac{1}{2\pi j} \int_W \frac{pI_s(p)e^{pt}}{(\Gamma + Gp + Cp^2)(1 - e^{-pt})}\, dp.$$

But

$$I_s(p) = \int_0^{T/2} Ae^{-pt}\, dt - \int_{T/2}^T Ae^{-pt}\, dt$$

$$= \frac{A}{p}(1 - e^{-\frac{pT}{2}}) - \frac{A}{p}e^{-\frac{pT}{2}}(1 - e^{-\frac{pT}{2}}) = \frac{A}{p}(1 - e^{-\frac{pT}{2}})^2.$$

Hence

$$e(t) = \frac{1}{2\pi j} \int_W \frac{A(1 - e^{-\frac{pT}{2}})^2 e^{pt}}{(\Gamma + Gp + Cp^2)(1 - e^{-pT})}\, dp. \tag{57}$$

Equation (57) is an integral taken around a closed contour and is determined by the sum of the residues of the poles lying within.   If this direct method of evaluating the integral is chosen, however, the response function will be expressed simply as a Fourier series and the whole point of introducing the $\mathfrak{S}$-transform would vanish.

Alternatively one can proceed as follows.[1]

Since the integrals over the rectangular arcs joining the end points of W_1 and W_3 are zero in the limit when these end points tend to infinity,

$$e(t) = \frac{1}{2\pi j} \int_{W_1} \frac{E_s(p)e^{pt}}{1 - e^{-pT}}\, dp - \frac{1}{2\pi j} \int_{W_3} \frac{E_s(p)e^{pT}}{1 - e^{-pt}}\, dp.$$

Now

$$\frac{1}{2\pi j} \int_{W_1} \frac{E_s(p)e^{pt}}{1 - e^{-pT}}\, dp = \frac{1}{2\pi j} \int_{W_1} E_s(p)e^{pt}\left(1 + e^{-pT} + e^{-2pT} + \cdots +\right) dp$$

$$= \frac{1}{2\pi j} \int_{W_1} E_s(p)e^{pt}\, dp + \frac{1}{2\pi j} \int_{W_1} E_s(p)e^{p(t-T)}\, dp + \cdots +, \tag{58}$$

where W_1 is the same path of integration as that of the $\mathcal{L}^{-1}$-transform; the integrals of Eq. (58), therefore, may be evaluated in the same way as those of Sec. 1·4.   In particular we have

$$\int_{W_1} E_s(p)e^{pt}\, dp = 0$$

for $t < 0$.   Hence for $t < T$,

$$e(t) = \frac{1}{2\pi j} \int_{W_1} E_s(p)e^{pt}\, dp - \frac{1}{2\pi j} \int_{W_3} \frac{E_s(p)e^{pt}}{1 - e^{-pT}}\, dp.$$

[1] This method of solution can be followed in all cases and is not peculiar to the problem at hand.

In the particular case with which we are dealing,

$$E_s(p) = \frac{A(1 - e^{-\frac{pT}{2}})^2}{\Gamma + Gp + Cp^2},$$

and, for $t < T/2$,

$$\frac{1}{2\pi j} \int_{W_1} E_s(p) e^{pt} \, dp = \frac{1}{2\pi j} \int_{W_1} \frac{A e^{pt}}{\Gamma + Gp + Cp^2} \, dp;$$

thus, for $0 < t < T/2$,

$$e(t) = \frac{1}{2\pi j} \int_{W_1} \frac{A e^{pt} \, dp}{\Gamma + Gp + Cp^2} - \frac{1}{2\pi j} \int_{W_3} \frac{A e^{pt}(1 - e^{-\frac{pT}{2}})^2}{(\Gamma + Gp + Cp^2)(1 - e^{-pT})} \, dp. \tag{59}$$

Since

$$\frac{A}{\Gamma + Gp + Cp^2} \qquad \text{and} \qquad \frac{A(1 - e^{-\frac{pT}{2}})^2}{[\Gamma + G(p) + Cp^2](1 - e^{-pT})}$$

are both $O(1/p)$, the two integrals of Eq. (59) can both be transformed into contour integrals consisting of the infinite left-hand circle and the straight lines W_1 and W_3, respectively. The only poles of these functions that lie inside these contours occur at the zeros of $\Gamma + Gp + Gp^2$; evaluating the residues at these poles gives

$$e(t) = \frac{4am}{n} e^{-nt} \left[\frac{\sin nt + e^{-\frac{mT}{2}} \sin n\left(t - \frac{T}{2}\right)}{1 + 2e^{-\frac{mT}{2}} \cos\left(\frac{nT}{2}\right) + e^{-mT}} \right]$$

for $t < T/2$, where

$$m = \frac{G}{2C}, \qquad n = \sqrt{\frac{\Gamma}{C} - m^2},$$

if $G^2 < 4\Gamma C$;

$$e(t) = 2A \left(\frac{a + b}{a - b}\right)\left(\frac{e^{bt}}{1 + e^{-\frac{bT}{2}}} - \frac{e^{-at}}{1 + e^{-\frac{aT}{2}}}\right)$$

for $t < T/2$ and where $a = m + jn$, $B = m - jn$, if $G^2 > 4\Gamma C$.

1·9. The Fourier Transform Method.—Before the Fourier transform or $\mathfrak{F}$-transform can be compared with the $\mathcal{L}$-transform as a tool in solving network problems, it is necessary to give a brief outline of the essentials of the theory. In the introduction (Sec. 1·1) it was stated that there are two alternative methods of deriving the transform theory; one based

upon the steady-state analysis and the superposition principle and the other, which was followed in the body of this chapter, formal and abstract. The former method will be utilized here, partly to illustrate the principle and partly because the simplicity of the discussion provides one of the chief arguments for the $\mathfrak{F}$-transform.

In Sec. 1·8, where the steady-state analysis was discussed, it was shown that a periodic function with period T could be expressed as a Fourier series:

$$f(t) = \sum_{-\infty}^{\infty} a_n e^{2\pi jnt/T},$$

where

$$a_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) e^{-\frac{2\pi jnt}{T}} \, dt,$$

provided that $f(t)$ satisfies certain conditions.

Now an aperiodic function can be regarded as the limiting case of a periodic function whose period T is allowed to extend to infinity. Formally, then, the extension to nonperiodic functions is made by allowing T to tend to infinity, in such a way that

$$\omega = \frac{2\pi n}{T} \qquad \text{and} \qquad \lim_{T \to \infty} \left(\frac{2\pi}{T} \right) = d\omega.$$

If the sum is replaced by an integral, one gets in the limit

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(j\omega) e^{j\omega t} \, d\omega,$$

where

$$F(j\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} \, dt. \tag{60}$$

$F(j\omega)$ is the simple double-sided[1] Fourier transform of $f(t)$, and is in fact usually just referred to as the Fourier transform of $f(t)$.

The basic Fourier integral theorem of Eq. (60) is not, of course, proved by an argument of the above kind, which, as presented here, has only dubious validity. But the foregoing discussion is valuable in that

---

[1] Double-sided because the limits of integration of the defining integral of $F(j\omega)$ go from $-\infty$ to ∞. In the single-sided definition used in the $\mathcal{L}$-transform case, the integration limits go from 0 to ∞. If $f(t) = 0$ for $t < 0$, it is, of course, immaterial whether the single- or double-sided transform be used.

it traces the direct connection with the steady-state analysis, which will now be used to give the transient response of the general linear network.[1]

In Sec. 1·8 it was shown that if the network was dissipative, the output response after a sufficiently long time to a driving function $E(j\omega)e^{j\omega t}$ is of the form

$$E(j\omega)e^{j\omega t}\Omega_{jk}(j\omega),$$

where $\Omega_{jk}(j\omega)$ is the appropriate system function.

Let it now be assumed that a driving function $f(t)$ be applied to the network, where $f(t)$ is defined for all real values of t, positive and negative, and is such that $\int_{-\infty}^{\infty} |f(t)|\, dt$ converges. Now by the Fourier integral theorem $f(t)$ may be regarded as a continuous sum of sinusoidal generators, each of which is applied to the network at time $t = -\infty$. To find the response of the network to this system the superposition principle is invoked, in a more powerful form than we have proved it valid in Sec. 1·3, to state that the network response may be regarded as the sum (or integral in this case) of the response to the individual component generators. Now for $t > 0$ the response of the network to a sinusoidal generator applied at time $t = -\infty$ is simply the steady-state response. Accordingly, if

$$E(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\, dt, \tag{61}$$

the transient response of the network may be written

$$i(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} E(j\omega)\,\Omega_{jk}(j\omega)e^{j\omega t}\, d\omega; \tag{62}$$

this integral can be evaluated by the theory of residues as given in Sec. 1·4, provided that the integral of the integrand of Eq. (62) taken around either the upper or lower infinite half circle is zero.

The most attractive point about this solution is its extreme simplicity. The integro-differential equations of the network have not been introduced at all, and there has been no need to discuss the condition of the system at time $t = 0$. Nevertheless a considerable price has been paid. The function $f(t)$ must now be defined for all real values of t, not only for positive ones. This might not appear a serious hardship, as it is usual to take $f(t) = 0$ for $t < 0$ in the $\mathcal{L}$-transform theory, but unfortunately it means that the initial conditions are fixed by the form of $f(t)$ and can no longer be chosen arbitrarily. Another disadvantage is that the restric-

[1] See E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, Oxford, New York, 1937, for a full theoretical treatment of this theory. It is shown there that $\int_{-\infty}^{\infty} |f(t)|\, dt$ must be convergent if the relationship is to hold.

tions on the form of $f(t)$ are much more severe, at least in the simple theory, because it is now necessary that $\int_{-\infty}^{\infty} |f(t)|\,dt$ converge, thus excluding a whole range of useful functions from discussion.

The second disadvantage can, however, be overcome at the cost of increased complexity. Thus the range of driving functions can be considerably extended if instead of a function

$$
\begin{aligned}
e(t) &= 0, & t &< 0, \\
e(t) &= f(t), & t &> 0,
\end{aligned}
$$

one considered the function

$$
\begin{aligned}
e(t) &= 0, & t &< 0, \\
e(t) &= f(t)e^{-ct}, & t &> 0.
\end{aligned}
$$

Then, even if $f(t) = O(t^n)$, $\int_{-\infty}^{\infty} |f(t)e^{-ct}|\,dt$ converges for all $c > 0$. The response to the modified function can be found; and then after the integrations have been performed, c can be allowed to tend to zero.

A more powerful attack, which will, in fact, do everything that can be done by the £-transform theory, can be based upon the "generalized Fourier transform" defined by

$$
F(j\omega) = \int_0^\infty f(t)e^{-(c+j\omega)t}\,dt, \tag{63a}
$$

where

$$
f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(j\omega)e^{(c+j\omega)t}\,d\omega, \qquad t > 0, \tag{63b}
$$

provided that $\int_0^\infty |f(t)|e^{-ct}\,dt$ converges.

This transformation[1] can be applied to the integro-differential equations exactly as was the £-transform, and the solution follows along virtually identical lines. Initial conditions can now, of course, be chosen arbitrarily. The generalized Fourier transform method has, however, no advantage over the Laplace transform method.

The above discussion brings out the fact, emphasized in Sec. 1·1, that although the simple Fourier transform can be used to handle a restricted class of network transient problems, the £-transform provides a compact analysis of greater power and scope.

This last statement must not, however, be interpreted as implying any essential mathematical difference between the two, for, on the contrary, they are closely interconnected. Thus formally, at least, the £-transform

---

[1] See *ibid.*, where this method is actually applied to solve a set of integro-differential equations.

can be obtained from the generalized $\mathfrak{F}$-transform by a shift of origin and a rotation of axes through $\pi/2$.

It is necessary to make this point because in some treatments the Fourier and the Laplace transforms are compared as if they were distinct mathematical tools. This attitude is inadmissible. As the previous discussion was intended to show, the difference between the two transforms lies in their scope and in their historical background, not in their mathematical nature.

1·10. Summary of the Use of £-transform Theory in Network Problems.—In order to place the theory upon a sound footing and at the same time display the fundamental properties of the Laplace transform in a clear light, care was taken to develop the mathematical analysis of Sec. 1·4 on a rigorous basis.

Many readers, however, will be interested in the Laplace transform only as a tool in the solution of practical network problems and be willing to take the validity of the theory for granted. Accordingly, this section has been written as a guide to the location of the more important results, which are scattered throughout the chapter, so that the reader may use the Laplace transform method with the minimum of theory.

The problem is that of finding the output response $v_k(t)$ of a linear network to a driving function $i_j(t)$ applied at time $t = 0$, when the initial charges on the condenser and the initial currents through the inductors are given.

The first step is to find the Laplace transform

$$I_j(p) = \mathcal{L}[i_j(t)] = \int_0^\infty e^{-pt} i_j(t)\ dt \tag{64}$$

of $i_j(t)$. This may be done either by direct integration as in Eq. (64) or by reference to Table 1·5 if $i_j(t)$ is one of the standard forms there tabulated.

The next step is to set up the transform network. In the case when all the initial conditions are zero, this can be done simply by replacing all the inductors, capacitors, and resistors by their p-impedances or p-admittances[1] and the driving function $i_j(t)$ by its Laplace transform $I_j(p)$. When the initial conditions are not all zero, the driving transform functions have to be suitably modified, as is discussed for the general mesh (nodal) case in Sec. 1·7. Applications to particular networks are in Sec. 1·5 above and in Sec. 1·11 below. The transform output response is found by the conventional steady-state analysis and will be of the form of Eq. (41).

It now remains only to evaluate $v_k(t)$. As shown in Sec. 1·4, $v_k(t)$ is given in terms of $V_k(p)$ by the explicit relation

[1] See Sec. 1·2 for a definition of these quantities.

$$v_k(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} V_k(p)e^{pt}\,dp, \tag{65}$$

where c is a real positive number such that all the poles of $V(p)$ lie to the left of c in the complex p-plane.

Integrals of the form of Eq. (65) were evaluated in Sec. 1·4 with the aid of the theory of residues, and a general expression was obtained for $v_k(t)$ when $V_k(p)$ was the product of an algebraic and a regular function. All the response transform functions considered in this chapter are of this type, and the principal results are restated here for convenience.

Two alternative cases exist:

1. $pV_k(p)e^{pt}$ is bounded for all $R(p) > c$. In this case $v(t) = 0$.
2. $pV_k(p)$ is bounded for all sufficiently large p in the half plane $R(p) < c$. In this case $v(t) = \Sigma$ residues of the poles of $V_k(p)e^{pt}$.

If $V_k(p)$ is the product of an algebraic and a regular function, it can be expressed in the form

$$V_k(p) = \prod_{s=1}^{n} \frac{H(p)}{(p - p_s)^{r_s}},$$

where $H(p)$ is bounded and regular for all finite p.

The poles of $V_k(p)$ occur at $p = p_s(1 = 1, \cdots, n)$, and the residue of $V_k(p)e^{pt}$ at $p = p_s$ is

$$\frac{1}{(r_s - 1)!} \frac{d^{r_s-1}}{dp^{r_s-1}} [V_{ks}(p)e^{pt}]_{p=p_s}, \tag{66}$$

where

$$V_{ks}(p) = V_k(p)(p - p_s)^{r_s}; \tag{67}$$

thus

$$v_k(t) = \sum_{s=1}^{n} \frac{1}{(r_s - 1)!} \frac{d^{r_s-1}}{dp^{r_s-1}} [V_{ks}(p)e^{pt}]_{p=p_s}. \tag{68}$$

The above procedure has been discussed in greater detail but along essentially the same lines in the main body of the text. In Sec. 1·11 it is applied to two practical examples.

1·11. Examples of Use of $\mathcal{L}$-transform Theory to Solve Practical Network Problems. *Response of an n-stage RC-coupled Pulse Amplifier to a Unit-step Function.*—The first problem to be considered is the voltage response developed across the anode of the last tube of the amplifier chain of Fig. 1·23, when the input voltage $e(t)$ is a step function of unit height. It will be assumed that all the stages are identical. Let $\imath_r(t)$ be the plate current of the rth tube, with $\mathcal{L}$-transform $I_r(p)$, and let $e_r(t)$ be

the voltage developed across the anode lead of the rth tube, with £-transform $E_r(p)$. Then the transform network assumes the form of Fig. 1·24. If the tubes are pentodes, the output plate conductance can be neglected in comparison with $1/R$, and

$$I_{r+1}(p) = g_m E_r(p) \qquad (r = 0, \cdots, n-1);$$
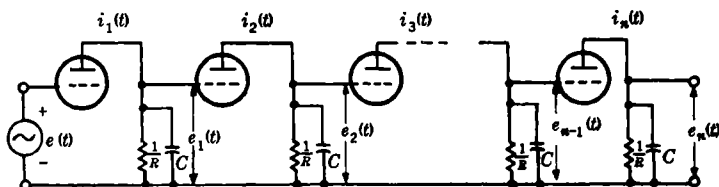
FIG. 1·23.—The n-stage RC-coupled pulse amplifier.
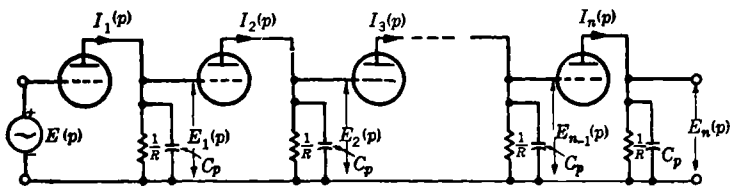
FIG. 1·24.—The n-stage RC-coupled pulse amplifier transform.

in addition one has the n equations

$$E_r(p) = \frac{I_r(p)}{\dfrac{1}{R} + Cp} \quad \cdots \quad (r = 1 \cdots n).$$

By a simple process of elimination one gets

$$E_n(p) = \frac{(g_m R)^n E(p)}{(1 + CRp)^n}.$$

Now, since

$$\begin{aligned} e(t) &= 0, \qquad t < 0, \\ &= 1, \qquad t > 0, \end{aligned}$$

one can write

$$E(p) = \int_0^\infty e^{-pt}\, dp = \frac{1}{p};$$

thus $e_n(t)$, the response function that equals $£^{-1}[E_n(p)]$, is given by

$$e_n(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{(g_m R)^n e^{pt}}{p(1 + CRp)^n}\, dp.$$

Normalizing by putting $t = t/RC$ and $e_n(t) = e_n(t)/(g_m R)^n$, one gets

$$e_n(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{e^{pt}}{p(1 + p)^n}\, dp.$$

Now $pE_n(p)$ is bounded for all n and all sufficiently large p; hence, by the results of Sec. 1·4,

$$e_n(t) = \sum \text{ residues of poles of } \frac{e^{pt}}{p(1 + p)^n}.$$

But $e^{pt}/p(1 + p)^n$ has a pole of order unity at $p = 0$ with residue unity and a pole of order n at $p + 1 = 0$ with residue

$$\frac{1}{(n - 1)!} \frac{d^{n-1}}{dp^{n-1}} \left(\frac{e^{pt}}{p} \right)_{p=-1}$$

Accordingly,

$$\begin{aligned}
e_n(t) &= 1 + \frac{1}{(n - 1)!} \frac{d^{n-1}}{dp^{n-1}} \left(\frac{e^{pt}}{p} \right)_{p=-1} \\
&= 1 + \sum_{r=1}^{n-1} \frac{1}{(n - 1)!} \binom{n - 1}{r} \left[\frac{d^r}{dp^r} \frac{1}{p} \frac{d^{n-r-1}}{dp^{n-r-1}} e^{pt} \right]_{p=-1} \\
&= 1 + \sum_{r=1}^{n-1} \frac{1}{(n - 1)!} \binom{n - 1}{r} \left[\frac{(-1)^r r!}{p^{r+1}} t^{n-r-1} e^{pt} \right]_{p=-1} \\
&= 1 - e^{-t} \sum_{r=1}^{n-1} \frac{t^{n-r-1}}{(n - r - 1)!}.
\end{aligned}$$

$e_n(t)$ is plotted as a function of t in Fig. 1·25 for values of n from 1 to 10.
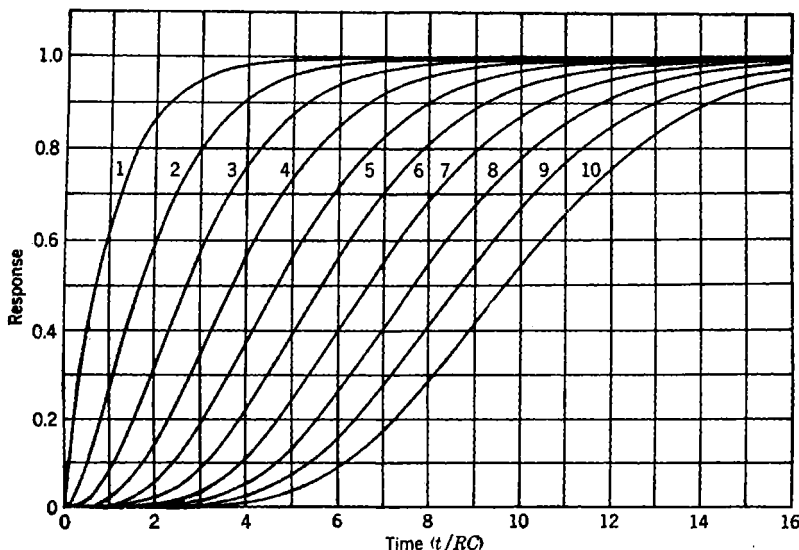
FIG. 1·25.—Response to a step function current of the n-stage RC-coupled pulse amplifier of Fig. 1·23.

Response of Four-terminal Network to a Unit-step Function.—To conclude this chapter, the ℒ-transform theory will be applied to find the output response $e_n(t)$ of the four-terminal network of Fig. 1·26 to a unit-step constant-current driving function $i(t)$. This circuit[1] has a very good rate
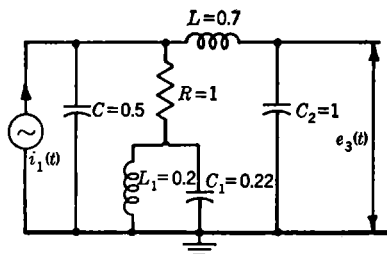
Fig. 1·26.—Four-terminal linear-phase coupling network.

of rise and extremely small overshoot. To evaluate the response numerically, the following normalized values for the circuit constants are taken:

$$C = 0.5, \quad C_1 = 0.22, \quad C_2 = 1,$$
$$L_1 = 0.2, \quad L_2 = 0.7, \quad R = 1.$$

$I_1(p)$, the transform-drive function, is $1/p$. The transform network is as shown in Fig. 1·27, and will be analyzed on the nodal basis. The circuit
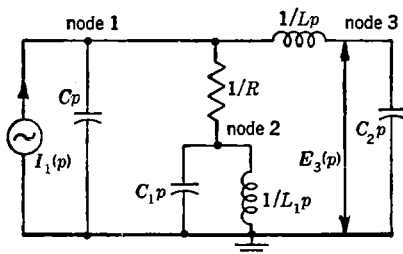
Fig. 1·27.—Transform of the network of Fig. 1·26.

has three independent node pairs as marked in Fig. 1·27, and, from the general transform Eq. (40) in the nodal case,

$$I_1(p) = Y_{11}E_1(p) + Y_{12}E_2(p) + Y_{13}E_3(p),$$
$$0 = Y_{21}E_1(p) + Y_{22}E_2(p) + Y_{23}E_3(p),$$
$$0 = Y_{31}E_1(p) + Y_{32}E_2(p) + Y_{33}E_3(p),$$

where the Y_{ij} are given by Eqs. (38a). If $\Delta(p)$ is the determinant of the array $Y_{ij}(p)$ and $\Delta_{13}(p)$ the cofactor of Y_{13} in $\Delta(p)$, then the equation can be solved to give

[1] R. L. Dietzold, Bell Telephone Laboratories, personal communication.

$$E_3(p) = \frac{I_1(p) \, \Delta_{13}(p)}{\Delta(p)}.$$

In the present case this may be expanded to give

$$E_3(p) = \frac{1}{p} \frac{\begin{vmatrix} -\dfrac{1}{R} & C_1 p + \dfrac{1}{R} + \dfrac{1}{L_1 p} \\[2mm] -\dfrac{1}{Lp} & 0 \end{vmatrix}}{\begin{vmatrix} Cp + \dfrac{1}{R} + \dfrac{1}{Lp} & -\dfrac{1}{R} & -\dfrac{1}{Lp} \\[2mm] -\dfrac{1}{R} & C_1 p + \dfrac{1}{R} + \dfrac{1}{L_1 p} & 0 \\[2mm] -\dfrac{1}{Lp} & 0 & C_2 p + \dfrac{1}{Lp} \end{vmatrix}}$$

$$= \frac{L}{p} \, (RL_1 C_1 p^2 + L_1 p + R) \{ RL^2 L_1 CC_1 C_2 p^5 + L^2 L_1 C_2 (C_1 + C_2) p^4$$
$$+ RL[LCC_2 + L_1 C_1 C_2 + L_1 CC_1] p^3 + L[LC_2 + L_1 (C + C_1 + C_2)] p^2$$
$$+ RL(C + C_2) p + L \}^{-1}.$$

Substituting the numerical values of the circuit elements,

$$E_3(p) =$$
$$\frac{2.8571 p^2 + 12.987 p + 64.9351}{p(p^5 + 6.5454 p^4 + 27.0130 p^3 + 67.7922 p^2 + 97.4026 p + 64.9351)}.$$

Now $pE_3(p)$ is bounded for all sufficiently large p, so that

$$e_3(t) = \sum \text{ residues of poles of } E_3(p) e^{pt}.$$

To find the poles p_r of $E_3(p)$, besides the simple pole at $p = 0$, it is necessary to solve for the roots of the 5th-degree polynomial in the denominator. By Newton's method of approximation, one of the roots, p_1 for example, is found to be -1.892.   Dividing the polynomial through by $p + 1.892$, we obtain the 4th-degree polynomial

$$p^4 + 4.6534 p^3 + 18.2088 p^2 + 33.3412 p + 34.3210,$$

the roots of which can be found by the standard formulas for quartic equations.   Then the poles of $E_3(p)$ lie at the points

$$p = 0,$$
$$p_1 = -1.892,$$

$$p_2 = -1.420 + 1.277j, \qquad p_4 = -0.907 + 2.931j,$$
$$p_3 = -1.420 - 1.277j, \qquad p_5 = -0.907 - 2.931j.$$

All these poles are of order unity, so that the residues

$$b_r = e^{pt}E_3(p_r)(p - p_r)$$

can be found by substitution.   This procedure is simplified by noting that since p_2, p_3 and p_4, p_5 are pairs of complex conjugates, b_2, b_3 and b_4, b_5 must
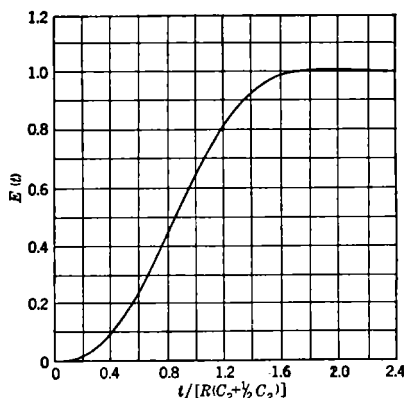
Fig. 1·28.—Response to step-function current of four-terminal linear-phase network.

also be pairs of complex conjugates.   The following values are obtained for the b's:

$$b_0 = 1,$$
$$b_1 = -1.509e^{-1.892t},$$
$$b_2 = +(0.1716 + 0.9698j)e^{-(1.420+1.277j)t},$$
$$b_3 = (0.1716 - 0.9698j)e^{-(1.420+1.277j)t},$$
$$b_4 = (0.083 - 0.044j)e^{-(0.907+2.931j)t},$$
$$b_5 = (0.083 + 0.044j)e^{-(0.907+2.931j)t};$$

thus, since $e_3(t) = \displaystyle\sum_{r=0}^{5} b_r,$

$$e_3(t) = 1 - 1.509e^{-1.892t} + e^{-0.907t}(0.1666 \cos 2.931t + 0.0885 \sin 2.931t)$$
$$+ e^{-1.420t}(0.3432 \cos 1.277t - 1.9396 \sin 1.277t).$$

This function is plotted against t in Fig. 1·28.

TABLE 1·5.—LIST OF LAPLACE TRANSFORM PAIRS

No.	$f(t)$	$F(p)$
1	1	$\dfrac{1}{p}$
2	$\delta(t)$	$\lim\limits_{\alpha \to \infty} \dfrac{\alpha}{p + \alpha}$
3	e^{at}	$\dfrac{1}{p - a}$
4	$\sin \beta t$	$\dfrac{\beta}{p^2 + \beta^2}$
5	$\cos \beta t$	$\dfrac{p}{p^2 + \beta^2}$
6	$\sinh \beta t$	$\dfrac{\beta}{p^2 - \beta^2}$
7	$\cosh \beta t$	$\dfrac{p}{p^2 - \beta^2}$
8	t	$\dfrac{1}{p^2}$
9	t^n	$\dfrac{n!}{p^{n+1}}$ (n a positive integer)
10	$t^n e^{-at}$	$\dfrac{n!}{(p + a)^{n+1}}$ (n a positive integer)
11	$f(t) = 0,\ t < 1$ $\quad = 1,\ t > 1$	$\dfrac{e^{-p}}{p}$
12	$f(t) = 1,\ t < 1$ $\quad = 0,\ t > 1$	$\dfrac{1 - e^{-p}}{p}$
13	$f(t) = t,\ t < 1$ $\quad = 0,\ t > 1$	$\dfrac{1 - (1 + p)e^{-p}}{p^2}$
14	$f(t) = 0,\ t < 1$ $\quad = t,\ t > 1$	$\dfrac{e^{-p}}{p} + \dfrac{e^{-p}}{p^2}$
15	$f(t) = t,\ t < 1$ $\quad = 1,\ t > 1$	$\dfrac{1 - e^{-p}}{p^2}$
16	$f(t) = t,\ t < 1$ $\quad = 2 - t,\ 1 < t < 2$ $\quad = 0,\ t > 2$	$\dfrac{(1 - e^{-p})^2}{p^2}$
17	$f(t) = 1 - t,\ t < 1$ $\quad = 0,\ t > 1$	$\dfrac{1}{p} - \dfrac{(1 - e^{-p})}{p^2}$

CHAPTER 2

HIGH-FIDELITY PULSE AMPLIFIERS

By Robert M. Walker and Henry Wallman

2·1. Introduction.[1]—Pulse amplifiers are employed in various branches of physical investigation, in radar and television receivers, and in certain new types of communication equipment (pulse-time modulation, facsimile, etc.).

The subject of the next two chapters is the amplification of direct pulses, the amplification of pulses of a carrier frequency being considered

(a)

(b)
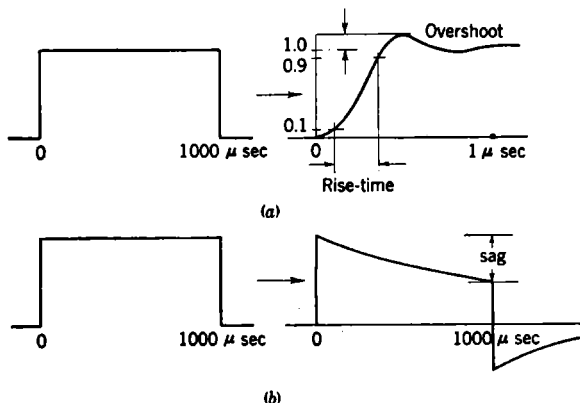
Fig. 2·1.—Reproduction of a rectangular pulse. (a) Reproduction of the leading edge of a rectangular pulse. Note the rise time and the overshoot. If the amplifier is linear and the pulse duration is large compared with the rise time, the reproduction of the trailing edge is the negative of the leading edge. (b) Reproduction of the flat top of a rectangular pulse. Note the sag. If the amplifier is linear, the amplitude of the undershoot following the pulse is equal to the sag.

in Chaps. 4 through 7. This chapter is concerned with pulse amplifiers of high fidelity; Chap. 3 describes pulse amplifiers of lower fidelity but much greater dynamic range.

Emphasis on the Time Domain.—The emphasis in this chapter is mainly on the time response of amplifiers, that is, upon the shape of the output waveform as a function of time for appropriate pulse input. This approach is chosen because it is really only the time response that is of any

[1] Sections 2·1, 2·2, and 2·3 are by Henry Wallman; the remainder of the chapter is by Robert M. Walker.

interest to the user of pulse amplifiers; the familiar amplitude- and phase frequency curves are merely means to the end of good pulse response, and should be regarded as matters of subordinate interest.

Reproduction of Rectangular Pulses.—There are many types of pulses— rectangular, delta-function, sawtooth, rounded, etc.—in common use, and in a linear amplifier the response to any one of them completely determines the response to any other.   In this discussion the rectangular pulse is used as the standard input signal.   A rectangular pulse is the sum of a positive step function and a delayed negative step function; hence its reproduction by a linear amplifier is the sum of the response to a positive step function and a delayed negative step function.

The reproduction of a step function can be divided into two distinct parts, namely, the reproduction of the leading and trailing edges and the reproduction of the flat top (Fig. 2·1).   These two aspects will be considered in order.

2·2. Leading Edge of Pulse; Rise Time and Overshoot.—The most important characteristics of the reproduction of the leading edge of a

FIG. 2·2.—Basic pentode amplifier.

rectangular pulse or step function are the rise time, usually measured from 10 to 90 per cent,[1] and the "overshoot" (see Fig. 2·1a).

A third characteristic is sometimes also of importance, namely the time duration over which the amplitude of the overshoot oscillations is appreciable.   The problem is to minimize these three parameters: rise time, overshoot, and overshoot-oscillation duration.

RC-coupling.—The basic pulse-amplifier stage is shown in Fig. 2·2. A tetrode or pentode is usually used if the gain is at all high, since triodes would have very high input capacity because of the Miller effect.

When the reproduction of the leading edge is of interest, the circuit may be simplified as shown in Fig. 2·3.

The capacity C is the total interstage shunt capacity and is made up of the sum of the output capacity of the first tube, the input capacity of the following tube, and stray wiring capacities.

[1] Other definitions of rise time are occasionally used, such as the intercept of the tangent drawn through the 50 per cent point of the step-function response.   For some applications it is desirable to measure rise time between the 5 and 95 per cent points or the 1 and 99 per cent points.

A voltage step function applied to the grid of the tube results in a current step function applied to the parallel RC-combination. It is the resulting voltage developed across the RC-combination that is the desired amplifier response. If the load resistor R is small compared with the plate resistance of the tube (this is always the case for high-speed pentode stages), the stage gain (= voltage amplification) is

$$\text{Gain} = g_m R, \qquad (1)$$

FIG. 2·3.—Simplified circuit of pulse-amplifier stage employing RC-coupling.

where g_m is the transconductance of the tube. The step-function response is an exponential curve, free from overshoot, as shown in Curve 1 of Fig. 1·25. Its rise time is given by

$$\text{Rise time} = 2.2RC. \qquad (2)$$

Dividing Eq. (1) by Eq. (2) yields

$$\frac{\text{Gain}}{\text{Rise time}} = \frac{g_m}{2.2C}. \qquad (3)$$

The right side of Eq. (3) is a figure of merit for the amplifier tube. The gain/rise time ratio has a value of about $200/\mu\text{sec}$ for type 6AK5 tubes if it is assumed that $g_m = 5000 \ \mu\text{mhos}$ and $C = 11.5 \ \mu\mu\text{f}$. Therefore with an RC-coupling between two type 6AK5 tubes, a gain of 10 with a rise time of $\frac{1}{20} \ \mu\text{sec}$ or a gain of 2 with a rise time of $\frac{1}{100} \ \mu\text{sec.}$, etc., can be obtained.

For laboratory purposes or whenever reduced tube life can be accepted, the figure of merit for type 6AK5 tubes can be increased to about $280/\mu\text{sec}$ by reducing the bias and thereby increasing g_m.

FIG. 2·4.—Shunt-peaked circuit.

Shunt Peaking.—The question arises as to what can be done to improve the gain/rise time ratio by use of circuits other than the RC-circuit. The next simplest circuit is the so-called shunt-peaked circuit shown in Fig. 2·4, for which the significant parameter is the ratio $m = L/R^2C$.

A family of step-function responses is shown in Fig. 2·5. There is no overshoot for $m \leq 0.25$. The performance of this circuit for various values of m is shown in Table 2·1. For $m = 0.41$, for example, Fig. 2·5 shows that a gain/rise time ratio between type 6AK5 tubes of $1.7 \times 200 = 340/\mu\text{sec}$ can be attained, accompanied, however, by an overshoot of 2.5 per cent.

TABLE 2·1.—PERFORMANCE OF SHUNT-PEAKED CIRCUIT

m	Relative speed (referred to RC-circuit)	Overshoot, %
0	1.0	0
0.250	1.4	0
0.414	1.7	3.1
0.500	1.9	6.7
0.600	2.1	11.4

Figure 2·5 demonstrates that the rise time can be reduced at the expense of increased overshoot. The proper compromise between speed and overshoot depends on various external factors. It is common to regard overshoots of 40 or 50 per cent as acceptable in servoamplifiers, whereas in television amplifiers perhaps 5 to 10 per cent is all that should be tolerated, and in certain measuring apparatus only about 1 per cent.
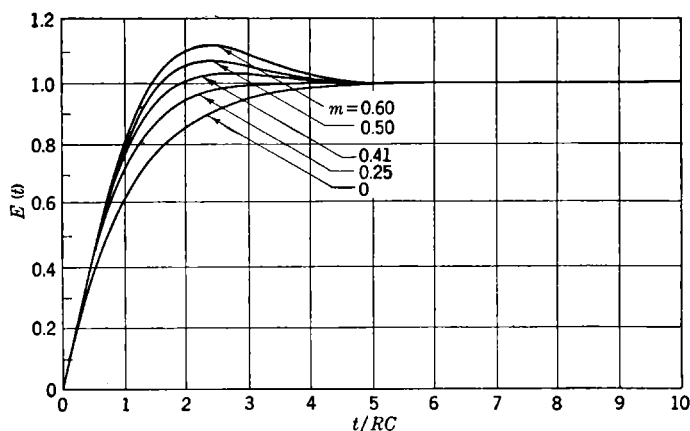
FIG. 2·5.—Response of shunt-peaked circuit to a step function of current.

Within the restriction of two-terminal coupling networks little improvement can be attained over shunt-peaking. However, there is one other two-terminal network (Fig. 2·6) that is worthy of mention. The steady-state performance of this network[1] gives it its name. Its step-function response (Fig. 2·7) has only 1 per cent overshoot.

Although the increase in performance of the circuit of Fig. 2·6 over that of the $m = 0.41$ shunt-peaked circuit is not very great, the extra complication is not great either; it is usually possible with a little ingenuity in layout to realize the capacity across the peaking coil as a stray capacity, so that no additional parts are required.

[1] S. Doba, Bell Telephone Laboratories, private communication.

Four-terminal Coupling Networks.—To get substantial improvement over the performance figures given in Table 2·1 it is necessary to go to

four-terminal networks, that is, to make use of the partition of the interstage capacity between an input and an output capacity.   A good example of such a circuit is the four-terminal linear-phase circuit shown in Fig. 1·26.   The step-function response of this circuit, shown in Fig. 1·28, exhibits an advantage in gain/rise time ratio of 2.48 as compared
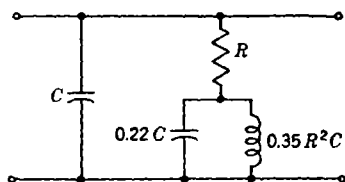
Fig. 2·6.—Two-terminal linear-phase network.
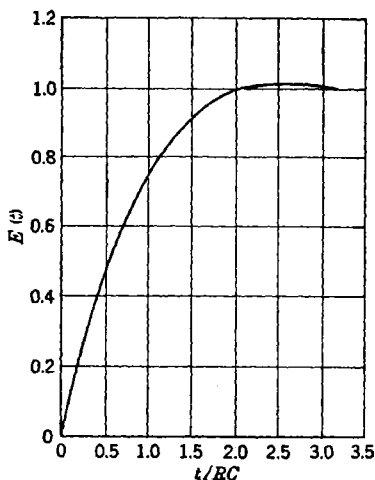
Fig. 2·7.—Response of two-terminal linear-phase network to a step function of current.   Gain/rise time advantage over the RC-circuit is 1.77; overshoot is 1 per cent.

pared with the RC circuit, and this ratio is accompanied by an overshoot of only 1 per cent.

The great speed and small overshoot of this circuit make it very attractive.   However, as with all four-terminal networks, a certain definite capacity ratio is assumed, in this case $\frac{1}{2}$.   By the reciprocity theorem it is possible to reverse the network, thereby accommodating a 2/1 capacity ratio; but for capacity ratios differing from either $\frac{1}{2}$ or 2, other circuit configurations must be employed.   No such consideration is involved in the use of two-terminal networks, and the need to employ a different configuration for a different capacity ratio may

Fig. 2·8.—Four-terminal network designed for 1/1 capacity ratio.

be regarded as the price of the increased performance of four-terminal networks.
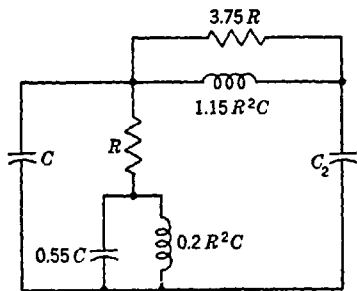
A four-terminal network designed for 1/1 capacity ratio[1] is shown in

[1] E. A. Schramm, Bell Telephone Laboratories, private communication from R. L. Dietzold.   The curves of Fig. 2·9 are due to A. J. Grossman of the Bell Telephone Laboratories.

Fig. 2·8, and its step-function response is shown in Curve *a* of Fig. 2·9. Curve *a* of Fig. 2·9 shows an advantage of 2.1 in gain/rise time ratio as
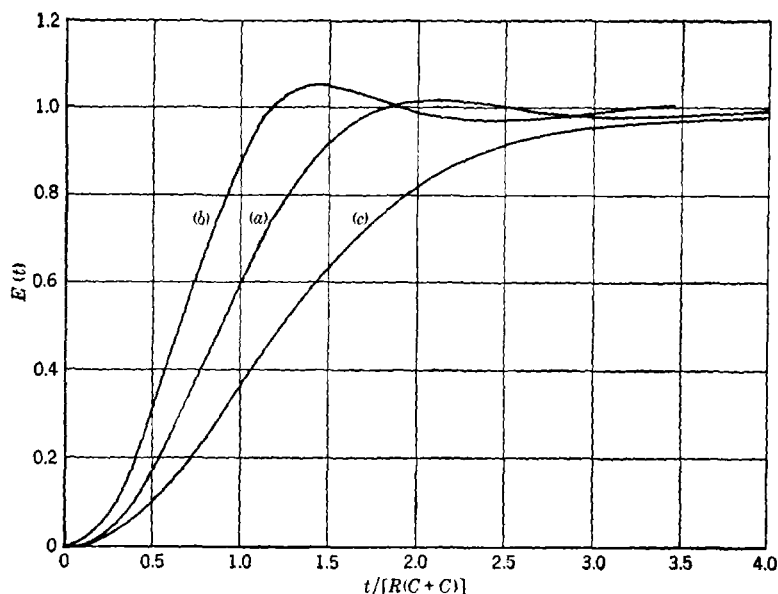
Fig. 2·9.—Response to current step function of network shown in Fig. 2·8. (*a*) $C_2 = C$; (*b*) $C_2 = C/2$; (*c*) $C_2 = 2C$.

compared with the *RC*-circuit and has 2 per cent overshoot. Curves (*b*) and (*c*) of Fig. 2·9 show the effect of designing the circuit for 1/1 capacity ratio when the actual capacity ratio is 2/1 or 1/2, respectively.
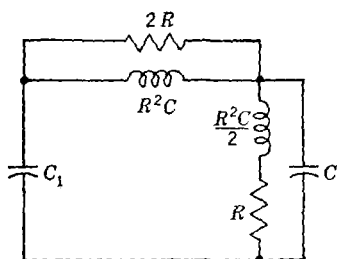
Fig. 2·10.—Series shunt-peaked circuit.

The somewhat simpler circuit shown in Fig. 2·10, called the series-shunt peaked circuit, was widely used at the Radiation Laboratory. Its step-function response is shown in Fig. 2·11*a*. Curve *a* of Fig. 2·11 shows an advantage of 2.06 in gain/rise time ratio compared with the *RC*-circuit and has 3 per cent overshoot. Curve *b* of Fig. 2·11 shows the effect of a left-hand capacity that is only half the right-hand capacity. For many applications the step-function responses of Curves *b* of Figs. 2·9 and 2·11 are entirely satisfactory. Although the circuits of Figs. 2·8 and 2·10 give substantially poorer performance than the circuit

shown in Fig. 1·26, the latter circuit is less tolerant with regard to capacity ratio.

Composition of Rise Times.—It is of interest to know how the rise times of the individual stages combine in multistage amplifiers.   A very
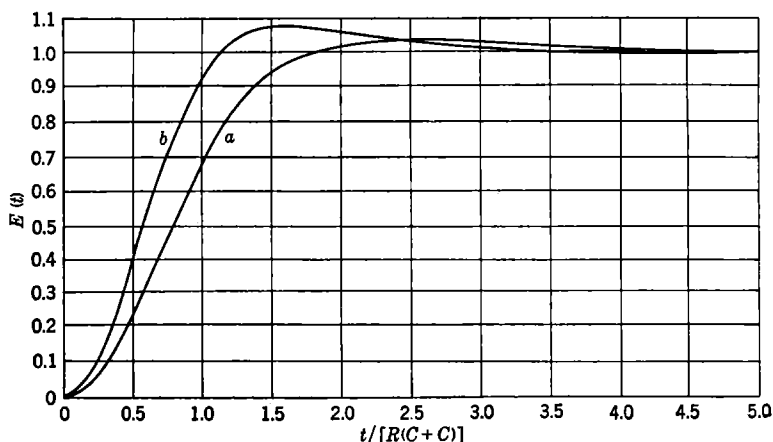
Fig. 2·11.—Response to current step function of series shunt-peaked circuit shown in Fig. 2·10.   (a) $C_1 = C$; (b) $C_1 = C/2$.

good answer can be given to this question when the individual stages are free from overshoot.   The result is as follows:

RULE 1.[1]   For an amplifier made up of n stages, each of which is free from overshoot, rise times add in the root-square, that is,

$$\tau = \sqrt{\tau_1^2 + \tau_2^2 + \cdots + \tau_n^2}, \tag{4}$$

where τ is the over-all rise time and $\tau_1, \tau_2, \ldots, \tau_n$ are the rise times of the individual stages.

For the special case in which $\tau_1 = \tau_2 = \cdots = \tau_n$, Eq. (4) reduces to

$$\tau = \tau_1 \sqrt{n}. \tag{5}$$

Equation (5) shows, for example, that if a nine-stage amplifier made up of identical stages free from overshoot is to have an over-all rise time of $\frac{1}{10}$ μsec, each stage must have a rise time of $\frac{1}{30}$ μsec.

[1] Rule 1 represents not the result of observation of many special cases but has a solid mathematical basis; it is in fact a translation into the language of transient response of the "central limit theorem" of probability, stated by Laplace in 1812.

Rules 2, 3, and 4 given below are essentially the results of observation of special cases.

All the rules given here, Rules 1 through 6, are safely applied only to minimum phase-shift networks.

Equation (4) is actually a statement of a trend in the limit, as n increases indefinitely, and for any finite value of n is only approximate. However, the approximation is very good, usually within 10 per cent, even for values of n as small as 2. This fact is illustrated in the following listing, taken from Fig. 1·25, of rise times of identical cascaded RC-coupled stages (Table 2·2).

TABLE 2·2.—RISE TIMES OF CASCADED RC-COUPLED STAGES

n	1	2	3	4	5	6	7	8	9	10
Relative rise time (in units of 2.2 RC)..........................	1.0	1.5	1.9	2.2	2.5	2.8	3.0	3.3	3.45	3.6

For stages having nonzero overshoot, there is no clean-cut rule corresponding to Rule 1. However the following rough statements can be made:

RULE 2. For stages having very small overshoot (1 or 2 per cent) the overshoot grows extremely slowly or not at all as the number of stages increases, and Eq. (4) still holds.

RULE 3. For stages having overshoots of about 5 to 10 per cent, the overshoot increases approximately as the square root of the number of stages, and the rise time increases substantially less rapidly than as the root-square.

Rule 3 can be illustrated by two tables. The first, Table 2·3, taken from Fig. 7·5, describes cascaded transitionally coupled double-tuned stages.

TABLE 2·3.—RISE TIME OF CASCADED TRANSITIONALLY COUPLED DOUBLE-TUNED STAGES

No. of stages	Overshoot, %	Relative rise time
1	4.30	1.00
2	6.25	1.32
4	8.40	1.69
6	10.00	1.95

The second table, derived from Figs. 2·13 and 2·14, describes cascaded two-terminal networks of maximum gain-bandwidth product (Table 2·4). These two tables illustrate the general principle that amplifier speed, especially in multistage amplifiers, can be substantially increased if overshoot requirements are relaxed; not only can the rise time per stage be reduced, but the over-all rise time is not very much larger than the rise time of one stage.

Table 2·4.—Rise Time of Cascaded Two-terminal Circuits of Maximum Gain-bandwidth Product

No. of stages	Overshoot, %	Relative rise time
1	9.4	1.00
3	15.2	1.45

A large number of examples of step-function responses of multistage amplifiers is given in two very valuable published papers.[1]

Rise Time Required in Measuring Apparatus.—When pulses are known to have a certain rise time, Eq. (4) can be used to determine how fast the response of the measuring apparatus should be in order to cause negligible slowing of the pulses being measured. Equation (4) shows that if the response of the measuring apparatus is twice as fast as the input pulse, the output pulse rise time is increased by only 11 per cent; if the measuring apparatus is three times as fast as the input pulse, the output pulse rise time is lengthened by only 5 per cent. Thus it can be said that for most purposes an amplifier two to five times faster than the pulses being measured can be regarded as infinitely fast.

Equivalent Rise Time of a Cathode-ray-tube Spot.—A cathode-ray tube does not have infinitely sharp focus; because of the coarseness of its spot, an intensity-modulated cathode-ray tube must be regarded as an example of a pulse-measuring instrument of nonnegligible rise time. The equivalent rise time is easily computed. If the spot is assumed to have radial symmetry with an approximately Gaussian-error-curve intensity distribution, the spot may be regarded as the reproduction of a delta-function impulse by an equivalent Gaussian-error-curve filter. The rise time of the step-function response turns out to be about equal to the time required for the electron beam to move a distance equal to the spot diameter at the writing speed in use (the spot diameter is measured between the 50 per cent points on its intensity-distribution curve). The equivalent rise time is thus given by

$$\frac{\text{Spot diameter}}{\text{Sweep length}} \times \text{time per sweep length.}$$

In a radar presentation employing a 10-mile sweep on a plan-position indicator (PPI), the time per sweep length is about 100 μsec and the ratio of spot diameter to sweep length is about 1/200. (The constants pertaining to a 5-in. tube employing a P7 persistent screen might be as follows: spot diameter = 0.3 mm; sweep length = 60 mm, both con-

stants being proportionately larger for a 12-in tube.)   Hence the equivalent rise time of the cathode-ray tube is 0.5 μsec.   The preceding paragraph indicates that there is little point in this case in providing the pulse amplifier with an over-all rise time less than about 0.2 μsec.

Relations between Steady-state and Transient Responses.—So far in this chapter nothing has been said about the amplitude-frequency or phase-frequency characteristics of pulse-amplifier stages, on the grounds that they are of no direct interest to the user of pulse amplifiers.   For two practical reasons, however, it is now desirable to discuss these matters.

The reasons are those of inadequacy of instrumentation, namely, the inadequacy at the present time of equipment for (1) experimentally measuring the pulse response of one stage of a really fast multistage amplifier and (2) mathematically computing the over-all pulse response of a multistage amplifier made up of complicated individual stages.   It can, however, be expected that both these deficiencies will be overcome in the next few years.

It is, of course, true that the pulse response of a linear amplifier can be exactly computed from its amplitude-vs.-frequency and phase-vs.-frequency curves, but the following qualitative rules are often useful:

RULE 4.   If τ is the rise time, 10 to 90 per cent, of the step-function response of a low-pass amplifier without excessive overshoot and having a 3-db bandwidth ℬ, then[1]

$$\tau\mathcal{B} = 0.35 \text{ to } 0.45.$$

RULE 5.   The following three characteristics of an amplifier go together:[2]

1. Small overshoot (not more than about 1 or 2 per cent).
2. Amplitude-vs.-frequency curve approximately gaussian.
3. Phase linear over the pass band.

An illustration of Rule 5 is given in Fig. 2·12, which shows curves of absolute value vs. frequency and phase error vs. frequency (i.e., phase deviation from linearity) for a number of circuits whose step-function responses display small overshoot.   The curves have been normalized to have their 3-db points at $f = 1$.   Especially noteworthy is the closeness of fit, down to -10 db, among the Gaussian error curve and the amplitude-vs-frequency curves for the two- and four-terminal linear-phase networks.   For contrast the absolute-value and phase-error vs. fre-

---

[1] For overshoots of less than 5 per cent the value 0.35 is the one to take.

A consequence of Rules 1 and 4 is that for circuits leading to zero or very small overshoot, the bandwidth inevitably decreases as the square root of the number of stages.

[2] Rule 5, like Rule 1, is related to the central limit theorem of probability.

quency curves are given for a circuit having 4.3 per cent overshoot (see Table 2·3).

Maximum Gain/Rise Time Ratio.—It would be very valuable to have theorems on the best possible gain/rise time ratio corresponding to Bode's
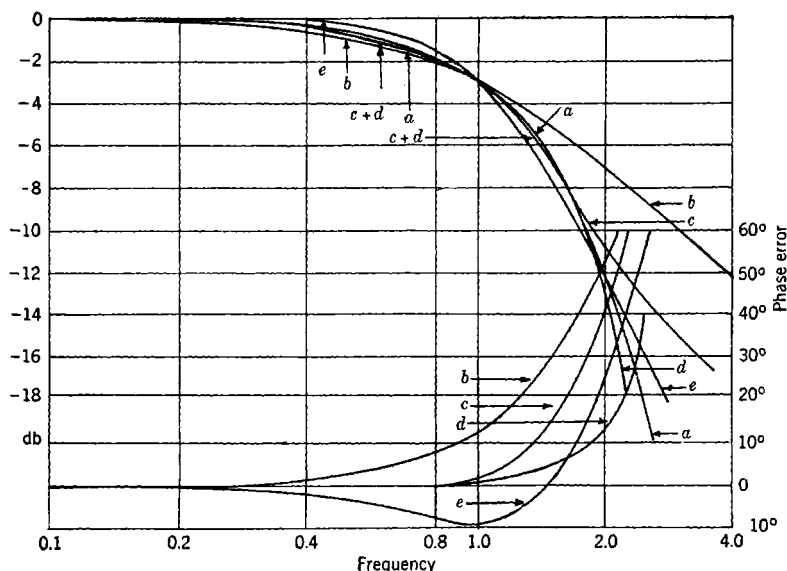
Fig. 2·12.—Normalized absolute value and phase error curves.   (a) Gaussian error curve, overshoot = 0 per cent (phase assumed linear); (b) RC-circuit, overshoot = 0 per cent; (c) two-terminal linear-phase network, overshoot = 1 per cent; (d) four-terminal linear-phase network, overshoot = 1 per cent; (e) low-pass equivalent of transitionally coupled double-tuned circuit, overshoot = 4.3 per cent.

theorems[1] on the best possible gain-bandwidth product.   It is clear that the problem is very difficult, even to formulate, particularly because the attainable gain/rise time ratio depends on the permissible overshoot. What is needed is a graph showing the largest possible gain/rise time ratio as a function of the fractional overshoot.   An especially significant

---

[1] H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945, Chap. 17.   The results are as follows: Consider low-pass interstage networks having input and output capacities $C/2$.   Then if a simple RC-circuit has its 3-db point at a frequency ω_0, interstage networks exist having flat gain out to ω_0 and with advantages in voltage gain over the RC-circuit of 2 and $\pi^2/2 = 4.93$ for the two-terminal and four-terminal cases respectively.

By using nonidentical stages, which are not individually flat in gain but which have an over-all gain curve that is flat, W. W. Hansen ["On Maximum Gain-band-width Product in Amplifiers," *Jour. Applied Physics*, **16**, 528–534 (1945)] has been able to obtain a mean stage gain times over-all bandwidth improvement factor for the four-terminal case of 5.06 as compared with the simple RC-circuit.

point on such a graph would, of course, be that corresponding to zero overshoot.

Figures 2·13 and 2·14 are of interest in this connection. The impedance of the two-terminal network of maximum gain-bandwidth product[1] is of the form $1/(p + \sqrt{1 + p^2})$, and the impulse response has the
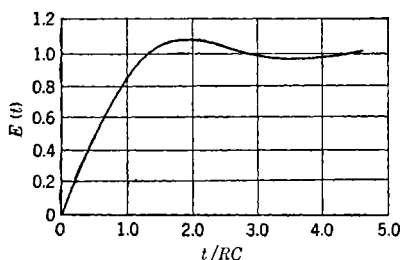
FIG. 2·13.—Response to step-function current of two-terminal flat-gain network of maximum gain-bandwidth product. The overshoot is 9.4 per cent, and the gain/rise time advantage over a parallel RC-circuit is 2.2.

equation[2] $J_1(t)/t$; the graph of Fig. 2·13 is derived by integrating $J_1(t)/t$ and modifying the time scale by a factor of 2 to take account of the advantage in gain-bandwidth factor. The impedance of the $n = 3$ approximation[3] to the four-terminal network of maximum gain-bandwidth product, which has itself a gain-bandwidth factor of 4.84, is of the form $1/(p + \sqrt{1 + p^2})^3$. The impulse response has the equation $3J_3(t)/t$, from which Fig. 2·14 is obtained.

The fractional overshoots shown in Figs. 2·13 and 2·14 are surprisingly small when one considers the sharpness of cutoff of the pass bands causing them; the overshoot oscillations decay extremely slowly, however.
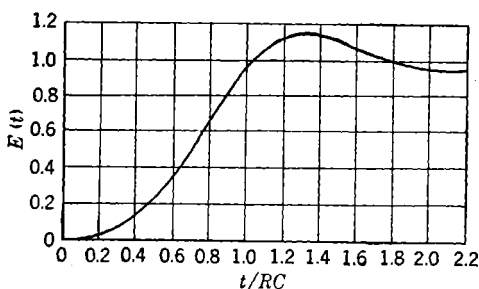
FIG. 2·14.—Response to step-function current of the $n = 3$ approximation to the four-terminal flat-gain network of maximum gain-bandwidth product. The overshoot is 15.2 per cent, and the gain/rise time advantage over a parallel RC-circuit is 3.75.

From Figs. 2·13 and 2·14 it can be conjectured that the possible improvement factor in gain/rise time ratio for two-terminal networks with not more than 9.4 per cent overshoot is about 2.2 and the corresponding figure for four-terminal networks (15.2 per cent overshoot) is about

[1] H. W. Bode, *op. cit.*, Eq. (17·13) and Fig. 17·4.

[2] G. A. Campbell and R. M. Foster, "Fourier Integrals for Practical Applications," *Bell System Tech. Monograph* **B584**, (1931), Formula 576.3.

[3] H. W. Bode, *op. cit.*, Eq. (17·31) and pp. 440 *et seq.*

3.75. These estimates are, for various reasons, subject to considerable question.

The advantage of four-terminal over two-terminal networks is considerably less in gain/rise time ratio than in gain-bandwidth product.

Future Trends.—It is clear that it is not possible to secure order-of-magnitude improvements in gain/rise time ratio beyond present practice by the use of conventional amplifier circuits. Various other possibilities exist, however.

One possible method employs vacuum tubes as negative-capacity elements. This method is illustrated in Fig. 2·15, where positive feedback through the small capacity C yields a negative capacitive input impedance for the first tube. Feedback chains employing circuits like this one may be of value; but because the

Fig. 2·15.—Negative-capacity circuit.

negative capacity tends to zero at higher frequencies, the scheme may be better adapted to achieving moderately fast rise times at a high impedance level than extremely fast rise times at a low impedance level.

The development of secondary-emission and beam-deflection tubes may yield substantial increases in transconductance.
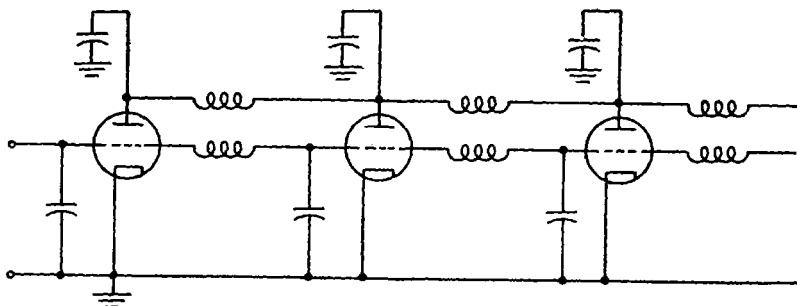
Fig. 2·16.—"Transmission-line" amplifier.

Schemes exist that make it possible, in principle, to achieve arbitrarily large gain/rise time ratios, even with present tubes, provided that the number of tubes is not limited. According to one proposal[1] the band to be amplified is partitioned into n subbands, each of which is separately amplified by a conventional amplifier, the resultant voltages being added at the output terminals. This method has the defect, resulting from its additive character, that the gains and relative phases of the various channels must be kept in accurate adjustment.

[1] C. W. Earp, British Patent 448113, accepted June 2, 1936.

A more ingenious proposal[1] is symbolically illustrated in Fig. 2·16. Transmission lines of the same propagation time are connected between the grids and the plates of a number of amplifier tubes. A voltage e_g, applied to the first grid, is transmitted along the grid line to the second grid, and at the same time produces a plate current $g_m e_g$ that is transmitted along the plate line. Because of the equal propagation rates along the grid and plate lines, the plate current $g_m e_g$ of the first tube arrives at the second plate in time to add to the plate current $g_m e_g$ caused by the signal voltage on the second grid. This process continues along the whole line and has the effect of producing an output current n times that of one tube, without a corresponding increase in shunt capacity. Thus the g_m/C ratio is multiplied by n.

In practice there may arise problems of line termination,[2] but the proposal is extremely promising.

In his patent, Percival suggests building the whole apparatus in a single evacuated envelope, in the form of a long cathode with one transmission line for the grid and another for the plate.

The crux of the matter is that because of the large area of the elements the tube can have an extremely high transconductance, while the equal propagation times along the grid and plate transmission lines eliminate the disadvantages of the large interelectrode capacitances resulting from the use of tube elements that are physically large.

2·3. Flat Top of Pulse.—This section is devoted to pulse distortion of the type shown in Fig. 2·1b, which illustrates the nonfaithful transmission of direct current by the amplifier.

Figure 2·2 shows that there are three RC-combinations that hinder the transmission of direct current. These are the series $C_g R_g$ grid circuit, the parallel $R_k C_k$ cathode circuit, and the parallel $R_s C_s$ screen circuit. The first of these circuits completely prevents transmission of direct current, and the other two reduce its transmission.

An exact analysis of the behavior of the amplifier shown in Fig. 2·2 would require examination of the interactions of these three circuits, but because of its complexity this procedure will not be followed, and the effect of each of the three circuits will be considered as if the others did not exist.

Grid Circuit.—If the plate-load resistor is small compared with the grid resistor (this assumption is made throughout this section), the series $C_g R_g$ circuit may be regarded as a voltage divider across which a step-

---

[1] W. S. Percival, British Patent 460562, accepted Jan. 25, 1937.

[2] In both the transmission line amplifier and the recently announced traveling wave tube there is a close integration of the electron stream with the load network; the vacuum tube and its circuit are, as it were, one.

function voltage is applied.   The voltage division ratio is

$$\frac{pR_gC_g}{1 + pR_gC_g},$$

where $p = j\omega$.   The step-function response, shown as Curve 1 in Fig.
2·17, is the exponential e^{-t/R_gC_g} whose time constant is R_gC_g.
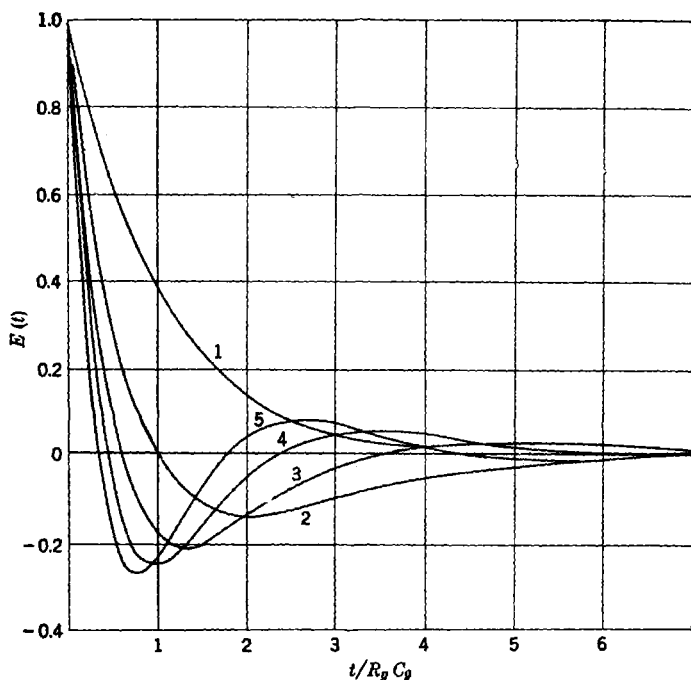
FIG. 2·17.—Step-function response of n capacitance-coupled stages each of time constant R_gC_g.

Observe that the tangent at $t = 0$ intersects the base line at a time
equal to one time constant.   Hence for a time that is a small fraction of a
time constant, the fractional sag is equal to the fraction of the time
constant; that is, the sag is 5 per cent at 5 per cent of a time constant.

Cascaded Grid Circuits.—The step-function response of n cascaded
identical grid circuits is easily calculated.   If the RC-product is normal-
ized to the value 1, the over-all voltage division ratio is

$$\left(\frac{p}{1 + p}\right)^n.$$

Therefore, the step-function response $f(t)$ is the Laplace transform of

$$\frac{1}{p}\left(\frac{p}{1+p}\right)^n = p^{n-1}\frac{1}{(1+p)^n};$$

hence $f(t)$ is the $(n-1)$st derivative of the Laplace transform of

$$\frac{1}{(1+p)^n};$$

that is[1]

$$f(t) = \frac{d^{n-1}}{dt^{n-1}}\frac{t^{n-1}e^{-t}}{(n-1)!}. \tag{6}$$

A graph of $f(t)$ is shown in Fig. 2·17 for $n = 1, \cdots, 5$.

The important features to observe in Fig. 2·17 and in Eq. (6) are that (1) the slopes at $t = 0$ increase directly with n, and (2) there is one additional crossing of the baseline for each additional capacitance-coupled stage.

For cascaded grid circuits with nonidentical time constants the graph of step-function response is very similar, although the mathematical expression involves only pure exponentials. For example, the step-function response of three capacitance-coupled stages having RC time constants $1/\alpha$, $1/\beta$, $1/\gamma$ respectively is

$$f(t) = \frac{\alpha^2(\gamma-\beta)e^{-\alpha t} + \beta^2(\alpha-\gamma)e^{-\beta t} + \gamma^2(\beta-\alpha)e^{-\gamma t}}{(\alpha-\beta)(\beta-\gamma)(\gamma-\alpha)},$$

The slope of $f(t)$ at $t = 0$ is $-(\alpha + \beta + \gamma)$. The following general statement is very useful:

Rule 6. For an amplifier made up of stages having imperfect transmission of direct current, the initial downward slope of the step-function response is the sum of the initial slopes of the component stages; that is, the slopes add arithmetically.

Rule 6 means that for amplifiers displaying small fractional sag in step-function response, the over-all sag can be computed simply by adding the sags from all individual causes. For example, if a certain amplifier has a step-function response displaying a sag in 100 μsec of 1 per cent for each of three coupling circuits, 2 per cent for a cathode circuit, and 2 per cent for a screen circuit, the over-all sag in 100 μsec is 7 per cent.

Correlation between Step-function Response and Low-frequency Cutoff.— Since the 3-db point of a single capacitance-coupled grid circuit is $f_0 = 1/(2\pi RC)$, there is a clear connection between the time constant T of exponential decay and the 3-db low-frequency point, namely, $T = 1/(2\pi f_0)$.

---

[1] G. A. Campbell and R. M. Foster, *op. cit.*, Formula 431. The coefficients of e^{-t} in the expansion of $f(t)$ are called Laguerre polynomials.

But for more than one stage there is no useful correlation between step-function response and low-frequency cutoff.

The reason is that the 3-db point increases as the square root of the number of stages whereas (Rule 6) the step-function slopes increase linearly with the number of stages. Unless the number of stages in an amplifier is stated, therefore, it does no good in determining sag to specify, as in common practice, the over-all low-frequency 3-db point.

Cathode Circuit.—An impedance Z_k connected between cathode and ground of an amplifier stage (see Fig. 2·2) causes inverse feedback; as is well known, the output current of the tube is

$$g_m e_g \frac{1}{1 + g_m Z_k}. \qquad (7)$$

FIG. 2·18.—(α) Plate current of tube having $R_k C_k$ cathode circuit when step-function voltage is applied to grid.

For the case in which Z_k is a parallel $R_k C_k$ combination, the fraction in expression (7) takes the form

$$\frac{1 + p R_k C_k}{1 + g_m R_k + p R_k C_k} = \frac{p + \alpha}{p + \alpha K}, \qquad (8)$$

where $p = j\omega$, $\alpha = 1/R_k C_k$, $K = 1 + g_m R_k$. The step-function response corresponding to Eq. (8) is

$$f(t) = \frac{K - 1}{K} \left(e^{-\alpha K t} + \frac{1}{K - 1} \right). \qquad (9)$$

A graph of Eq. (9) is shown in Fig. 2·18.

The slope of Eq. (9) at $t = 0$ is $-\alpha(K - 1) = -g_m/C_k$. Hence if there is a 1 per cent sag allowed in a time t_1, it is necessary that

$$C_k = 100 g_m t_1.$$

For a type 6AK5 tube, with $g_m = 5000$ μmhos,

$$C_k = \frac{t_1}{2}.$$

If $t_1 = 500$ μsec a 250-μf condenser would thus be required. Capacity values as large as this are usually impractical; in pulse amplifiers required to display small sag in step-function response it is therefore usual to leave the cathode resistor unbypassed. In that case the tube plate current, when a step-function voltage $-e_g$ is applied to the grid, is a step function of amplitude

$$g_m e_g \frac{1}{1 + g_m R_k}.$$

The gain-reduction factor $1/(1 + g_m R_k)$ is the price that must be paid for faithful flat-top reproduction.

Sometimes, however, the cathode resistor is bypassed by a small condenser chosen for its effect in reducing rise time (see Sec. 2·4).

Screen Circuit.—An impedance Z_s in the screen lead causes a drop in screen voltage and hence in plate current. The effect may be analyzed as follows, assuming $Z_k = 0$ and a constant ratio k between screen and plate current (k is usually about $\frac{1}{5}$ or $\frac{1}{4}$): Let i_p be the actual plate current. Then ki_p is the screen current; this produces a screen voltage drop ki_pZ_s, which, in turn, leads to a plate-current reduction $ki_pZ_sg_{m,s}$, where

$$g_{m,s} = \frac{\partial i_p}{\partial e_s}$$

is the screen-to-plate transconductance. Hence

$$i_p = g_m e_g - ki_p Z_s g_{m,s},$$

or

$$i_p = \frac{g_m e_g}{1 + Z_s k g_{m,s}}.$$

But $kg_{m,s} = 1/r_s$, where $r_s = \partial e_s/\partial i_s$ is the dynamic screen resistance ($\approx$ plate resistance of the tube connected as triode with screen strapped to plate). Hence

$$i_p = g_m e_g \frac{1}{1 + \dfrac{Z_s}{r_s}}. \tag{10}$$

If Z_s is a parallel $R_s C_s$ combination, the fraction in Eq. (10) may be written as

$$\frac{p + \beta S}{p + \beta(S + 1)},$$

where $\beta = 1/r_s C_s$ and $S = r_s/R_s$. The step-function response has the equation

$$f(t) = \frac{1}{S + 1} (e^{-\beta(S+1)t} + S). \tag{11}$$

A graph of Eq. (11) has the same general form as Fig. 2·18.

The slope of Eq. (11) at $t = 0$ is $-\beta = -1/r_s C_s$. Hence for a 1 per cent sag allowed in time t_1 it is necessary to have

$$C_s = \frac{100t_1}{r_s}.$$

For a type 6AC7 tube, with r_s approximately equal to 20,000 ohms,

$$C_s = \frac{t_1}{200}.$$

If $t_1 = 500$ μsec, $C_s = 2.5$ μf. Since capacitors having such values are not too bulky, screen leads are usually bypassed.

Flat-top Compensation.—The plate supply decoupling network $R_d C_d$ (see Fig. 2·2) permits compensation for the flat-top distortion due to $R_g C_g$.

The proper relation among R_l, C_d, R_g, C_g is given by

$$R_l C_d = R_g C_g. \tag{12}$$

If Eq. (12) is satisfied, then the voltage across R_g is given by Eq. (13), provided the tube of Fig. 2·2 is assumed to be a constant-current generator and the loading effect of R_g upon the plate circuit is neglected:

$$e_{R_g} = g_m e_g R_l \frac{p(p + \alpha + \delta)}{(p + \alpha)(p + \delta)}, \tag{13}$$

where $p = j\omega$, $\alpha = 1/R_g C_g = 1/R_l C_d$, and $\delta = 1/R_d C_d$. The step-function response of Eq. (13) is

$$f(t) = \frac{\alpha e^{-\delta t} - \delta e^{-\alpha t}}{\alpha - \delta}, \quad \alpha \neq \delta, \tag{14a}$$

or

$$f(t) = e^{-\delta t}(1 + \delta t), \qquad \alpha = \delta. \tag{14b}$$
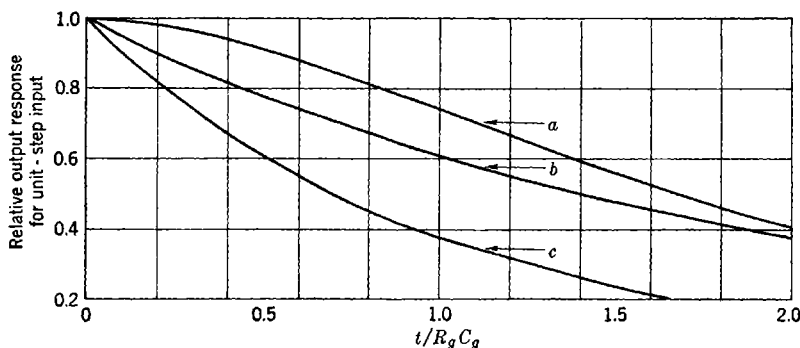
Fig. 2·19.—Flat-top compensation for $R_g C_g$ with $R_d C_d$. Curves (a) and (b) for the case $R_d = R_l$.   (a) $R_l C_d = R_g C_g$; (b) $[R_l R_d/(R_l + R_d)]C_d = R_g C_g$; (c) $R_d = 0$.

Equations (14a) and (14b) display zero slope at $t = 0$, and it is in this sense that Eq. (12) was said to be the proper criterion for the decoupling constants. The somewhat more usual relation

$$\frac{R_l R_d}{R_l + R_d} C_d = R_g C_g \tag{15}$$

leads to the step-function response $e^{-\delta t}$. Figures 2·19a and b compare criteria (12) and (15) for the case in which $\alpha = \delta$, i.e., $R_l = R_d$. It is seen

that even for a sag as large as 10 per cent, criterion (12) (Fig. 2·19a) permits 2.5 times as long a time duration as does criterion (15) (Fig. 2·19b). For comparison the uncompensated case is shown also, Fig. 2·19c.

If $R_d \gg R_l$, the distinction between criteria (12) and (15) tends to disappear, because of the inexactness of commercial resistor and capacitor values.   In principle, however, even for R_d/R_l as large as 5, a 1 per cent sag occurs at about seven times greater time duration with criterion (12) than with criterion (15).

Schemes exist for overcompensating the flat top of a pulse; such procedures are sometimes useful with pulses of known duration.   There are also methods for compensating the screen circuit by means of the decoupling circuit, but these are not discussed here.

2·4. Inverse Feedback.—The main advantages of inverse feedback for pulse amplifiers are gain stabilization and improved linearity in the ratio of output to input voltage.   Inverse feedback is also employed to reduce the rise time of unpeaked amplifiers.

Cathode Resistor.—The plate current of an amplifier tube with unbypassed cathode resistor R_k is

$$\frac{g_m e_g}{1 + g_m R_k}. \tag{16}$$

Equation (16) shows that the sensitivity to changes in g_m is divided by $1 + g_m R_k$; hence distortion caused by variation of g_m with signal amplitude is similarly reduced, and the amplifier linearity is improved.   For the case where $g_m R_k = 1$, the distortion is cut in half.

If R_k is bypassed, there is no reduction in distortion for frequencies for which the bypassing is effective.   Nevertheless, the variability in quiescent plate current among tubes of the same type is reduced; and because there is a good correlation between transconductance and quiescent plate current for tubes of a given type, the transconductance variability is also reduced.   This is an important argument in favor of self-bias, in view of the usual 2/1 range in g_m permitted by the JAN-1A tube specifications under fixed-bias conditions.

Cathode Peaking.—As mentioned in Sec. 2·3 it is usually difficult to bypass the cathode resistor with a condenser large enough to give good flat-top reproduction.   In that case and provided also that the plate circuit is constrained to be a simple *RC*-circuit, it is worth while to choose a value of cathode bypass condenser according to the following discussion, in order to increase amplifier speed.

The analysis of this cathode compensation scheme is most simply carried out by normalizing the plate-circuit constants to consist of a 1-ohm resistor in parallel with a 1-farad condenser, setting $1 + g_m R_k = K$ and denoting the ratio of cathode-to-plate time constants by ρ.   The output

voltage is

$$\frac{g_m e_g}{1 + g_m Z_k} \frac{1}{1 + p} = g_m e_g \frac{1 + p\rho}{K + p\rho} \frac{1}{1 + p}. \tag{17}$$

If the cathode is left unbypassed, $\rho = 0$, and the response of Eq. (17) to a unit step-function voltage e_g is

$$\frac{g_m}{K} (1 - e^{-t}). \tag{18}$$

If, on the other hand, the cathode bypass condenser is chosen so that

$$\text{Cathode time constant} = \text{plate time constant,} \tag{19}$$

then $\rho = 1$ and the response of Eq. (17) to unit step-function voltage e_g is

$$f(t) = \frac{g_m}{K} (1 - e^{-Kt}). \tag{20}$$

Equation (20) displays K times as much speed as Eq. (18) and fully as much gain.
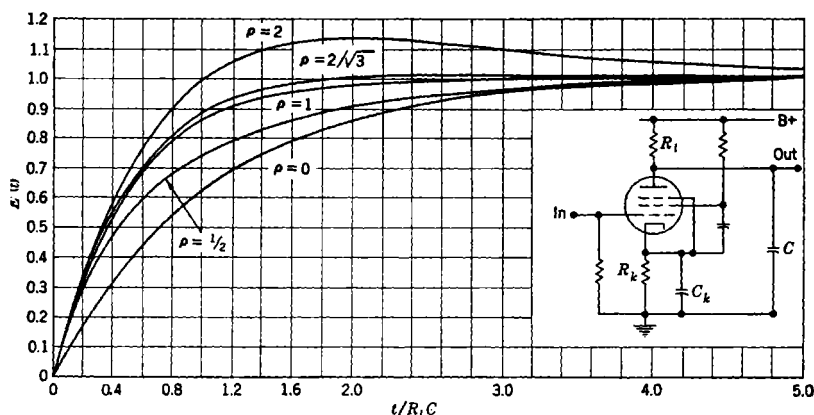
Fig. 2·20.—Step-function response of cathode-peaked pulse-amplifier stage,

$$K = 1 + g_m R_k = 2.$$

Ratio of cathode-to-plate time constants $= \rho = R_k C_k / R_l C$.

It must be pointed out that although the cathode compensation scheme increases the gain/rise time ratio as compared with the unbypassed cathode resistor case, it yields no advantage in gain/rise time ratio as compared with the case in which the cathode is grounded or bypassed completely. This conclusion follows from Eq. (20), which shows that the K-fold increase in speed over the grounded-cathode case comes at the cost of a K-fold decrease in gain.

Cathode compensation can be employed with constants other than those prescribed by Eq. (19). In Fig. 2·20 is shown a family of step-

function-response curves calculated for $K = 2$ and various values of ρ. As these curves show, the step-function response is overcompensated for too large values of ρ.

It is possible to combine cathode compensation with plate-circuit peaking of various sorts, but the details are complicated and are not given here.   The gain/rise time ratio never exceeds that of the grounded-cathode case.

Screen Resistor.—A resistor between screen and screen voltage supply stabilizes the quiescent screen current and hence the plate current.   The effect is very much like that of a cathode resistor and for the same reason reduces the transconductance variability among tubes of the same type.

The number of ohms in a screen resistor that produces the same degenerative effect as a 1-ohm cathode resistor can be determined by an
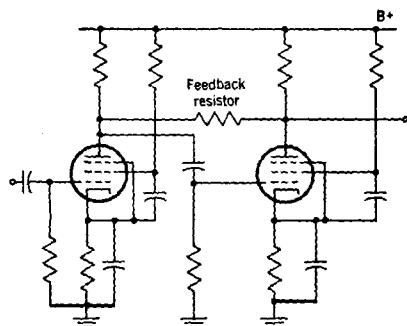
Fig. 2·21.—Feedback from plate to preceding plate.

argument like that used in Sec. 2·3 under "Screen Circuit."   It turns out that a screen resistor of

$$g_m r_s \tag{21}$$

ohms is equivalent to a 1-ohm cathode resistor, where, as before, r_s is the dynamic screen resistance.

A typical value of Expression (21) for a type 6AC7 tube with $g_m \approx 10{,}000$ μmhos and $r_s \approx 20{,}000$ ohms is 200; hence a 60,000-ohm screen resistor has the current-stabilizing effect of a 300-ohm cathode resistor.

Grid-plate Resistive Feedback.—Local feedback can be applied from the plate of an amplifier stage back to its own grid as shown in Fig. 2·21. This circuit can be extended from stage to stage, in which case it is customarily called a "feedback chain."   This subject is treated at length in Chap. 6.   In particular, as shown in Sec. 6·6, it is possible by such means to attain. with an amplifier employing simple interstage

networks (except for the first and last) the gain and bandwidth performance that would otherwise require complicated interstage networks.

Over-all Negative Feedback.—Moderate amounts of over-all negative feedback can be employed without much difficulty in two- or three-stage pulse amplifiers that use two-terminal coupling networks. For proper design of the loop gain characteristic the reader is referred to the exhaustive treatise of H. W. Bode.[1]

Figure 2·22 illustrates a simple application of negative feedback to a two-stage amplifier and cathode follower. Amplifiers of this sort can easily be built with inverse-feedback loop gains of 100, thereby reducing the sensitivity to g_m variations by a factor of 100. In particular, variation of gain with respect to heater voltage is made negligible.
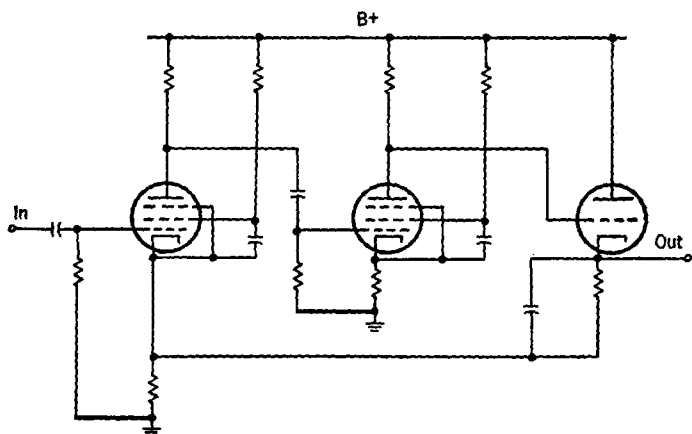
FIG. 2·22.—Simple illustration of over-all feedback.

2·5. Gain Control of Pulse Amplifiers. *Attenuators.*—Potentiometer-type step attenuators of the capacitance-compensated type, as shown in Fig. 2·23, may be used at the input terminals of a pulse amplifier or between stages. For pure attenuation with no waveform distortion, all the RC-products must be the same. If the impedance level at the attenuator can be kept low enough, say 100 ohms, so that the maximum RC-product (maximum resistance times stray capacity) is negligibly small compared with the rise time of the waveforms used, the addition of the compensating capacitances is unnecessary and a simple potentiometer can be used.

For intermediate cases where an extremely low impedance potentiometer is not permissible, a partially compensated potentiometer

[1] H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945.

scheme can sometimes be used. This involves adding from input
terminal to arm a condenser having a value equal to the capacity existing
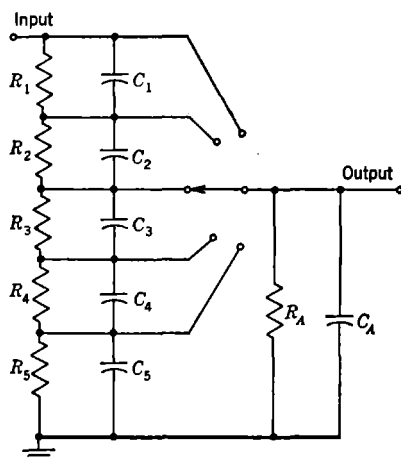
$$R_1C_1 = R_2C_2 = R_3C_3 = R_4C_4 = R_5C_5 = R_AC_A$$

Fig. 2·23.—Compensated step attenuator.
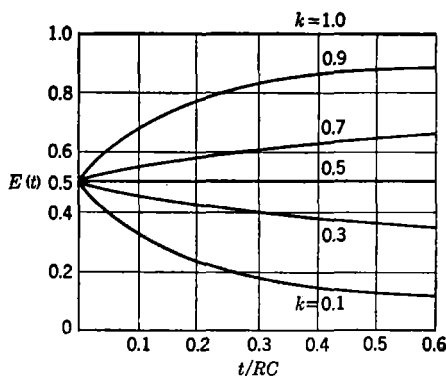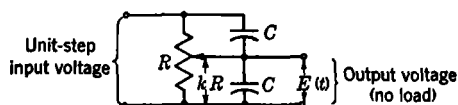
Fig. 2·24.—Step-function response for partially compensated potentiometer.

from arm to ground. The distortion of step-function response for
different potentiometer adjustments is shown in Fig. 2·24. For voltage
divisions up to 3 ($k > 0.33$), adding the extra condenser is an improve-

ment; but for higher attenuation ($k < 0.33$), it is better to omit the condenser.

For use at line impedance levels (50 to 100 ohms), any of the standard types of ladder, T, or bridged-T attenuators may be used if the resistors are sufficiently noninductive (composition resistors, for example) and stray capacitances are kept down to a minimum.

A very interesting type of constant-impedance attenuator has been reported by H. E. Kallmann.[1] Use of "Kelvin line" permits a linear relation between decibels and shaft rotation, and the attenuation can be made independent of frequency up to about 70 Mc/sec or more.

Control of Tube Characteristics.— Since the gain of a pulse-amplifier stage is a direct function of g_m, variation of g_m by changing d-c voltages on the tube elements is a possible way of controlling gain. This method is practical in stages where the signal amplitude is small compared with the grid base. By simultaneously applying the control voltage to several successive stages, an appreciable range of control may be achieved. The variable voltage may be applied to the control grid (through R_g), the screen grid (through R_s), or the suppressor grid. In the third case it is desirable to use a tube with good suppressor control, such as the type 6AS6.

FIG. 2·25.—Variation of gain by feedback (cathode degeneration).

This type of control is better suited to adjusting the output amplitude with a fairly constant input signal than it is for correcting for large input level variations where there is danger of overloading the first stage.

Variable Negative Feedback.—One s mple method of gain control is shown in Fig. 2·25. A variable cathode resistor is so arranged that it varies the feedback but not the tube bias. Gain variations up to 5/1 can usually be accomplished in this way without difficulty. The limit depends on the speed of the stage, since, for higher speeds, the maximum size of the variable resistor must be held down. This method is good for variable input levels because the input voltage overload level increases as gain is reduced.

For other types of feedback amplifiers, the gain can also be changed

[1] H. E. Kallmann, "Portable Equipment for Observing Transient Response of Television Apparatus," *Proc. I.R.E.*, **28**, 351–359 (1940).

by changing the feedback ratio.   This may not be easy if the amount of feedback is large but can usually be accomplished with a compensated attenuator in the feedback path.

2·6. D–c Restoration.—This widely accepted but misleading designation refers to a process of clamping either the most positive excursion or the most negative excursion of a signal to a specified voltage level. For a constant-amplitude periodic signal it is desired to accomplish this without significant waveform distortion.

D-c restoration is most often used where the signals are "single-sided," as they are in television and radar pulse systems.   When such asymmetrical signals have been passed through capacitive couplings, their d-c component is lost, and for this particular type of signal the use of the term "d-c restorer" is fairly well justified.

A typical d-c restorer circuit for positive-goin   signals is shown in Fig. 2·26.   For negative-going signals the diode is reversed.   Good d-c restoration requires signal amplitudes at least as large as 1 volt, since diode conductance is small for small signals.
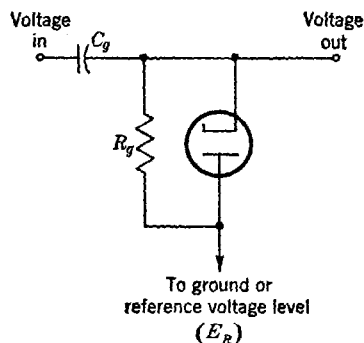
Fig. 2·26.—Typical d-c restorer circuit (shown for positive-going signals).

A danger with the d-c restorer shown in Fig. 2·26 for positive-going signals is that diode heater-to-cathode hum current develops a hum voltage across R_g.   For type 6AL5 diodes with $R_g = 1$ megohm, and 6.3 volts alternating current applied to the diode heaters, with one side of the heaters grounded, the hum voltage in 12 out of 98 tubes tested was in excess of 0.25 volts, and 1 tube produced 1.4 volts of hum.   This hum voltage may be very disturbing.   It may be avoided or reduced by use of one of the following: (1) A nonthermionic diode, such as a germanium crystal (but the back resistance of such crystals is small, tending to increase the sag in pulse response), (2) direct current on the diode heaters, (3) a separate heater winding for the diode heaters, suitably biased, or (4) smaller values of R_g (this also tends to increase sag).   The heater-hum difficulty does not arise in the case of a d-c restorer for negative-going signals.

The action of the circuit of Fig. 2·26 upon signals is shown in Fig. 2·27.   Although not visible in this illustration, waveform distortion occurs whenever the amplitude of the envelope of the waveform increases, i.e., while the coupling condenser is being charged to its new value. When the amplitude of the envelope decreases, there is a delay in restoring

the level of the baseline until the condenser has had time to discharge to the new value.

This discussion is based on the assumptions of (1) voltage generator impedance low compared with R_g, (2) an ideal diode, and (3) a zero
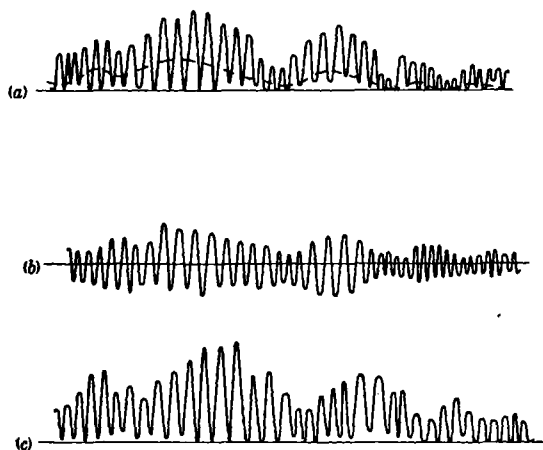
Fig. 2·27.—Effect of d-c restorer (time scale long compared with time constants involved). (a) Original signal, d-c component shown dotted; (b) signal after d-c component has been removed; (c) signal at output terminals of d-c restorer.
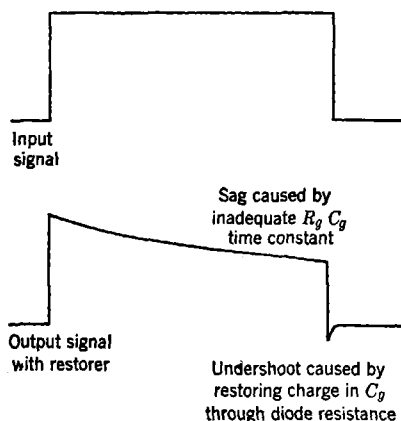
Fig. 2·28.—Clipping caused by d-c restorer (with rectangular input pulse having a time duration of the same order of magnitude as R_gC_g).

impedance supply for E_R. Ordinarily, the parameters are such that these approximations are justifiable. Also, it is assumed that the time constant R_gC_g is large compared with the duration of individual pulses within the signal envelope but short enough to follow envelope varia-

tions.   Here caution must be exercised, for these last assumptions are
frequently not true.   Figure 2·28 shows the result when a rectangular
pulse having a duration of the same order as $R_g C_g$ is passed through the
circuit.   In this case the flat top of the signal has been distorted because
of the inadequate grid time constant $R_g C_g$, and the undershoot at the
end of the signal has been radically altered in form by the action of the
diode.   The residual undershoot shown is due to the finite resistance
of the actual diode; the ratio of undershoot to sag is approximately
$R_{\text{diode}}/(R_{\text{diode}} + R_{\text{source}})$.

For negative-going signals it is possible to use the grid of a zero-bias
amplifier as its own d-c restorer, but the results are often not so good
as the use of a separate diode for the purpose, because the grid-cathode
conductance may not be large.

Where a d-c restorer is used, it must be remembered that the $R_g C_g$
coupling circuit can no longer be exactly compensated by the decoupling
network (see Sec. 2·3), since a nonlinear network with a nonlinear ele-
ment cannot be exactly compensated by a linear network.   Therefore,
where a d-c restorer is required, the $R_g C_g$ coupling circuit should be
adequate to preserve flat top by itself.

2·7. Limiting Amplifiers.—For a good many purposes, particularly to
avoid driving an intensity-modulated cathode-ray tube into the "bloom-
ing" region, it is necessary to place an absolute limit on the magnitude
of the output signal.   This limit can be specified either with respect to
(1) the a-c axis of the input signal or (2) the peak-to-peak excursion.

Both cases are customarily treated by adjusting the bias of a nega-
tive-going pulse-amplifier stage so that it is driven beyond plate-current
cutoff for signals larger than the specified value.   Thus a definite limit is
established.

A common way to set the limit is to increase the screen dropping
resistor R_s until the desired limit is obtained, leaving R_k at its normal
value.   This reduces the quiescent-plate-current variability among tubes
and hence the limit-level variability, as discussed in Sec. 2·4.

If R_s is bypassed, there is variation, however, of limit level with duty
ratio, i.e., with the average value of plate current.   The reason is that
average plate current determines average screen current; this, in turn,
in the presence of a bypassed screen resistor, determines quiescent screen
voltage and hence limit level.   The only way to combine limit-level
stabilization with regard to both duty-ratio and tube variability is to
leave R_s unbypassed.

Unfortunately, leaving R_s unbypassed is not practical; it not only
greatly reduces gain but also increases the input capacity of the tube
by the amount $\mathcal{G}C_{gs}$, where $\mathcal{G}$ is the voltage gain from grid to screen and
C_{gs} is the grid-to-screen capacity; this is an example of a somewhat
unconventional Miller effect.   The conclusion is that an engineering

compromise has to be made between limit-level stabilization with regard to duty ratio and limit-level stabilization with regard to tube variability.

If limiting of the peak-to-peak type is desired, d-c restoration should be used at the grid of the limiter stage. This method may be economically used by making $R_k = 0$, so that the grid performs its own restoring action to prevent excursions in a positive direction from ground level. (R_s is then the sole current-stabilizing element.) It may be better to use a separate diode as the restorer, with its plate tied to the amplifier grid and its cathode tied to the desired potential. Certain precautions regarding the use of d-c restorers are noted in Sec. 2·6.

The output voltage from such a limiter has its positive excursion limited if the output signal is taken from the plate. Therefore, if it is desired to limit the negative excursion instead, the limiting stage should be followed by another that inverts the polarity.

2·8. The Mixing of Multiple Input Signals. *Resistance Networks.—* Almost any of the types of resistance mixers used in the communication field can be employed for pulses if noninductive resistors are used, capacities are kept down, and the impedance level is reasonably low.

Tube Mixing.—Because a pentode is essentially a constant-current generator, one may parallel the plates of two or more pentode amplifiers, using a common load impedance. The effect of increased output capacity on the

Fig. 2·29.—Feedback-amplifier mixer.

speed of the plate circuit must be taken into account, and allowance must be made for the effect of the increased direct current in the load and decoupling resistors.
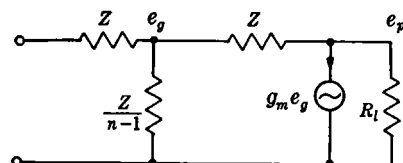
Fig. 2·30.—Nodal analysis of feedback-type mixer.

This system has the distinct advantage that it is unilateral. For cases where a maximum of isolation is desired among the input signals to be mixed, it is an excellent method. The linearity of mixing is not very good, however.

Under particular conditions tube mixing may be accomplished by using more than one grid as a signal-injecting element. For example, the suppressor of the type 6AS6 may be used in this way.

Mixing Network with Feedback Amplifier.—Figure 2·29 shows a type of mixer whose output voltage is accurately additive. In addition, coupling among input channels is very low if the amplifier gain is high.

The behavior of this mixer circuit may be understood from the nodal analysis of Fig. 2·30, drawn for a single input channel. Setting $Y = 1/Z$, $G = 1/R_l$, the determinant Δ is obtained:

$$\Delta = \begin{vmatrix} Y & -Y & 0 \\ -Y & (n+1)Y & -Y \\ 0 & g_m - Y & Y + G \end{vmatrix}.$$

Hence

$$\frac{e_p}{E_1} = \frac{\Delta_{13}}{\Delta_{11}} = \frac{-Y(g_m - Y)}{(n+1)Y(Y+G) + Y(g_m - Y)}$$

$$= -\frac{1}{1 + \dfrac{(n+1)(Y+G)}{g_m - Y}} = -\frac{1}{1 + \dfrac{(n+1)\left(1 + \dfrac{Y}{G}\right)}{g_m/G - Y/G}}$$

$$\approx -\frac{1}{1 + \dfrac{n+1}{g_m R_l}},$$

for $Y \ll G$, $g_m \gg G$.

If[1] $g_m R_l \gg n + 1$, the output voltage is accurately equal to the negative of the input voltage, regardless of variations in g_m. By the principle of superposition the mixer as a whole is accurately additive in its input voltages, except for a sign reversal, i.e.,

$$e_p = -(E_1 + E_2 + \cdots + E_n).$$

Figure 2·31 shows a three-channel feedback mixer. The condenser C is included to block direct current. Because of the inverse-feedback action, the resistors R can be made much larger than usual for a given speed of response.

Nonadditive Mixing.—In some applications, particularly where the pulse output voltage is used to modulate the intensity of a cathode-ray tube, it may be desirable to arrange matters so that if two or more input signals occur at the same time, the output voltage will be that due to the largest signal only; this process is called nonadditive mixing. For example, if a cross-hatched pattern is to be presented on an intensity-modulated cathode-ray tube, it is desirable to suppress the extra brilliance at the intersections of the lines that would occur if the two signals were added.

---

[1] For a multistage amplifier $g_m R_l$ should be replaced by the gain $\mathcal{C}$ of the amplifier with the feedback loop open.

This can be done for any number of input channels of positive polarity
signals by putting the signals
through separate cathode followers
operating with a common cathode-
load impedance R_k and biased in
their quiescent conditions approxi-
mately to plate-current cutoff
(Fig. 2·32). If R_k can be made
large enough so that the gain is
nearly unity, then a signal of 10
volts in one input channel intro-
duces approximately 10 volts addi-
tional bias on the other cathode
followers. Thus a signal of less
than about 10 volts in any of the
other input channels cannot over-
come the grid bias and hence does
not appear in the output voltage.
Therefore the largest input signal
at any instant effectively masks
all smaller ones.

FIG. 2·31.—Typical three-channel feedback-
amplifier mixer.

For cases where the value of R_k is such that the gain is substantially
less than unity the biasing-off action is not complete, and some addition
occurs for signals that are smaller than the largest one.
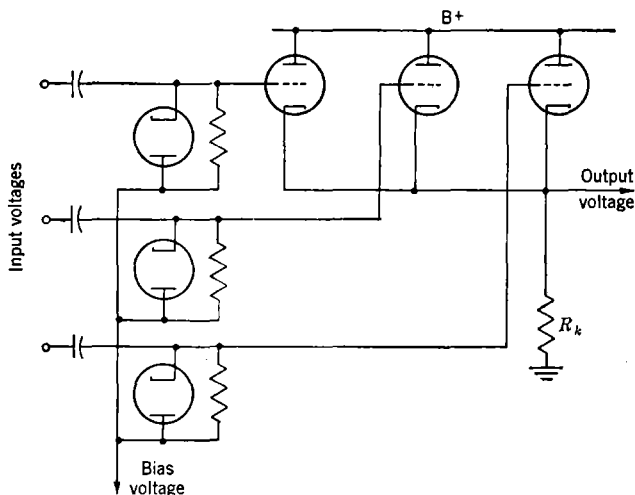
FIG. 2·32.—Nonadditive mixer for positive-polarity signals. Any number of input channels
can be used, with one tube for each input channel.

When the initial quiescent bias is insufficient for plate-current cutoff, there are two significant effects: (1) The gain is less for small signals than for larger ones (a nonlinearity is introduced in the gain characteristic), and (2) the gain is reduced, again resulting in some addition of signals. Effects (1) and (2) are both due to the loading of the output circuit by the additional cathode impedances when the tubes are not operated at initial cutoff. Usually some additive effect can be tolerated, so that complete initial cutoff is not required.
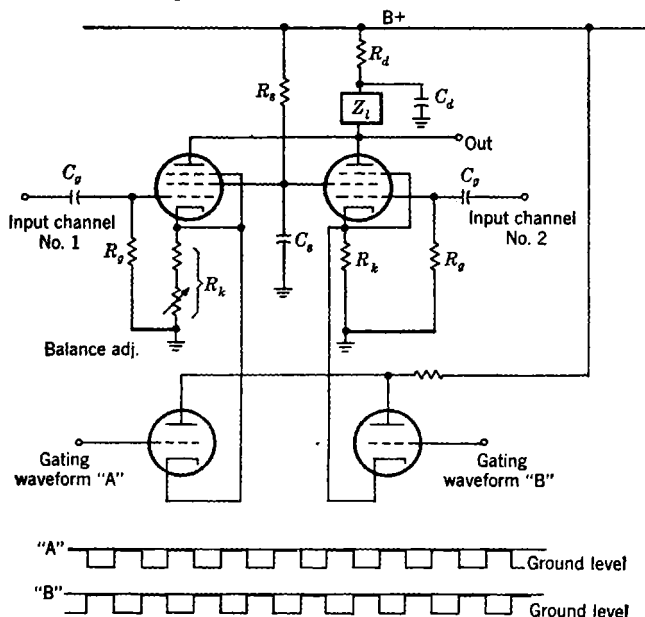
Fig. 2·33.—Electronic input-channel switching. Input signal No. 1 is passed through when waveform "A" is below ground and waveform "B" is at ground.

In principle the same action can be obtained with diodes instead of cathode followers if the input signals come from sufficiently low impedance sources and the common load impedance of the diodes is large compared with the forward impedance of any one of the diodes. By poling the diodes properly, either polarity of input signal may be used, provided that all input signals have the same polarity. The diode method is not likely to be practical in most cases.

For either of these methods the d-c level of the input signals must be held constant; this ordinarily requires d-c restoration if the signals have appreciable duty ratio.

2·9. Electronic Switching of Pulse Amplifiers.—The need often arises to switch the input signal to a pulse amplifier so that it will be applied

only at a desired time. It is seldom permissible merely to turn off the plate or screen voltage or cut off the suppressor of an amplifier stage, since this action causes a severe switching transient, which is passed on to the rest of the amplifier.

The switching transient can be eliminated or at least reduced to a very short duration by the method shown in Fig. 2·33. The additional pentode is used to compensate for the change of current with switching in the amplifier pentode. An adjustment is provided for balancing the currents.

If it is desired to switch the amplifier input terminals from one input channel to another, the second input signal may be applied to the grid of the second pentode.

The device for generating the gating waveform (rectangular waveform) may be a multivibrator or trigger circuit as the switching requirements demand. For periodic switching controlled by a sine-wave generator (such as the power-line current), a "squaring" amplifier can be used. The essential requirement is that the d-c level of the top of the gating waveforms be fixed. If necessary, d-c restorers can be used at the grids of the cathode followers, but often direct coupling can be arranged.

2·10. Output Stages. *High-impedance Load.*—Many of the applications for pulse amplifiers concern a load that is essentially a high resistance shunted by a small capacity, such as the input impedance of a cathode-ray tube. For a tube used as a voltage output stage of a pulse amplifier of this sort, a logical figure of merit is the ratio of maximum output voltage to rise time with resistance-capacitance load. (This figure of merit disregards the important question of gain.)

For signals that are either symmetrical or of negative polarity at the grid of the tube the maximum output voltage is $E_{max} = I_p R_l$, where I_p is the quiescent plate current of the tube and R_l is the load resistor. If the total output circuit shunt capacity, including that of the load, is C and if τ denotes the 10 to 90 per cent rise time, then

$$\frac{E_{max}}{\tau} = \frac{I_p}{2.2C} \qquad \text{volts/sec} \tag{22}$$

[compare Eq. (22) with Eq. (3)].

For signals of positive polarity at the grid of the tube, the same formula holds if I_p is redefined as the incremental plate current from the quiescent point to some allowable maximum. It is difficult to evaluate the figure of merit for the general case, since the allowable value of I_p is sometimes determined by plate dissipation, sometimes by emission, and sometimes by the requirement of avoiding grid current.

The following table has been compiled by considering the case of

symmetrical signals, making reasonable assumptions about the limiting factors on I_p and assuming 20-$\mu\mu$f capacity beyond the amplifier tube itself.

TABLE 2·5.—RATIO OF MAXIMUM OUTPUT VOLTAGE TO RISE TIME FOR VARIOUS TUBE TYPES

Tube type	I_p, ma	C	E_{max}/τ (volts/μsec)
6AK5	10	23	200
6AC7	15	25	270
2-6AK5's (in parallel)	20	26	350
6AG7	30	27.5	500
6V6GT/G	45	27.5	750
6L6G	75	30	1130

The listings in Table 2·5 are sensitive to the assumed external capacity.

The proposal is sometimes made that the figure of merit [Eq. (22)] of an output stage can be increased by using cathode-peaking (Sec. 2·4). Consider a type 6AK5 tube, for example, with $g_m = 5000$ μmhos and $R_k = 200$ ohms, so that $1 + g_m R_k = 2$. Suppose the plate-load resistor is 2000 ohms, and suppose the cathode bypass condenser is chosen so that the cathode and plate time constants are equal. Then, as follows from Eq. (20), the response to a step-function grid voltage is as fast as normally results from the parallel combination of a 1000-ohm load resistor and the output capacity.

However, it is argued, the maximum output voltage is what is achieved from a 2000-ohm load resistor, i.e., 20 volts for a maximum change in plate current of 10 ma. Hence the ratio E_{max}/τ is thought to be twice what it is in the grounded-cathode case. The fallacy in this argument is that the inverse feedback is able to double the speed of output-circuit response only when there is enough plate current available to drive the output circuit with a sharply peaked current waveform, having an initial current twice its steady-state value. If this doubled plate current is not available, the inverse-feedback action fails. Hence the improvement in speed pertains to small signals only and not to limiting signals.

What happens in the cathode-peaked amplifier stage above is that the maximum output signal is equal to 20 volts, corresponding to the combination of 10 ma and 2000 ohms, but the speed of response of a limiting signal is no greater than that of a grounded-cathode stage having the same plate circuit.

Push-pull Amplifiers.—When driving the deflection plates of a cathode-ray tube, considerable output voltage is required, and it is better
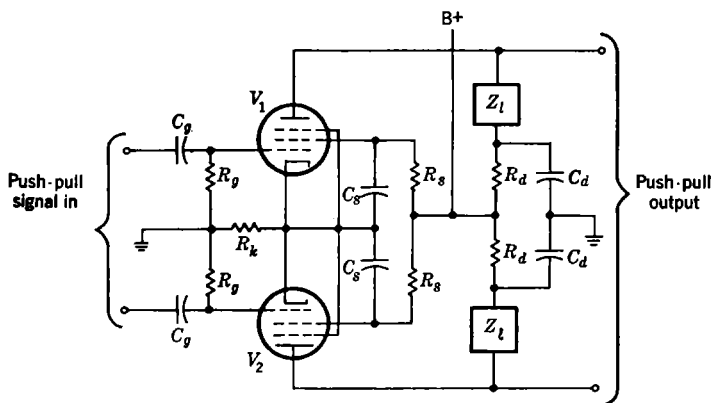
FIG. 2·34.—Typical push-pull pulse-amplifier stage.

$R \approx 80$ ohms (for 6AC7's)

(b)
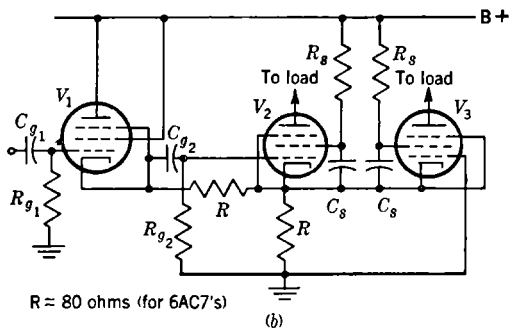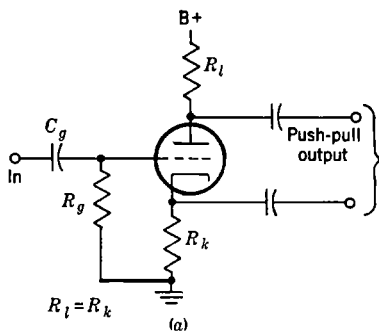
FIG. 2·35.—Phase inverter.   (a) Single-tube phase inverter; (b) balanced-output imped-
ance phase inverter.

from the standpoint of focus to apply the signals in push-pull. A push-pull output stage for such purposes is usually constructed like two identical single-ended stages, each driving a single deflection plate. The only difference is that the two cathodes are tied together and go through a common bias resistor. Because there is in principle zero signal current flowing through the cathode resistor, leaving the cathode resistor unbypassed entails no loss in gain, in contrast with the single-sided amplifier case. Such a stage is shown in Fig. 2·34.

Phase Inverters.—The transition between a single-sided stage and a push-pull stage is accomplished by means of a phase-inverter stage. It is usually impossible to employ a center-tapped pulse transformer because of its inadequate flat-top performance.

The circuit diagrams of two types of phase-inverter stages are shown in Fig. 2·35. The type shown in Fig. 2·35a is best used at fairly low level, where R can be kept small without requiring excessive plate current. At frequencies above about 10 Mc/sec a pentode should be employed in the circuit of Fig. 2·35a to avoid the signal transmission path through the grid-plate capacitance. When a pentode is employed, the plate resistor should be larger than the cathode resistor in the ratio of the cathode current to the plate current.

For operation at still higher frequencies it is wise to proportion the voltage divider consisting of the grid-cathode capacitance and the cathode-ground capacitance so that its loss is equal to the loss through the tube from grid to cathode, i.e., so that

$$\frac{C_{gk}}{C_{gk} + C_{ko}} = \frac{g_m R}{1 + g_m R},$$

where C_{gk} is the grid-cathode capacitance and C_{ko} is the cathode-ground capacitance. This requirement usually necessitates adding grid-cathode capacitance.

The phase inverter shown in Fig. 2·35b has low gain but has the advantage of equal and low output impedances. Moreover, both heater-cathode and plate-supply hum are canceled in this arrangement, whereas neither type of hum is canceled in the circuit of Fig. 2·35a.

The action of the circuit of Fig. 2·35b may be explained by denoting by i_s the signal current in the cathode follower, by i_1 the branch current (from the left-hand push-pull tube) that flows in the resistor R between push-pull cathode and ground, and by i_2 the corresponding branch current from the right-hand push-pull tube. Then, if the push-pull tubes are assumed to have equal transconductance g_m,

$$i_1 = g_m R i_s,$$
$$i_2 = -g_m R(i_s + i_1 + i_2).$$

Hence

$$i_2 = -i_1(1 + g_m R) - g_m R i_2,$$
$$i_2(1 + g_m R) = -i_1(1 + g_m R);$$

consequently

$$i_2 = -i_1.$$

Low-impedance Load.—A pulse amplifier is often required to drive a low-impedance line of character-istic impedance between 50 and 100 ohms. If such a line is termi-nated in its characteristic imped-ance, it can be treated as if it were a load resistor. But the necessity frequently arises for isolating the d-c level of the amplifier plate from the line, and in order to pre-serve good reproduction of pulse flat top an extremely large block-ing condenser would then be re-quired. The customary way of avoiding these difficulties is to use

FIG. 2·36.—Use of cathode follower to drive low-impedance line.

a cathode follower between the last amplifier stage and the line, as shown in Fig. 2·36.

Cathode Follower.—The cathode follower is a form of voltage feedback amplifier having a voltage gain less than unity. For a load impedance R, the gain is

$$\mathcal{G} = \frac{g_m R}{1 + g_m R \dfrac{\mu + 1}{\mu}}. \tag{23}$$

Since $\mu \gg 1$ for pentodes and most triodes, Eq. (23) can be written

$$\mathcal{G} \approx \frac{g_m R}{1 + g_m R}. \tag{23a}$$

The output resistance is

$$\frac{R}{1 + g_m R \dfrac{\mu + 1}{\mu}}, \tag{24}$$

or approximately

$$\frac{R}{1 + g_m R} = R(1 - \mathcal{G}). \tag{24a}$$

The input capacitance of a triode cathode follower is

$$C_{in} \approx C_{gp} + (1 - \mathcal{G})C_{gk}, \tag{25}$$

where $\mathcal{G}$ is the gain of the cathode follower, C_{gp} is the direct grid-plate capacity, and C_{gk} is the direct grid-cathode capacity. For a pentode cathode follower with screen bypassed to cathode,

$$C_{in} \approx (1 - \mathcal{G})(C_{gs} + C_{gk}), \tag{26}$$

where C_{gs} is the direct grid-screen capacity.

As an example consider a type-6AC7 pentode cathode follower with a g_m of 9000 μmhos and a 1000-ohm cathode load resistor. Then from Eqs. (23a), (24a), and (26) it follows that the gain $\mathcal{G}$ is $\frac{9}{10}$, the output impedance is 100 ohms, and the tube input capacity is about 1.3 $\mu\mu$f. To this 1.3 $\mu\mu$f must be added the usual wiring and socket capacity of about 4 $\mu\mu$f; the total input capacity at zero frequency is thus reduced to about a third the usual 6AC7 value. It must be remembered, however, that for a cathode follower with RC load the input capacity increases somewhat with frequency because of the lower gain of the cathode follower at the higher frequencies; unfortunately, it is precisely at the higher frequencies that reduction in capacity is important.
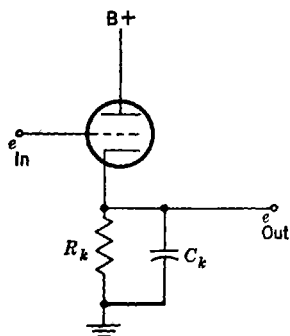
FIG. 2·37.—Circuit for analysis of effect of capacity in shunt with load of cathode follower.

The low output resistance of a cathode follower is its most valuable feature. If a condenser C_k is connected across the resistor R_k (see Fig. 2·37), then the rise time is determined by the combination of C_k and the output impedance of the cathode follower; this rise time is only $1 - \mathcal{G}$ as long as the rise time of the $R_k C_k$ combination; i.e.,

$$\tau = \frac{2.2 R_k C_k}{1 + g_m R_k}. \tag{27}$$

For a given external capacity C_k considerably greater speed can therefore be obtained.

Equation (27) was derived on the assumption of a constant value of g_m; that is for small signals. However, it holds for positive-going step functions even of extremely large amplitude, since in the positive direction the g_m of a tube does not decrease until the grid is driven far into the positive region.

In the case of negative-going signals, Eq. (27) holds only for small signals. A large negative step-function grid voltage moves the grid voltage instantaneously; because the cathode voltage cannot follow instantaneously, the instantaneous grid bias is then large. (It must be remembered that the trailing edge of a positive pulse is a negative-going

step, to which the above remarks apply.) The tube is cut off if the amplitude of the step is larger than the difference between the quiescent grid-cathode bias and the bias value for cutoff. The condenser C_k then discharges through R_k only, until the cathode potential falls to the point where the tube again conducts.

The conclusions for pulses of amplitude about equal to the grid base of the tube are that (1) the cathode follower is an unsymmetrical device that is faster on the rising edge of the pulse than on the falling edge, (2) the fall time is a function of amplitude and may be as large as that of the external circuit. The fall time can be reduced by increasing the quiescent voltage drop across R_k, which ordinarily means increasing the quiescent plate current of the tube.

2·11. Examples. *Pulse Amplifier for Deflection-modulated Cathode-ray Tube.*—The amplifier whose circuit diagram is shown in Fig. 2·38 was designed for use with the Radiation Laboratory Model P-5 Synchroscope[1] and is intended to drive the deflection plates of a type 5JP1 cathode-ray tube.

The over-all rise time of the amplifier is $\frac{1}{50}$ μsec, with about 1 per cent overshoot over-all, and the sag in a 200-μsec rectangular pulse is about 1 per cent. The gain is 200, and the deflection sensitivity at the amplifier input terminals is about $\frac{1}{3}$ volt/inch. The maximum peak-to-peak output voltage is about 220 volts, or $3\frac{1}{2}$ inches.

The amplifier is operated in Class A, so that it accepts signals of either polarity and its gain is independent of duty ratio.

The input step attenuator is of the compensated type discussed in Sec. 2·5, the semiadjustable condensers being set to make the RC products equal. The voltage ratio per step is about 3/1.

Fidelity of flat-top reproduction is achieved by leaving the cathode resistors unbypassed and employing large screen bypass condensers and large $R_g C_g$ time constants. The coupling condenser between the type 6AC7 phase-inverter tube and the left 6AG7 tube is only 0.001 μf, but flat-top compensation is used in that stage (see Sec. 2·3). The compensating network is placed in the grid lead rather than the plate-supply lead; this reversal permits a very small value of δ [see Eq. (14)] and is a good idea where the plate-load resistor is small, as is the case in very high-speed amplifiers.

The peaking circuits are of the four-terminal linear-phase type (Fig. 1·26) whose step-function response is shown in Fig. 1·28.

The phase inverter has the form of Fig. 2·35a. For the reasons given in Sec. 2·10 it is operated as a pentode. Hence the plate-load resistor (390 ohms, = 4300 in parallel with 430) is 20 per cent larger than the

---

[1] Vol. 22, Chap. 7, of the Radiation Laboratory Series.
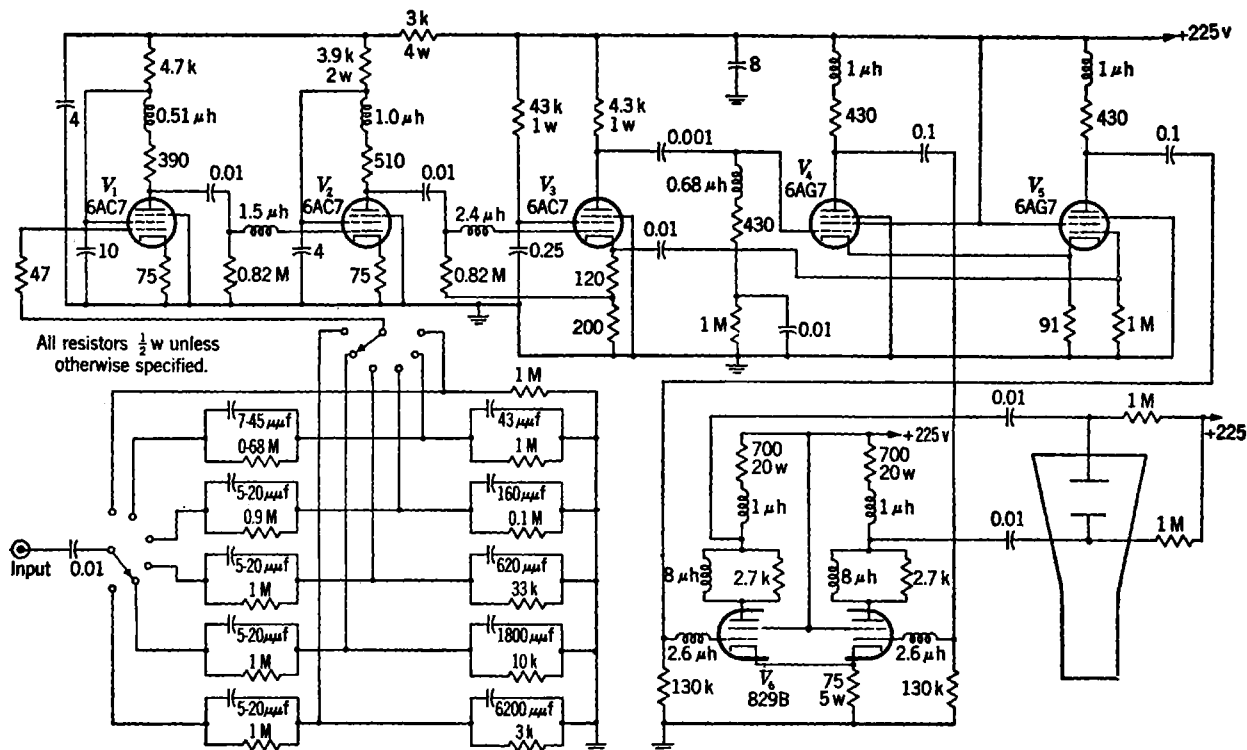
FIG. 2·38.—Pulse amplifier for deflection-modulated cathode-ray tube.

cathode-load resistor (320 ohms), because the cathode current is about 20 per cent larger than the plate current.

Only the plate circuit of the phase inverter is peaked; the low output impedance of the cathode circuit is adequately fast without peaking.

Pulse Amplifier for Intensity-modulated Cathode-ray Tube.—The amplifier whose circuit diagram is shown in Fig. 2·39 was designed to drive the cathode of an intensity-modulated cathode-ray tube.
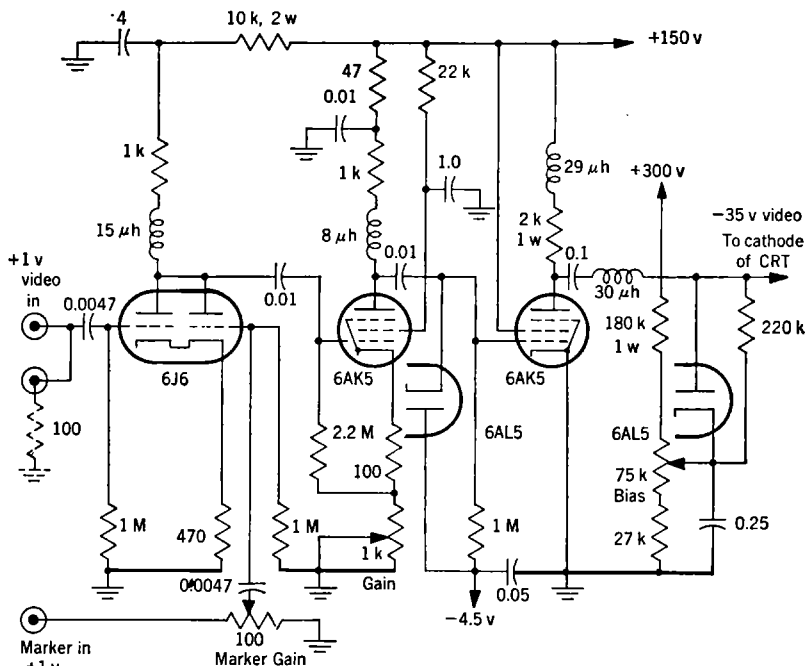
FIG. 2·39.—Pulse amplifier for intensity-modulated cathode-ray tube.

The over-all rise time of the amplifier is about $\frac{1}{20}$ μsec, with about 5 per cent overshoot over-all, and the sag in a 1000-μsec rectangular pulse is about 2 per cent. The gain is about 20, with very conservative ratings of tube transconductances. The output voltage is about 35 volts.

The amplifier is intended for one-sided signals of positive polarity at the amplifier input terminals, and d-c restorers are used at the input terminals of the last amplifier tube and at the cathode of the cathode-ray tube.

Positive signal pulses, limited to an amplitude of 1 volt, are brought to the amplifier terminals on a 100-ohm coaxial line, whose terminating resistor is located on the amplifier chassis.

Tube-mixing (Sec. 2·8) of the signal and market pulses is accomplished in the 6J6 plate circuit.   The gain of the mixer stage is about unity.

Gain is adjusted by varying the amount of degeneration in the cathode lead (Sec. 2·5) of the first type 6AK5 tube.

The peaking of the 6J6 mixer stage must take into account the shunting effect of the triode plate resistance. A larger value of peaking inductance is therefore required for the same increase in speed.   The second stage employs simple shunt-peaking, with $m = 0.41$ (Sec. 2·2). In the last stage, where it is necessary to achieve large output voltage as well as high speed, a four-terminal network is employed; this is a variant of the network of Fig. 1·26.

A 47-ohm, 0.01-μf decoupling network will be observed in the plate-supply lead of the second stage.   Its purpose is to decouple the amplifier from the particular power supply in actual use and has no general significance.

CHAPTER 3

PULSE AMPLIFIERS OF LARGE DYNAMIC RANGE

BY HARRY J. LIPKIN

3·1. Introduction.—Certain special applications require high-gain video, or pulse, amplifiers that must handle a large dynamic range of incoming signals, with voltage ratios between largest and smallest signals of 10^5 or more. One such application is the "crystal-video receiver,"[1] in which an r-f pulse signal from the antenna is immediately detected and the resulting video signal amplified. The type of amplifier required for this purpose is quite different from the high-fidelity pulse amplifiers discussed in Chap. 2, as is made clear by the fact that not one of the amplifiers discussed in this chapter employs high-frequency peaking. The purpose of this chapter is to present the special features and design considerations of pulse amplifiers in which the emphasis is on high gain and, above all, very large dynamic range.

Gain.—To obtain maximum sensitivity, the amplifier should have sufficient gain to bring pulses up from the amplifier noise level to an amplitude suitable for viewing on an indicator or for triggering an auxiliary circuit. The rms open-circuit thermal-agitation noise level $e = \sqrt{4kTRB}$ is approximately 7 μv for an amplifier source resistance $R = 3000$ ohms and bandwidth $B = 1$ Mc/sec; if the amplifier noise figure is 3 db, the equivalent input noise level is 10 μv. To obtain a noise level output of 3 volts, a gain of 110 db is required. For other bandwidths, noise figures, output levels, and source resistances the required gain varies somewhat, but the order of magnitude remains the same.

Dynamic Range.—Signals appearing at the input terminals of the amplifier may have amplitudes as small as that of the noise, that is, several microvolts, or as large as several volts. If no special care is taken to handle the large signals, they will either drive successive stages of the amplifier to cutoff or cause heavy grid current to flow, causing the amplifier to block for a considerable period after each strong pulse and thereby rendering it completely useless for the amplification of weak signals during that time. Some applications require that the period of insensitivity following strong signals be reduced below an acceptable minimum; others require further that the width of strong pulses be preserved despite limiting. In either case the problem of overload is the central

---

[1] See Vol. 23 of the Radiation Laboratory Series.

one in this type of amplifier and requires an entirely different approach from that of Chap. 2.

Microphonics.—A problem peculiar to high-gain pulse amplifiers is that of microphonics. Mechanical shock or vibration of the first amplifier tube causes fluctuations in the plate current which appear as low-frequency signals at the output of the amplifier. If the low-frequency response of the amplifier is good enough by conventional standards to pass the pulse satisfactorily, experience shows that microphonic signals are considerably stronger than saturation level, in spite of all mechanical precautions that may have been taken, such as shock mounting. Elimination of microphonic signals can therefore be accomplished only by rejecting low frequencies somewhere in the amplifier. This rejection results in distortion of the pulse and in low-frequency overshoots which must be minimized in the amplifier design.

Small Amplifiers.—High-gain video amplifiers are often useful for ultraportable applications in which performance is secondary to light weight and low power consumption. Design of these amplifiers requires a different emphasis. Very often the circuits actually used are inefficient according to normal criteria such as gain-bandwidth product or number of stages.

Theoretical Approach.—The classic method of using the amplitude and phase response of a network in order to determine its transient behavior is very useful in linear cases; this treatment can be applied to high-gain video amplifiers, however, only with reservations. In the consideration of signals below saturation level linear theory is useful to determine rise time or optimum bandwidth for good signal-to-noise ratio. However, in regard to large signals and overload effects the amplifier is primarily nonlinear, and the linear theory is of little use.

Use with Square-law Detector.—High-gain video amplifiers are frequently preceded by a crystal detector to make a crystal video receiver. Since the detector is a square-law device, quantities such as dynamic range and signal-to-noise ratio are different if specified at the input terminals of the crystal from what they are at the input terminals of the amplifier. For example, a 100-db dynamic range at the amplifier input terminals corresponds to only a 50-db dynamic range at the crystal input terminals. A more extended treatment of crystal video receivers is presented in Vol. 23, Chap. 19, of this series.

3·2. Theory of Overshoots.—In video amplifiers condensers are widely used for coupling and bypass purposes. If these condensers are not to cause pulse distortion, they must behave as a perfect short circuit for any frequency other than zero. This restriction requires that the voltage across the condenser and therefore its charge remain constant regardless of the presence of signals. However, all signal voltages tend

to charge or discharge such condensers, each pulse causing a small change
in the charge of each condenser.   After the end of the signal pulse, each
condenser tends to restore itself to its normal charge, thereby producing
transients that appear as spurious signals following the pulse.   These
spurious signals are referred to as low-frequency overshoots, since they
arise from the poor low-frequency response of the amplifier, i.e., the
failure of the condensers to have zero impedance at low frequencies.
In the remainder of this chapter, they are referred to simply as over-
shoots and are the only type of such phenomena considered here.

In the type of pulse amplifier to which this chapter is devoted, the
large dynamic range makes the matter of overshoots the main problem.
It is not sufficient for the overshoots to be reduced to 10 per cent or even 1
per cent of the signals producing them, because even 1 per cent of a very
strong signal is still considerably stronger than the very weak signals to
which the amplifier should be sensitive.   In order that the overshoot on a
1-volt signal be less in amplitude than a 10-μv signal, the overshoot must
be less than 0.001 per cent.   It is for this reason that the problem of how
overshoots are produced and how they can be controlled is given the
extensive analysis that follows.

Generation of Overshoots.—Consider the circuit of Fig. 3·1, which
represents a simplified coupling
circuit.   In the absence of signal,
all voltages are zero, there is no
current flow, and there is no charge
on the condenser.

Let a rectangular pulse of cur-
rent I_a be applied at the input ter-
minals.   The condenser appears as
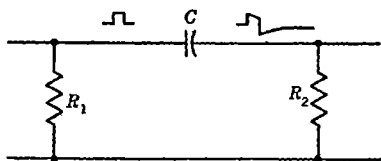
Fig. 3·1.—Coupling circuit.

a short circuit to the leading edge of the pulse.   The voltages across R_1 and
R_2 are equal and, if shunt capacity is neglected, rise instantaneously to

$$E = I_a \frac{R_1 R_2}{R_1 + R_2}. \tag{1}$$

During the flat portion of the pulse, the condenser C charges as
though through the resistances R_1 and R_2 in series.   If this flat portion
continued indefinitely, a steady-state condition would finally be reached
where all the current flowed through R_1, the voltage across it being
$I_a R_1$.   The voltage across R_2 would be zero, and the voltage across C
would therefore be $I_a R_1$.   The time constant determining the rate at
which this condition is approached is $(R_1 + R_2)C$.   Therefore the
voltage e_C across the condenser is

$$e_C = I_a R_1 (1 - e^{\overline{(R_1 + R_2)C}}). \tag{2}$$

If the pulse length τ is short in comparison to the time constant $(R_1 + R_2)C$, the voltage across the condenser at the end of the pulse is approximately

$$e_C \approx I_a R_1 \frac{\tau}{(R_1 + R_2)C}. \tag{3}$$

At the end of the pulse the current ceases suddenly, leaving this voltage on the condenser. The voltage due to the discharge current divides between R_1 and R_2, which are in series across C, in proportion to the resistances. Therefore the voltage across R_2 is

$$e_{R_2} = \frac{I_c R_1 \tau}{(R_1 + R_2)C} \frac{R_2}{(R_1 + R_2)} = \frac{I_a R_1 R_2}{R_1 + R_2} \frac{\tau}{(R_1 + R_2)C}. \tag{4}$$

Comparison of Eqs. (4) and (1) shows that the ratio of the overshoot to the signal, called the fractional overshoot, is

$$a = \frac{\tau}{(R_1 + R_2)C}. \tag{5}$$

This overshoot decays exponentially to zero with the time constant $(R_1 + R_2)C$. R_1 is usually a plate-load resistor connected between the plate and the B^+ voltage supply, and R_2 is a grid resistor connected between grid and ground.

It is standard practice to make R_2 as large as possible in order to reduce the fractional overshoot to a minimum without affecting the other characteristics of the amplifier, which are controlled by the plate-load resistor. However, this is a dangerous procedure if grid current flows. To see this, assume as a rough approximation that when the grid is driven positive, it acts to cause a short circuit between grid and ground but that when it is driven negative, it acts as an open circuit. During the pulse, therefore, R_2 is zero, and Eq. (3) becomes

$$e_C \approx \frac{I_a \tau}{C}. \tag{6}$$

At the end of the pulse, the grid circuit is open, and the resistance across the terminals of R_2 regains its original value. Thus the division of the condenser voltage takes place across the two resistors, making the over-shoot amplitude

$$e_{R_2} = \frac{I_a R_2 \tau}{C(R_1 + R_2)}, \tag{7}$$

and the fractional overshoot is

$$a = \frac{\tau}{CR_1}. \tag{8}$$

Thus the fractional overshoot is what it would be if the grid resistor were zero. However, the decay time constant of this overshoot is $(R_1 + R_2)C$. Therefore, the conventional design, with R_1 small and R_2 large, is worthless in this condition of grid-current flow. The fractional overshoot is large because the high grid resistor has no effect, but the recovery is very slow because the high grid resistor is then effective.

In this analysis it has been assumed that the plate-load resistor R_1 has been returned to ground instead of to B^+ as is actually the case in an amplifier. This fact has no effect on the overshoot calculation in that it merely adds a d-c component to the voltage across R_1.

FIG. 3·2.—Decoupling circuit.

Similar calculations can be made for the decoupling, screen bypass, and cathode bypass circuits shown in Figs. 3·2, 3·3, and 3·4. The results are shown in Table 3·1.

It should be noted that the overshoot in the decoupling circuit is of the same sign as the signal, instead of the opposite sign as in the case in coupling circuits. Its amplitude depends on the plate-load resistor
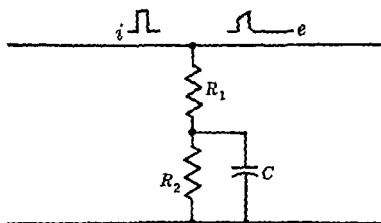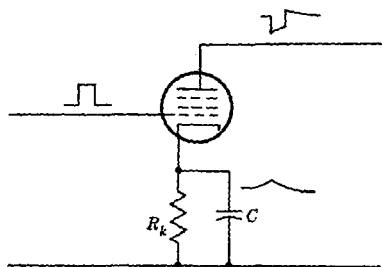
FIG. 3·3.—Cathode bypass circuit.
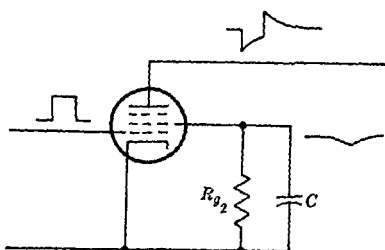
FIG. 3·4.—Screen bypass circuit.

and is independent of the decoupling resistor. The decay time constant is, however, proportional to the decoupling resistor.

The screen and cathode circuits behave like the coupling circuit, except that in both of these cases tube resistances shunt the resistor determining the time constant. Also, since the coupling condenser blocks direct current completely, whereas the screen and cathode bypass condensers merely degenerate the d-c, the fractional overshoot is multiplied by a factor of slightly less than unity. This factor becomes unity if the degeneration is complete. In the case of the screen bypass condenser this factor is complicated and involves the control character-

istics of the screen grid. Since its effect in most practical cases is negligible, it is omitted in Table 3·1. The result is, therefore, a somewhat conservative approximation. Similar simplifying assumptions have been made in the cathode bypass case.

TABLE 3·1.—FRACTIONAL OVERSHOOT AND RECOVERY-TIME CONSTANT FOR COMMON
CIRCUITS

Circuit	Fig.	Fractional overshoot, $\tau \ll T$	Recovery time constant T
Coupling...............	3·1	$\dfrac{\tau}{(R_1 + R_2)C} = \dfrac{\tau}{T}$	$(R_1 + R_2)C$
Decoupling............	3·2	$\dfrac{\tau}{R_1 C} = \dfrac{R_2}{R_1}\dfrac{\tau}{T}$	$R_2 C$
Cathode bypass........	3·3	$\dfrac{g_m \tau}{C}$	$\dfrac{C}{g_m + \dfrac{1}{R_k}}$
Screen bypass.........	3·4	$\dfrac{\tau}{T}$	$\dfrac{R_{g_2} r_{g_2}}{R_{g_2} + r_{g_2}} C$

All of these calculations are subject to the approximation $\tau \ll T$.

Cascaded Overshoots.—Since each condenser produces an overshoot, it it necessary not only to understand how a single overshoot is generated but also to ascertain the effect of several overshoot-producing circuits in cascade. Consider first the response of the simple coupling circuit of Fig. 3·1 to a signal that has already passed through such a circuit in a previous stage. The input pulse then has a slightly drooping top and is followed by an overshoot as shown in Fig. 3·5. The action during the pulse itself is very similar to the case of the flat pulse, except that the charge on the condenser is slightly less because of the decrease in the driving current during the pulse. This is a second-order effect and can generally be disregarded.
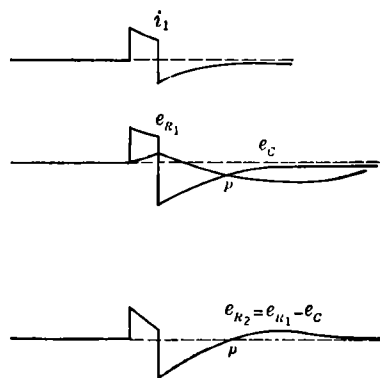
FIG. 3·5.—Waveform produced by coupling circuit whose input contains a pulse that has already passed through one such stage.

Thus, the voltage across the condenser at the end of the pulse can be calculated from Eq. (3).

At the trailing edge of the pulse, the input current drops not to zero, as in the previous case, but to the negative value of the overshoot produced by the previous stage. This current produces an output voltage that is the same in per cent as the input overshoot. To this voltage is

added the voltage across the condenser, and an overshoot is thus produced whose amplitude is the sum of the original overshoot and the one that would be produced by this circuit if there were no previous overshoot present.

The decay of this double overshoot is more complicated than that of the single case. The sudden change in sign of the signal at the end of the pulse is followed more slowly by the condenser, which discharges and then charges in the opposite direction, aiming always for the point at which the condenser voltage equals the input voltage and the output voltage is zero. The waveforms for this case are shown in Fig. 3·5. Since the input signal is decreasing exponentially while the condenser voltage is increasing, a point (point p in Fig. 3·5) is reached where the two curves cross and the input voltage crosses the baseline. The condenser now discharges, its voltage being always greater than the input voltage. This discharge action produces a secondary overshoot in which the voltage rises to a maximum in the same direction as the original signal and then drops exponentially to zero, as shown in Fig. 3·5.
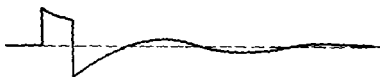
FIG. 3·6.—Waveform of pulse obtained from several successive coupling-circuit stages. There is one crossing of the baseline for each coupling circuit.

Each successive condenser adds another overshoot. Thus after a number of such stages a waveform similar to that shown in Fig. 3·6 is obtained.

Although the preceding analysis presents a good qualitative picture of the secondary overshoot, in order to obtain quantitative information about its amplitude and duration it is necessary to go to a more rigorous mathematical treatment. Consider then the differential equation for the circuit of Fig. 3·1:

$$\frac{de_2}{dt} + \frac{e_2}{T} = \frac{R_1 R_2}{(R_1 + R_2)} \frac{di_1}{dt}, \tag{9}$$

where e_2 is the output voltage, i_1 is the input current, and T is the time constant $(R_1 + R_2)C$. The solution of Eq. (9) is

$$e_2 = A e^{-\frac{t}{T}} + \frac{R_1 R_2}{(R_1 + R_2)} e^{-\frac{t}{T}} \int \frac{di_1}{dt} e^{\frac{t}{T}} dt, \tag{10}$$

where A is a constant of integration.

If i_1 is the result of the overshoots produced in m previous stages, it can be considered as the sum of exponential terms, each term representing an overshoot. Thus

$$i_1 = \sum_{j=1}^{m} I_j e^{-\frac{t}{T_j}}, \tag{11}$$

where T_j is the time constant of the jth circuit, all the T_j's being assumed, for the moment, to be unequal. Substituting Eq. (11) in Eq. (10), the result is

$$e_2 = A e^{-\frac{t}{T}} + \sum_{j=1}^{m} -\frac{R_1 R_2}{(R_1 + R_2)} \frac{I_j}{T_j} e^{-\frac{t}{T}} \int e^{\frac{t}{T} - \frac{t}{T_i}} dt$$

$$= A e^{-\frac{t}{T}} + \sum_{j=1}^{m} \frac{E_j}{1 - \frac{T_j}{T}} e^{-\frac{t}{T_i}}, \qquad (T_j \neq T), \tag{12}$$

where $E_j = R_1 R_2 / (R_1 + R_2) I_j$ is the output voltage that would be produced by an input signal I_j, if the effect of C were neglected.

If any of the T_j's are equal to one another, there are terms in the expression for i_1 of the form $(t/T_j)^k e^{-\frac{t}{T_i}}$, which also appear in the expression for e_2. If any of the T_j's are equal to T, terms of the form $(t/T)^k e^{-\frac{t}{T}}$ appear in the expression for e_2. The exact values of these terms can be determined by substitution in Eq. (10). Because of the complicated algebra involved in this substitution, only the results for the general case are given here.

Consider the general case of an amplifier in which, preceding the circuit being considered, there are r circuits of time constant T and s circuits of time constant T_j, where there are m different values of the time constant T_j and n_j circuits for each value T_j. Then $s = \sum_{j=1}^{m} n_j$, and the total number of circuits preceding the one considered is therefore $r + \sum_{j=1}^{m} n_j$. Then

$$i_1 = \sum_{k=0}^{l-1} I_k \left(\frac{t}{T}\right)^k e^{-\frac{t}{T}} + \sum_{j=1}^{m} \sum_{k=0}^{n_j-1} I_{jk} \left(\frac{t}{T_j}\right)^k e^{-\frac{t}{T_i}} \tag{a}$$

and

$$e_2 = A e^{-\frac{t}{T}} + \sum_{k=0}^{l-1} E_k \left(\frac{t}{T}\right)^k e^{-\frac{t}{T}} \tag{13}$$

$$+ \sum_{j=1}^{m} \sum_{k=0}^{n_j-1} E_{kj} e^{-\frac{t}{T_i}} \left[\frac{\left(\frac{t}{T_j}\right)^k}{1 - \frac{T}{T_j}} - \frac{1}{\left(1 - \frac{T_j}{T}\right)} \sum_{i=1}^{k} \frac{k! \left(\frac{T}{T_j}\right)^{k-i}}{(k-i)! \left(1 - \frac{T}{T_j}\right)^{i-1}} \right], \quad (b)$$

where E_k and E_{ki} are the products of I_k and I_{ki} by $R_1 R_2/(R_1 + R_2)$.

Equation (13) is quite complicated for use in amplifier design.   If the presence of equal time constants is disregarded in an amplifier and Eq. (12) is used instead, an error is introduced which is, in general, not sufficient to alter the order of magnitude of the overshoot.   Hence Eq. (12) is generally used for all rough calculations.

Examination of Eq. (12) shows that each overshoot entering the input terminals appears at the output terminals modified in amplitude by the factor $1/1 - (T_j/T)$, but with the exponent unchanged.   There is added a term corresponding to the overshoot that would be produced by this particular coupling circuit in the absence of other overshoots. The amplitude of this term is dependent upon the initial value of the overshoot.   Since this has been shown by the previous analysis to be the sum of the individual overshoots each acting independently, the value of A can easily be obtained.

Using this analysis it is simple to trace each overshoot through an amplifier.   At each successive coupling circuit an overshoot of given time constant is changed in amplitude, but not time constant, by a factor $1 - T_j/T$.   If the time constant T of a coupling circuit is much greater than the time constant T_j of the overshoot, $1 - T_j/T$ is close to unity, and the overshoot amplitude is unchanged by this circuit.   This conclusion is reasonable, since the charge on a condenser can not change appreciably during a time short in comparison to the time constant of the circuit.   If on the other hand T is much less than T_j, the factor $1 - T_j/T$ is very nearly equal to $-T_j/T$.   A long overshoot is therefore reduced in amplitude by a coupling network having a short time constant by a factor equal to the ratio of the time constants involved. Furthermore, the sign is changed, making the long time constant overshoot appear on the other side of the baseline.   It is this change of sign which gives rise to secondary and higher-order overshoots.

If T is nearly equal to T_j, very large values are obtained for the overshoot amplitude.   However, in order to satisfy the initial conditions, A is also large and opposite in sign to the other overshoot amplitude. Since T is very nearly equal to T_j, the exponents of these terms are very nearly equal for small values of t/T.   Thus the two large terms very nearly cancel.   To avoid taking the difference between two nearly equal large terms a good approximation can be made for small values of t/T by assuming that T is equal to T_j.

Nonlinear Effects.—Although ordinary linear-circuit analysis breaks down in cases where amplifier stages are overdriven, the theory of overshoots as described here can still be applied.   It is merely necessary to consider the effects of limiting on each signal wherever limiting occurs. The charging of the coupling condenser by grid current has already been considered.

The first effect of limiting is to clip the top of the pulse, as shown in Fig. 3·7, reducing the amplitude considerably. When limiting occurs, the term fractional overshoot, as used in the preceding analysis, becomes ambiguous. This term was defined in reference to the signal amplitude at the point where the overshoot is produced. Thus the actual amplitude of an overshoot of 1 per cent, produced before limiting, will be greater at the amplifier output terminal than a 10 per cent overshoot produced after limiting has reduced the signal amplitude by a factor greater than 10. Therefore, although the use of fractional overshoot is convenient for the calculation of a single circuit, it is better to convert to absolute values in volts before combining the effects of several circuits. If this is done, there is no difficulty in using the methods previously outlined.

FIG. 3·7.—Limiting.

The shape of the top of the pulse depends upon whether the limiting is caused by driving the grid of a tube beyond cutoff or into the positive region. If the tube is cut off, the top of the pulse appearing on the plate must be flat, as shown in Fig. 3·7, because changes in the grid voltage below cutoff have no effect on the plate current. However, if limiting is produced by driving the grid positive, an increase in the grid voltage still increases the plate current, even though considerable power is required from the driving source. The top of a pulse limited by grid current still has a nonzero slope, although it is very much less than the slope of the input pulse.

As long as the pulse alone, and not the overshoot, is clipped, Eq. (13) holds without reservations. The limiting affects only the value of the constant A, which is chosen to meet the initial conditions as modified by the clipping of the pulse. In most cases, the overshoot amplitudes are kept sufficiently low so that they are not clipped. If as in Fig. 3·8 this is not true, Eq. (13) still applies, but allowance must be made for the discontinuity. The limited portion of the overshoot (between points

FIG. 3·8.—Limiting of an overshoot.

a and b in Fig. 3·8) can be considered as a single overshoot, having an infinite time constant. The constant A of Eq. (13) can be found from the initial conditions (Point a). The portion of the overshoot following

Point b is the same as if there were no limiting, except that a new value of the constant A, determined by the conditions at Point b, is again required.

3·3. Circuit Design for Minimum Overshoot.—Since the magnitude of the overshoot produced by a given circuit is inversely proportional to the time constant of the circuit, it is theoretically possible to make all overshoots negligible simply by making all the time constants sufficiently large. Practically, however, there is a limit to the maximum usable values of resistance and capacitance, and it is generally impossible to design a high-gain amplifier with no overshoots at all. Furthermore, even if it were possible, some form of low-frequency-rejection filter to eliminate microphonics would still be needed. This filter would necessarily introduce an overshoot, since it would effectively introduce a short time constant.

Although it is impossible to eliminate overshoots completely, it is advantageous wherever possible to design circuits so that the overshoots produced by them are negligible at the output terminals of the receiver. The magnitude of the overshoot following a very strong signal must therefore be negligibly less than the magnitude of a very weak signal, or else the fractional overshoot must be negligibly less than the ratio of the weakest signal to the strongest.

Because of limiting, the dynamic range of signals actually present at any given point in the amplifier varies from point to point in the amplifier. At each point of limiting, all signals above a certain amplitude are clipped. The dynamic range beyond this point is thus reduced, and therefore the allowable fractional overshoot is increased. Thus the first few circuits have the most severe requirements on the time constants in order to eliminate overshoots. The nearer a given circuit is to the output end the shorter its time constant can be without introducing appreciable overshoot.

If the dynamic range of an amplifier is one million, the fractional overshoot produced by any circuit preceding the first point of limiting must be less than one-millionth if the overshoot is to be negligible at the output end. This condition can generally be achieved only if there is a short time constant later on in the amplifier. If all time constants were long, the fractional overshoot produced by the first circuit would be approximately equal to the ratio of the pulse length to the time constant. To make this negligibly less than 10^{-6}, a time constant of the order of seconds would be required for a 1-μsec pulse. It is impossible to obtain this value of time constant in most practical cases. As discussed in Sec. 3·2, use of a short time constant T_2 later in the amplifier reduces the magnitude of the overshoot of time constant T_1 due to the first circuit by $1 - T_1/T_2$. In most practical cases this is approximately the ratio

of the time constants. The fractional overshoot for the overshoot of time constant T_1 is then given by

$$a = \frac{\tau}{T_1} \frac{T_2}{T_1} = \frac{\tau T_2}{T_1^2}. \tag{14}$$

Solving Eq. (14) for T_1 yields

$$T_1 = \sqrt{\frac{\tau T_2}{a}}. \tag{15}$$

Assuming that the overshoot must be one-tenth of the weakest signal in order to be negligible, a is 10^{-7} for a dynamic range of one million. Then, assuming as typical values $\tau = 2$ μsec and $T_2 = 5$ μsec, T_1 is found to be equal to 10,000 μsec, which is not an unreasonably high value for some circuits.

Overshoots in Conventional RC-circuits.—The expressions for the fractional overshoots of the common RC coupling, decoupling, cathode-bypass, and screen-bypass circuits are given in Table 3·1.

For coupling circuits, the upper limit on grid resistors for most tubes is about 1 megohm. If no grid current is drawn, a condenser of 0.01 μf gives a time constant of 10,000 μsec. If the value of the resistor is closer to 0.1 megohm, the condenser required is 0.1 μf. These are reasonable values for use in most amplifiers, except those where extremely small size is required. However, the use of a "postage-stamp" size 0.01-μf condenser with a 1-megohm grid resistor allows a fairly small amplifier to be built.

If grid current is drawn, the grid resistor is shunted by the input conductance of the tube and no longer affects the time constant, which is now dependent upon the plate-load resistor and coupling condenser only. Since the plate-load resistor must carry the plate current of the tube, values higher than 50,000 ohms are not generally used. Although some special applications allow the use of a higher resistor, triodes are used in these cases for certain reasons discussed below (see Sec. 3·5). The plate resistance of the triode, therefore, effectively shunts the load resistor, and the parallel combination is generally less than 50,000 ohms. The coupling condenser required for a time constant of 10,000 μsec is therefore at least 0.2 μf and is generally higher. Elimination of overshoots therefore becomes impractical with simple RC-coupling if grid current is drawn. Note that the use later in the amplifier of several small time constants instead of only one introduces another factor in Eq. (14) reducing the fractional overshoot a. This procedure allows the use of a time constant smaller than 10,000 μsec in the first circuit.

Decoupling circuits can generally be designed for negligible overshoot. The plate load and decoupling resistors can commonly be made

as high as 10,000 ohms. This value of resistance requires a capacity of 1 μf to get a time constant of 10,000 μsec. For a large amplifier this capacity is allowable, but for small amplifiers some other scheme must be used. Generally several small time constants later in the amplifier alleviate the problem.

Screen-bypass circuits require components of the same order of magnitude as decoupling circuits and are therefore subject to the same limitations. For this reason, it is often advantageous, especially where small size is important, to use triodes and eliminate the screen-bypass circuit.

Table 3·1 shows that the resistance determining the time constant of the cathode circuit is equal to the g_m of the tube. Enormous cathode-bypass condensers are required to obtain a time constant of 10,000 μsec; e.g., a tube having a g_m of 3000 μmhos requires a cathode-bypass condenser of 30 μf. For this reason it is common practice to leave cathode resistors unbypassed. If the cathode resistor is kept small, the loss in gain due to degeneration is not appreciable. In some cases, the cathode resistor is bypassed by a small condenser in order to improve the high-frequency response. In this case the time constant is so short that the overshoot is negligible and is obscured by the slow trailing edge of the pulse caused by poor high-frequency response.

Secondary Overshoots in Circuits with Two Short Time Constants.— When there are two short time constants, it can be shown by considering Eq. (12) that there will, in general, be a secondary overshoot. The factor $1 - T_i/T$ is negative for the term involving the larger of the two time constants. After a sufficiently long time the term having the shorter time constant becomes negligible with respect to this term. Since its sign is negative, it represents an overshoot in the opposite direction from the original overshoot and therefore a secondary overshoot. The time at which the crossing of the baseline occurs is always equal to some value between those of the two time constants and

FIG. 3·9.—Elimination of secondary overshoot by drawing grid current.

can be found by equating the two terms. Hence, the secondary overshoot can be considered to be of negligible magnitude only if the longer time-constant overshoot is negligibly small.

*Elimination of Secondary Overshoot by Means of Grid Current.—*By use of grid current, the secondary overshoot normally produced by two short time constants can be avoided. As shown in Fig. 3·9, the charging of the coupling condenser by grid current produces an overshoot so much

larger than that in the ordinary case that it completely obliterates the secondary overshoot normally present. The recovery in this case is, of course, poorer than in the absence of grid current; however, in some applications elimination of the secondary overshoot is more desirable than quick recovery.

This effect can very easily be considered quantitatively by using different values for the charging time constant and the discharging time constant of the circuit drawing grid current.

Let T_{2a} be the time constant of charge of this circuit and T_{2b} the time constant of recovery. The expression for the instantaneous overshoot voltage, given by Eq. (12), involves T_{2b} only, since there is no grid current during the overshoot. (This assumption supposes no secondary overshoot, which would drive the tube into the grid-current region again.) Let T_1 be the other short time constant in the amplifier. Then, substituting in Eq. (12),

$$e_2 = Ae^{-\frac{t}{T_{2b}}} + \frac{\tau e^{-\frac{t}{T_1}}}{T_1\left(1 - \frac{T_1}{T_{2b}}\right)}.$$ (16)

A is determined by the initial value of the overshoot. Since the initial value is determined by the charging of the coupling condenser, T_{2b} is not involved. Thus the fractional overshoot is the sum of those overshoots produced by T_{2a} and T_1 acting independently, or

$$a = \frac{\tau}{T_1} + \frac{\tau}{T_{2a}}.$$ (17)

Substituting this equation into Eq. (16) at $t = 0$ and solving for A,

$$A = \tau \frac{T_{2b} - T_1 - T_{2a}}{T_{2a}(T_{2b} - T_1)}.$$ (18)

Substituting this into Eq. (16),

$$e_2 = \frac{\tau_1(T_{2b} - T_1 - T_{2a})}{T_{2a}(T_{2b} - T_1)}\, e^{-\frac{t}{T_{2b}}} + \frac{\tau e^{-\frac{t}{T_1}}}{T_1\left(1 - \frac{\tau}{T_{2b}}\right)}.$$ (19)

If $T_{2b} \gtreqless T_1 + T_{2a}$, both terms have the same sign, and there is no secondary overshoot. Optimum recovery occurs when $T_{2b} = T_1 + T_{2a}$, and the first term drops out. In this case, the recovery is the same as if the second coupling circuit were absent.

Choke-coupled Circuit.—It has been shown that when grid current is drawn, the plate-load resistor must be made as high as possible to reduce the charging time constant of the coupling condenser. Since this

produces a large voltage drop across the resistor, considerable power is wasted. A high supply voltage and a high-wattage resistor are necessary. To avoid these complications, a choke can be used in place of the plate-load resistor. The choke acts as a high impedance during the charging of the coupling condenser but has a negligible d-c voltage drop. The circuit is shown in Fig. 3·10.

In a circuit of this type, the inductance of the choke may resonate either with the stray shunt capacity C_s or with the coupling condenser C_c, producing a train of damped oscillations following each pulse. This train of oscillations is effectively a series of overshoots and is therefore highly undesirable. To avoid this effect, both circuits must be damped by the grid resistor so that the transient following the pulse is not more than a single or at the most a double overshoot. Since the value of the grid resistor is determined by the amount of gain

FIG. 3·10.—Choke-coupling circuit.

desired, the inductance of the choke is determined by the condition for greater than critical damping. Thus

$$\left. \begin{array}{l} L \leqq \dfrac{C_c R^2}{4}, \\[2mm] L \geqq 4C_s R^2. \end{array} \right\} \tag{20}$$

The overshoot produced by the charging of the coupling condenser can be analyzed in the same manner as that for the coupling circuit of Sec. 3·2. At the start of the pulse there is no current flowing in the choke. Therefore all the current must be charging the coupling condenser. If the pulse length is short compared with the resonant period of the tuned circuit formed by the choke and the coupling condenser, the current flow through the inductance is negligible and the approximation can be made that all the current flows through the condenser during the pulse. The voltage across the condenser at the end of the pulse is then

$$e_{c_s} = \frac{I_a \tau}{C_c}. \tag{21}$$

The discharging of the condenser through the choke and the grid resistor can be broken up into two parts: (1) the building up of the current in the choke and (2) the discharging of the condenser. This simplification is possible because the time constant of the inductance circuit is very short compared with the discharge time constant of the condenser. The condenser can therefore be considered as a source of constant voltage while

the current builds up in the inductance; the inductance can be considered as a short circuit during the discharge of the condenser.

At the instant of the trailing edge of the pulse, there is no current flow because of the inductance of the choke. The full voltage of the condenser appears across the choke. The current then increases exponentially, approaching a steady-state condition where the full voltage across the condenser appears across the resistor. The time constant is L/R. The condenser discharges with a time constant RC as in the conventional RC-coupled circuit. This is shown in Fig. 3·11. Thus the overshoot differs from that produced by RC-coupling only in the slow rise time caused by the charging of the inductance. The magnitude of the overshoot is equal to the original voltage across the condenser, which is given by Eq. (21). Since the magnitude of the signal is

$$e_s = I_a R_g, \tag{22}$$

the fractional overshoot is

$$a = \frac{\tau}{R_g C_c}. \tag{23}$$

Typical values for a choke-coupled circuit are $R_g = 27{,}000$ ohms, $L = 85$ mh, and $C_c = 0.02$ μf.

The time constant of the inductive circuits is then 3.15 μsec, which is small compared with 540 μsec, the time constant of discharge of the condenser. This value is far from the 10,000 μsec needed for elimination of the overshoot. Choke coupling is therefore not used in applications where overshoots are not allowable.

"Smearer" Circuit.—The type of circuit commonly called a "smearer" is similar in principle to the method of using grid current to eliminate a secondary overshoot. This circuit makes use of the sharp drop in plate resistance that occurs when a tube is driven positively from a point where the plate current is low. Figure 3·12 shows a smearer. The tube is operated with a very high plate-load resistor shunted by a condenser C_s. This condenser can be an actual capacitor, but in many cases the stray capacity is sufficient.
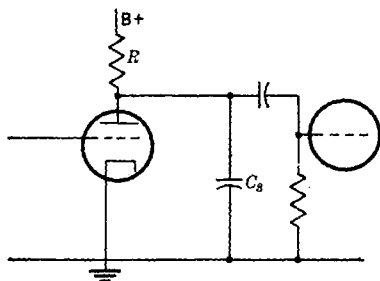
Fig. 3·11.—Overshoot produced by choke-coupling circuit.

Fig. 3·12.—Smearer circuit.

A positive signal on the grid of this tube drives it into a highly conducting region where the plate resistance is low. The condenser C_s is therefore charged very rapidly, as the time constant is short. At the end of the pulse, the negative overshoot on the grid cuts the tube off, thereby raising the plate resistance of the tube to infinity. The condenser C_s must therefore discharge through the high plate-load resistor R. Since this time constant is long, the negative signal at the plate comes back exponentially to the baseline. If this time constant is sufficiently long compared with the time constants of the overshoots, the overshoots are completely eliminated, at the cost, however, of stretching the pulse considerably and obscuring any weak signals that may be present during the decay of the pulse.

Since the decay of the pulse in this type of circuit is exponential, the previous analysis of cascaded overshoots applies to this circuit as well, and an analytic treatment of it similar to that for the double short-time-constant circuit is possible.

Diode Clippers.—An obvious method of eliminating overshoots before they have an opportunity to be amplified and to produce secondary overshoots is to clip them by means of diodes as they occur. A diode, introduced into the circuit after each overshoot is generated, should pass the signal but reject the overshoot, which is in the opposite direction. Unfortunately, this simple procedure is not practical for several reasons:

1. Since there is no perfect diode having a sharp discontinuity in its resistance, diodes act as ideal diodes only at high levels. Thus, if there are two overshoot-producing circuits in the early stages of the amplifier, diodes are of little use, since the secondary overshoot produced at low levels is in the same direction as the signal and cannot be eliminated at high levels by a diode.

2. Because the diode does not act as a perfect short circuit in one direction and a perfect open circuit in the other, overshoots are not completely clipped by the diode but are merely attenuated. Thus a strong overshoot may come through a diode in sufficient magnitude to cause trouble later on.

3. The impedance of a diode in the backward direction, particularly at low levels, is far from infinite. Thus a diode loads the circuit in which it is introduced. This is particularly serious when the stage gain is high and all extraneous loading must be avoided.

4. In clipping an overshoot, a diode also obliterates all weak signals occurring during the overshoot. This effect is illustrated in Fig. 3·13.

Despite these disadvantages, diode clippers can be used advantageously in numerous applications as long as their limitations are kept in

mind.   The new germanium crystals,[1] which are considerably better than most common vacuum tubes for this purpose, may make possible the use of diode clippers in applications for which a vacuum-tube diode could not be considered.

There are several types of circuits in which diodes can be used.   The series diode, shown in Fig. 3·14a, presents a low impedance to the signal and a high impedance to the overshoot, thus attenuating it by a factor equal to the ratio of the diode load resistance to the diode back resistance.   Since the diode forward resistance is not zero, there is some attenuation of the signal as well.
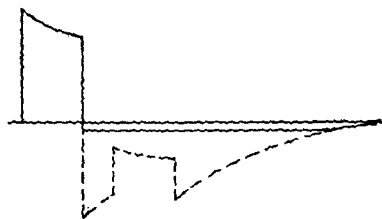
FIG. 3·13.—Loss of weak signal when clipping overshoot.

The shunt diode, shown in Fig. 3·14b, presents a high impedance to the signal and shunts the overshoot with a low resistance.   The reduction of
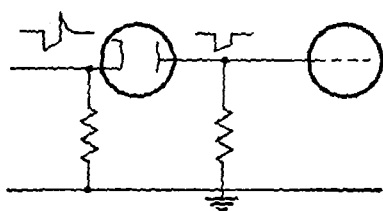
FIG. 3·14a.—Series diode.
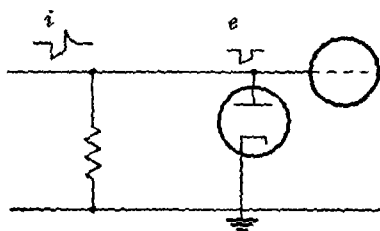
FIG. 3·14b.—Shunt diode.

overshoots resulting from use of a shunt diode is a function of the resistance of the circuit in which it is inserted and is equal to the ratio of the circuit resistance to the diode back resistance.   Here, again, there is some attenuation of the signal itself, this time because the back resistance is finite.

In using diodes, care must be taken not to produce unintentional "smearing" of the signal by the slow discharge of stray capacitances through the back resistance of the

FIG. 3·15.—Delay-line coupling circuit.

diode.   In the series diode, shown in Fig. 3·14a, smearing can happen unless the diode resistor is small.

----

[1] See Vol. 17, Chap. 5, of this series.

Diodes used directly in the circuit producing the overshoot do more than merely clip the overshoot. If a shunt diode, for example, is used in place of a grid resistor, the coupling condenser charges through a high resistance. The charging time constant is therefore long, producing a small overshoot. Moreover, the condenser discharges through a low

resistance, thus recovering much more rapidly than would normally be expected. This same effect can be obtained without an auxiliary diode in a stage where the grid is driven negative, and any overshoot produces grid current. The grid-cathode circuit here acts as a diode, but only when there is no overshoot present before this stage. Any earlier overshoot is not shortened

FIG. 3·16.—Response of delay-line circuit to pulse.
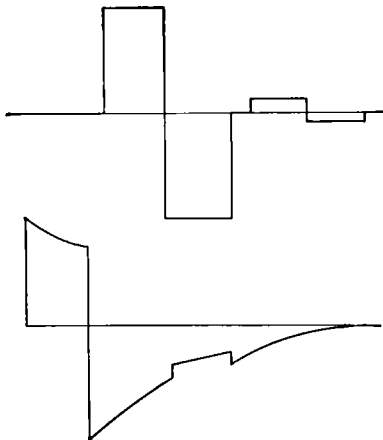
FIG. 3·17.—Advantage of rectangular overshoot.

by this short time constant; instead it charges the coupling condenser in the opposite direction through the low-impedance path of grid circuit or diode and produces a serious secondary overshoot.

Delay-line Grid Circuit.—If a delay line is used in the grid circuit of an amplifier, as shown in Fig. 3·15, a pulse at the input terminal travels to the end of the line and is reflected back in opposite phase. This reflected pulse in turn produces a second pulse, equal and opposite in sign to the original, and delayed by a time which is twice the length of the line, as shown in Fig. 3·16. This reflected signal can be considered as an overshoot, which, instead of being exponential in shape, is rectangular. If the duration of an overshoot is defined as the time required for it to decay to a given absolute level, the rectangular overshoot has the great advantage of being constant in duration, regardless of signal strength, whereas the duration of the usual overshoot with exponential decay increases with signal strength. Thus the weak signals following a rectangular overshoot are preserved, whereas those following an exponential can be lost, as shown in Fig. 3·17.

The effect of the delay line upon a long overshoot is illustrated in Fig. 3·18. The delayed signal is subtracted from the undelayed signal to give the resultant. The difference in amplitude between the delayed and the

undelayed signals is the amount by which the undelayed signal has decayed during the delay and is equal to the amplitude of the overshoot:

$$a = -a_0(1 - e^{-\frac{t_d}{T}}) \approx a_0 \frac{t_d}{T}, \tag{24}$$

where a_0 is the fractional overshoot of the original signal, t_d is the delay time, and T is the time constant of the overshoot. This relation is very much like that for the reduction of overshoots by the factor of $1 - T_1/T_2$ by circuits having short time constants. Thus, the delay-line circuit behaves with respect to previous overshoots like a circuit having a time constant equal to the delay time.

The main disadvantage of the delay-line circuit is that the characteristic impedances of common delay lines are considerably lower than the load impedances generally desired in a high-gain amplifier. The use of this type of circuit, therefore, causes a loss of gain in all amplifiers except wide-band amplifiers, which use low load resistors. Also, there is a small loss in signal-to-noise ratio because the delayed noise adds to the undelayed noise, increasing the noise power.

This analysis has assumed ideal behavior of the line circuit. Actually there is attenuation in the line, and the reflected signal is not so large as the original. This effect modifies Eq. (24), replacing the 1 by a number

Fig. 3·18.—Effect of delay line upon signal with overshoot (delay time equal to pulse width).

less than unity. Note that if the attenuation of the line is correctly adjusted, this number can be made equal to $e^{-\frac{t_d}{T}}$, in which case the overshoot completely cancels out. This adjustment is generally not practical for large-scale production but is useful in the laboratory.

If the line is not terminated in its characteristic impedance, there are multiple reflections, producing multiple overshoots. Even if these over-

shoots are small, they are important for a sufficiently large signal.   The inherent advantage of the rectangular overshoot over the exponential is then lost.   If the terminating impedance is less than the characteristic impedance, there is a reversal of sign on reflection, and a signal of the
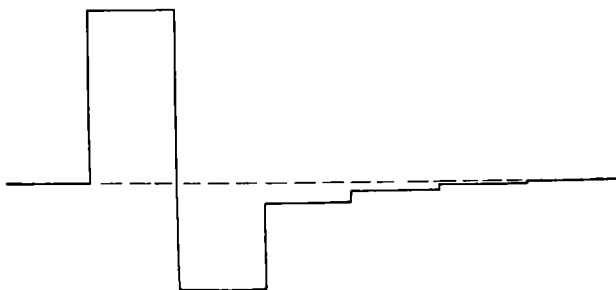
FIG. 3·19a.—Pulse response of delay-line circuit terminated in too low an impedance.

same sign as the original signal is sent down the line.   All the reflections are therefore of the same sign as the first, and the result is a lengthening of the overshoot in steps, as shown roughly in Fig. 3·19a.   If the termination is higher than the characteristic impedance, there is no sign reversal, and the signal is reversed in sign once each trip, thus producing reflections
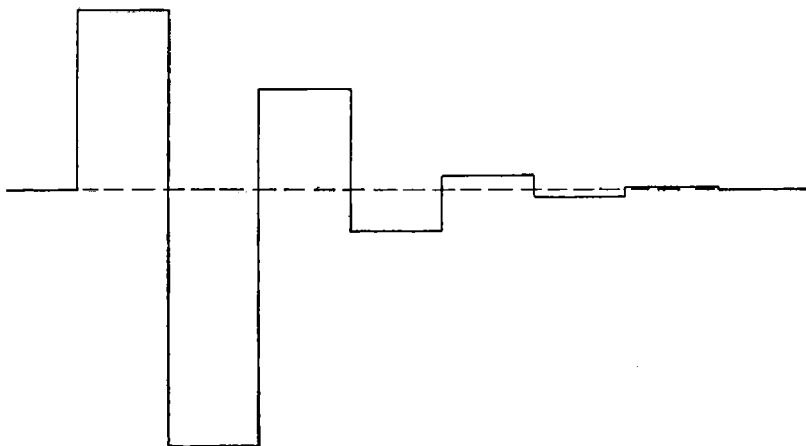
FIG. 3·19b.—Pulse response of delay-line circuit terminated in too high an impedance.

of successively alternating sign.   This produces multiple overshoots, such as are roughly shown in Fig. 3·19b.

Exact termination is difficult, especially with commercial tolerances on resistors.   The maximum per cent overshoot that can be expected is roughly the tolerance of the resistor used.   The difficulty of termination

is accentuated by any nonlinearity in the grid circuit, such as would occur with a positive signal strong enough to draw grid current.   For this reason a certain degree of mismatch is to be expected; and since it is preferable to have a lengthened single overshoot rather than multiple overshoots, it is desirable to design the termination slightly lower than the characteristic impedance.   This still gives a result sufficiently better than simple RC-coupling to justify its use for some applications.

FIG. 3·20.—Delay-line circuit using crystal clippers.

The combination of germanium-crystal clippers and delay lines is very effective.   A crystal in series with the termination prevents loading of the original signal by the termination but still acts as a match for the reflection, which is opposite in sign.   The nonlinear effect of the grid circuit on the termination can be eliminated by use of a series crystal, which allows the original signal to pass but acts as an open circuit for the reflected signal.   This crystal also automatically clips the overshoot.   A circuit of this kind is shown in Fig. 3·20.

Inverse-feedback Pairs.—The inverse-feedback-pair circuit described in Chap. 6 can be used in a high-gain video amplifier to improve the overshoots.   Because of two effects, considerably better performance can be obtained from a properly designed feedback pair: (1) If a feedback-pair amplifier is designed so that negative signals appear at the input terminals of the pair, the constants of the pair may be adjusted so that without excessive loss of gain the positive signal appearing at the second grid is limited to a value that does not cause grid current to flow.   (2) Feedback also has the effect of reducing the overshoot of the interstage coupling circuit by a considerable factor from what it would be in the absence of feedback.

It is shown in Chap. 6 that the relative values of the load resistors in a feedback pair can be adjusted without changing the over-all response of the pair, provided that the feedback resistor is properly varied.   This adjustment has the effect of varying the relative gains of the two stages without changing the over-all response.   Overloading of the second grid can therefore be minimized by putting as much gain as possible in the second stage and as little as possible in the first.   The maximum signal appearing at the second grid is then much less than in a circuit with the same over-all gain and no feedback, and the overshoot on the maximum signal is not so great.   In some cases, it may even be possible to design

the pair so that no grid current is drawn at all; this is a considerable improvement.

The theory of Chap. 6 cannot be applied to this problem without reservations, because large signals drive the tubes over a sufficiently wide range to make the assumptions of linearity invalid. Some theory using a nonlinear characteristic or making certain approximations can perhaps be worked out, but this has not yet been done. The procedure for designing this type of amplifier has been one of cut-and-try, using the tube-characteristic curves given in the tube manuals.

The effect of feedback upon overshoots can be calculated by using the same fundamental approach as was used in Sec. 3·2. If the time constant of the circuit is assumed to be long in comparison with the pulse length, it is possible to calculate from the charging current and the pulse length the volt-

FIG. 3·21.—Inverse feedback pair.

age appearing across the coupling condenser at the end of the pulse. Thus, for the circuit of Fig. 3·21, the voltage across the condenser at the end of the pulse is

$$e_C = \frac{i\tau}{C} = \frac{E_g}{R_g}\frac{\tau}{C},\tag{25}$$

where E_g is the amplitude of the signal pulse appearing at the grid of the second stage. The portion of this that appears across the grid resistor at the end of the pulse can be shown to be

$$\frac{e_g}{e_C} = \frac{R_g}{R_g + \dfrac{(R_2 + R_{12})R_1}{R_1 + R_2 + R_{12}} + \dfrac{g_m R_1 R_2 R_g}{R_1 + R_2 + R_{12}}}.\tag{26}$$

Substituting Eq. (25) in Eq. (26) one finds the fractional overshoot to be

$$a = \frac{e_g}{E_g} = \frac{e_g}{e_C}\frac{e_C}{E_g} = \frac{\tau}{C\left[R_g + \dfrac{g_m R_1 R_2 R_g}{R_1 + R_2 + R_{12}} + \dfrac{(R_2 R_{12})R_1}{R_1 + R_2 + R_{12}}\right]}\tag{27}$$

$$= \frac{\tau}{CR_g\left(\dfrac{g_m R_1 R_2}{R_1 + R_2 + R_{12}}\right)\left(1 + \dfrac{1 + \dfrac{R_{12}}{R_2}}{g_m R_g} + \dfrac{R_1 + R_2 + R_{12}}{g_m R_1 R_2}\right)}$$

$$\approx \frac{\tau}{CR_g\,\dfrac{g_m R_1 R_d}{R_1 + R_2 + R_{12}}}.$$

Thus for an amplifier where no grid current is drawn, so that R_g can safely be made large, the fractional overshoot is reduced by a factor of $g_m R_1 R_2 / (R_1 + R_2 + R_{12})$ over that of an amplifier without feedback. This effect is very useful in the early stages of high-gain amplifiers where huge capacitors would otherwise be necessary to eliminate overshoots.

As an example, consider the following typical values for such a circuit:

$$R_1 = R_2 = 10,000 \quad \text{ohms.}$$
$$R_{12} = 50,000 \quad \text{ohms.}$$
$$g_m = 3000 \quad \mu\text{mhos.}$$

For these values, the fractional overshoot is improved by the appreciable factor of 6.

The recovery time constant, as might be expected, is given by the denominator of Eq. (27);

$$T = C R_g \left(\frac{g_m R_1 R_{12}}{R_1 + R_2 + R_{12}} \right) \left(1 + \frac{1 + \dfrac{R_{12}}{R_2}}{g_m R_g} + \frac{R_1 + R_2 + R_{12}}{g_m R_1 R_2} \right)$$
$$\approx C R_g \left(\frac{g_m R_1 R_2}{R_1 + R_2 + R_{12}} \right). \quad (28)$$

Direct-coupled Inverse-feedback Pair.—The obvious method for elimi-

FIG. 3·22.—Direct-coupled inverse-feedback pair.

nation of the problem of charging of the coupling condenser by grid current is the elimination of the coupling condenser. Any conventional direct-coupled circuit can be used, carrying along with it the usual disadvantage of direct-coupled circuits, in particular, the need for a more complicated power supply. Use of an inverse-feedback pair in conjunction with direct coupling, as shown in Fig. 3·22, allows the use of an ordinary power supply of the type used for other video amplifiers.[1]

In this circuit, the first tube of the pair is generally operated at a low plate voltage, at which the tube still performs satisfactorily as an amplifier. The grid of the second tube is directly coupled to the first plate, and the large second cathode resistor furnishes sufficient cathode bias to bring the cathode potential above the grid potential by the desired

[1] The unique advantages of the direct-coupled inverse-feedback pair were demonstrated by R. J. Grambsch at the Radiation Laboratory.

amount. Because of the two types of d-c degeneration present in this circuit, it is unusually stable with respect to power-supply and tube variations. The unique feature that enables its use as a practical amplifier is the effect of the feedback in reducing overshoot caused by the cathode bypass condenser, thus making possible bypassing of the large cathode resistor, which would otherwise cause a high loss in gain due to degeneration.

By an analysis similar to that for the inverse-feedback pair, it can be shown that feedback reduces the cathode-circuit overshoot by the same factor as it reduces the coupling-circuit overshoot in the capacity-coupled feedback pair. This factor is given by Eq. (26), with R_g taken as infinite. Therefore, the factor by which the fractional overshoot is reduced is $1 + (g_m R_1 R_2 / R_1 + R_2 + R_{12})$.

Thus it is possible to design a direct-coupled feedback pair in which the cathode of the second stage is bypassed. This has the advantage over all other types of circuits considered here in that it eliminates the problem of grid current. Note, however, that it is restricted to the case where negative signals only are impressed upon the first grid of the pair; for amplifiers that must handle both positive and negative signals this type of circuit is useless.

Effect of Shunt Capacitance.—The shunt capacitance to ground always present in coupling circuits has been neglected thus far because, in common *RC*-coupled circuits, its effect is only to slow up the leading and trailing edges of the pulse, since this capacitance must be
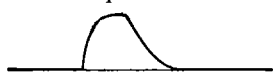
Fig. 3·23.—Usual effect of shunt capacity.
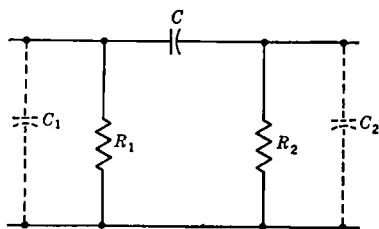
Fig. 3·24.—Coupling circuit including shunt capacity.

charged before any voltage change takes place. Thus the usual effect of shunt capacitance is that shown in Fig. 3·23.

With some of the special circuits described here, such as the choke-coupled or feedback circuits, another effect of shunt capacitance may be present which leads to overshoots. This effect can occur in all circuits where the full voltage across the coupling condenser does not appear as a signal on the grid after the pulse but divides between several branches.

Consider the typical coupling circuit of Fig. 3·24, with shunt capacitances C_1 and C_2. At the end of the pulse, before there is current flow in the resistors to bring about the distribution of voltages discussed in Sec. 3·2, there will be a brief flow of current charging up the capacitances C_1 and C_2. The voltage on the coupling condenser is thus divided between the grid and plate circuits inversely as the associated shunt capacitances.

Therefore the magnitude of the overshoot depends upon the ratio of the capacitances rather than on the ratio of the resistances, as was mentioned in Sec. 3·2.   As current begins to flow through the resistors, however, the charge on these shunt capacitances changes, and the conditions of the previous analysis hold.   The time required for this change to take place depends upon the time constants of the shunt capacitance and the load resistors.

The initial magnitude of the overshoot as caused by this shunt capacitance can differ considerably from the values obtained from the previous analysis.   Thus for the case where R_1 is very much greater than R_2 or for the analogous cases of choke coupling and inverse feedback, which depend for their effectiveness upon having only a small portion of the condenser

voltage appearing on the grid, the actual initial overshoot may be considerably greater than would be expected from Table 3·1 or from Eqs. (23) or (27).   This effect is of comparatively short duration, as the ratio of its time constant to that of the expected overshoot is equal

Fig. 3·25.—Spike often produced by shunt capacity.

to the ratio of the shunt capacitance to the coupling condenser.   Thus it appears as a sharp spike at the beginning of the overshoot, as shown in Fig. 3·25.

This effect may be troublesome in cases where the circuit has been designed to have a negligible overshoot according to previous considerations.   The spike may be enough larger than the overshoot to come through the amplifier as an appreciable signal.   Because it depends to a large extent upon stray capacitances, the effect is difficult to track down experimentally.   One obvious cure is to transfer the spike to the plate by loading the offending grid circuit with a small condenser; this changes the distribution of the coupling-condenser voltage between the plate and grid circuits.

3·4. Design Considerations.—In the design of a complete amplifier, the individual circuits are chosen and combined in order to give the desired over-all performance.   Because of the serious overload problem, the main consideration is usually one of overshoots; however, there are also other factors that must be considered.

Recovery.—A strong signal passing through an amplifier produces overshoots obscuring following signals.   The amplifier is ready for normal operation only after these overshoots have decayed to a level less than the weakest signals handled by the amplifier.   The recovery time of an amplifier can then be defined as the time required for the decay to occur after the strongest signal that the amplifier can handle.

Overshoots that are reduced to negligible amplitude at the output terminals of the amplifier can still be damaging to recovery.   Overshoots produced by long-time-constant circuits are often easily reduced to negligible magnitude by a short time constant later in the amplifier. However, if the overshoot attains a sufficient magnitude to overdrive any stage before it encounters the short time constant, it causes a variation in gain of this stage over a period of time during this overshoot, even though the overshoot itself is later reduced to negligible amplitude. This condition is illustrated in Fig. 3·26, which shows a possible block diagram with waveforms of the last stages of an amplifier.   The signal at the grid of the first of these stages has a long negative overshoot which has been produced by an earlier circuit.   This overshoot effectively drives the grid negatively to a point below its normal operating level and holds it there for a considerable time.   During this period the transconductance of the tube is below normal, and the gain of the stage for any signal occur-
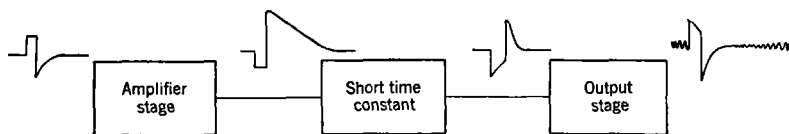
FIG. 3·26.—Effect of long overshoot on recovery.

ring during this period is low.   The output voltage of this stage is coupled to the output stage through a short-time-constant coupling, which produces an overshoot of short duration and eliminates the long overshoot.   However, this does not help the recovery, which has been affected by the loss of gain of the previous stage during the overshoot. Thus, at the output of the amplifier, no long overshoot is visible, but poor recovery is indicated by the absence of noise immediately following the signal.   (The output noise level is, of course, proportional to the gain of the amplifier because the input noise level is constant.)

Equation (5) shows that the amplitude of the overshoot produced by a coupling circuit is inversely proportional to the time constant of recovery. Thus for good recovery either the time constant can be made very long, producing an overshoot of negligible amplitude, or it can be made very short, producing an overshoot of comparatively large amplitude but decaying very rapidly.   Both of these approaches must be used in the design of any amplifier.   Between these possibilities lies one value of the time constant which gives the worst possible result, as can be shown by considering the expression for the instantaneous value of the overshoot voltage

$$e = \frac{\tau}{T} e^{-\frac{t}{T}}. \tag{29}$$

This equation assumes that there is only one circuit producing the overshoot.   If t is set equal to t_r, the recovery time desired for a given amplifier, then differentiation of Eq. (29) shows that the overshoot at this time is a maximum for $T = t_r$.   Thus for good recovery the time constant should be chosen either very much greater or very much less than the desired recovery time.

If several overshoots are present in an amplifier, the recovery time of the amplifier is determined by the worst individual overshoot.   Since all short time constants are made as short as possible, several equal short time constants are often found in amplifiers.   The recovery is then longer than if only one short time constant were present but is of the same order of magnitude.

Closely related to the problem of recovery is the problem of the allowable number of overshoots in a given amplifier.   Since there is no possibility of eliminating overshoots completely, high-gain video amplifiers can be divided into two groups.   Those amplifiers in which there is no secondary overshoot in the same direction as the original signal will be referred to as single-overshoot amplifiers.   Those in which a secondary overshoot and further overshoots are present will be referred to as multiple-overshoot amplifiers.   Whether or not a given amplifier should have single or multiple overshoots is determined by its application.

Single-overshoot Amplifiers.—Single-overshoot amplifiers have an advantage over multiple-overshoot amplifiers where the recovery time desired is of the same order of magnitude as the pulse length.   Thus where extremely rapid recovery is necessary, the single-overshoot amplifier must be used, as it must also in applications where a secondary overshoot appearing as a spurious signal would cause difficulty.   In some cases it may be possible to eliminate the secondary overshoot in the output circuit, as discussed later in this section, but this may involve a sacrifice in recovery time.

There are several approaches to the design of a single-overshoot amplifier.   The obvious method is to make one time constant short to give the desired recovery and low-frequency rejection, while all the other time constants are very long.   Since it has been shown that a sufficiently long time constant cannot be obtained in the coupling circuit in the presence of grid current, it is necessary to prevent the charging of the coupling condenser by grid current either by "brute force" or by use of the special inverse-feedback-pair circuits previously described, coupled either capacitively or directly.

In the brute-force method of preventing grid current, the operating potentials of the stage are chosen so that the maximum output voltage of each stage having a positive output signal will not drive the succeeding stage into grid current.

If the amplifier handles signals of only one polarity, the stages can be divided into two types: those driven by negative signals and those driven by positive signals.   The maximum output voltage of a negatively driven stage is that produced by cutting off the tube; this is equal to the voltage drop in the load resistor.   To keep this drop low, a tube that draws a relatively low plate current for a given g_m should be used and the tube should be operated at a low bias.   In order to obtain satisfactory amplification from the positively driven stage while operating it at a high bias, the tube used must be one having a reasonable value of transconductance at a high grid bias.   Tubes of high power consumption such as the 6AG7 must therefore be used.   If the amplifier must handle both polarities of signals, high-current tubes must be used all the way through, with each tube operated at a bias and a current low enough to give a satisfactorily limited output.   Because increasing bias and decreasing plate current tend to decrease the transconductance, the gain obtainable per stage in an amplifier required to handle both polarities of signals is considerably less than in an amplifier handling only one polarity.

The use of inverse-feedback pairs allows the use of common amplifier tubes operated at comparatively low current.   This circuit permits a considerable saving in weight and power consumption over the brute-force circuit.   Note, however, that this type of circuit is subject to the spike type of overshoot due to shunt capacitance described in Sec. 3·3. The brute-force circuit, where all the voltage across the coupling appears as overshoot across the grid, does not suffer from this effect. The inverse-feedback pair is restricted to circuits where only one polarity of signal is used, and a negative signal is applied to the first grid of the pair.   For this type of amplifier, performance as good as that of the brute-force circuit can be obtained at a considerable saving of weight and power. For amplifiers required to handle both polarities of signals, the only approaches discussed here that can be used are (1) direct coupling and (2) the prevention of grid current by brute force.

Another possibility for an amplifier required to handle only one polarity of signal is the double short-time-constant circuit discussed in Sec. 3·3 which eliminates the secondary overshoot by making use of grid current. Because there are two short time constants to reduce the overshoots produced in the previous circuits, the time constants of these circuits need not be so long as in amplifiers having only a single short time constant. Thus in this circuit a small amount of grid current is allowable in an early stage, since the overshoot produced can be rendered negligible by the two short time constants following.   This possibility allows even further simplification of the amplifier than was obtainable with the use of inverse-feedback pairs and is particularly useful in low-voltage applications where the power supply does not permit the use of brute force, direct coupling,

or inverse feedback.   However, the recovery of the double-short-time-constant circuit is generally not so good as that of circuits using only a single short time constant, inasmuch as there are two short time constants to determine the recovery, instead of one.[1]

Special circuits, such as the delay-line coupling circuit and germanium-crystal clippers, can be used to advantage in single-overshoot amplifiers. The delay-line circuit can be used in place of the short time constant; the germanium crystal can be used wherever a clipper can serve a useful purpose (see Sec. 3·3).

Multiple-overshoot Amplifiers.—For applications where multiple overshoots are allowable and the desired recovery time is moderately long in comparison with the pulse length, a multiple-overshoot amplifier is the simplest solution.   Simple *RC*-coupled circuits can be used with moderately short time constants in the coupling circuits.   Grid current is allowable provided that the grid resistors are not more than twice the plate-load resistors, thus making the time constants for positive and negative signals approximately equal.   Bypass and decoupling circuits can use much shorter time constants than are permitted in a single-overshoot amplifier, since there are more short time constants in the amplifier to reduce the overshoots produced by these circuits.   For some applications the use of circuits more complicated than the simple *RC*-circuit is advantageous.   The choke-coupled circuit previously described can be used to some advantage in a multiple-overshoot amplifier to obtain a high gain per stage.   The smearer circuit can be used at some points to improve the recovery.   Delay lines can be used in place of any or all short time constants with the improvement in recovery discussed in Sec. 3·3.   They are, however, subject to the disadvantages also outlined in that section, mainly the one of low impedance, which reduces the gain available per stage.   Recovery can also be improved by clipping overshoots with germanium crystals, also with a possible loss in gain.

As will be shown later in this section, multiple-overshoot amplifiers are much less subject to interference from microphonics and extraneous low-frequency signals than are single-overshoot amplifiers because of the greater low-frequency rejection afforded by the larger number of short time constants.

Output Circuits.—If the application for which a high-gain video amplifier is used requires that the output signal be viewed on an indicator, standard output circuits, such as have been described in Chap. 2, can be

---

[1] According to Eq. (19), it is possible to adjust the time constants of the double-short-time-constant circuit to eliminate the first term and obtain recovery equivalent to that of a single circuit.   In actual practice the recovery is always worse than that of a single circuit because of the difficulty of exact compensation.   Overcompensation is usually necessary to ensure that there be no secondary overshoot.

used. However, there are many applications where the amplifier output is never viewed but is merely used as a trigger for some other circuit. For these applications, it is often advantageous to make the amplifier output circuit a trigger circuit of some sort.

The most common type of output trigger circuit is the multivibrator, one form of which is shown in Fig. 3·27. The various types of circuits, the types of waveform obtainable, and other features of the multivibrator are discussed elsewhere in this series.[1] However, there are certain specific characteristics of these circuits, regarded as output circuits for high-gain video amplifiers, which will now be examined.

It is generally advantageous to trigger the multivibrator with a negative signal on the normally conducting tube; in this way the amplification of this tube is utilized. Furthermore, the multivibrator action can prevent the charging of the input coupling condenser by grid current. If there were no multivibrator action, the positive overshoot would drive the input grid into the positive region and charge the coupling condenser with grid current, thus producing a long secondary overshoot. The multivibrator action, however, causes a large negative pulse, having a duration determined by the constants of the circuit, to appear on this grid. If this pulse is greater in amplitude than the positive overshoot on the input signal, as may well be, the grid will remain negative, and no grid current will be drawn. Care must be taken, however, that the positive overshoot is not greater than the nega-

FIG. 3·27.—Multivibrator output stage.

tive multivibrator pulse; otherwise the overshoot will serve as a trigger to the multivibrator, causing the pulse to end too soon.

If a multivibrator output circuit is used with a multiple-overshoot amplifier, it is often desirable to make the length of the multivibrator pulse slightly greater than the time required for all the overshoots following a strong signal to decay. There is then only one output signal for each input signal to the amplifier, regardless of the number of overshoots produced in the amplifier. This eliminates the possibility of overshoots appearing as spurious signals and providing false triggers to later circuits.

One variation of this circuit can be used in cases where the original pulse length is to be preserved, but all overshoots must be eliminated.

[1] Vol. 19, Chap. 9, "Rectangular waveform generators," of the Radiation Laboratory Technical Series.

In this case, the input signal to the grid of the normally conducting tube in the multivibrator is of positive sign, with a negative overshoot as shown in Fig. 3·28. The positive signal is amplified by the first tube, producing a negative signal which has no effect on the second tube, which is already nonconducting. The negative overshoot, however, is in the proper direction to cause multivibrator action, and the resulting multivibrator pulse can be made sufficiently long to cover all the overshoots. If the output signal is taken from the plate of the first tube, the result is a negative signal followed by a large positive overshoot, equal in length to the duration of the multivibrator pulse. If this overshoot is objectionable, it can be clipped by a diode.

Another special output circuit is shown in Fig. 3·29. It is used in amplifiers operating at a low voltage, where it is desirable to get a high-
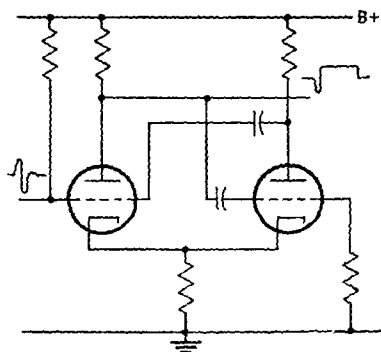
FIG. 3·28.—Multivibrator output circuit.   FIG. 3·29.—High-output voltage circuit.
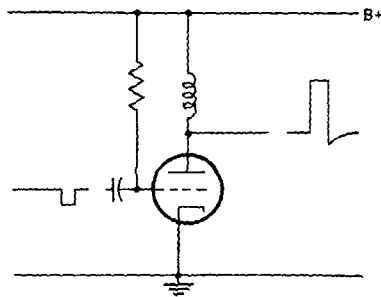
voltage output signal, usually considerably higher than the supply voltage. The output stage is a power-amplifier tube, operated to draw a high plate current through an inductive load. When a negative signal on the grid of this tube shuts off the current, a high voltage $L \, di/dt$ is generated across the inductance; this can be many times greater than the supply voltage. It is also possible to combine this circuit with the multivibrator by using an inductive plate load in the tube that is normally conducting.

Microphonics and Low-frequency Interference.—In any amplifier that has a high gain in the audio range, extraneous signals may appear as the result of mechanical shock or vibration or of pickup from adjacent power equipment. The problem of reducing this effect to a negligible level is different for each individual amplifier, depending upon the performance requirements of the amplifier and the particular conditions producing extraneous signals under which the amplifier will be used. However, certain general approaches to the problems are applicable in all cases.

Despite all precautions that may be taken to prevent extraneous signals from getting into the amplifier, it is almost certain that they will be present with sufficient amplitude to necessitate their removal by some sort of filtering in the amplifier. This filtering must be inserted at some point before the signals reach sufficient amplitude to cause variation in the transconductance of the tubes, as in the case of long overshoots discussed at the beginning of this section. The simplest form of such a filter is the short time constant used in single-overshoot amplifiers, which
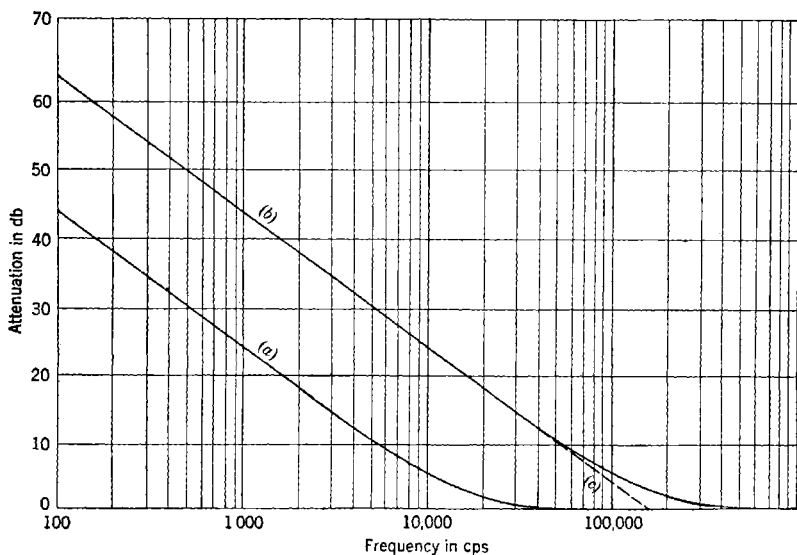
FIG. 3·30.—Attenuation vs. frequency for microphonic rejection circuits. (a) *RC*-circuit, $T = 10$ μsec; (b) *RC*-circuit, $T = 1$ μsec; (c) delay line, $t_d = 1$ μsec.

will reject low frequencies to a certain extent. This is shown in Fig. 3·30, which plots attenuation vs. frequency for two values of the time constant. The shorter the time constant the better the low-frequency rejection; but if the time constant is too short relative to the pulse length, the pulse will drop sharply in amplitude during its length; this drop is generally undesirable. If several short time constant circuits are present in an amplifier, the low-frequency rejection is much better than with only one, but there are more overshoots. Thus, in the design of an amplifier using *RC*-coupling, the time constants are made as short as possible without making the pulse response unsatisfactory.

The delay-line grid circuit described above can be used to reject low frequencies. If the line is lossless, the reflected signal, for any low-frequency sine wave, is equal in amplitude and almost 180° out of phase with

the original signal. Because of the phase delay introduced by the line, the cancellation is not complete. The cancellation can easily be calculated, since the resultant signal is the difference between two equal sine waves differing in phase by $2\pi f t_d$, where f is the frequency and t_d is the delay of the reflected signal with respect to the original signal. The attenuation-vs.-frequency curve for the delay line is also plotted in Fig. 3·30, showing that the low-frequency rejection of this circuit is very nearly equivalent to that of an RC-circuit having a time constant equal to t_d.

The choke-coupled circuit has particularly good low-frequency rejection because it has two frequency-sensitive elements, the choke and the coupling condenser. It is roughly equivalent to two RC-circuits in cascade.

As was shown for the case of long overshoots, it is insufficient merely to make the amplitude of low-frequency signals negligible at the output terminals of the receiver. The low-frequency rejection circuit must be inserted before the amplitude of the extraneous signals is large enough to change the transconductance of any tube.

The other approach to the problem of extraneous low-frequency signals is to prevent their occurrence initially. Interference brought in by the power supply can be reduced by inserting appropriate filters in the B-supply and heater leads or in the power line. Stray pickup can be reduced by proper shielding and grounding. Microphonics can be reduced by shock-mounting the amplifier, particularly the first tube, and by using tubes that are less microphonic. The exact extent to which these methods must be used and the amount of filtering necessary depend upon the relative difficulty that these schemes impose. Each amplifier presents an individual problem.

Pulse Stretching.—There are some applications for high-gain video amplifiers requiring that the pulse length be preserved in passing through the amplifier, regardless of signal strength. This problem is serious because of the large dynamic range of the amplifier. A signal that is several volts at the input terminal is repeatedly amplified and limited throughout the amplifier, thus the signal appearing at the output terminals is only that portion of the input pulse below the saturation level of the amplifier, i.e., about one-millionth of the original signal, amplified up to saturation level. The width of the output pulse is therefore the width of the input pulse measured at a point 120 db down. Therefore, any slope in the trailing edge of the pulse due to shunt capacitance, as shown in Fig. 3·23, results in stretching of a strong pulse after repeated amplification and limiting.

Pulse stretching can very easily be analyzed quantitatively if a few simplifying assumptions are made. Consider a pulse, such as that shown in Fig. 3·23, consisting of an exponential rise followed by an exponential

decay, as might occur if a rectangular pulse were distorted by the presence of shunt capacitance in a single circuit, e.g., the amplifier input circuit. Assume that this pulse is applied to an extremely wide-band amplifier, which has a negligible effect on the slopes of the leading and trailing edges of the pulse and merely amplifies and limits the signal.   The output signal of this amplifier is then applied to a device measuring the length of the pulse at the level corresponding to the amplitude of the minimum usable signal, so that the portion of the signal above the minimum usable level is of no interest.   Since the amplifier does not distort the portion of the pulse below this level, the apparent pulse length at the output terminals is equal to the length of the pulse at the input terminals, measured at the minimum usable level.   The pulse therefore appears to be longer than its true length by an amount equal to the time required for the pulse to decay from full amplitude to the minimum usable level.

Since the decay is exponential, the expression for the signal voltage during the decay is given by

$$e_s = E_a e^{-\frac{t}{T}}, \tag{30}$$

where T is the time constant of the circuit, and E_a is the pulse amplitude. This equation can be reduced to

$$\frac{20 \log_{10} \dfrac{E_a}{e_s}}{t} = \frac{8.68}{T} = 53.6b, \tag{31}$$

where b is the 3-db video bandwidth of an RC-circuit having a time constant T, b therefore being equal to $1/2\pi T$.

If e_s is taken as the minimum usable signal, the quantity $20 \log_{10} (E_a/e_s)$ is the signal amplitude expressed in decibels above the minimum usable level.   Equation (31) shows that the ratio of this quantity to the pulse stretching caused by this signal is, for a given single-stage RC-coupled amplifier, a constant independent of the signal strength, depending only on the bandwidth of the amplifier.   The ratio, expressed in decibels per microsecond, is often used to specify the pulse stretching for an amplifier.[1]

In an actual amplifier, the simplifying assumptions may not hold. The circuits within the amplifier may lengthen the decay time.   Since limiting occurs between stages, using the over-all bandwidth of the amplifier as b in Eq. (31) does not give the correct result.   The effects of each circuit cannot be considered individually, because the amount of pulse stretching produced by several circuits is not the sum of the stretching

---

[1] When a square-law detector precedes the amplifier, as mentioned in Sec. 4·1, the value of the pulse stretching in decibels per microseconds for the *receiver*, consisting of the combination of detector and amplifier, is half that for the amplifier.   In this chapter it is the value for the *amplifier* alone that is considered.

produced by the individual circuits. Thus the exact calculation of the pulse stretching is somewhat complicated. However, the approximation of Eq. (31) is often adequate if cautiously interpreted.

The obvious method of reducing pulse stretching is by decreasing the rise time of the circuits causing it, i.e., increasing the bandwidth of the stages causing the pulse stretching. This method involves a certain sacrifice in gain per stage because the minimum shunt capacity is determined by the wiring and tube interelectrode capacitances, and the only method of reducing the rise time is therefore the reduction of the load resistors with consequent loss of gain.

Another possible method of reducing pulse stretching is by introduction of an overshoot early in the amplifier before appreciable limiting takes place. This method has the effect of making the signal cross the baseline at a definite point, as shown in Fig. 3·31, instead of approaching the baseline asymptotically, as in the previous case. This means that the maximum amount of pulse stretching that

(a)

(b)

Fig. 3·32.—Effect of integration on long overshoot.

Fig. 3·31.—Stretching limited by overshoot.

can occur in the amplifier is the distance *ab* in Fig. 3·31, regardless of the amount of amplification and limiting following this point.

The use of pulse-length discrimination may in some cases impose a restriction on permissible fractional overshoot. One type of circuit frequently used in pulse-length discriminators is a limiter followed by an integrating circuit, which gives an output proportional to the area of the pulse which is, since the amplitude is limited, proportional to the pulse length. If there is an overshoot of small peak amplitude but long duration present, its area may be sufficient to make it an appreciably large signal after integration, as shown in Fig. 3·32. In this case, therefore, precautions must be taken to ensure that long-time-constant overshoots are negligible, not only before integration but after integration as well.

3·5. Small Amplifiers.—High-gain video amplifiers find considerable application in portable and airborne equipment, where space, weight, and power drain must be kept as low as possible. These requirements completely change the design of the amplifier. Circuits are chosen not only for their performance but for their ability to make use of small components and to require a minimum of power.

Types of Small Amplifiers.—The various applications of small amplifiers may be grouped into several general classes according to the power available for the amplifier.   There are applications where sufficient power is available to meet the needs of any circuit chosen.   The problem is then merely the design of a lightweight, compact amplifier, using tubes with a-c-operated heaters, operating at moderately high plate voltages and currents.

There are applications where the available power supply limits the plate current, the heater current, or both.   In this case tubes with a-c heaters are again used, but they may have to be types selected for low heater drain, operated at lower plate current.   It is quite possible that this type of amplifier may use more tubes than one in which B-supply current consumption is no consideration.

Where the only source of power is a storage battery or generator of 28 volts, it is advantageous to design an amplifier that does not require a plate voltage higher than 28 volts, thus eliminating the need for dynamotors, vibrators, or similar equipment.   Here the primary concern is to obtain the desired gain and performance with a 28-volt B-supply, size and heater power being of secondary importance.   The B-supply current, negligible in comparison with heater current, is of no concern.

For some portable applications, the only source of power available is dry batteries.   Since the operating time of such equipment is inversely proportional to the power drain, it is important that the power drain be kept as low as possible.   Filament-type tubes must be used wherever possible because of the enormous saving in power over tubes using indirectly heated cathodes.

Choice of Tubes.—In small amplifiers, the prime consideration in choosing tubes and circuits is gain rather than gain-bandwidth product, as in most other pulse amplifiers.   The best tube is the one furnishing the most gain for a given space or power requirement.   In many applications, the only requirement on amplifier response is that the output waveform indicate the presence of signals, and no other information regarding the nature of such a signal is necessary.   Thus very large distortion of the pulse is allowable.

From this point of view it becomes evident that triodes are generally superior to pentodes for small amplifiers, despite the advantage of pentodes in having a higher amplification factor.   In order to utilize the high pentode amplification factor, high plate-load resistors must be used. Their use produces a high d-c voltage drop across the load resistor, thus necessitating the waste of considerable power and the use of high-wattage load and decoupling resistors.   The use of the choke-coupled circuit to circumvent this problem introduces added complexity to the circuit and increases its size by adding the chokes, which may well be larger than the

tubes used. By choosing suitable compromise values, an amplifier in which the gain per stage is higher than that obtainable with triodes can still be built using pentodes without excessive power consumption. However, the use of twin-triode tubes makes possible a considerably higher gain per *envelope* than that obtainable with pentodes. Furthermore, the elimination of the screen-dropping resistor and bypass condenser may allow a considerable saving in size, especially where miniature and sub-miniature tubes are used, and the size of the amplifier may well be determined more by the number and size of the other components than by the tubes. The screen-bypass condenser for the first tube in particular presents a serious problem since it must be fairly large to ensure good recovery.

The main disadvantage of triodes for normal pulse amplifier use, namely, the Miller effect, is of little significance for this type of amplifier. Although the grid-plate capacitance appears across the input terminals multiplied by the gain of the stage, thus causing the input capacity of the stage to be unusually high, this merely has the effect of reducing the bandwidth, thereby slowing the pulse rise and fall time. With triodes having low grid-plate capacitance, such as the 7F8 and 6J6, the rise time is still fast enough for a 2-μsec pulse to reach a flat top, so that the Miller effect does not reduce the pulse gain. A reduction in pulse gain may occur, however, for shorter pulses or with tubes such as the 6SN7 and 6SL7 having higher grid-plate capacitances. The coupling condenser must be kept considerably larger than the equivalent input capacitance, since the two capacitances are effectively in series across the plate load and act as a capacitance voltage divider. If the coupling condenser is too small, a considerable portion of the signal is lost across it. There is, therefore, a limitation on the shortest possible time constant that can be used without loss of gain.

Choice of Circuit.—The choice of the circuit to be used for a small amplifier depends upon the size and power requirements and the desired recovery. Brute-force prevention of grid current, with large time constants for all circuits except one, is not suitable for a small amplifier because of the large components required. Inverse-feedback pairs, direct- or capacity-coupled, can be used for amplifiers using heater-cathode type tubes, with a high-voltage plate supply. Because of the absence of separate cathodes in filament-type tubes, direct coupling cannot be used for them. Neither type of inverse feedback can be used under low voltage conditions without excessive loss of gain. Thus, the double short time constant is the only scheme outlined here usable for a single overshoot amplifier having a low plate-supply voltage.

For multiple-overshoot amplifiers, straight *RC*-coupling is simplest and best. Choke coupling can be used, but the advantages scarcely

justify the added weight and size.   In applications with a battery or 28-volt power supply, more stages are necessary for a given gain than with heater-cathode type tubes operated with a high B-supply voltage, because the transconductance of the latter is much higher.

For some special applications in which a high B-supply voltage of approximately 300 volts is available but the allowable current drain is low, a reduction of the current drain by a factor of 2 can be achieved by connecting the tubes in pairs in series across the B-supply, as shown in Fig. 3·33.   For pulses, the pair represents an ordinary pair of amplifier stages, since the cathode

FIG. 3·33.—Series-fed circuit.

of the second stage is effectively grounded by the large condenser C. For direct current, the two tubes are in series across the B supply.   Thus
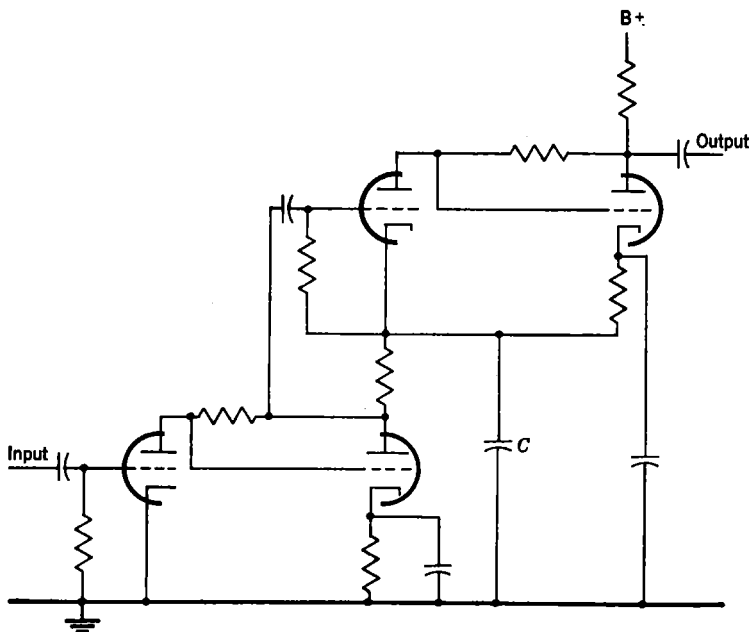
FIG. 3·34.—Series-fed direct-coupled inverse-feedback pairs.

the voltage across each tube is half the *B* supply (150 volts, if the total is 300), which is still enough for good amplification, whereas the current drawn from the supply for the two tubes is only the plate current of one tube.   This same idea can be extended to the direct-coupled inverse-feedback pair, as shown in Fig. 3·34.

The characteristics of the amplifier are not determined completely by the circuit diagram.   As has been discussed previously, stray capacitance can play an important part in the pulse response of the amplifier.   Even more important in small amplifiers is the possibility of stray feedback, which may have a peculiar effect on the pulse response of the amplifier; if there is enough feedback, the result is oscillation.   These effects can be avoided by proper layout and by properly grounding all ground points. Techniques for improving stability are discussed elsewhere in this series in great detail.[1]   Briefly, the following points must be considered: The stages should be laid out in a line from input to output terminals; common ground points on the chassis should not be used for several stages, heaters should be grounded to the chassis at points where there are no other connections.   These points will be illustrated in Sec. 3·6.

3·6. Examples.—In this section, six amplifiers are described; these present a good cross section of the types of amplifiers that can be used. For convenience they are numbered from one to six and are referred to by number.   A summary of the characteristics of these amplifiers is given in Table 3·4, which specifies the fidelity with which the pulse is reproduced, the recovery time for full sensitivity after a 1-volt signal, the

TABLE 3·2.—SUMMARY OF AMPLIFIER CHARACTERISTICS

No.	Fidelity	Approximate recovery, μsec	No. of over-shoots	Circuit	Output circuit	Tubes used	Power required	Size
1	Good	15	Single	Brute force	Simple	Large pen-todes	Considerable	Large
2	Fair	30	Single	Direct-coupled pairs	Multivi-brator	Submin. triodes	A-c heater low B drain	Tiny
3	Fair	30	Single	Double short time con-stant	Choke	Loktal twin-triode	28 volts	Small
4	Poor	100	Multiple	Straight *RC*	Special 28 volts	Loktal twin-triode	28 volts	Small
5	Poor	100	Multiple	Choke coupled	Smearer	Miniature pentode	A-c heater low B drain	Small
6	Poor	100	Multiple	Straight *RC*	Multivi-brator	Miniature twin-triode fila-ment type	Batteries	Small

[1] *Cf.* Chap. 8.

number of overshoots, the type of circuit used, the type of output circuit used, the general class of tubes used, the power requirements, and the relative size.

Amplifier 1.—Figures 3·35 to 3·38 show a high-gain video amplifier using the brute-force single-overshoot circuit, consisting of six stages, three 6AC7's and three 6AG7's. The amplifier[1] is designed for negative pulses of about 2 μsec, and holds pulse stretching[2] to 120 db/μsec.

The stages of the amplifier are designed in pairs, a typical one of which is shown in Fig. 3·35. In order to drive the 6AG7 into grid current, the grid voltage must rise above the cathode potential. At zero grid bias the plate current of the 6AG7, as operated in this amplifier, is about 64 ma.
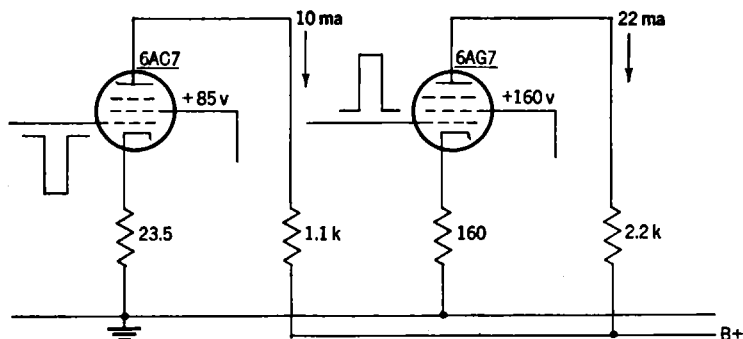
FIG. 3·35.—Typical pair of stages for amplifier No. 1.

Adding to this the zero-signal screen current of 6 ma, there is a total of 70 ma flowing through the cathode resistor during a pulse that barely drives the amplifier into grid current. The cathode voltage is therefore $0.070 \times 160 = 11.2$ volts. If no grid current is to be drawn, the signal on the grid of this tube must therefore be limited to less than 11.2 volts. The limited output voltage of the 6AC7 is $0.010 \times 1100 = 11$ volts, which is less than 11.2 volts and therefore does not drive the 6AG7 into grid current. In the circuit diagram of Fig. 3·36 the values are slightly different and indicate on the basis of the previous calculation that grid current will occur. Actually, there are some discrepancies in the tube characteristics, which make the circuit of Fig. 3·36 practical without the drawing of grid current.

The time constants of all circuits are long except for that of the coupling circuit between the fourth and fifth stages, which is made short to

----

[1] M. F. Crouch, "The Design of High-gain Video Amplifiers for Pulse Reception," Radiation Laboratory Internal Group Report 61-7/14/43, pp. 29–31.

[2] 60 db/μsec for a receiver consisting of a square-law detector preceding an amplifier of this type.

All resistors are $\frac{1}{2}$ watt unless otherwise specified
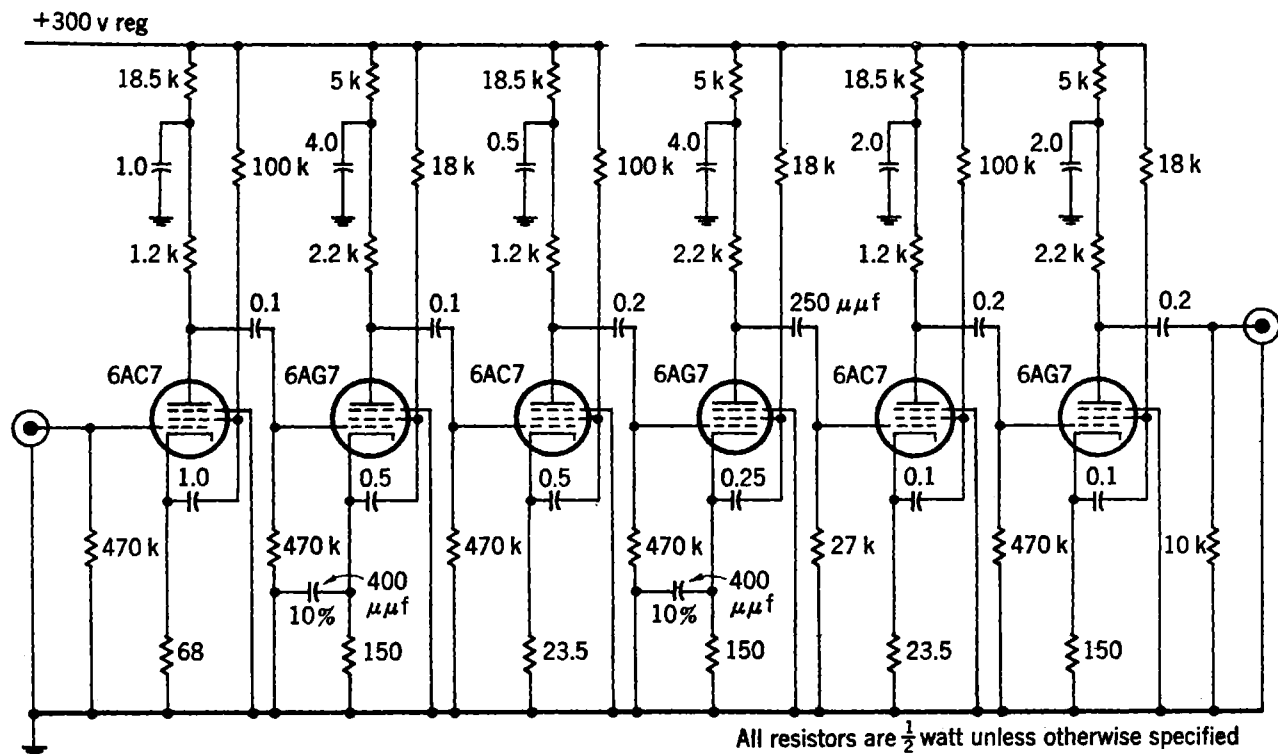
FIG. 3·36.—Complete circuit diagram for amplifier No. 1.

introduce a single overshoot and to reduce microphonics.   Because the
application requires that there be a sag of not more than 20 per cent in a
2-μsec pulse, a time constant of 10 μsec is used for this circuit.   Since the
signal on this grid is negative, the overshoot produced drives the tube
into grid current, thus shortening the time constant of recovery for this
circuit.

Cathodes are left unbypassed, except for the small condensers used for
high-frequency peaking in the second and fourth stages.   The screen
bypass condensers are returned to the cathodes rather than to ground so
that the signal component of the screen current may not flow through the
cathode resistor and produce added degeneration.   The degeneration is
kept as low as possible by the use of small cathode resistors.
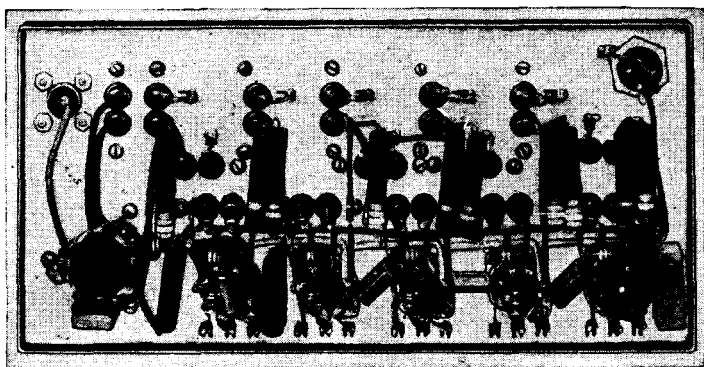
Fig. 3·37.—Layout of amplifier No. 1.

To ensure that the pulse stretching be held to 120 db/μsec, the band-
width of the amplifier, according to Eq. (31), must be about 2 Mc/sec,
which is the actual case.

The layout of the amplifier is straightforward and is shown clearly
in Fig. 3·37.   The row of condensers nearer the tubes are the screen bypass
condensers, the others being decoupling condensers.   Of particular impor-
tance are the ground connections.   There are three ground points for
each tube, connected to pins 1, 2, and 3 of the tubes.   For the 6AC7 these
correspond to the shell, heater, and suppressor.   For the 6AG7 these are
the shell and suppressor, heater, and internal shield.   In general, to
ensure stability all circuit grounds should be connected to the suppressor
pin and none to the heater.   In this case, it has been found permissible
to violate this principle to a small extent by returning the cathode circuit
to the heater pin.   Any further violation of this principle will definitely
cause trouble.   In the form of Fig. 3·37 the amplifier is stable with a small
cover over the circuits of the first tube and the remainder of the amplifier
open.   For convenience, a cover fitting the entire amplifier is used.

To reduce microphonics the first tube is shock-mounted.   The leads to this tube, as shown in Fig. 3·37, are not connected to the socket itself but to tie points mounted on a bakelite ring, which are connected by flexible leads to the shock-mounted tube.

The amplifier requires an electronically regulated power supply furnishing 120 ma at 300 volts.   Filtering in the heater line is necessary to prevent transients on the power line from being picked up within the amplifier and being amplified as signals.

Figure 3·38 shows the top view of the chassis of this amplifier, showing the large condensers used for screen bypass and decoupling.

Amplifier 2.—Figures 3·39 to 3·41 show a small amplifier made of subminiature tubes in a circuit of direct-coupled inverse-feedback pairs
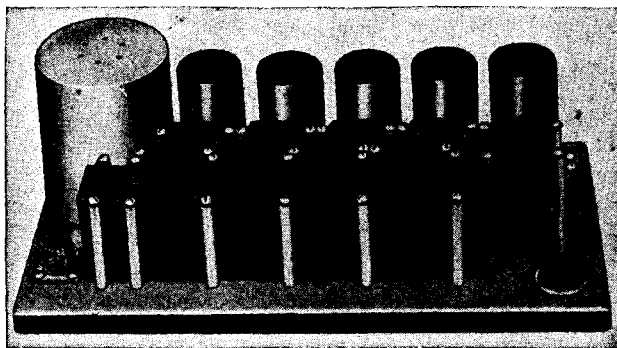
FIG. 3·38.—Amplifier No. 1.

with a multivibrator output circuit.   The inverse-feedback pairs are designed as described in Sec. 3·3.   The last two tubes form a multivibrator which is triggered by the negative signal appearing on the grid.   The short time constant for the elimination of long overshoots and low-frequency interference is the 1-μsec coupling time constant in the multivibrator, determined by 50 μμf and 20,000 ohms.   In addition, the time constant in the grid circuit of the first tube of the multivibrator is shorter than is apparent, since the 1-megohm grid resistor is shunted by the conductance of the grid, which is normally positive.

Overshoots from other sources are minimized by the use of direct coupling to the positively driven stages and of long time constants in the negatively driven stages.   Decoupling has been eliminated, as the amplifier is quite stable if the *B*-supply impedance is low.   This low impedance can generally be achieved by using a condenser of moderate size (between 0.1 and 1 μf) across the supply.   The circuit diagram for this amplifier is shown in Fig. 3·39,

The unique feature of this amplifier is the small size obtained by the use of subminiature tubes. The details of layout and construction are shown in Figs. 3·40 and 3·41. The tube clamps are mounted on a bakelite strip, along with turret lugs, which are used for all connections. Slots are put into the strip to allow for the mounting of the postage-stamp-
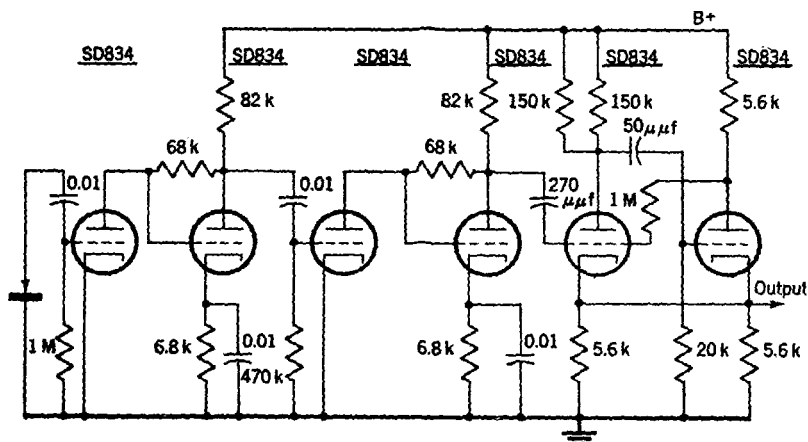
Fig. 3·39.—Circuit diagram of amplifier No. 2.

size 0.01-μf condensers. The power leads are strips of copper foil running along the back side of the strip, as shown in Fig. 3·40a, the upper strip being the *B* bus and the bottom strip being the heater bus. This arrangement serves the purpose of keeping the power leads close to the chassis, out of the way, and avoids congestion of leads. The strips are distant from circuit components and, having high capacity to ground and low

Fig. 3·40a.—Back of mounting strip.
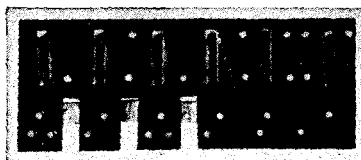
Fig. 3·40b.—Front of mounting strip.

inductance, are not so likely to cause feedback trouble as are ordinary wire leads.

The strip is mounted in the chassis, a similar bakelite strip being used to insulate the copper strips from the chassis. All connections are made to the turret-lug terminals on the strip, except for ground connections, which are made to turret lugs mounted directly into the chassis below the strip. There is a separate ground lug for each stage, plus other ground

lugs where the heaters are grounded.   Some lugs serve as grounds for two heaters; others for one, depending upon convenience, but no lug is used both for heaters and for other circuits.   With the precaution of separate
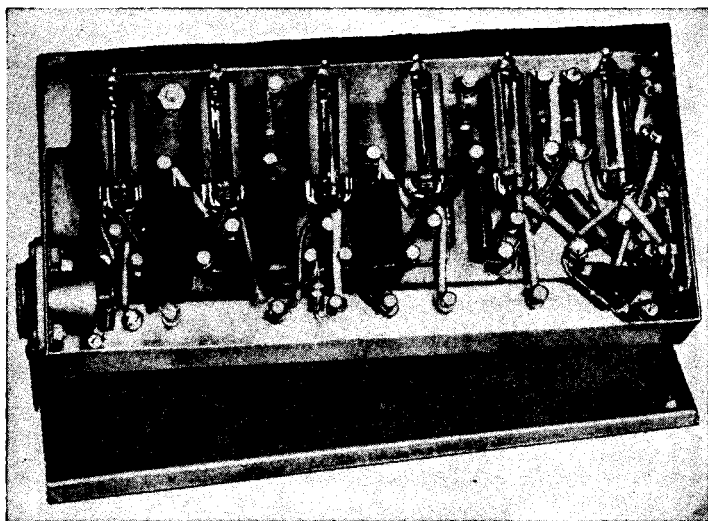
FIG. 3·41.—Layout of amplifier No. 2.
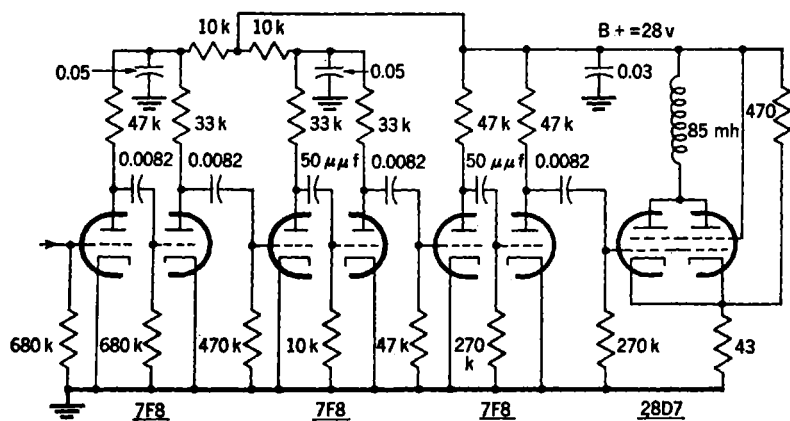
FIG. 3·42.—Amplifier No. 3.

grounds and the copper strip as the heater lead, there is no need for further filtering of the heater current.

Amplifier 3.—Figure 3·42 shows an amplifier circuit that operates from a 28-volt supply.   Because the double-short-time-constant circuit is used, there is only a single overshoot.   The type 7F8 twin triode is used

for all stages except the output and performs quite well with a plate sup-
ply of only 28 volts, although the gain per stage is naturally not so high as
it would be with a higher plate voltage. Six stages of amplification are
required before the output stage.

The amplifier is designed for a negative input signal. The two short
time constants are in the fourth and sixth grid circuits, with all other
time constants made very long. The time constant in the sixth grid cir-
cuit is short when grid current is drawn but increases to 16 μsec during the
overshoot. (For strong signals triodes 3 and 5 are cut off, and their plate
resistance is then infinite.) The time constant in the fourth stage is about
2 μsec. Thus, in Eq. (19), T_1 is 2 μsec, T_{2b} is 13.5 μsec, and T_{2a} is very
short. This circuit clearly meets the condition for no secondary over-
shoot, namely, that T_{2b} be greater than the sum of T_{2a} and T_1, with a con-
siderable factor of safety.
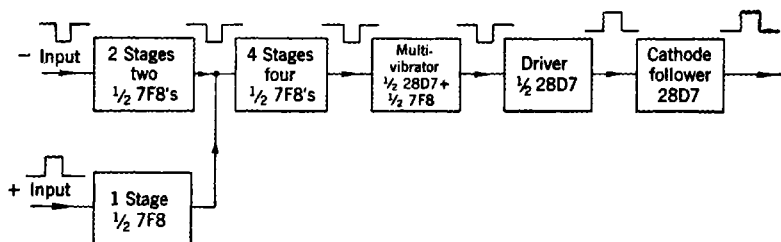
FIG. 3·43.—Block diagram of amplifier No. 4.

The recovery of the amplifier after strong signals is therefore gov-
erned by the 16-μsec time constant. Since this coupling circuit occurs in
the sixth stage after considerable limiting has taken place, the recovery is
acceptable for most applications.

The overload problem in general is not so severe in 28-volt amplifiers
as in those having a high plate voltage. Since the maximum possible
signal that can appear upon any plate is about 10 volts instead of 100,
an improvement of 20 db is automatically present.

The output circuit used is the one previously described in Sec. 3·4.
The tube used is the twin pentode 28D7, which can draw a fairly heavy
current with only 28 volts on the plate and screen.

The layout of this amplifier is not shown but is quite straightforward.
The tubes are laid out in a line from input to output. The general
placement of components is similar to that of Amplifier No. 4, as illus-
trated in Fig. 3·45b, except that the added complications of folding the
strip back on itself and of using twin input circuits are not necessary.

Amplifier 4.—The amplifier shown in Figs. 3·43 to 3·45 is also designed
to operate from a 28-volt power supply. The special requirement for
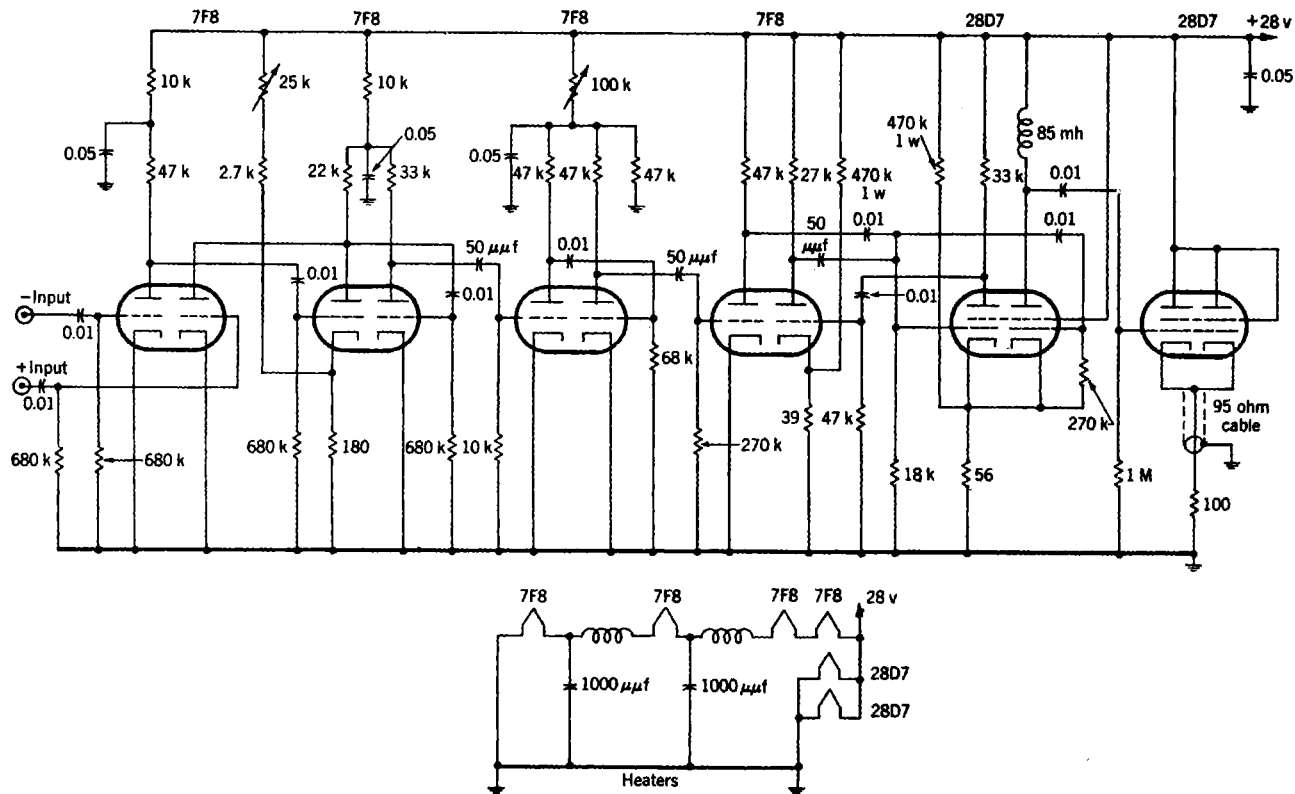
FIG. 3·44 —Circuit diagram of amplifier No. 4.

this amplifier is that signals from either of two input channels, with opposite signs of input signal, cause the amplifier to furnish a 12-volt trigger into a 95-ohm cable. The exact nature of this amplifier can be understood from the block diagram of Fig. 3·43. In order to get the necessary relative inversion, the two input channels are mixed after one stage of amplification for the positive signal and two stages for the negative. This mixing stage is followed by four stages of amplification and a multivibrator. In parallel with the normally conducting tube of the multivibrator, except for the plate circuit, is a driver stage using the high-voltage output circuit described in Sec. 3·4. The large output pulse is used to drive a cathode-follower output stage. Considerable output is obtained by use of type 28D7 twin-beam power-amplifier tubes as the driver and the cathode follower. The 7F8 twin triode, although not designed for 28-volt operation, performs well as an amplifier under this condition and is used for all stages except the output.

The circuit shown in Fig. 3·44 is one of straight *RC*-coupling with multiple overshoot. The time constants are chosen to make the recovery as rapid as possible. Short time constants are introduced into the amplifier as late as possible, so that the signal is limited before reaching these circuits and produces smaller overshoots. However, it is important that long overshoots do not reach sufficient amplitude to drive tubes into grid current. Hence, the short time constants are used in the coupling circuits of the first negatively driven stages, where the long positive overshoot would drive into grid current.

The time constants in the early stages are all made long. The problem of grid current is disregarded in the first positively driven stage (the maximum possible signal is still so small that any overshoot produced by grid current is made negligible by the short time constants following). The following stage, however, although negatively driven, is driven into grid current by the long overshoot. Since this would produce an extremely long paralysis of the amplifier, a short time constant is used for coupling at this point. The next stage is driven into grid current by the signal; hence the grid resistor is made approximately equal to the plate resistor to prevent a long recovery. The coupling condenser is made as large as possible (0.01 μf) to reduce the amplitude of the overshoot produced here to a minimum. At the output of this stage, the long overshoot has been amplified to a level where it is again troublesome, and another short-time-constant coupling is inserted. The time constant of coupling into the multivibrator is long, but the grid resistor is kept low because the requirements of the multivibrator, which is a straightforward circuit. All decoupling-circuit time constants are large; thus there are only two short time constants in the circuit.

The layout of this amplifier presents a peculiar problem because the

space requirements do not permit the use of a narrow strip.   The layout as shown in Fig. 3·45 solves the problem.

Figure 3·45a is a diagram showing the pin numbers of the tubes and numbering them for identification, and Fig. 3·45b is a photograph of the
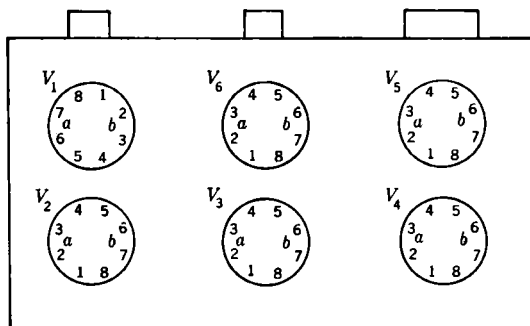
FIG. 3·45a.—Layout of amplifier No. 4.
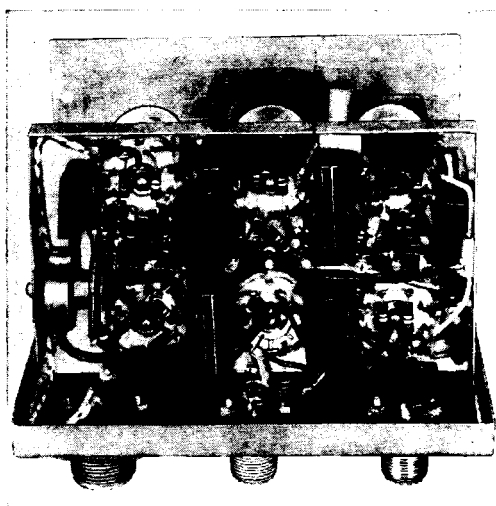
FIG. 3·45b.—Photograph of layout of amplifier No. 4.

layout.   The functions of the tubes, as numbered in Fig. 3·45a are as follows:

Triodes V_{1a} and V_{2a} are the first two stages of amplification in the negative signal channel, the input lead coming in from a connector on top of the chassis, located at the center of Tubes V_1, V_2, V_3, and V_6.   This connector is obscured by the parts covering it, but the input coupling condenser can be seen running alongside Tube V_1, next to the large decoupling condenser.

Triode V_{1b} is the input stage for the positive signal channel, the input connector being clearly visible on the front panel in front of Tube V_1.

The two channels are mixed at the plate of V_{2a}, and the following four stages of amplification are Triodes V_{2b}, V_{3a}, V_{3b}, and V_{4a}.

Beam-power Tube V_{5a} and Triode V_{4b} form the multivibrator, V_{5a} being the normally conducting tube; Tube V_{5b} is the driver, and Tube 6, both halves connected in parallel, is the cathode follower.

The input and high-level amplifier stages are at opposite corners of the amplifier. Although the output stage is next to the input, the feedback loop is broken by the normally nonconducting half of the multivibrator. Therefore, there is no transmission of signal around the loop,
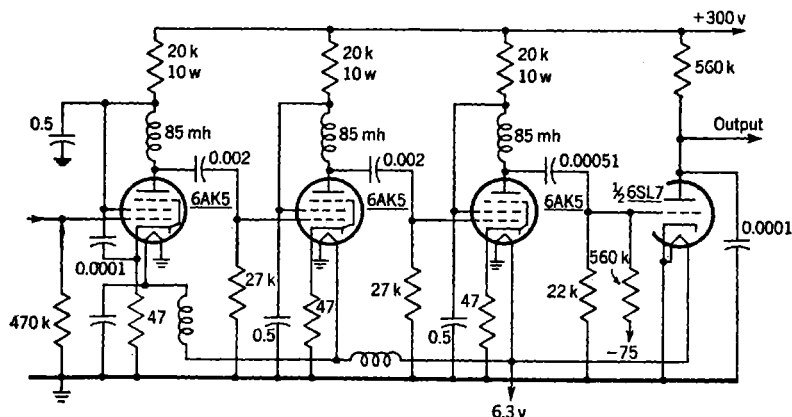
FIG. 3·46.—Amplifier No. 5.

except immediately after a signal, when the recovery time of the multivibrator prevents continuous ring-around. Stray feedback is reduced still more by the 0.05-μf decoupling condensers, which, standing on their sides, act as interstage shields, since they are at ground potential for pulses.

One difficulty encountered when 28D7 tubes are used is their heater power requirement of 11.2 watts. This requirement means that unless proper precautions are taken, there will be a considerable temperature rise in any equipment using these tubes. With the amplifier enclosed in a metal box with no ventilation, the temperature at points in the chassis rises to above 90°C. The maximum temperature can be reduced by providing for ventilation by making holes in the box and installing a fan.

Amplifier 5.—Figure 3·46 shows the circuit diagram of an amplifier using miniature pentodes with the choke-coupled circuit. This amplifier was designed to get the most gain from the fewest tubes with a recovery time of 100 μsec. For this reason the 6AK5-pentode choke-coupled circuit was chosen. For most purposes, this amplifier is inferior to one that could be designed using twin triodes, such as the miniature 6J6 or the

loktal 7F8. This is because the twin triodes, even with less gain per stage, permit more gain per envelope and also permit elimination of the chokes, which in this case occupy more space than the tubes. Furthermore, the power consumption of the triodes can be reduced, with a reasonable gain being maintained, by using high load resistors.

The chokes chosen for this amplifier, 85 mh, represent the largest readily available air-core chokes. The coupling condensers are not critical so long as the circuit is more than critically damped. The load resistors are chosen to give a gain of 40 db per stage, so that the over-all gain is about 120 db.

To keep the amplifier stable, extreme care is required in layout and grounding, because of the high gain in so small a space. To avoid
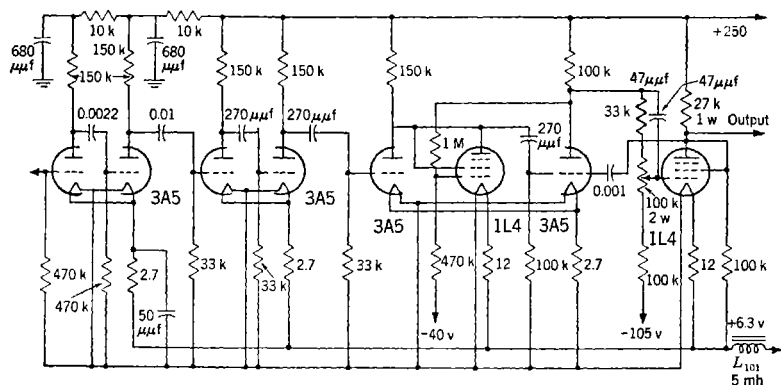
Fig. 3·47.—Amplifier No. 6.

instability, a small condenser (100 $\mu\mu$f) is necessary to bypass the screen of the first tube to its cathode directly at the tube socket. The heaters must be grounded properly, and heater chokes are required.

To remove the necessary overshoot in the output of the third amplifier stage, a smearer is used. Any triode is suitable for the purpose. In this case a 6SL7 is used, simply because the other half of the tube is in use elsewhere in the equipment.

Amplifier 6.—Figure 3·47 shows the circuit diagram of an amplifier designed to operate from a battery supply for portable use. In this type of application, an enormous saving in heater power is realized by use of filament-type tubes, such as the 3A5 twin triode and the 1L4 pentode. Because there is no cathode in these tubes to isolate the filament supply from the signal path, the filament circuit is much more critical than it is in amplifiers using tubes with indirectly heated cathodes.

Because the filament carries signal current, the filtering of the filament circuit is very critical. A large condenser, about 25 μf, is generally

necessary across the filament of the first tube in order to ensure stability. With proper care in layout and grounding and separate filament-voltage dropping resistors for each stage, the 25-μf filament condenser should be sufficient for stability. However, the 25-μf condenser alone would not be sufficient to eliminate low-frequency interference that could enter the amplifier through the filament circuit if a vibrator or dynamotor were used. For this purpose a series iron-core choke of low d-c resistance would be required.

Except for the filament circuit, the amplifier is quite straightforward, up to the multivibrator. Since there are two short 9-μsec time constants in the latter stages, the long time constants in the first few stages need not be very long. Thus the time constant between the first two stages is about 1000 μsec, and that between the second and third is about the same, depending upon the signal level that determines the plate resistance. The amplifier is designed for a positive input signal; hence the input to the third stage is positive, and the grid resistor must be less than the plate resistor to prevent blocking by grid current. To make the time constant sufficiently large a condenser larger than that between the first two stages is required at this point. The next two stages have short time constants, 9 μsec, and low grid resistors, to prevent blocking by grid current.

The sixth stage is the conducting half of a multivibrator, triggered by the negative input signal. This stage is d-c coupled to the normally nonconducting tube of the multivibrator. Because of the very wide variation in the cutoff characteristics of the 3A5 tube, a 1L4 is used triode-connected, with its bias set by the 100k potentiometer in the grid circuit. This controls the firing level of the multivibrator and is generally set so that the triggering by noise is sufficiently infrequent.

Multivibrator operation is not so reliable when using these filament-type tubes as it is for heater-type tubes. In particular, all impedances when the tubes are conducting are considerably higher than for corresponding heater-type tubes. Also, because of the absence of cathodes, cathode coupling is impossible. For this reason, d-c coupling is necessary in the multivibrator to ensure good recovery. In addition an extra tube is necessary to prevent overshoots and extraneous signals from turning the gate off too soon. This prevention of overshoots is accomplished by connecting the plate of a tube to the plate of the stage that drives the multivibrator. This tube is normally biased off but is d-c coupled to the plate of the normally conducting stage of the multivibrator. When the multivibrator fires, the positive gate is impressed upon this tube, rendering it conducting. This loads the fifth amplifier stage heavily so that no signals or overshoots pass this point during the duration of the gate.

CHAPTER 4

SYNCHRONOUS AND STAGGERED SINGLE-TUNED HIGH-FREQUENCY BANDPASS AMPLIFIERS

By Henry Wallman

4·1. Introduction.[1]—Typical of the wide-band bandpass amplifiers to be discussed in Chaps. 4 to 8 of this volume are amplifiers 2 Mc/sec wide at 30 Mc/sec of 110-db gain, 16 Mc/sec wide at 16 Mc/sec of 80-db gain, and 20 Mc/sec wide at 200 Mc/sec of 100-db gain.

The design of amplifiers of this type poses two main problems:

1. The theory of the interstage coupling elements needed to achieve such bandwidths and gains.
2. The practical questions of obtaining freedom from regeneration.

This chapter contains a detailed discussion of two schemes of interstage coupling; other schemes are considered in Chaps. 5 and 6. Chapter 7 contains comparisons of the various amplifier-design methods and considerations of transient response, and Chap. 8 discusses measurement and alignment procedures, and means of avoiding regeneration.

Although Chap. 7 contains a fairly extended treatment of the reproduction of pulses of a carrier frequency, the questions principally analyzed in Chaps. 4 to 6 are those involved in the design of amplifiers of large steady-state bandwidths. The reasons for placing the main emphasis on steady-state rather than transient-response considerations are the following:

1. For a tube type of given g_m and C, speed of response is cut in half when one goes from a low-pass circuit to its centered bandpass analogue, because of the double-sideband nature of carrier-frequency signals. Thus when 6AK5's are used in a bandpass amplifier that uses a single-tuned circuit (which is the bandpass analogue of RC-coupling), the gain/rise time ratio is only $100/\mu\text{sec}$ instead of $200/\mu\text{sec}$ as in the low-pass case (see Sec. 2·2).
2. The circuits that are practical in bandpass amplifiers have to be much simpler and consequently less efficient than the bandpass analogues of the high-speed, low-overshoot circuits of Chap. 2. Thus the bandpass analogue of the four-terminal linear-phase net-

work (Fig. 1·26), although entirely possible in theory, is quite unreasonable in practice, as Fig. 4·1 shows.

3. Because of reasons 1 and 2 and the g_m/C limitations of present tube types, it is necessary in high-gain carrier-frequency pulse amplifiers to accept fairly large overshoots (10 per cent or more) in order to achieve rise times as short as $\frac{1}{10}$ μsec. Fortunately for radar applications, such overshoots are tolerable.

4. With the variability in tube capacities permitted by the present JAN specifications, the tuning changes caused by replacing tubes make it pointless to design a bandpass interstage network for the utmost performance with regard to small overshoot.
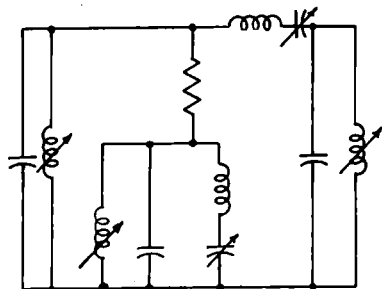
Fig. 4·1.—Bandpass analogue of four-terminal linear-phase network. Each of the inductors (capacitors) has to be tuned with its associated capacitor (inductor) to the desired band center.

[For applications requiring carrier-frequency pulse amplifiers having high gain (10^5 or more), fast rise time ($\frac{1}{10}$ μsec or less), and very small overshoot (1 or 2 per cent), it will be necessary to have tube-types of better g_m/C ratio than the type 6AK5 and/or circuits accurately adjusted to the capacities of the individual tubes in use.]

The Different Interstage-coupling Schemes.—The most common types of interstage coupling methods are

1. Synchronous single-tuned.
2. Stagger-tuned.
3. Double-tuned (including stagger-damped).
4. Inverse feedback.

Each of these types has its advantages and disadvantages with respect to

1. Efficiency, i.e., gain-bandwidth product.
2. Constructional simplicity.
3. Noncriticalness of adjustment.
4. Ease of gain control and gain stability.

These various characteristics will be examined for each of the types.

Roughly speaking it can be said that the synchronous single-tuned scheme can be put in one class of maximum simplicity and minimum efficiency, whereas the other types are in another class having greater complication but also considerably greater efficiency.

The coupling schemes to which this chapter is devoted are those of the synchronous single-tuned amplifier and the stagger-tuned amplifier.

Double-tuned circuits are discussed in Chap. 5 and inverse feedback in Chap. 6.

4·2. One Single-tuned Circuit.
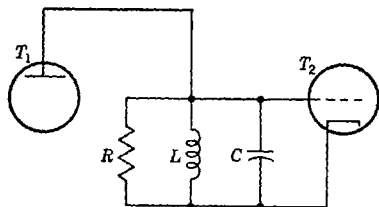
Fig. 4·2.—A-c diagram of single-tuned amplifier stage.

The Equation.—If d-c returns are neglected, the diagram of a single-tuned amplifier stage is as shown in Fig. 4·2, where C is the total circuit capacity (the output capacity of T_1 plus the input capacity of T_2 plus the wiring capacity) and R is the total circuit resistance (the parallel resistance of the load resistor, the plate resistance of T_1, the input resistance of T_2, and the equivalent shunt loss resistance of L and C). The impedance is found to be

$$Z(f) = R \frac{d}{d + j\left(\dfrac{f}{f_0} - \dfrac{f_0}{f}\right)}, \tag{1}$$

where

$$f_0 = \frac{1}{(2\pi \sqrt{LC})}$$

is the *resonant frequency* or *band center* and

$$d = \frac{1}{2\pi f_0 RC} = \frac{1}{Q} = \frac{2\pi f_0 L}{R} = \frac{1}{R}\sqrt{\frac{L}{C}}$$

is the *dissipation factor* (the reciprocal of the Q).

If the transconductance of T_1 is g_m, the voltage gain from the grid of T_1 to the grid of T_2 is

$$\mathcal{G} = g_m R \tag{2}$$

at band center and

$$g_m Z(f) = g_m R \frac{d}{d + j\left(\dfrac{f}{f_0} - \dfrac{f_0}{f}\right)} \tag{3}$$

as a function of frequency. The gain is a maximum at band center, and this maximum gain will be denoted by $\mathcal{G}$.

Geometric Symmetry.—The complex function in Eq. (1) displays "geometric symmetry"; i.e., for any two frequencies f and f_0^2/f having f_0 as their *geometric* mean, the absolute values are equal and the phase angles opposite in sign.

Bandwidth.—Throughout Chaps. 4 to 8 of this book, "bandwidth" is denoted by $\mathfrak{B}$ and means full 3-db bandwidth, i.e., the bandwidth included between the left and right half-power or 0.707-voltage points.

The reasons for this choice of 3-db bandwidth are first that the mathematics is easier and second that the noise bandwidth[1] is, except for a single single-tuned stage, quite accurately equal to the 3-db bandwidth.[2]

From Eq. (1) it follows that a 3-db point $f = \bar{f}$ occurs where

$$\frac{\bar{f}}{f_0} - \frac{f_0}{\bar{f}} = d,$$

or, substituting $d = 1/(2\pi f_0 RC)$, where

$$\bar{f} - \frac{f_0^2}{\bar{f}} = \frac{1}{2\pi RC}.$$

Because of the geometric symmetry about f_0 the other 3-db point is $f_0^2/\bar{f}$; hence the bandwidth is

$$\mathfrak{B} = \frac{1}{2\pi RC}. \tag{4}$$

Thus, the bandwidth of the circuit of Fig. 4·2 is independent of the center frequency.

Another consequence of Eq. (4) is

$$d = \frac{\text{bandwidth}}{\text{band center}} = \frac{\mathfrak{B}}{f_0}. \tag{5}$$

Approximate Form; Arithmetic Symmetry.—One may write

$$\frac{f}{f_0} - \frac{f_0}{f} = \frac{f^2 - f_0^2}{ff_0} = \frac{(f + f_0)(f - f_0)}{ff_0}.$$

---

[1] The noise bandwidth of an amplifier, the B in the noise power formula kTB, is the width of an idealized bandpass filter having the same total area as that under the power-vs.-frequency curve of the amplifier and having the same peak value. For synchronous single-tuned amplifiers the ratio of noise bandwidth to 3-db bandwidth is 1.57, 1.122, 1.155, 1.13, 1.11, 1.10, . . . , 1.06, for 1, 2, 3, 4, 5, 6, . . . , ∞ stages, and for a single staggered n-uple $(1/\sqrt{(1 + x^{2n})}$ voltage curve) the ratio is 1.11, 1.05, 1.025, 1.02, 1.01, . . . , 1.00 for $n = 2, 3, 4, 5, 6, . . . , \infty$.

[2] In addition there is the following fact about transient response and bandwidth: For a large variety of common coupling arrangements, e.g. one, two, or infinitely many synchronous single-tuned circuits, one or two transitionally coupled double-tuned circuits, a transitionally coupled triple-tuned circuit, etc., the rise time is very closely a certain fixed constant (namely, 0.7) times the reciprocal of the 3-db bandwidth. Rise time here denotes the time for the step-function response to go from 10 to 90 per cent of its steady-state value (see Fig. 7·7).

Now for values of f close to f_0 it is approximately true that $(f + f_0)/f = 2$; hence

$$\frac{f}{f_0} - \frac{f_0}{f} \approx 2\frac{f - f_0}{f_0}.$$

For circuits whose bandwidth is small compared with the resonant frequency, i.e., low-dissipation circuits, a good approximation to the right side of Eq. (1) is therefore

$$R\,\frac{d}{d + j2\,\dfrac{f - f_0}{f_0}},$$

or, using Eq. (5),

$$Z(f) \approx R\,\frac{\tfrac{1}{2}\mathfrak{B}}{\tfrac{1}{2}\mathfrak{B} + j(f - f_0)}. \tag{6}$$

In contrast to the right side of Eq. (1), Expression (6) displays "arithmetic symmetry," i.e., for any two frequencies f and $2f_0 - f$
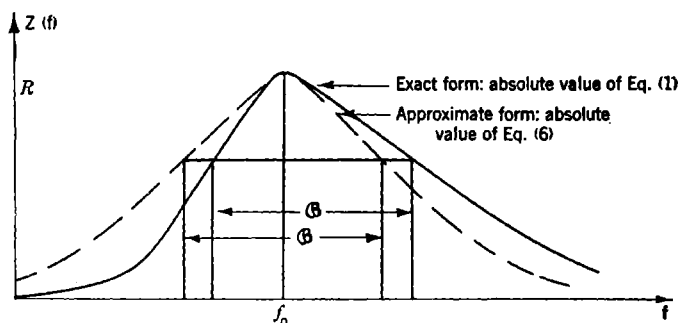
Fɪɢ. 4·3.—Resonance curve of single-tuned circuit.

having f_0 as their *arithmetic* mean the absolute values are equal and the phase angles opposite in sign (see Fig. 4·3).

It is true of the approximation in Expression (6), as of the exact form in Eq. (1), that

$$\mathfrak{B} = \frac{1}{2\pi RC}.$$

The exact form of Eq. (1) would, of course, show arithmetic symmetry if plotted on a logarithmic frequency scale.

Denoting $f - f_0$ in Expression (6) by x, so that x represents frequency off resonance, Expression (6) is proportional to

$$\frac{\tfrac{1}{2}\mathfrak{B}}{\tfrac{1}{2}\mathfrak{B} + jx}. \tag{7}$$

Taking $\circledR = 2$ yields the normalized selectivity function of a single-tuned circuit

$$\frac{1}{1 + jx},$$

whose absolute value

$$\frac{1}{\sqrt{1 + x^2}}$$

is shown in Fig. 4·4.

4·3. Amplifier Figures of Merit. *Gain-bandwidth* $(\mathcal{G}\circledR)$ *Product for a Single Stage.*—The evident figure of merit for a one-stage amplifier is the
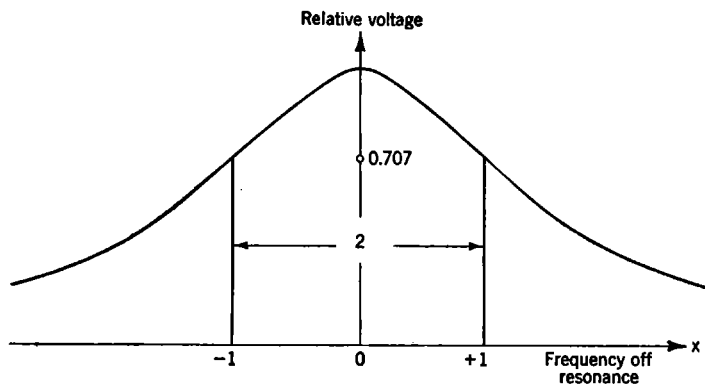
Fig. 4·4.—Normalized resonance curve of single-tuned circuit; arithmetic symmetry.

product of voltage gain at band center by bandwidth.[1]   For any given circuit configuration it is always possible to increase the bandwidth at the cost of a proportionate reduction in gain, thereby preserving the product of gain and bandwidth.   This fact follows from the impedance-level transformation applied to the given circuit, which consists of multiplying the value of each of its resistances by a factor k, leaving the capacity values unaltered, and multiplying the value of each inductance by k^2. It is not hard to see that the impedance at any frequency f is k times the impedance of the original network at frequency kf; hence the gain at each frequency is multiplied by k, and the bandwidth divided by k.

[1] On a steady-state basis; on a transient basis the appropriate criterion is gain/rise time (see Chap. 2).   Sometimes the amplifier designer is concerned with steady-state behavior (wide-band amplifiers), sometimes with transient behavior (fast amplifiers). Fortunately, as previously mentioned, for the more common interstage circuits the product of bandwidth and rise time is approximately independent of the particular coupling circuit employed.   Rise time is, however, not the only index of transient response; the percentage of overshoot is also very important.   A full characterization of transient response can be given only by a graph such as that of the step-function response (see Fig. 7·7).

$\mathcal{G}\mathcal{B}$ *Product and* $\mathcal{G}\mathcal{B}$ *Factor for One Single-tuned Stage.*—From Eqs. (2) and (4) it follows that the gain-bandwidth product for one single-tuned stage is[1]

$$\mathcal{G}\mathcal{B} = \frac{g_m}{2\pi C}. \tag{8}$$

In order to focus attention on the circuit rather than on the tube the $\mathcal{G}\mathcal{B}$ product will be normalized by expressing it in units of $g_m/(2\pi C)$; so normalized this quantity will be called the $\mathcal{G}\mathcal{B}$ "factor." Thus, Eq. (8) is equivalent to the statement for one single-tuned stage that

$$\mathcal{G}\mathcal{B}\ \text{factor} = 1. \tag{9}$$

$\mathcal{G}\mathcal{B}$ *Product for Mult'stage Amplifiers.*—When dealing with an n-stage amplifier, the appropriate figure of merit is not over-all gain[2] times over-all bandwidth, but rather

$$(\text{Over-all gain})^{1/n} \times \text{over-all bandwidth}.$$

As before, one can show that for a given scheme of interstage coupling this product is constant, whether the over-all bandwidth is large or small.

The following notation will therefore be adopted: whenever one speaks of the "$\mathcal{G}\mathcal{B}$ product" of an n-stage amplifier, the $\mathcal{G}$ will denote the nth root of the over-all gain and the $\mathcal{B}$ will denote the over-all bandwidth. Further, the "$\mathcal{G}\mathcal{B}$ factor" will mean the $\mathcal{G}\mathcal{B}$ product expressed in units of $g_m/2\pi C$.

It is sometimes convenient to give the name "mean stage gain" to the nth root of over-all gain.

4·4. Cascaded Synchronous Single-tuned Circuits.—*The* $\mathcal{G}\mathcal{B}$ *Factor for Synchronous Single-tuned Amplifiers.*—If n identical single-tuned stages with selectivity functions $1/(1 + jx)$ (see Fig. 4·4) are cascaded, the over-all selectivity function is

$$\left(\frac{1}{1 + jx}\right)^n \tag{10}$$

because of the isolating action of the tubes, and the over-all bandwidth, derived by setting the absolute value of Expression (10) equal to $1/\sqrt{2}$, is

$$\text{Over-all bandwidth} = \text{one-stage bandwidth} \times \sqrt{(2^{1/n} - 1)}.$$

For $n > 1$ a good approximation to $2^{1/n} - 1$ is $(\ln 2)/n$, from the power series expansion of $2^{1/n}$; moreover $1/\sqrt{\ln 2} = 1.2$ very closely; hence

---

[1] Note that Eq. (8) is independent of frequency; for a given gain it is no easier to get a wide bandwidth at a high frequency than at a low.

[2] "Gain" always means voltage gain at the center of the over-all pass band.

one may write

$$\text{Over-all bandwidth} = \frac{\text{one-stage bandwidth}}{1.2 \sqrt{n}} \qquad \text{(approx.)}.$$

Therefore, for an amplifier of cascaded synchronous single-tuned circuits,

$$\text{G\textcircled{B} factor} = \frac{1}{1.2 \sqrt{n}} \qquad \text{(approx.)}. \tag{11}$$

Equation (11) approaches zero as n increases, and the rapidity with which it approaches zero represents the principal weakness of synchronous single tuning.

TABLE 4·1.—SHRINKING OF OVER-ALL BANDWIDTH IN A SYNCHRONO $_{I}$S SINGLE-TUNED
AMPLIFIER

n	$\sqrt{(2^{1/n} - 1)}$ (exact)	$\dfrac{1}{1.2 \sqrt{n}}$ (approx.)
1	1.00	
2	0.64	0.59
3	0.51	0.48
4	0.44	0.42
5	0.39	0.37
6	0.35	0.34
7	0.32	0.32
8	0.30	0.29
9	0.28	0.28

Table 4·1 shows, for example, that if a nine-stage synchronous single-tuned amplifier is to have an over-all bandwidth of 4 Mc/sec, then each individual stage must be $4/0.28 = 14.3$ Mc/sec wide. Further, assuming $g_m/2\pi C = 57.3$ Mc/sec, as for 6AC7's with $g_m = 9000$ μmhos and $C = 25$ $\mu\mu f$, the G\textcircled{B} product is, from Eq. (11), only

$$16 \text{ Mc/sec} = 0.28 \times 57.3 \text{ Mc/sec};$$

hence the stage gain is only 4.

Maximum Bandwidth Possible at a Given Over-all Gain.—The circumstance that the G\textcircled{B} factor tends to zero implies that there is a maximum bandwidth for cascaded amplifiers of a given over-all gain and employing a given tube type no matter how many stages are permitted.[1]

For synchronous single-tuned stages the maximum bandwidth occurs when the gain per stage is $\sqrt{e}$ ($= 4.34$ db), and the use of any larger number of stages yields only a smaller over-all bandwidth. This state-

[1] This theorem appears to have been first noted by Alan Hazeltine, "Discussion on 'The Shielded Neutrodyne Receiver,' by Dreyer and Manson," *Proc. I.R.E.*, **14**, 395–412, (1926), particularly p. 406. The author is indebted to Mr. H. A. Wheeler for this reference.

ment is proved as follows: Denote by $\mathcal{G}$ the required over-all gain, so that for an n-stage cascaded amplifier the stage gain is $\mathcal{G}^{1/n}$; denote by $\mathcal{B}(n)$ the over-all bandwidth of the n-stage amplifier; the problem is to find the value of n maximizing $\mathcal{B}(n)$.

From Eq. (11) one has

$$\mathcal{G}^{1/n}\mathcal{B}(n) = \frac{g_m}{2\pi C}\frac{1}{1.2\sqrt{n}} \tag{12}$$

or

$$\mathcal{B}(n) = \frac{g_m}{2\pi C}\frac{1}{1.2\sqrt{n}}\frac{1}{\mathcal{G}^{1/n}}.$$

Setting $(d/dn)\mathcal{B}(n) = 0$ shows that $\mathcal{B}(n)$ is a maximum for $n = 2\ln\mathcal{G}$, or $\mathcal{G}^{1/n} = \sqrt{e} = 4.34$ db, completing the proof.

Unfortunately, the limitation on over-all bandwidth represented by this theorem is by no means an academic one; for present tubes the maximum possible bandwidth often turns out to be considerably less than is needed. Consider, for example, a synchronous single-tuned 100-db amplifier. The maximum bandwidth occurs with 23 stages (100 db/4.3 db). The $\mathcal{G}\mathcal{B}$ factor is $1/(1.2\sqrt{23}) = 0.174$. If 6AC7's are used and if $g_m/2\pi C$ is assumed to be equal to 57.3 Mc/sec, the $\mathcal{G}\mathcal{B}$ product is 10 Mc/sec for 23 stages. Since the stage gain is $\sqrt{e} = 1.65$, the over-all bandwidth is only 6 Mc/sec, which is entirely inadequate for many purposes. (It is worth repeating that the use of any number of stages greater than 23 would only *reduce* the over-all bandwidth.)

4·5. Example of a Synchronous Single-tuned Amplifier. *Six-stage* 110-*db* 6AC7 *Synchronous Single-tuned Amplifier Centered at* 30 *Mc/Sec.*— For bandwidths a third or less of the maximum bandwidth possible with synchronous single-tuned circuits, the use of such circuits is not too uneconomical, and in such cases this method has been widely employed, principally because of its simplicity. The following is a typical example:

If an interstage capacity of 25 $\mu\mu$f and, to be conservative, a 6AC7 transconductance of 8000 μmhos are assumed, the ratio $g_m/2\pi C$ is 51 Mc/sec.

The stage gain is $\frac{110}{6} = 18.3$ db or 8.25; the bandwidth per stage is therefore 6.17 Mc/sec (51/8.25). The over-all bandwidth is then, according to Table 4·1, 2.16 Mc/sec (6.17 × 0.35).

The individual interstages have the a-c diagram of Fig. 4.2 where $C = 25$ $\mu\mu$f, L is chosen to resonate with C at 30 Mc/sec (hence $L = 1.13$ μh), and $R = 1030$ ohms as follows from either

$$\text{Stage gain} = g_m R = 8.25$$

or

$$\text{Stage bandwidth} = \frac{1}{2\pi RC} = 6.17 \text{ Mc/sec.}$$

Practical Embodiment.—The circuit of an actual amplifier of this sort, designed by P. R. Bell, is shown in Fig. 4·5, and a photograph is shown
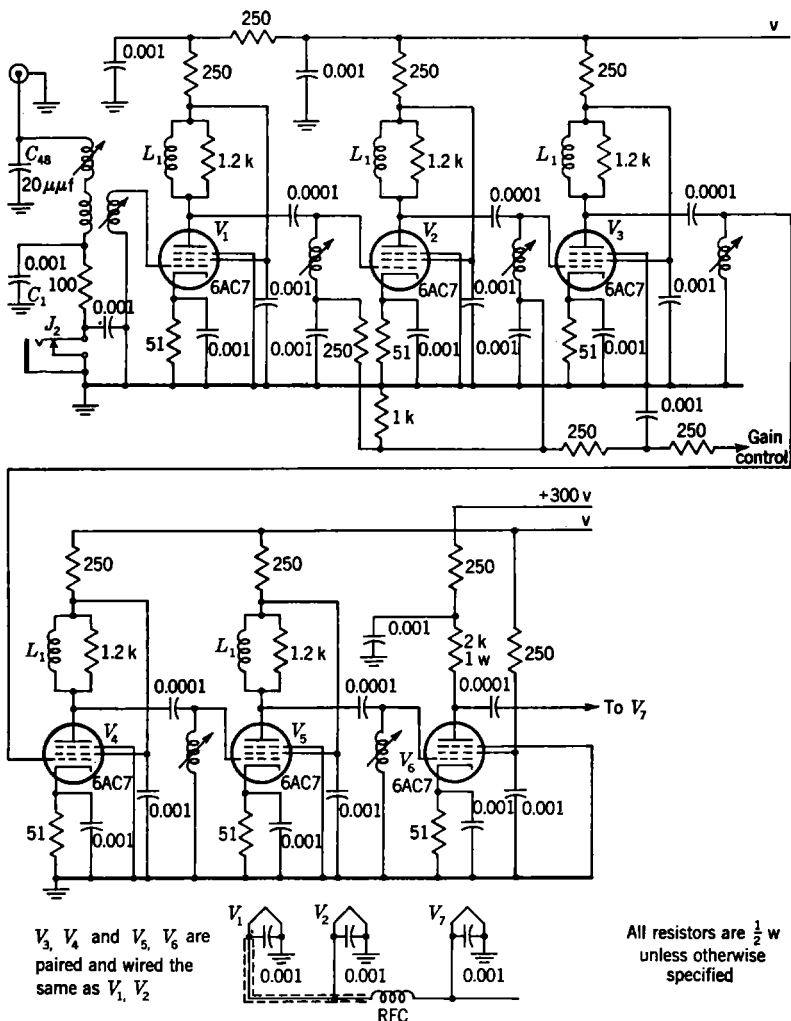
FIG. 4·5.—Circuit diagram of 6-stage 110-db 6AC7 synchronous single-tuned amplifier centered at 30 Mc/sec.

in Fig. 4·6. The coils marked L_1 in Fig. 4·5 are self-resonant chokes at 30 Mc/sec, wound on the load resistors, and are used to keep the plate voltage as high as the screen voltage.

FIG. 4·6.—Photograph of 6-stage 110-db 6AC7 synchronous single-tuned amplifier centered at 30 Mc/sec.

The tuning coils are tuned by copper plugs that can be moved into the field of the coils, thereby reducing their inductances. The alignment of such amplifiers can be carried out by tuning the individual coils so as to give maximum second-detector output signal when a 30 Mc/sec sine-wave signal is supplied to the input circuit. To assure that there are no gross errors, such as an incorrect damping resistor, it is, however, advisable to explore the amplifier pass band with a frequency-modulated signal generator, as described in Chap. 8.

4·6. Staggered *n*-uples. Arithmetic Symmetry.—In this section single-tuned circuits will be considered, in the approximation of Sec. 4·2, to have arithmetic symmetry. The exact case of geometric symmetry will be taken up in Sec. 4·7.

Definition, Advantages, and Disadvantages of a Staggered n-uple.—The goal of this section is to demonstrate that

(1) it is possible to stagger the tuning of *n* suitably damped single-tuned stages so as to get a $1/\sqrt{(1 + x^{2n})}$ selectivity curve (such an arrangement is called a flat-staggered *n*-uple[1]), as depicted in the upper half of Fig. 7·7, and

(2) the $\mathcal{GB}$ factor for a flat-staggered *n*-uple is 1.[2]

It is important to have a clear picture of the relative standing of points (1) and (2). It is often thought that getting the "maximally flat" selectivity curve $1/\sqrt{(1 + x^{2n})}$ is a good thing in itself, i.e., that there is

[1] Of bandwidth 2; other bandwidths are then obtained trivially (see Table 4·2).

[2] Another way of expressing (2) is this: A flat-staggered *n*-uple has as great an *over-all* bandwidth as does just one single-tuned stage of the same stage gain.

something inherently desirable in a flat selectivity curve. Only under extremely unusual conditions is this the case, however. If one is concerned with steady-state considerations of covering a large range of frequencies ("wide" amplifiers), it would be better to have a gain characteristic with small dips, covering a substantially wider band (see Sec. 7·6). If, on the other hand, transient considerations are controlling, as in television ("fast" amplifiers), then the flat-topped curve is not ideal because it leads to overshoots.

The only virtues of the $1/\sqrt{(1 + x^{2n})}$ curve [point (1) above] are its simple mathematical character and its easily recognized shape when viewed on a cathode-ray tube with a swept-frequency generator; it is here adopted only as a means to an end, the end being the high $g\mathcal{B}$ factor [point (2)]. In a sense, a flat selectivity curve is the *penalty* for a good $g\mathcal{B}$ factor.[1] A really useful accomplishment, from the point of view of transient response, would be a stagger-tuning scheme with a good $g\mathcal{B}$ factor and yet a rounded (Gaussian-error) over-all selectivity curve.

Complex Impedance Having $1/\sqrt{(1 + x^{2n})}$ as Its Absolute Value.— The (minimum phase shift) complex impedance having $1/\sqrt{(1 + x^{2n})}$ as its absolute value will now be determined.

This procedure turns out to be very useful in the treatment of the exact single-tuned circuit, and the complex impedance is needed in any case for the determination of transient response in Chap. 7.

Factoring,[2] $1 + x^{2n} = x^{2n} - (-1) =$

$$[(x^2-\epsilon_1)(x^2-\bar{\epsilon}_1)][(x^2-\epsilon_2)(x^2-\bar{\epsilon}_2)] \cdots [(x^2-\epsilon_{n/2})(x^2-\bar{\epsilon}_{n/2})], \; n \text{ even,} \\ [(x^2-\epsilon_1)(x^2-\bar{\epsilon}_1)][(x^2-\epsilon_2)(x^2-\bar{\epsilon}_2)] \cdots [(x^2+1)], \qquad\qquad n \text{ odd,} \tag{13}$$

where the ϵ are the nth roots of -1 and overlining indicates complex conjugate. One now seeks complex expressions of the form $(jx - \kappa_m)(jx - \bar{\kappa}_m)$ whose absolute values squared are equal to the brackets of Eq. (13). Equating coefficients of

$$|(jx - \kappa_m)(jx - \bar{\kappa}_m)|^2 = (x^2 + \kappa_m^2)(x^2 + \bar{\kappa}_m^2) = x^4 + (\kappa_m^2 + \bar{\kappa}_m^2)x^2 + \kappa_m^2\bar{\kappa}_m^2$$

and $(x^2 - \epsilon_m)(x^2 - \bar{\epsilon}_m) = x^4 - (\epsilon_m + \bar{\epsilon}_m)x^2 + \epsilon_m\bar{\epsilon}_m$ shows that $\kappa_m^2 = -\epsilon_m$, or

$$\kappa_m = (-\epsilon_m)^{1/2}, \tag{14}$$

where in order to make the resulting impedances realizable, the value of κ_m is selected that lies in the left half-plane. It then follows that the κ_m are those of the $2n$th roots of $(-1)^{n+1}$ that lie in the left half-plane.

---

[1] Fortunately the overshoot that goes with a $1/\sqrt{(1 + x^{2n})}$ curve is not too great for moderate values of n, so that the penalty is not too severe; for $n = 1, 2, 3, 4, 5, 6, 7$, the percentage overshoots in the step-function responses are 0, 4.3, 8.1, 10.9, 12.8, 14.3, 15.4, respectively (see Fig. 7·7).

[2] Such a factoring was used by V. D. Landon, "Cascade Amplifiers with Maximal Flatness," *RCA Rev.*, **1941**, 347–362, in particular p. 350; a similar factoring was earlier used by S. Butterworth, "On the Theory of Filter Amplifiers," *Wireless Eng.*, **1930**, 536–541, in particular p. 537.

Equations (13) and (14) show that $1/\sqrt{1 + x^{2n}}$ is the absolute value of

$$
\begin{aligned}
&1/\{(jx - \kappa_1)(jx - \bar{\kappa}_1)(jx - \kappa_2)(jx - \bar{\kappa}_2) \cdots (jx - \kappa_{n/2})(jx - \bar{\kappa}_{n/2})\}, \; n \text{ even,} \\
&1/\{(jx - \kappa_1)(jx - \bar{\kappa}_1)(jx - \kappa_2)(jx - \bar{\kappa}_2) \cdots (jx + 1)\}, \qquad\qquad n \text{ odd;}
\end{aligned} \quad (15)
$$

see Fig. 4·7.

Equation (15) is thus a complex function having $1/\sqrt{(1 + x^{2n})}$ as absolute value.   Equation (15), moreover, has no poles in the right half of the $\lambda = y + jx$ plane and hence[1] is a realizable impedance function. Finally Eq. (15) has no zeros in the right half of the plane and hence, as a minimum phase-shift impedance, is *uniquely* determined by its absolute value.[2]   In other words, Eq. (15) is *the* complex impedance function of minimum phase shift having $1/\sqrt{(1 + x^{2n})}$ as absolute value.
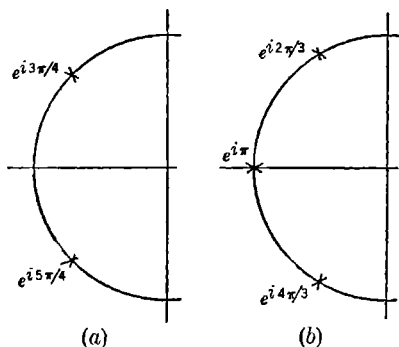
FIG. 4·7.—Location of poles for a flat staggered *n*-uple.   (a) *n* = 2; the 4th roots of −1 in the left half-plane.   (b) *n* = 3; the 6th roots of +1 in the left half-plane.

By writing κ_m (Eq. (14)) in the form $a + jb$, one sees that

$$\frac{1}{\sqrt{(1 + x^{2n})}}$$

is the absolute value of

$$
\frac{1}{\left[\sin \dfrac{\pi}{2n} + j\left(x \pm \cos \dfrac{\pi}{2n}\right)\right]\left[\sin \dfrac{3\pi}{2n} + j\left(x \pm \cos \dfrac{3\pi}{2n}\right)\right] \cdots \left[\sin \dfrac{(n-1)\pi}{2n} + j\left(x \pm \cos \dfrac{(n-1)\pi}{2n}\right)\right]}, \; n \text{ even,}
$$

$$
\frac{1}{\left[\sin \dfrac{\pi}{2n} + j\left(x \pm \cos \dfrac{\pi}{2n}\right)\right]\left[\sin \dfrac{3\pi}{2n} + j\left(x \pm \cos \dfrac{3\pi}{2n}\right)\right] \cdots (1 + jx)}, \; n \text{ odd.} \quad (16)
$$

Each of the n linear factors of Eq. (16) is the selectivity function of a staggered single-tuned stage; for example, the first two factors

[1] O. Brune, "Synthesis of a Finite Two-terminal Network," *Jour. Math. Phys.*, **10**, 191–236 (1931).

[2] Y. W. Lee, "Synthesis of Electrical Networks," *Jour. Math. Phys.*, **11**, 83–113 (1932); H. W. Bode, U.S. Patent 2123178 (1938), "Relations between Attenuation and Phase in Amplifier Design," *Bell System Tech. Jour.*, **1940**, 421–454, and *Network Analysis and Feedback Amplifier Design*, van Nostrand, New York, 1945, Chap. 14.

$$\frac{1}{\sin \dfrac{\pi}{2n} + j\left(x \pm \cos \dfrac{\pi}{2n}\right)}$$

are the selectivity functions of single-tuned stages with [see Expression (7)] bandwidths $2 \sin \pi/2n$ and resonance peaks $\cos \pi/2n$ to the left and right respectively of band center.   Clearly, therefore, the entire complex impedance in Eq. (16), and with it the absolute value $1/\sqrt{(1 + x^{2n})}$, can be synthesized by means of a cascade of n amplifier stages coupled by suitable single-tuned circuits.   These stages are arranged in symmetrical pairs (Fig. 4·8), except that in the case where n is odd the last factor represents a centered stage.

The point of this synthesis by factoring is that the isolating action of vacuum tubes permits the building up of complicated impedances a

(a)   (b)   (c)
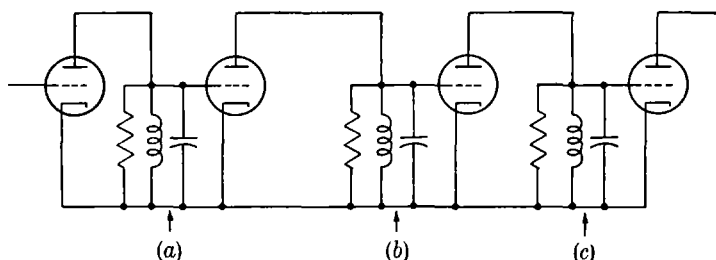
Fig. 4·8.—Synthesis of $1/\sqrt{(1 + x^{2n})}$ curve.

factor at a time, each simple factor being extremely easy to adjust and with no interaction among the various adjustments.

Synthesis by factoring can, of course, be used to obtain absolute-value curves more complicated than $1/\sqrt{(1 + x^{2n})}$.   Indeed, the method appears to be of general application in all cases where one desires an amplifier of specified pass band and more or less high gain—"filter amplifiers," as Butterworth would call them.   The amplifying action of several tubes is needed in order to secure the gain.   One then might as well make use of their isolating action also, thereby making it possible to employ only very simple circuits throughout.

Proof That the Ϙℬ *Factor of a Flat Staggered n-uple Is* 1.—Equation (16) has demonstrated point (1) of Sec. 4·6, that it is possible to stagger n single-tuned stages so as to get a $1/\sqrt{(1 + x^{2n})}$ selectivity curve; it remains to show (2) that the Ϙℬ factor is 1.   This follows very easily, however, by associating the numerator $\frac{1}{2}$ with each of the n factors of Eq. (16).   So modified, each of these factors represents a single-tuned stage of Ϙℬ factor 1.   The absolute value of the product,

$$\frac{\left(\frac{1}{2}\right)^n}{\sqrt{(1 + x^{2n})}},$$

has an over-all gain of $(\frac{1}{2})^n$ and therefore a stage gain of $\frac{1}{2}$. Because its over-all bandwidth is 2, the $\mathcal{G}\mathcal{B}$ factor is 1, as desired. This completes the proof of the assertions of Sec. 4·6.

Staggered n-uple Table. Arithmetic Symmetry.—Table 4·2, derived from Eq. (16), shows how to make up a flat staggered *n*-uple, $n = 1$, . . . , 7, centered at f_0 and with over-all bandwidth $\mathcal{B}$, under the assumption of arithmetic symmetry, i.e., $\mathcal{B}/f_0$ small.

Table 4·2 shows that the component single-tuned stages of a staggered *n*-uple have smaller bandwidths than the over-all bandwidth of the *n*-uple. This case contrasts with the case of synchronous-tuned amplifiers, where, as has been seen, the individual bandwidths are considerably larger than the over-all bandwidth.

TABLE 4·2.—FLAT STAGGERED *n*-UPLES

Approximate case: Arithmetic symmetry,

Band center = f_0, over-all bandwidth = $\mathcal{B}$, and $\mathcal{B}/f_0$ small

n	Component single-tuned stages
2. Staggered-pair.........	Two stages staggered at $f_0 \pm 0.35\mathcal{B}$ of bandwidth $0.71\mathcal{B}$
3. Staggered-triple.......	Two stages staggered at $f_0 \pm 0.43\mathcal{B}$ of bandwidth $0.5\mathcal{B}$ One stage centered at f_0 of bandwidth $\mathcal{B}$
4. Staggered-quadruple...	Two stages staggered at $f_0 \pm 0.46\mathcal{B}$ of bandwidth $0.38\mathcal{B}$ Two stages staggered at $f_0 \pm 0.19\mathcal{B}$ of bandwidth $0.92\mathcal{B}$
5. Staggered-quintuple...	Two stages staggered at $f_0 \pm 0.48\mathcal{B}$ of bandwidth $0.31\mathcal{B}$ Two stages staggered at $f_0 \pm 0.29\mathcal{B}$ of bandwidth $0.81\mathcal{B}$ One stage centered at f_0 of bandwidth $\mathcal{B}$
6. Staggered-sextuple.....	Two stages staggered at $f_0 \pm 0.48\mathcal{B}$ of bandwidth $0.26\mathcal{B}$ Two stages staggered at $f_0 \pm 0.35\mathcal{B}$ of bandwidth $0.71\mathcal{B}$ Two stages staggered at $f_0 \pm 0.13\mathcal{B}$ of bandwidth $0.97\mathcal{B}$
7. Staggered-septuple....	Two stages staggered at $f_0 \pm 0.49\mathcal{B}$ of bandwidth $0.22\mathcal{B}$ Two stages staggered at $f_0 \pm 0.39\mathcal{B}$ of bandwidth $0.62\mathcal{B}$ Two stages staggered at $f_0 \pm 0.22\mathcal{B}$ of bandwidth $0.90\mathcal{B}$ One stage centered at f_0 of bandwidth $\mathcal{B}$

4·7. Staggered *n*-uples: Geometric Symmetry.—Consideration will now be given to the exact treatment of single-tuned circuits, a matter which, as seen in Sec. 4·2, involves geometric symmetry.[1]

Definition of an Exact Flat-staggered n-uple.—The goal of this section is to demonstrate that

1. It is possible to stagger n single-tuned stages so as to get a

$$\frac{1}{\sqrt{\delta^{2n} + \left(f - \dfrac{1}{f}\right)^{2n}}}$$

[1] The exact case appears to have been first treated, but in a cumbersome way, by Rudolf Schienemann, "Trägerfrequenzverstärker groszer Bandbreite mit gegeneinander verstimmten Einzelkreisen," *Telegraphen Fernsprech Technik*, **1939**, 1–7. Schienemann was apparently the first writer to point out the main advantage of stagger-tuning over synchronous-tuning, i.e., its larger gain-bandwidth product, although others had earlier noted the possibility of synthesizing complicated networks from simple circuits separated by tubes ("filter amplifiers" in Butterworth's designation).

selectivity curve (such an arrangement is called an exact flat-staggered *n*-uple, whose bandwidth/band center ratio[1] is δ).

2. The $G\mathcal{B}$ factor for a flat-staggered *n*-uple is 1.

Two Staggered Single-tuned Stages of Equal Dissipation Factor.—Substituting $d = 1/(2\pi f_0 C)$ in Eq. (3) one gets

$$\frac{g_m}{2\pi C} \frac{\dfrac{1}{f_0}}{d + j\left(\dfrac{f}{f_0} - \dfrac{f_0}{f}\right)} \tag{17}$$

as the gain function.   Expressing gain in units of $g_m/2\pi C$ one can write

$$\frac{\dfrac{1}{f_0}}{d + j\left(\dfrac{f}{f_0} - \dfrac{f_0}{f}\right)}. \tag{18}$$

It is seen that the bandwidth of Expression (18) is df_0 and the gain is $1/(df_0)$, in accord with the $G\mathcal{B}$ factor of 1 for a single-tuned circuit.

Now suppose a single-tuned stage peaked at frequency α is followed by a stage peaked at $1/\alpha$, so that $f = 1$ is the geometric mean of their resonant frequen-

FIG. 4·9.—Two exact single-tuned circuits of same dissipation factor.

cies.   Suppose further that the two stages have the same dissipation factor d (Fig. 4·9).   Then the complex gain function of the product is [see Eq. (18)]

$$\frac{\dfrac{1}{\alpha}}{d + j\left(\dfrac{f}{\alpha} - \dfrac{\alpha}{f}\right)} \frac{\alpha}{d + j\left(f\alpha - \dfrac{\alpha}{f}\right)}.$$

Multiplying the denominators together yields

$$d^2 + jdf\alpha - j\frac{d}{f\alpha} + j\frac{df}{\alpha} - f^2 + \frac{1}{\alpha^2} - j\frac{d\alpha}{f} + \alpha^2 - \frac{1}{f^2}$$

$$= d^2 + \alpha^2 + \frac{1}{\alpha^2} + jd\alpha\left(f - \frac{1}{f}\right) + \frac{jd}{\alpha}\left(f - \frac{1}{f}\right) - \left(f^2 + \frac{1}{f^2}\right). \tag{19}$$

---

[1] And whose band center is $f = 1$; other center frequencies are then obtained trivially (see Table 4·3).   The value of δ may be either less or greater than 1.

Now

$$\alpha^2 + \frac{1}{\alpha^2} = \left(\alpha - \frac{1}{\alpha}\right)^2 + 2,$$

and

$$-\left(f^2 + \frac{1}{f^2}\right) = -\left(f - \frac{1}{f}\right)^2 - 2.$$

Hence the right side of Eq. (19) can be written as

$$d^2 + \left(\alpha - \frac{1}{\alpha}\right)^2 + j\left(f - \frac{1}{f}\right) d \left(\alpha + \frac{1}{\alpha}\right) - \left(f - \frac{1}{f}\right)^2,$$

and one has

$$\frac{\dfrac{1}{\alpha}}{d + j\left(\dfrac{f}{\alpha} - \dfrac{\alpha}{f}\right)} \frac{\alpha}{d + j\left(f\alpha - \dfrac{1}{f\alpha}\right)}$$
$$= \frac{1}{\left[d^2 + \left(\alpha - \dfrac{1}{\alpha}\right)\right]^2 + j\left(f - \dfrac{1}{f}\right)\left[d\left(\alpha + \dfrac{1}{\alpha}\right)\right] + \left[j\left(f - \dfrac{1}{f}\right)\right]^2}. \quad (20)$$

The point to observe about Eq. (20) is that the only variable in the right-hand side is the single combination $j\left(f - \dfrac{1}{f}\right)$, displaying geometric symmetry about $f = 1$, which is the geometric mean of α and $1/\alpha$.

Flat-staggered n-uples.—Multiplying the two terms in each bracket of Eq. (16) shows that $1/\sqrt{1 + x^{2n}}$ is the absolute value of

$$\frac{1}{\left[1 + jx2\sin\dfrac{\pi}{2n} + (jx)^2\right]\left[1 + jx2\sin\dfrac{3\pi}{2n} + (jx)^2\right]\cdots}$$
$$\left[1 + jx2\sin\dfrac{(n-1)\pi}{2n} + (jx)^2\right], \ n \text{ even,}$$

$$\frac{1}{\left[1 + jx2\sin\dfrac{\pi}{2n} + (jx)^2\right]\left[1 + jx2\sin\dfrac{3\pi}{2n} + (jx)^2\right]\cdots}$$
$$(1 + jx), \ n \text{ odd.} \quad (21a)$$

By analogy with Eq. (21a) it follows that

$$\frac{1}{\sqrt{\delta^{2n} + \left(f - \dfrac{1}{f}\right)^{2n}}}$$

is the absolute value of

$$\dfrac{1}{\left\{\delta^2 + j\left(f - \dfrac{1}{f}\right)2\delta \sin \dfrac{\pi}{2n} + \left[j\left(f - \dfrac{1}{f}\right)\right]^2\right\} \cdots}$$
$$\left\{\delta^2 + j\left(f - \dfrac{1}{f}\right)2\delta \sin \dfrac{(n-1)\pi}{2n} + \left[j\left(f - \dfrac{1}{f}\right)\right]^2\right\}, \; n \text{ even,}$$

$$\dfrac{1}{\left\{\delta^2 + j\left(f - \dfrac{1}{f}\right)2\delta \sin \dfrac{\pi}{2n} + \left[j\left(f - \dfrac{1}{f}\right)\right]^2\right\} \cdots}$$
$$\left\{\delta + j\left(f - \dfrac{1}{f}\right)\right\}, \; n \text{ odd.} \quad (21b)$$

It is now possible to synthesize the absolute value

$$\dfrac{1}{\sqrt{\delta^{2n} + \left(f - \dfrac{1}{f}\right)^{2n}}}$$

by equating the coefficients of $j\left(f - \dfrac{1}{f}\right)$ in Expressions (20) and (21b). For the first factor of Expression (21b), for example, one needs to know the dissipation factor d_1 and resonant frequency α_1 satisfying

$$d_1^2 + \left(\alpha_1 - \dfrac{1}{\alpha_1}\right)^2 = \delta^2, \quad (22a)$$

$$d_1\left(\alpha_1 + \dfrac{1}{\alpha_1}\right) = 2\delta \sin \dfrac{\pi}{2n}. \quad (22b)$$

Squaring Eq. (22b) and replacing

$$\left(\alpha_1 + \dfrac{1}{\alpha_1}\right)^2 \text{ by } \left(\alpha_1 - \dfrac{1}{\alpha_1}\right)^2 + 4$$

one gets

$$d_1^2\left(\alpha_1 - \dfrac{1}{\alpha_1}\right)^2 + 4d_1^2 = 4\delta^2 \sin^2 \dfrac{\pi}{2n};$$

substituting Eq. (22a) yields

$$d_1^4 - d_1^2(4 + \delta^2) + 4\delta^2 \sin^2 \dfrac{\pi}{2n} = 0;$$

solving,

$$d_1^2 = \dfrac{4 + \delta^2 - \sqrt{16 + 8\delta^2 + \delta^4 - 16\delta^2 \sin^2 \dfrac{\pi}{2n}}}{2};$$

using the double-angle formula $1 - 2 \sin^2 \pi/2n = \cos \pi/n$, one concludes that

$$d_1^2 = \frac{4 + \delta^2 - \sqrt{16 + 8\delta^2 \cos \dfrac{\pi}{n} + \delta^4}}{2}. \qquad (22c)$$

Equations (22a) and (22c) completely express d_1 and α_1 in terms of δ.

Proceeding similarly with the other factors of Eq. (21) it is seen that

$$\frac{1}{\sqrt{\delta^{2n} + \left(f - \dfrac{1}{f}\right)^{2n}}}$$

is the absolute value of

$$\left[\frac{\dfrac{1}{\alpha_1}}{d_1 + j\left(\dfrac{f}{\alpha_1} - \dfrac{\alpha_1}{f}\right)} \frac{\alpha_1}{d_1 + j\left(f\alpha_1 - \dfrac{1}{f\alpha_1}\right)}\right]$$

$$\left[\frac{\dfrac{1}{\alpha_3}}{d_3 + j\left(\dfrac{f}{\alpha_3} - \dfrac{\alpha_3}{f}\right)} \frac{\alpha_3}{d_3 + j\left(f\alpha_3 - \dfrac{1}{f\alpha_3}\right)}\right]$$

$$\cdots \left[\frac{\dfrac{1}{\alpha_{n-1}}}{d_{n-1} + j\left(\dfrac{f}{\alpha_{n-1}} - \dfrac{\alpha_{n-1}}{f}\right)} \frac{\alpha_{n-1}}{d_{n-1} + j\left(f\alpha_{n-1} - \dfrac{1}{f\alpha_{n-1}}\right)}\right], \; n \text{ even},$$

$$["] \cdots ["] \left[\frac{1}{\delta + j\left(f - \dfrac{1}{f}\right)}\right], \; n \text{ odd} \qquad (23)$$

where, for $k = 1, 3, \cdots, n - 1$,

$$d_k^2 = \frac{4 + \delta^2 - \sqrt{16 + 8\delta^2 \cos k\pi/n + \delta^4}}{2},$$

$$\left(\alpha_k - \frac{1}{\alpha_k}\right)^2 + d_k^2 = \delta^2. \qquad (24)$$

The factors in Eq. (23) are the impedances of single-tuned stages of dissipation factors d_k, resonant in pairs at frequencies α_k and $1/\alpha_k$. This establishes assertion (1) of Sec. 4·7; assertion (2) is evident from the fact that each of the factors of Eq. (23) has the $\mathcal{GB}$ factor 1, as does

$$\frac{1}{\sqrt{\delta^{2n} + \left(f - \dfrac{1}{f}\right)^{2n}}}$$

itself.

Table 4·3.—Flat-staggered n-uples

Exact case: geometric symmetry

Band center $= f_0$; over-all bandwidth $= \mathcal{B}$; and $\mathcal{B}/f_0 = \delta$

n	Component single-tuned stages

2. Staggered-pair........... Two stages staggered at $f_0\alpha$ and f_0/α of dissipation factor d, where

$$d^2 = \frac{4 + \delta^2 - \sqrt{16 + \delta^4}}{2}$$

$$\left(\alpha - \frac{1}{\alpha}\right)^2 + d^2 = \delta^2$$

3. Staggered-triple......... Two stages staggered at $f_0\alpha$ and f_0/α of dissipation factor d, one stage centered at f_0 of bandwidth $\mathcal{B}$, where

$$d^2 = \frac{4 + \delta^2 - \sqrt{16 + 4\delta^2 + \delta^4}}{2}$$

$$\left(\alpha - \frac{1}{\alpha}\right)^2 + d^2 = \delta^2$$

4. Staggered-quadruple..... Two stages staggered at $f_0\alpha_1$ and f_0/α_1 of dissipation factor d_1, two stages staggered at $f_0\alpha_3$ and f_0/α_3 of dissipation factors d_3, where

$$d_1^2 = \frac{4 + \delta^2 - \sqrt{16 + 5.656\delta^2 + \delta^4}}{2}$$

$$\left(\alpha_1 - \frac{1}{\alpha_1}\right)^2 + d_1^2 = \delta^2,$$

$$d_3^2 = \frac{4 + \delta^2 - \sqrt{16 - 5.656\delta^2 + \delta^4}}{2}$$

$$\left(\alpha_3 - \frac{1}{\alpha_3}\right)^2 + d_3^2 = \delta^2$$

5. Staggered-quintuple..... Two stages staggered at $f_0\alpha_1$ and f_0/α_1 of dissipation factor d_1, two stages staggered at $f_0\alpha_3$ and f_0/α_3 of dissipation factor d_3, one stage centered at f_0 of bandwidth $\mathcal{B}$, where

$$d_1^2 = \frac{4 + \delta^2 - \sqrt{16 + 6.472\delta^2 + \delta^4}}{2}$$

$$\left(\alpha_1 - \frac{1}{\alpha_1}\right)^2 + d_1^2 = \delta^2,$$

$$d_3^2 = \frac{4 + \delta^2 - \sqrt{16 - 2.472\delta^2 + \delta^4}}{2}$$

$$\left(\alpha_3 - \frac{1}{\alpha_3}\right)^2 + d_3^2 = \delta^2$$

The various stages of an exact flat-staggered n-uple do not have the same gain at band center [$f = 1$ in Eq. (23)], but instead, as follows from

$$d_k^2 + \left(\alpha_k - \frac{1}{\alpha_k}\right)^2 = \delta^2,'$$

the gain of each stage at band center is inversely proportional to its own resonant frequency. Each pair of staggered stages has as much gain at band center as any other pair, however, and for the odd n case the gain of the centered stage is exactly the mean gain of the n-uple. Further, for the odd n case the bandwidth of the centered stage is exactly the over-all bandwidth of the n-uple.

TABLE 4·4.—FLAT-STAGGERED n-UPLES

Asymptotic case: for $\delta = \mathcal{B}/f_0$ less than 0.3

Bandwidth $= f_0$; over-all bandwidth $= \mathcal{B}$; and $\mathcal{B}/f_0 = \delta$

n	Component single-tuned stages
2. Staggered-pair.........	Two stages staggered at $f_0 \pm 0.35\mathcal{B}$ of dissipation factor 0.71δ
3. Staggered-triple........	Two stages staggered at $f_0 \pm 0.43\mathcal{B}$ of dissipation factor 0.5δ, one stage centered at f_0 of bandwidth $\mathcal{B}$
4. Staggered-quadruple....	Two stages staggered at $f_0 \pm 0.46\mathcal{B}$ of dissipation factor 0.5δ, two stages staggered at $f_0 \pm 0.92\mathcal{B}$ of dissipation factor 0.19δ
5. Staggered-quintuple....	Two stages staggered at $f_0 \pm 0.48\mathcal{B}$ of dissipation factor 0.31δ, two stages staggered at $f_0 \pm 0.29\mathcal{B}$ of dissipation factor 0.81δ, one stage centered at f_0 of bandwidth $\mathcal{B}$
6. Staggered-sextuple......	Two stages staggered at $f_0 \pm 0.48\mathcal{B}$ of dissipation factor 0.26δ, two stages staggered at $f_0 \pm 0.35\mathcal{B}$ of dissipation factor 0.71δ, two stages staggered at $f_0 \pm 0.13\mathcal{B}$ of dissipation factor 0.97δ
7. Staggered-septuple......	Two stages staggered at $f_0 \pm 0.49\mathcal{B}$ of dissipation factor 0.22δ, two stages staggered at $f_0 \pm 0.39\mathcal{B}$ of dissipation factor 0.62δ, two stages staggered at $f_0 \pm 0.22\mathcal{B}$ of dissipation factor 0.90δ, one stage centered at f_0 of bandwidth $\mathcal{B}$

Exact Case.—Table 4·3, derived from Eqs. (23) and (24), shows how to make up an exact flat staggered n-uple, $n = 1, \cdots, 5$, centered at f_0 and with over-all bandwidth $\mathcal{B}$. The ratio $\mathcal{B}/f_0$ is denoted by δ. Because nearly equal numbers are subtracted in these formulas, calculations should be carried out to three decimals.

Asymptotic Values.—The asymptotic values of the expressions in Table 4·3 for small values of δ are easily derived, taking into account the approximations

$$\alpha = 1 + \frac{1}{2}\left(\alpha - \frac{1}{\alpha}\right),$$

and

$$\frac{1}{\alpha} = 1 - \frac{1}{2}\left(\alpha - \frac{1}{\alpha}\right)$$

for small values of $\alpha - (1/\alpha)$. These asymptotic values are listed in Table 4·4; they are accurate for $\delta = \mathcal{B}/f_0$ less than 0.3; and because of their simplicity and accuracy in that range, they are to be preferred to the values derived from Table 4·2.

4·8. Flat-staggered Pairs, in Detail. *Graphs.*—Figure 4·10 was prepared from Table 4·3 for the case $n = 2$. The use of Fig. 4·10 will be clear from the following example:

Suppose a flat-staggered pair of 8 Mc/sec bandwidth is to be designed with band center at 10 Mc/sec. Then $f_0 = 10$ Mc/sec, $\mathcal{B} = 8$ Mc/sec, and $\delta = 0.8$, so that from Fig. 4·10 one finds that $\alpha = 1.33$ and $d = 0.535$. Therefore, the pair is to be constructed of one stage staggered at

$$10 \times 1.33 = 13.3 \text{ Mc/sec},$$

of dissipation factor 0.535 and hence of bandwidth

$$0.535 \times 13.3 = 7.1 \text{ Mc/sec},$$

and one stage staggered at $10/1.33 = 7.5$ Mc/sec, of dissipation factor 0.535 and hence of bandwidth $0.535 \times 7.5 = 4.0$ Mc/sec.

Cascaded Pairs and Comparison with Synchronous Single-tuned Stages.—The selectivity function of a flat staggered pair has the form $1/\sqrt{(1 + x^4)}$ in the low dissipation case and $1\Big/\sqrt{\delta^4 + \left(f - \dfrac{1}{f}\right)^4}$ in the exact case. These curves are squarer than those for single-tuned stages; and because they are squarer, the bandwidth goes down less rapidly;[1] it is easy to see that the over-all bandwidth of m pairs is the bandwidth of one pair times $\sqrt[4]{(2^{1/m} - 1)}$. For $m > 1$, a good approximation is

$$\text{Over-all bandwidth of } m \text{ pairs} = \frac{\text{bandwidth of one pair}}{1.1 \sqrt[4]{m}}$$

(see Table 4·5).

TABLE 4·5.—SHRINKING OF OVER-ALL BANDWIDTH, m-CASCADED FLAT-STAGGERED PAIRS

m	$\sqrt[4]{2^{1/m} - 1}$ (exact)	$\dfrac{1}{(1.1 \sqrt[4]{m})}$ (approx.)
1	1.00	. . . .
2	0.80	0.76
3	0.71	0.69
4	0.66	0.64
5	0.62	0.61

[1] And the overshoot is larger: 4.3 per cent for one pair, 6.25 per cent for two pairs, 7.7 per cent for three pairs, 8.3 per cent for four pairs, 9.9 per cent for six pairs (see Fig. 7·5).
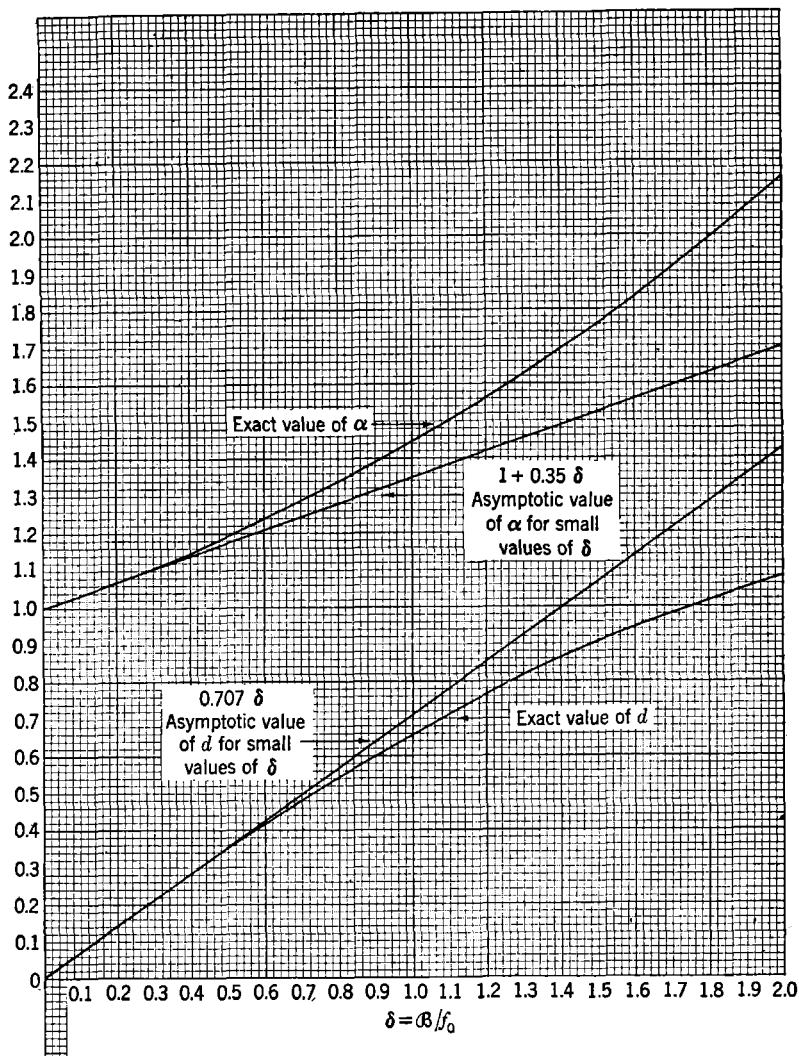
Fig. 4·10.—Design curves for an exact flat staggered pair.

Note:
1. An exact flat staggered pair of stage gain G has as great an over-all bandwidth $\mathcal{B}$ as does one single-tuned stage of gain G, i.e., $G\mathcal{B} = g_m/2\pi C$.
2. An exact flat staggered pair of over-all bandwidth $\mathcal{B}$, geometrically centered at f_0 consists of two single-tuned stages staggered at $f_0\alpha$ and f_0/α of dissipation factor d, α being given in the upper graph and d in the lower graph as functions of $\delta = \mathcal{B}/f_0$.
3. An amplifier made up of m pairs has $1/(1.1 \sqrt[4]{m})$ the bandwidth of one pair.

For an n-stage amplifier, n even, made up of $n/2$ flat-staggered pairs, the $\mathcal{G}\mathcal{B}$ factor is therefore

$$\frac{1}{1.1\sqrt[4]{\dfrac{n}{2}}}. \tag{25}$$

Dividing Expression (25) by Expression (11) one sees that the advantage in $\mathcal{G}\mathcal{B}$ factor of an n-stage flat-staggered-pair amplifier over an n-stage synchronous single-tuned amplifier is

$$\frac{1.2\sqrt{n}}{1.1\sqrt[4]{\dfrac{n}{2}}}. \tag{26}$$

Three elements contribute to this advantage: the $n/2$ instead of n, because there are only half as many $1/\sqrt{(1+x^4)}$ selectivity curves as there are stages; the fourth root instead of the square root, because of the $1/\sqrt{(1+x^4)}$ curve rather than the $1/\sqrt{(1+x^2)}$ curve; and the 1.1 instead of the 1.2.

Expression (26) has the value 2.0 for $n = 6$. This means that a six-stage amplifier in the form of three flat-staggered pairs has twice the $\mathcal{G}\mathcal{B}$ factor of a six-stage synchronous single-tuned amplifier; hence for the same over-all gain the six-stage flat-staggered-pair amplifier has twice the over-all bandwidth.

4·9. Flat-staggered Triples, in Detail. *Graphs.*—Figure 4·11 was prepared from Table 4·3 for the case $n = 3$. The use of Fig. 4·11 will be clear from the following example:

Suppose a flat-staggered triple with a 20.6 Mc/sec bandwidth and band center at 14.3 Mc/sec is to be designed. Then $f_0 = 14.3$ Mc/sec, $\mathcal{B} = 20.6$ Mc/sec, and $\delta = 1.44$, so that from Fig. 4·11, $\alpha = 1.84$ and $d = 0.60$.

Therefore, the triple is to be constructed from

One stage staggered at $14.3 \times 1.84 = 26.3$ Mc/sec, of dissipation factor 0.60 and hence of bandwidth $0.60 \times 26.3 = 15.8$ Mc/sec.

One stage staggered at $14.3/1.84 = 7.8$ Mc/sec, of dissipation factor 0.60 and hence of bandwidth $0.60 \times 7.8 = 4.7$ Mc/sec.

One stage centered at 14.3 Mc/sec of bandwidth 20.6 Mc/sec.

Cascaded Triples and Comparison with Synchronous Single-tuned Stages.—The selectivity function of a flat staggered triple has the form $1/\sqrt{(1+x^6)}$ in the low-dissipation case and $1\Big/\sqrt{\delta^6 + \left(f - \dfrac{1}{f}\right)^6}$ in
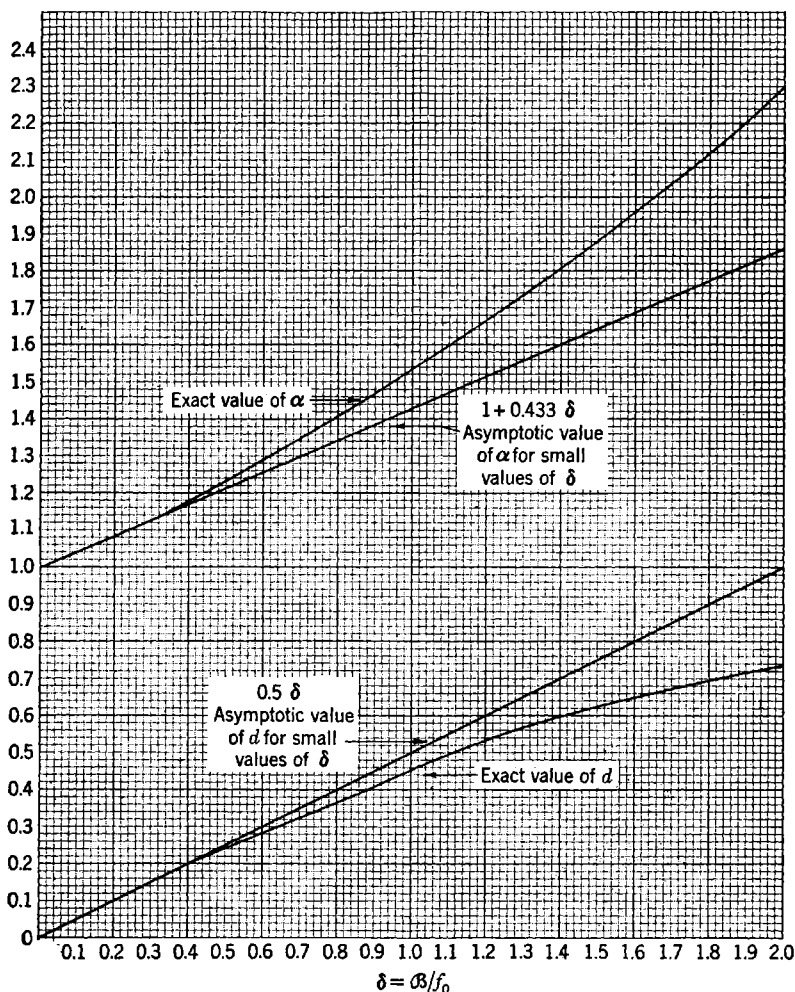
FIG. 4·11.—Design curves for an exact flat staggered triple.

NOTE:

1. An exact flat staggered triple of stage gain $\mathcal{G}$ has as great an over-all bandwidth $\mathcal{B}$ as does one single-tuned stage of gain $\mathcal{G}$, i.e., $\mathcal{G}\mathcal{B} = g_m/2\pi C$.

2. An exact flat staggered triple of over-all bandwidth $\mathcal{B}$, geometrically centered at f_0, consists of two single-tuned stages staggered at $f_0\alpha$ and f_0/α of dissipation factor d, and one single-tuned stage centered at f_0 of bandwidth $\mathcal{B}$, α being given in the upper graph and d in the lower graph as functions of $\delta = \mathcal{B}/f_0$.

3. An amplifier made up of m triples has $1/(1.06 \sqrt[6]{m})$ the bandwidth of one triple.

the exact case. These curves are squarer than those for a staggered pair; and because they are squarer, the bandwidth goes down less rapidly;[1] it is easy to see that the over-all bandwidth of m triples is the bandwidth of one triple times $\sqrt[6]{2^{1/m} - 1}$. For $m > 1$ a good approximation is

$$\text{Over-all bandwidth of } m \text{ triples} = \frac{\text{bandwidth of one triple}}{1.06 \sqrt[6]{m}}$$

(see Table 4·6).

TABLE 4·6.—SHRINKING OF OVER-ALL BANDWIDTH, m-CASCADED FLAT-STAGGERED TRIPLES

m	$\sqrt[6]{2^{1/m} - 1}$ (exact)	$1.06 \sqrt[6]{m}$ (approx.)
1	1.00	. . . .
2	0.86	0.84
3	0.80	0.79
4	0.76	0.75
5	0.73	0.72

For an n-stage amplifier, n divisible by 3, made up of $n/3$ flat-staggered triples, the $\mathcal{GB}$ factor is therefore

$$\frac{1}{1.06 \sqrt[6]{\frac{n}{3}}}. \tag{27}$$

Dividing Expression (27) by Expression (11), the advantage in $\mathcal{GB}$ factor of an n-stage flat-staggered-triple amplifier over an n-stage synchronous single-tuned amplifier is found to be

$$\frac{1.2 \sqrt{n}}{1.06 \sqrt[6]{\frac{n}{3}}}. \tag{28}$$

Expression (28) has the value 2.5 for $n = 6$. This means that a six-stage amplifier in the form of two flat-staggered triples has 2.5 times the $\mathcal{GB}$ factor of a six-stage synchronous single-tuned amplifier; hence for the same over-all gain the six-stage flat-staggered-triple amplifier has 2.5 times the over-all bandwidth.

4·10. Gain Control of Stagger-tuned Amplifiers. *First-order Effects.* In a cascaded linear amplifier without feedback the over-all gain at any frequency is the product of all the g_m's by the product of all the impedances at that frequency. Because of this multiplicative property, if one

---

[1] And the overshoot is larger: 8.15 per cent for one triple, 11.2 per cent for two triples, 14.2 per cent for four triples (see Fig. 7·6).

of the stages has its g_m reduced but its selectivity characteristic left unchanged, the over-all gain is reduced in proportion, whereas the over-all selectivity curve is completely unaffected. Consequently, to the extent to which tube capacities and loadings do not vary with g_m, it is possible to gain-control a stagger-tuned amplifier in any stage or combination of stages. In particular, there is no need to gain-control in pairs.

Second-order Effects.—To a certain extent it is however the case that reducing the g_m of a stage does affect its input capacity and input loading. Variations in input capacity can often be held within narrow limits by the well-known artifice of leaving unbypassed part of the cathode-bias resistor of the gain-controlled stages, and variations in input resistance can be held within narrow limits by choice of a suitable cathode-bypass condenser. In practice, these measures have not proved necessary.

Because of grid-plate capacity there is a variation in input loading and capacity due to the Miller effect. Rather bad offenders in this respect are 6AK5's, which have a ratio of transconductance to grid-plate capacity, socket included, that is about a third that of 6AC7's. Miller effect in 6AK5's is somewhat vexing, but it is not fatal in wide-band amplifiers even at frequencies as high as 200 Mc/sec.

4·11. Examples of Stagger-tuned Amplifiers. *Example* 1. *Nine-stage 80-db 6AC7 Flat-staggered Triple Amplifier Covering the Band 8.25 to 24.75 Mc/Sec.*—The geometric mean of 8.25 and 24.75 is 14.3 Mc/sec, which is therefore the geometric center of the band.

Preliminary estimates show that it is sufficient to use nine 6AC7 stages arranged in the form of three flat-staggered triples. From Table 4·6 it follows that the bandwidth per triple must be 20.6 Mc/sec in order to obtain the 16.5 Mc/sec over-all bandwidth.

The mean stage gain is most conveniently determined from the relation that the $\mathcal{GB}$ factor for a flat-staggered triple is equal to 1. Assuming a 6AC7 g_m of 9000 μmhos and a C of 25 $\mu\mu$f, the ratio

$$\frac{g_m}{2\pi C} = 57.3 \text{ Mc/sec};$$

consequently the mean stage gain is $57.3/20.6 = 2.78$, or 8.9 db.

The center frequencies and bandwidths of the stages making up the three triples have already been worked out in Sec. 4·9. The load resistors are determined [from the relation that stage bandwidth $= 1/(2\pi RC)$] to be 403, 1360, and 309 ohms respectively; the practical values are 390, 1500, and 330.

A block diagram of an amplifier of this sort is shown in Fig. 4·12.[1]

[1] This amplifier was built at the Radiation Laboratory, 1943, and required the theory of the case of large fractional bandwidth. The widest stagger-tuned amplifier constructed at the Radiation Laboratory had a bandwidth of 35 Mc/sec; the center

The order of frequencies is 14.3, 7.8, 26.3; 7.8, 26.3, 14.3; 26.3, 7.8, 14.3.   No special significance is to be attached to this order, although it seems wise to center both the first stage and the stage driving the detector. The tuning coils are fixed-tuned and wound on $\frac{1}{4}$-in. bakelite rods to resonate with 25 $\mu\mu$f at the appropriate frequencies, as measured with a

(a)

(b)
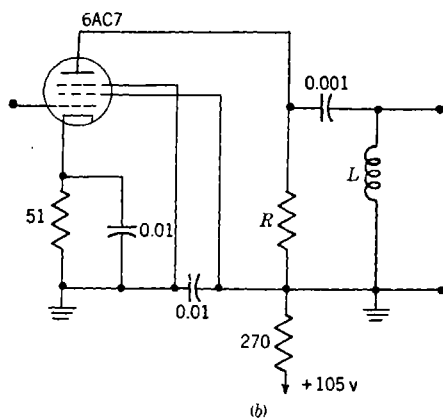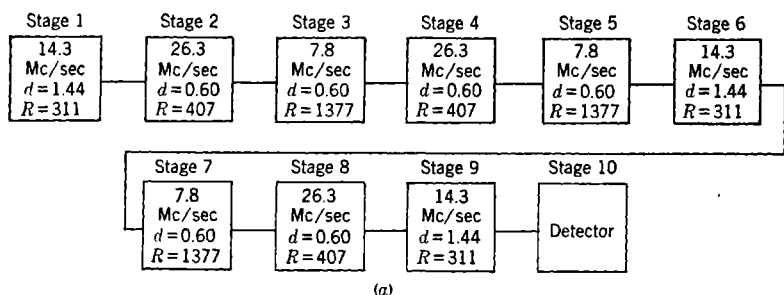
Fig. 4·12.—(a) Block diagram of nine-stage 80-db 6AC7 flat-staggered triple amplifier covering the band 8.25 to 24.75 Mc/sec.   (b) Typical stage.   Apart from the values of *R* and *L* the various amplifier stages have the same constants.

Q-meter.   Because of the large bandwidth/band center ratio, the tuning is very uncritical.   In all likelihood the coils could be wound from information contained in the inductance tables, and it would thus be possible to build amplifiers of this type with no equipment other than a soldering iron and a pair of pliers.

Example 2.   Four-stage 80-db 6AK5 Flat-staggered Quintuple, 10 Mc/Sec wide at 30 Mc/Sec.—The purpose of this amplifier, shown in

frequency was 80 Mc/sec, and 15 6AK5's were employed in three flat-staggered quintuples.   It was used as i.f. amplifier of an experimental radar system employing $\frac{1}{30}$ μsec pulses.   The gain was 100 db.

Figs. 4.13 and 4.14, is to get high performance in a small space. The amplifier is folded back on itself, with the input connection and the first three amplifier stages located on one half of the chassis and the fourth amplifier stage, detector, and a pulse amplifier stage on the other half. The two halves are separately wired, each half having great accessibility because of its open L-shaped chassis. The two halves are connected with metal bushings, and the cover is then put over the combination.

There is only one connection between the two halves of the chassis, a wire connecting the tuning coil between Stage 3 and Stage 4 to the grid of Stage 4. This wires goes through one of the metal bushings, the

Fig. 4.13.—Circuit diagram of four-stage 80-db 6AK5 flat-staggered quintuple, 10 Mc/sec wide at 30 Mc/sec. The coils are unity-coupled. The chassis shown in Fig. 4.14 contains a pulse amplifier stage also.

return path for the signal currents being conducted along the inside surface of the second chassis, through the inside surface of the metal bushing, to the inside surface of the first chassis. In folded amplifiers it is essential in connecting the two chassis together to adhere to the principle of providing for the signal currents between the chassis a return path that lies entirely on the inner surfaces of the chassis. No difficulties are experienced in building folded amplifiers of over 100-db gain when such precautions are taken.

The use of "feed-through" button bypass condensers,[1] mounted in the wall of the chassis, makes it possible to have only two components per amplifier stage in the interior of the chassis, namely, a tuning coil and a damping resistor. The decoupling resistors are situated on the other

[1] Obtainable from the Erie Resistor Manufacturing Corporation, Erie, Pa.

Fig. 4·14a.—Bottom chassis view of four-stage 80-db 6AK5 flat-staggered quintuple, 10 Mc/sec wide at 30 Mc/sec.

side of the vertical chassis wall, as can be seen in the outer view of Fig. 4·14.

The tuning coils are on powdered iron cores, fixed-tuned and bifilar-wound, i.e., unity-coupled, thus eliminating the need for blocking condensers. Although there are only four stages, there are five tuning coils, the input coil, which is the center-tuned coil of the quintuple, being the fifth.

The amplifier is designed according to the entry for $n = 5$ in Table 4·3, which leads to the values of center frequency and bandwidth shown in Fig. 4·13. To achieve a gain of 80 db in four stages requires a mean
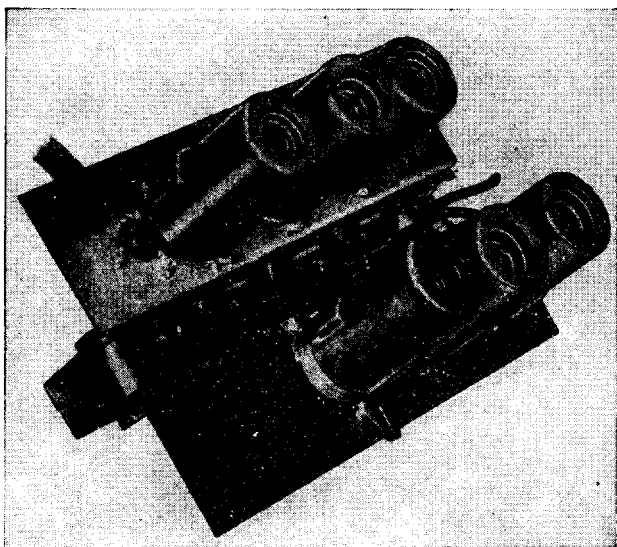
Fig. 4·14b.—Top chassis view of four-stage 80-db 6AK5 flat-staggered quintuple, 10 Mc/sec wide at 30 Mc/sec.

voltage gain of 10 per stage. Since the ᏀᏔ-product for a flat-staggered quintuple is equal to $g_m/2\pi C$, an over-all bandwidth of 10 Mc/sec requires the 6AK5's to be operated with a $g_m/2\pi C$ ratio of 100 Mc/sec. In the amplifier of Fig. 4·13 the interstage capacity C is 10 $\mu\mu$f; hence the tubes have to be operated with a g_m of 6280 μmhos. This value is above the normal rated 6AK5 value, but the special conditions of use of this amplifier make it legitimate to overrun the tubes.

Example 3. *Twelve-stage* 100-db 6AK5 *Amplifier, Made Up of Four Flat-staggered Triples,* 20 *Mc/Sec Wide at* 200 *Mc/Sec.*—This amplifier was designed and built by M. T. Lebenbaum of Radio Research Laboratory. The rough design of such an amplifier will first be worked through, to show how one determines whether or not a given number of stages is

enough to provide the desired gain and bandwidth.   For a 20-Mc/sec over-all bandwidth in four triples, the bandwidth per triple has to be 26.3 Mc/sec (20/0.76), from Table 4·6.   To achieve a gain in 12 stages of 100 db the required stage gain is 8.33 db = 2.61.   The required $g \mathfrak{G}$-product for each triple is then $2.61 \times 2.63 = 68.5$ Mc/sec.   Therefore a $g_m/2\pi C$ ratio of 68.5 Mc/sec is needed; and if $C = 11 \ \mu\mu f$, a g_m of 4750 μmhos is required.   This is a reasonable value, slightly less than the nominal g_m for 6AK5's, and hence validates the rough design.

The problem is now reduced to obtaining a flat-staggered triple of 26.3 Mc/sec bandwidth at 200 Mc/sec.   The value of $\delta = \mathfrak{G}/f_0$ is 0.132, which is small enough to warrant use of Table 4·4.   From this table, one finds that the triple should be composed of one stage at

$$200 + (0.43 \times 26.3) = 211.3 \ \text{Mc/sec},$$

of dissipation factor $0.5 \times 0.132 = 0.066$ and hence of bandwidth $0.066 \times 211.3 = 13.9$ Mc/sec; one stage at

$$200 - (0.43 \times 26.3) = 188.7 \ \text{Mc/sec},$$

of dissipation factor $0.5 \times 0.132 \approx 0.066$ and hence of bandwidth $0.066 \times 188.7 = 12.4$ Mc/sec; and one stage at 200 Mc/sec of bandwidth 26.3 Mc/sec.

The damping resistances for these bandwidths and an interstage capacity of 11 $\mu\mu$f are 1040, 1160, and 550 ohms.   A distinction has to be made, however, between the damping resistances and the actual plate-load resistors, especially at frequencies as high as 200 Mc/sec.   In parallel with the tangible load resistors are three types of tube input resistance due to (1) transit-time effects, (2) cathode-lead inductance, and (3) "Miller-effect" feedback through the grid-plate capacitance from the plate circuits.   The first two of these resistances are of positive sign in the amplifiers discussed here and vary inversely as the square of frequency.   Since with 6AK5's the input resistance from these two effects is about 100,000 ohms at 30 Mc/sec, its influence is negligible at that frequency; at 200 Mc/sec, however, the input resistance is only about 2000 ohms and its influence is considerable.   Even more important in stagger-tuned amplifiers is the Miller-effect input resistance, because it may have either sign.   This input resistance measured at the grid-circuit resonant frequency is positive for a tube whose plate circuit is tuned to a frequency lower than that of the grid circuit and negative in the opposite case.   At 200 Mc/sec this effect is serious in 6AK5's because of the large grid-plate susceptance.   In such a case, moreover, the input resistance varies rapidly over the band, and the individual circuits are not really single-tuned.   An exact analysis is extremely complicated, but it has been found possible to get satisfactory results, even at 200 Mc/sec, by

experimental determination of the plate-load resistor values needed to give the required individual bandwidths. It would probably be difficult to go much beyond 200 Mc/sec with 6AK5's, however.

In the case at hand the constants of a triple are shown in Fig. 4·15; observe that the only nontube loading in the 211 Mc/sec circuit is the loss resistance, because of finite Q, of the coil connecting the plate to $B+$.

The unbypassed 10-ohm resistors between the $B+$ lead and the screen terminals suppress a tendency of 6AK5's to oscillate parasitically at about 500 or 600 Mc/sec when one employs high-Q bypass condensers such as the mica button type. Although these parasitic oscillations
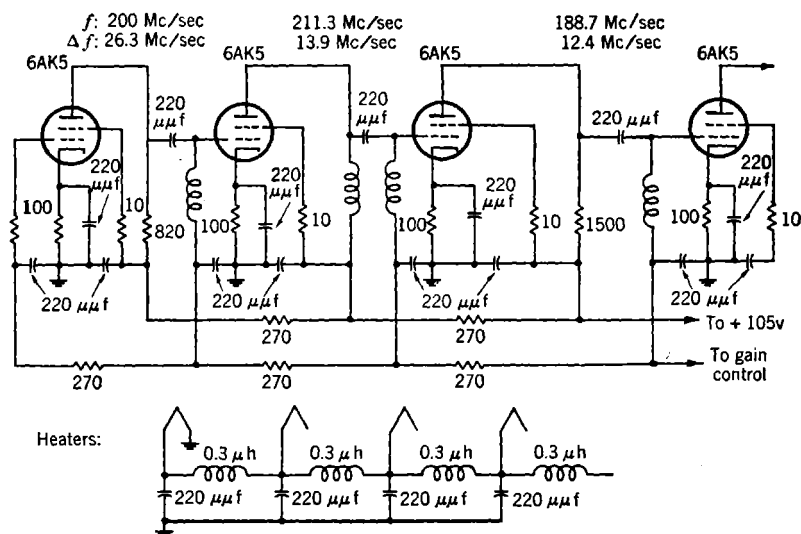
Fig. 4·15.—Circuit diagram of one triple of 12-stage 100-db 6AK5 amplifier made up of four flat-staggered triples, 20 Mc/sec wide at 200 Mc/sec.

occur only when the $B+$ voltage is at least 30 per cent above its rated value, it seems wise to be cautious. This matter of parasitic oscillations constitutes, incidentally, an argument against bypass condensers of high-Q.

Figure 4·16 is a photograph of a few stages of this amplifier. The amplifier is tunable, and the tuning coils are of special interest. The coils are essentially small springs of phosphor-bronze wire which are wound, as shown in Fig. 4·16, on $\frac{1}{8}$-in. diameter linen-bakelite rods. These rods are backed by other springs, not shown, and connected to screws on the upper side of the chassis so as to permit extension or retraction of the bakelite rod into the underside of the chassis. This action has the effect
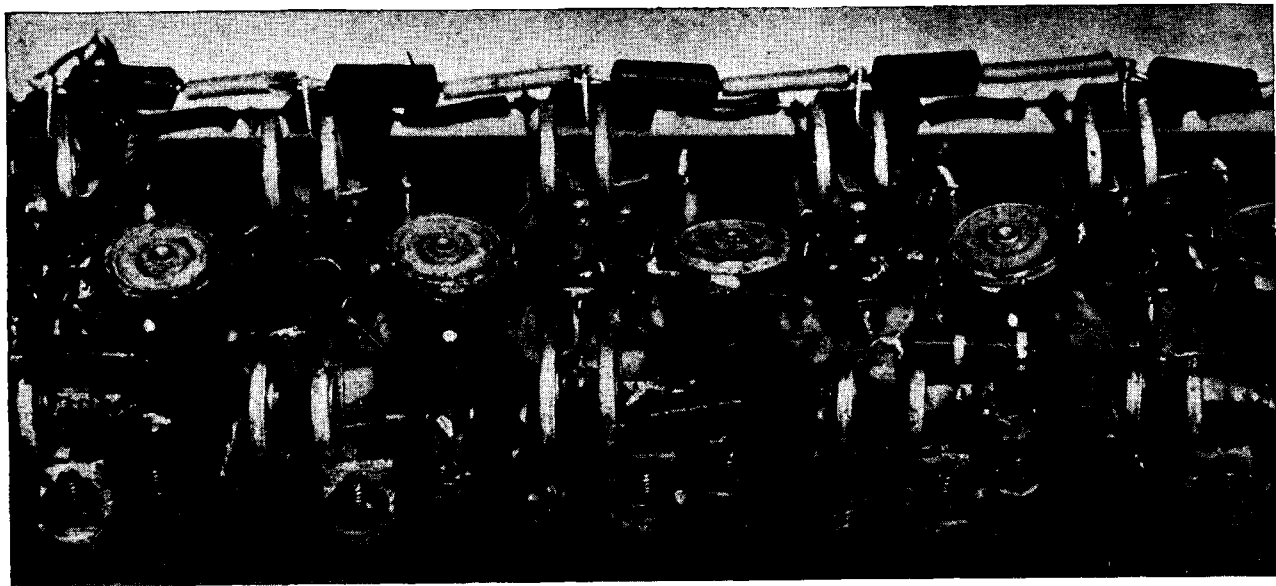
Fig. 4·16.—Bottom view of a few stages of a 12-stage 100-db 6AK5 amplifier made up of four flat-staggered triples, 20 Mc/sec wide at 200 Mc/sec.

of increasing or decreasing the spacing between coil turns, thereby varying the coil inductance.

The alignment procedure is very simple and consists of connecting a signal generator to the amplifier input terminals and a voltmeter across the detector output terminals and then peaking each stage for maximum meter deflection, with the signal generator set to the frequency appropriate to that stage. The order in which the peaking is done is immaterial. It is easy to see how this procedure could be further simplified by a scheme of matching a color code of the adjustment screws to a color code of the signal-generator settings.

The reason for making this amplifier tunable is the wide variation at 200 Mc/sec of the resonant frequency of the individual circuits, principally because of the variability in 6AK5 interelectrode capacities. The sum of the variability in 6AK5 input and output capacities permitted by the JAN-1A specifications is ± 0.9 $\mu\mu$f. If the average interstage capacity is 11 $\mu\mu$f, there is a fractional variability in capacity of about $\pm \frac{1}{12}$; this means a variability in center frequency of about $\pm \frac{1}{24}$, which is about ± 8 Mc/sec at 200 Mc/sec. Although such extreme variations in capacity, and hence in tuning, are very unlikely, variations half as large, i.e., ± 4 Mc/sec at 200 Mc/sec, are entirely likely. Such variations are too large to be neglected in a 20-Mc/sec wide amplifier; hence the provision for tuning.

In summary one can say that a 20-Mc/sec wide amplifier at 200 Mc/sec is more difficult to build than a 20-Mc/sec wide amplifier at 60 Mc/sec in only one respect, namely, the need for tuning the individual stages of the 200-Mc/sec amplifier.

CHAPTER 5

DOUBLE-TUNED CIRCUITS

BY RICHARD Q. TWISS

5·1. Introduction.—The theory of the double-tuned interstage coupling has been extensively discussed by a large number of writers[1] for the broadcast receiver case, where the fractional bandwidth is small (less than 0.1) and the Q's of the primary and secondary circuits are large (greater than 20). In this chapter, however, attention will be confined
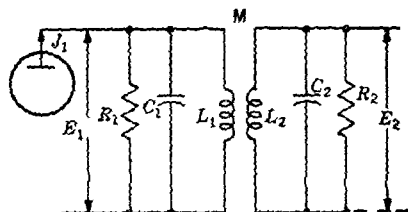
FIG. 5·1a.—Inductance-coupled double-tuned circuit.
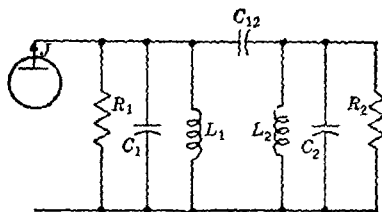
FIG. 5·1b.—Capacity-coupled double-tuned circuit.

to the wide-band fixed-tuned case (bandwidths in excess of 4 Mc/sec, fractional bandwidths in excess of 0.1), and only two types of coupling will be considered, the inductance-coupled circuit of Fig. 5·1a and the capacity-coupled circuit of Fig. 5·1b. Of these the inductance-coupled circuit is the more useful, for the reasons given in Sec. 5·2, and unless otherwise stated it will be assumed that this circuit is being discussed.

---

[1] A very useful and complete account is given by C. B. Aiken, "Two Mesh Tuned Coupled Circuit Filters," *Proc. I.R.E.*, **35**, February 1937. Considerable use will be made of the results of this paper in the present chapter; and as far as possible, the same symbols will be used. Reference may also be made to F. E. Terman, *Radio Engineers Handbook*, 1st ed., McGraw-Hill, New York, 1943, Sec. 3.

The approximate theory, mentioned above and hereafter called the "high-Q" theory, still yields very useful information when applied to circuits whose Q may be as low as 1. This is fortunate, since the more exact low-Q case is very cumbersome mathematically and has been fully worked out only for the case of transitional coupling. The approximate theory yields almost exact expressions for the gain-bandwidth factor and midband gain of a double-tuned circuit. In addition it gives values for the Q's and the coefficient of coupling needed to achieve a desired bandwidth and shape of amplitude response that are useful first approximations. The exact low-Q theory must, however, be used to find the correct resonant frequencies of the tuned circuits, the exact values of Q, and the coefficient of coupling. Alternatively, the high-Q theory can be used to find the approximate values of the circuit constants, and the correct values can then be found experimentally by a cut-and-try process.

Some of the more important results of the high-Q theory are given in Sec. 5·2, and an outline of the low-Q theory is given in Sec. 5·5. The latter section consists of a set of a design data for the transitionally coupled case.

In Sec. 5·6 is given a brief discussion of the so-called "stagger-damped" bandpass amplifier; this is a scheme for obtaining very wide-band amplifiers by over- and underdamping successive stages of double-tuned circuits. The chapter concludes with a number of examples chosen to illustrate the theory.

5·2. The General High-Q Case.—The basic circuit upon which the analysis is based is that of Fig. 5·1a. The results can also be applied to the circuit of Fig. 5·1b by suitable modification of the basic formulas.

The following symbols will be used, the notation being essentially that of Aiken.[1]

f_0, the midband frequency.

$$2\pi f_0 = \omega_0 = \frac{1}{\sqrt{L_1 C_1}} = \frac{1}{\sqrt{L_2 C_2}}.$$

$k = \dfrac{M}{\sqrt{L_1 L_2}}$, the coefficient of coupling.

$s = k\sqrt{Q_1 Q_2}$, the coupling index.

$\rho = Q_1/Q_2$, the Q ratio.

$b = \rho + 1/\rho = (Q_1/Q_2) + (Q_2/Q_1)$.

$\beta = (s^2 - \tfrac{1}{2}b)/(s^2 + 1)$, the shape index.

$v = \sqrt{Q_1 Q_2}\,(\omega/\omega_0 - \omega_0/\omega)$, the frequency variable.

J, the primary current.

[1] C. B. Aiken, "Two Mesh Tuned Coupled-circuit Filters," *Proc. I.R.E.*, **25**, No. 2, February 1937. In the remainder of this chapter this paper will be referred to simply as Aiken.

E_1, the primary voltage.

E_2, the secondary voltage.

It is shown by Aiken that Z_{12}, the transfer impedance, is given approximately by

$$Z_{12} = \frac{E_2}{J} = \frac{js \sqrt{R_1 R_2}}{(1 + s^2) + jv \sqrt{b + 2} + (jv)^2}, \tag{1}$$

so that the absolute value of Z_{12} is given by

$$|Z_{12}| = \left|\frac{E_2}{J}\right| \frac{s \sqrt{R_1 R_2}}{\left[(1 + s^2)^2 - 2\left(s^2 - \frac{b}{2}\right)v^2 + v^4\right]^{\frac{1}{2}}}, \tag{2}$$

when Q_1, $Q_2 > 20$ and $k < 0.05$.

One of the most common uses for the double-tuned circuit is as interstage network in a bandpass amplifier. If J is the output current of a tube whose dynamic resistance r_p is included in R_1, then J is related to the grid-cathode voltage of the tube by the equation

$$J = -g_m e_g,$$

where g_m is the transconductance and e_g the grid-cathode voltage. The expression for the absolute value of the stage voltage gain G can therefore be written

$$\mathsf{G} = \left|\frac{E_2}{e_g}\right| = \frac{g_m \sqrt{R_1 R_2}\, s}{\left[(1 + s^2)^2 - 2\left(s^2 - \frac{b}{2}\right)v^2 + v^4\right]^{\frac{1}{2}}}. \tag{3}$$

Equations (2) or (3), which give the amplitude response of the network, are of fundamental importance and yield most of the useful information needed in the design of amplifiers employing high-Q double-tuned circuits.

Universal Resonance Curves.—Aiken has utilized Eq. (1) to plot universal resonance curves for $2|Z_{12}|/\sqrt{R_1 R_2}$ for four different values of b, namely $b = 2, 10, 50,$ and 200. In this chapter universal curves are plotted for the two cases $Q_1 = Q_2 (b = 2)$ and $Q_1 = \infty (b = \infty)$ as discussed in Secs. 5·3 and 5·4.

By suitable transformation of the variables it is possible to provide a single set of resonance curves that provide all the essential data for the amplitude characteristic. If a new variable u and a new parameter β defined by

$$\left.\begin{aligned} u &= \frac{v}{(1 + s^2)^{\frac{1}{2}}} = \sqrt{Q_1 Q_2}\left(\frac{\omega}{\omega_q} - \frac{\omega_0}{\omega}\right)\frac{1}{(1 + s^2)^{\frac{1}{2}}} \\ \beta &= \frac{s^2 - \frac{1}{2}b}{1 + s^2} \end{aligned}\right\} \tag{4}$$

are introduced,[1] then Eq. (1) can be put in the form

$$|Z_{12}| = \frac{s \sqrt{R_1 R_2}}{\sqrt{1 + s^2}} \frac{1}{(1 - 2\beta u^2 + u^4)^{\frac{1}{2}}};$$ (5)

and if $|Z_{12}| \sqrt{1 + s^2}/s \sqrt{R_1 R_2}$ is plotted as a function of u for various values of β, the required resonance curves are obtained.

It may be noted in passing that these curves also give the amplitude response of staggered pairs and negative feedback pairs if the parameters are suitably redefined.

Equation (5) has not been used to plot universal resonance curves in this chapter because the increased difficulty of interpreting the curves outweighs the advantages gained from displaying the information on a single family of curves, but Eq. (5) makes it clear that the shape, as opposed to the scale of the amplitude response, is determined by the single parameter β.

Critical and Transitional Coupling.—From Eq. (2) it may be seen that the midband gain $\mathcal{G}_0$ is given by

$$\mathcal{G}_0 = \frac{g_m \sqrt{R_1 R_2}\, s}{1 + s^2},$$ (6)

which depends only upon the coupling index s and not upon the Q ratio ρ. The value of k that, for fixed Q's, makes the gain $\mathcal{G}_0$ a maximum is called the *critical coefficient of coupling;* and when k has this value, the circuit is said to be *critically coupled.* From Eq. (6) the corresponding value of s is

$$s = 1,$$

so that, when the circuit is critically coupled, the midband gain is

$$\frac{g_m \sqrt{R_1 R_2}}{2}$$ (7)

---

[1] It may be seen from Eq. (4) that as $s^2 \rightarrow \infty$ (the overcoupled case), $\beta \rightarrow 1$ for all b, that is, for all Q ratios. As $s^2 \rightarrow 0$ (the undercoupled case), however, $\beta \rightarrow -\frac{1}{2}b$, which does depend on the Q-ratio. The explanation for this is simple. As k, and therefore s, tends to zero the selectivity curve approaches that of the product of the two single-tuned circuits whose Q's are the primary and secondary Q's respectively. In the equal-Q case this selectivity curve is the product of the selectivity curves of two single-tuned circuits with the same Q, and it is not possible to get a sharper response curve than this. In the unequal-Q case the selectivity curve can become much sharper; if the product of the Q's is kept constant, then when one of the Q's is infinite, the over-all bandwidth is zero. In this chapter attention is confined to the case $-1 \leqq \beta \leqq 1$, because cases where $\beta < -1$ are of little practical interest.

and the coefficient of coupling is

$$k = \frac{1}{\sqrt{Q_1 Q_2}} = \text{geometrical mean of } \frac{1}{Q_1} \text{ and } \frac{1}{Q_2}. \tag{8}$$

Equation (7) gives the maximum attainable value for the midband gain.

In practice it is customary to use not that value of k which gives maximum midband gain but instead that value which gives the flattest selectivity curve. From Eq. (2) the corresponding value of s is given by

$$s^2 = \tfrac{1}{2}b, \tag{9}$$

since in this case

$$\frac{d|Z_{12}|}{dv} = \frac{d^2|Z_{12}|}{dv^2} = \frac{d^3|Z_{12}|}{dv^3} = 0$$

when $v = 0$. When Eq. (9) is satisfied, the circuit is said to be *transitionally coupled*, because if the coupling coefficient is increased beyond the transitional value, the curve has two peaks, whereas for values of k below transitional the curve has a single peak. The transitional value of the coupling coefficient is

$$k = \frac{1}{\sqrt{Q_1 Q_2}} \sqrt{\frac{b}{2}} = \sqrt{\frac{1}{2}\left(\frac{1}{Q_1^2} + \frac{1}{Q_2^2}\right)} = \text{rms of } \frac{1}{Q_1} \text{ and } \frac{1}{Q_2}. \tag{10}$$

When the Q's are equal, $b = 2$, and transitional and critical coupling coincide. For all other values of Q, however, transitional coupling is greater than critical, since the rms of two unequal quantities is greater than the geometric mean.

The bandwidth $\mathfrak{B}$ between 3-db points in the transitionally coupled case, as determined from Eq. (3), is

$$\mathfrak{B} = \frac{\sqrt{1 + s^2}}{2\pi \sqrt{C_1 C_2} \sqrt{R_1 R_2}}. \tag{11}$$

Gain-bandwidth Factor.—It has been shown above that the general shape of the selectivity curve depends only upon a single parameter β, defined by Eq. (4). Accordingly when comparing circuits with different Q-ratios but with the same values of β (that is, the same peak to midband ratios) and the same bandwidths, it is sufficient to compare the midband gains. The ratio of the responses at midband is then equal to the ratio of the responses at any other frequency. A more general basis of comparison is given by the ratio of the product of midband gain and bandwidth when the circuits have the same values of β. Because of the simplicity of the resulting expressions, the bandwidth is usually taken at the half-power points. The product of midband gain and bandwidth

is, in this case, called the "gain-bandwidth product" of the circuit. In the important case of transitional coupling, this gain-bandwidth product is given by

$$\mathcal{G}\mathcal{B} = \frac{g_m}{2\pi 2 \sqrt{C_1 C_2}} \frac{2s}{\sqrt{1 + s^2}}, \tag{12}$$

where $s^2 = \frac{1}{2}b = \frac{1}{2}(Q_1/Q_2 + Q_2/Q_1)$.

Now $g_m/(2\pi 2 \sqrt{C_1 C_2})$ is the gain-bandwidth product of a single-tuned circuit if a tapped coil is used to match the plate capacity to the grid
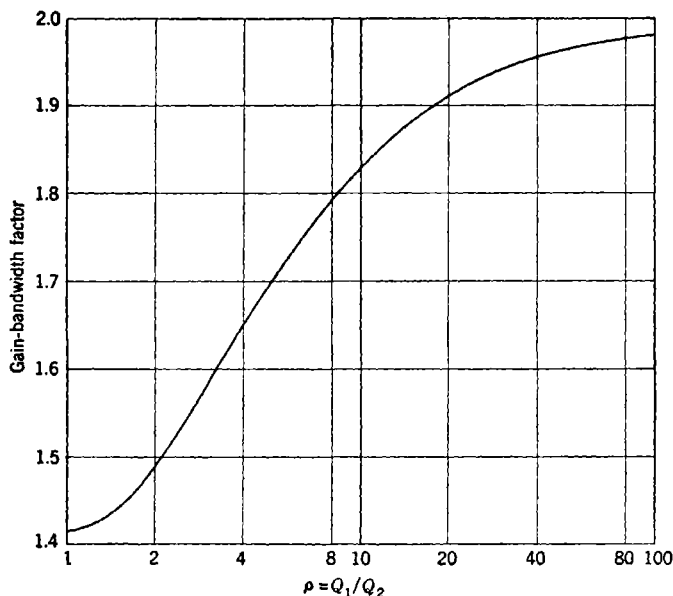
Fig. 5·2.—Gain-bandwidth factor of transitionally coupled double-tuned circuit as function of Q-ratio.

capacity, thus a convenient dimensionless figure of merit is the so-called "gain-bandwidth factor," the ratio of the gain-bandwidth product of the particular coupling circuit to the gain-bandwidth product of a single-tuned coupling circuit. On this basis the gain-bandwidth factor for a single-tuned circuit is unity and for a transitionally coupled double-tuned circuit is

$$\frac{2s}{\sqrt{1 + s^2}}, \tag{13}$$

where s is given by Eq. (2). If s is eliminated between Eqs. (2) and (13), the gain-bandwidth factor can be shown to be equal to

$$\frac{2\sqrt{1+\rho^2}}{(1+\rho)}. \tag{14}$$

When $\rho = 1$, the equal-Q case, the gain-bandwidth factor is $\sqrt{2}$; as ρ tends to infinity, this factor increases monotonically to 2. Expression (14) is plotted as a function of ρ in Fig. 5·2.

When the coupling is not transitional, the expression for the bandwidth is complicated and the general expression for the gain-bandwidth product will not be computed. It is easy, however, to compare the effect of different Q-ratios on midband gain of circuits having the same shape index β and the same bandwidth.

The midband gain in the general case is

$$\mathcal{G}_0 = \frac{g_m s \sqrt{R_1 R_2}}{1 + s^2}. \tag{15}$$

The shape index β is assumed constant where

$$\beta = \frac{s^2 - \frac{1}{2}b}{s^2 + 1}; \tag{16}$$

and since the bandwidth is also assumed constant, it is necessary that

$$\gamma^2 = \frac{Q_1 Q_2}{1 + s^2} \tag{17}$$

be constant, as may be seen from Eqs. (4) and (5).

From Eqs. (15), (16), and (17) it can be shown that

$$\mathcal{G}_0 = \frac{g_m \gamma}{2\pi f_0 \sqrt{C_1 C_2}} \left(\frac{\beta + \frac{1}{2}b}{1 + \frac{1}{2}b} \right)^{\frac{1}{2}}. \tag{18}$$

It has been shown above that β is never greater than unity, so that $\mathcal{G}_0$ increases monotonically with $\frac{1}{2}b$. When $\frac{1}{2}b = 1$, the equal-Q case,

$$\mathcal{G}_0 = \frac{g_m \gamma}{2\pi f_0 \sqrt{C_1 C_2}} \left(\frac{1 + \beta}{2} \right)^{\frac{1}{2}},$$

and when Q_1 is infinite, so that $\frac{1}{2}b = \infty$,

$$\mathcal{G}_0 = \frac{g_m \gamma}{2\pi f_0 \sqrt{C_1 C_2}}.$$

The ratio of the midband gain in the two cases is simply

$$\left(\frac{1 + \beta}{2} \right)^{\frac{1}{2}}. \tag{19}$$

For strongly overcoupled circuits, where $\beta \approx 1$, Expression (19) shows that the unequal-Q case gives no larger gain than the equal-Q case.

For very weakly coupled circuits, on the other hand, where $\beta \approx -1$, the unequal-Q circuit has a considerable advantage.[1] At transitional coupling, where $\beta = 0$, the ratio is $1/\sqrt{2}$, a result already obtained above.

The increased gain-bandwidth factor offered by the unequal-Q transitionally coupled case is not obtained without having to pay a considerable price, for it will now be shown that the amplitude response of the unequal-Q coupling is liable to develop serious asymmetry under conditions of slight mistuning.

The Effect of Mistuning.—If the secondary circuit is mistuned from midband by an amount $+\Delta f$, where Δf is small, then from Eq. (116) in Aiken the transfer impedance can be written

$$Z_{12} = \frac{s\sqrt{R_1 R_2}}{\left[(1 + s^2 + \omega_d^2)^2 + \omega_d^2(b-2) + 2\omega_d v\left(\dfrac{Q_2}{Q_1} - \dfrac{Q_1}{Q_2}\right) \right.}$$
$$\left. - 2v^2\left(s^2 + \omega_d^2 - \dfrac{b}{2}\right) + v^4 \right]^{1/2}, \quad (20)$$

where $\omega_d = \sqrt{Q_1 Q_2}\, \Delta f/f_0$.

It will be noted that there is now a term in the first power of v in the denominator, so that the resonance curve is asymmetrical unless $Q_1 = Q_2$. The greater the Q-ratio the more serious is the asymmetry produced by mistuning.

Aiken gives a number of curves illustrating the magnitude of this effect in particular cases.

Variation of k with Q-ratio When the Bandwidth Is Fixed.—Another feature of the unequal-Q case, which may prove undesirable, is that the coefficient of coupling required for a given fractional bandwidth increases with the Q-ratio, according to the same law as does the gain-bandwidth factor. From Eq. (11)

$$\frac{\mathcal{B}}{f_0} = \frac{k\sqrt{1 + s^2}}{s},$$

so that $k = (\mathcal{B}/f_0)(s/\sqrt{1 + s^2})$, which, but for the factor $\mathcal{B}/(2f_0)$, is equal to the gain-bandwidth factor given by Eq. (13). Accordingly as ρ increases from 1 to ∞, k increases monotonically from $\mathcal{B}/(f_0\sqrt{2})$ to $\mathcal{B}/f_0$.

Despite the more critical dependence upon tuning and the larger coefficient of coupling, the unequal-Q coupling has a number of important applications, particularly as an interstage coupling circuit in wide-band amplifiers where the greater gain-bandwidth factor is important and as the input network of an amplifier where the addition of extra resistance (to equalize the primary and secondary Q's) would make the noise figure seriously worse.

[1] See footnote to Eq. (4).

In these applications it is usually desirable to make the Q-ratio as high as possible. Hence the special case where one of the Q's is infinite is discussed in Sec. 5·4. The equal-Q case is discussed in Sec. 5·3.

Gain-bandwidth Factor of an n-stage Amplifier.—The gain-bandwidth factor of an n-stage amplifier is defined as $2\pi 2\sqrt{C_1 C_2}/g_m$ times the product of the *over-all* 3-db bandwidth and the *mean* stage gain. For the transitionally coupled double-tuned circuit the ratio of over-all 3-db bandwidth to 3-db bandwidth of a single stage is

$$(2^{1/n} - 1)^{1/4} \approx \frac{(\ln 2)^{1/4}}{n^{1/4}} = \frac{1}{1.1 n^{1/4}};$$

the approximate expression is very accurate for large n and is within 10 per cent of the correct value even for $n = 2$.

Accordingly the gain-bandwidth product of an n-stage transitionally coupled double-tuned circuit is

$$\frac{g_m}{2\pi 2\sqrt{C_1 C_2}} \frac{2s}{\sqrt{1 + s^2}} (2^{1/n} - 1)^{1/4}, \tag{21}$$

and the gain-bandwidth factor is

$$\frac{2s}{\sqrt{1 + s^2}} (2^{1/n} - 1)^{1/4} \approx \frac{2s}{\sqrt{1 + s^2}} \frac{1}{1.1 n^{1/4}}, \tag{22}$$

where s is given in terms of the Q-ratio by Eq. (9).

The Driving-point Admittance.—A quantity of particular importance in noise-figure calculations is the driving point admittance Y_{11} defined by

$$Y_{11} = \frac{J}{E_1}.$$

Aiken has shown that

$$Y_{11} = \frac{(1 + s^2 - v^2) + jv\sqrt{b + 2}}{R_1 \left(1 + j\sqrt{\dfrac{Q_2}{Q_1}}\, v\right)}, \tag{23}$$

which can be expressed as the sum of a conductance G_{11} and a susceptance B_{11}, where

$$\left.\begin{aligned}
G_{11} &= \frac{1}{R_1} \frac{(1 + s^2) + \dfrac{Q_2 v^2}{Q_1}}{1 + \dfrac{Q_2 v^2}{Q_1}} = \frac{1}{R_1} + \frac{s^2}{R_1 \left(1 + \dfrac{Q_2 v^2}{Q_1}\right)}, \\
B_{11} &= \frac{1}{R_1} \frac{\sqrt{\dfrac{Q_2}{Q_1}}\, v \left[1 + \dfrac{Q_1}{Q_2} + v^2 - (1 + s^2)\right]}{1 + \dfrac{Q_2 v^2}{Q_1}}.
\end{aligned}\right\} \tag{24}$$

The absolute value of Y_{11} is given by

$$|Y_{11}| = \frac{1}{R_1} \frac{\left[(1 + s^2)^2 - 2\left(s^2 - \frac{b}{2}\right)v^2 + v^4 \right]^{\frac{1}{2}}}{\left(1 + \frac{Q_2 v^2}{Q_1} \right)^{\frac{1}{2}}} \tag{25}$$

$$= \frac{s R_2}{\sqrt{R_1 R_2}\, |Z_{12}| \left[1 + Q_2^2 \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right)^2 \right]^{\frac{1}{2}}}; \tag{26}$$

the absolute value of the primary driving-point admittance is thus equal to $s/\sqrt{R_1 R_2}$ times the absolute value of the ratio of the transfer admittance Y_{12} to the driving-point admittance of a single-tuned circuit (the secondary circuit). This last result can be used to provide universal resonance curves for the driving-point admittance if so desired. In Aiken curves of the driving-point impedance are given for the equal-Q case, and it is there shown, as can be deduced by inspection of Eq. (25). that the absolute value of the driving-point impedance is double-humped even when the circuit is transitionally coupled.

Capacity-coupled Circuit.—In the small fractional-bandwidth case the gain-bandwidth factor of the circuit of Fig. 5·1b is the same as that of the inductively coupled circuit, and the results of this section apply to this case if the coefficient of coupling is defined by

$$k = \frac{C_{12}}{\sqrt{(C_1 + C_{12})(C_2 + C_{12})}} \approx \frac{C_{12}}{\sqrt{C_1 C_2}}.$$

The chief drawback to the capacity-coupled circuit, however, is that the inductances L_1 and L_2 must resonate with capacities larger than C_1 and C_2, thereby reducing the gain-bandwidth factor. It is approximately the case that L_1 must resonate with C_1 in parallel with the series combination of C_{12} and C_2 and L_2 must resonate with C_2 in parallel with the series combination of C_{12} and C_1. The gain-bandwidth factor for this circuit is therefore multiplied by a factor approximately equal to

$$\sqrt{\frac{C_1 C_2}{(C_1 + C_{12})(C_2 + C_{12})}} \approx \frac{1}{1 + k}.$$

For fractional bandwidths with $k > 0.1$ the loss in gain-bandwidth is appreciable and is the chief theoretical reason for rejecting the capacity-coupled circuit in favor of the inductively coupled circuit.

5·3. The High-Q, Equal-Q Case.—When $Q_1 = Q_2 = Q$ the transfer impedance is

$$|Z_{12}| = \frac{s \sqrt{R_1 R_2}}{[(1 + s^2)^2 + 2(s^2 - 1)v^2 + v^4]^{\frac{1}{2}}}$$

where $R_1 C_1 = R_2 C_2$, and $v = Q(\omega/\omega_0 - \omega_0/\omega)$.

Universal resonance curves for this case are given in Fig. 5·3, where $2|Z_{12}|/\sqrt{R_1 R_2}$ is plotted, as a function of v, for various values of k. It will be noticed that all curves for coupling coefficients greater than unity have the same maximum value, a feature peculiar to the equal-Q case.

The more important properties of the circuit are stated below. They can be obtained from the general results of Sec. 5·2 by insertion of the special values $Q_1 = Q_2$, $b = 2$.
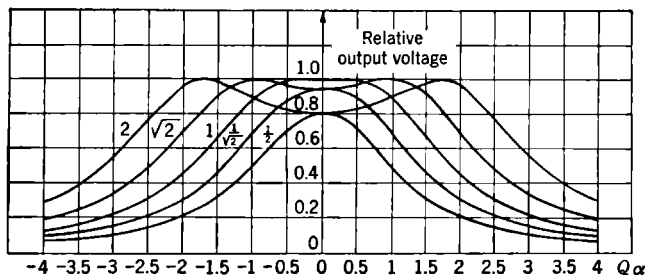
FIG. 5·3.—Absolute value vs. frequency curves of high-Q equal-Q double-tuned circuits as function of coefficient of coupling. The parameters associated with the various curves are the values of kQ.

1. Gain-bandwidth product of single stage

$$= \sqrt{2} \, \frac{g_m}{2\pi 2 \, \sqrt{C_1 C_2}}. \tag{27}$$

2. The transitional coefficient of coupling $= 1/Q =$ critical coefficient of coupling.

3. Gain-bandwidth factor of single stage $= \sqrt{2}$.

4. Gain-bandwidth factor of n stages

$$= \sqrt{2} \, (2^{1/n} - 1)^{1/4} \approx \frac{\sqrt{2}}{1.1 n^{1/4}}. \tag{28}$$

5. Bandwidth at half-power points with transitional coupling:

$$\mathfrak{B} = \frac{f_0 \sqrt{2}}{Q} = \frac{\sqrt{2}}{2\pi RC}, \tag{29}$$

where $C = \sqrt{C_1 C_2}$, $R = \sqrt{R_1 R_2}$; hence

$$k = \frac{1}{Q} = \frac{1}{\sqrt{2}} \frac{\mathfrak{B}}{f_0}. \tag{30}$$

6. The voltage gain at midband with transitional coupling:

$$\mathfrak{G} = \frac{g_m R}{2}.$$

7. Midband value of driving-point impedance:

$$Z_{11} = \frac{R_1}{2}.$$

Design of Equal-Q Double-tuned Circuit to Meet Given Specifications.—
The design of a transitionally coupled equal-Q circuit to have a given
bandwidth is very simple. From Eq. (29) k and Q are found at once. If
the primary and secondary capacities are known, it is possible to write
down the values for the damping resistors.

In the more general case where what is given is the bandwidth at an
arbitrary level and the height of the peaks, the design is best performed
from the universal resonance curves of Fig. 5·3. The value of s cor-
responding to the given ratio of peak response to midband response is
found from Fig. 5·3, as is also the value of v at which the curve, with this
value of s, intersects the level at which the bandwidth is specified. Since
v and the bandwidth are known, Q can be determined; and because s is
known, k can then be found. With the determination of the damping
resistors the design is complete.

5·4. The High-Q Case When One of the Q's Is Infinite.—When Q_1 is
infinite, R_1, s, b, and v all become infinite, so that the formulas given in
Sec. 5·2 for the general case yield expressions that are indeterminate
forms. Instead of considering these directly it is better to eliminate R_1
from the expression for the transfer impedance so that this quantity is
given solely in terms of finite quantities.

If this is done, Z_{12} is given by

$$Z_{12} = \frac{jk}{2\pi f_0 \sqrt{C_1 C_2} \left[k^2 + \frac{j\alpha}{Q_2} + (j\alpha)^2 \right]} \tag{31}$$

hence

$$|Z_{12}| = \frac{k}{2\pi f_0 \sqrt{C_1 C_2} \left[k^4 + \alpha^2 \left(\frac{1}{Q_2^2} - 2k^2 \right) + \alpha^4 \right]^{1/2}} \tag{32}$$

where $\alpha = \omega/\omega_0 - \omega_0/\omega$. It is clear from Eq. (32) that transitional
coupling k_t is given by

$$k_t = \frac{1}{Q_2 \sqrt{2}}.$$

When the circuit is transitionally coupled, the bandwidth at half-power
points is given by

$$\frac{B}{f_0} = k_t,$$

and the midband gain is

$$\mathcal{G} = \frac{g_m}{k_t 2\pi f_0 \sqrt{C_1 C_2}} = \frac{g_m \sqrt{2}\, Q_2}{2\pi f_0 \sqrt{C_1 C_2}} = \sqrt{2}\, R_2 \sqrt{\frac{C_2}{C_1}}\, g_m. \qquad (33)$$

The gain-bandwidth product is

$$2\, \frac{g_m}{2\pi 2 \sqrt{C_1 C_2}},$$

and the gain-bandwidth factor is 2.   All these results have already been obtained in Sec. 5·2 above.

Universal Resonance Curves.—The shape of the response curve, in the case when one of the Q's is infinite, may be varied either by changing k or by changing Q.   Accordingly, the most useful means of presenting
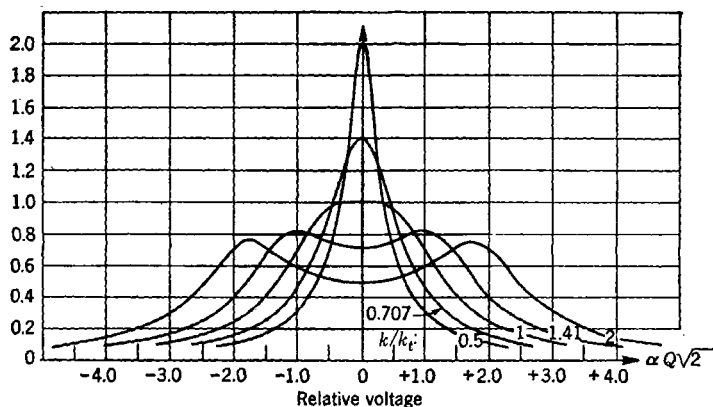
Fig. 5·4.—High-Q double-tuned circuit, loaded one side only, Q fixed, k variable.

universal resonance curves for this case is to give one set of response curves with Q_2 fixed and k variable and another set with k fixed and Q_2 variable.

1. Q Fixed, k Variable.   For fixed-Q, transitional coupling is given by $k = k_t = 1/(Q \sqrt{2})$.   In Fig. 5·4 the expression

$$\frac{Z_{12} 2\pi f_0 \sqrt{C_1 C_2}}{\sqrt{2}\, Q_2} = \frac{\dfrac{k}{k_t}}{\left\{ \left(\dfrac{k}{k_t}\right)^4 - 2\left(\dfrac{\alpha}{k_t}\right)^2 \left[\left(\dfrac{k}{k_t}\right)^2 - 1 \right] + \left(\dfrac{\alpha}{k_t}\right)^4 \right\}^{1/2}} \qquad (34)$$

is plotted as a function of $\alpha/k_t = \sqrt{2}\, Q_2\alpha$, for $k/k_t = \frac{1}{2}$, $1/\sqrt{2}$, 1, $\sqrt{2}$, 2.

In contrast to the behavior of equal-Q double-tuned circuits, discussed in Sec. 5·3, maximum secondary voltage does *not* occur at transitional

coupling, nor does the height of the peaks remain constant when the circuit is overcoupled.

Maximum secondary voltage occurs at critical coupling; if Q_1 is really infinite, critical coupling means zero coupling, infinite primary voltage, zero bandwidth, and zero gain-bandwidth product.

2. k FIXED, Q VARIABLE.   For fixed k, transitional coupling is yielded by $Q = Q_t = 1/(k\sqrt{2})$.   In Fig. 5·5 the expression

$$Z_{12}2\pi f_0 \sqrt{C_1 C_2}\, k = \frac{1}{\left[1 - 2\left(\dfrac{\alpha}{k}\right)^2 \left(1 - \dfrac{1}{q^2}\right) + \left(\dfrac{\alpha}{k}\right)^4\right]^{1/2}} \tag{35}$$

is plotted as a function of $\alpha/k = \sqrt{2}\,Q_t\alpha$, for $q = Q/Q_t = \frac{1}{2}$, $1/\sqrt{2}$, 1, $\sqrt{2}$, 2.
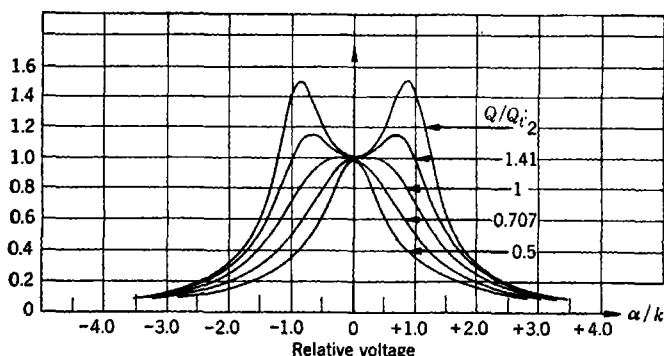
FIG. 5·5.—High-Q double-tuned circuit, loaded one side only, k fixed, Q variable.

It may be noted that with k fixed and Q variable the voltage at band center is constant, a result that is evident from Eq. (32).

The bandwidth at half-power points is

$$\mathcal{B} = \frac{1}{2\pi R_2 C_2 \sqrt{2}}. \tag{36}$$

Input Driving-point Admittance.—From Eq. (24) it follows that the conductive and susceptive components of the driving-point admittance are given by

$$\left.\begin{aligned} G_{11} &= \frac{k^2 Q_2^2 \left(\dfrac{C_1}{C_2}\right)}{R_2(1 + Q_2^2 \alpha^2)}. \\[2mm] B_{11} &= \frac{Q_2 \left(\dfrac{C_1}{C_2}\right)(1 + Q_2^2 \alpha^2 - k^2 Q_2^2)}{R_2(1 + Q_2^2 \alpha^2)}, \end{aligned}\right\} \tag{37}$$

and the absolute value of the driving point admittance is given by

$$|Y_{11}| = \frac{2\pi f_0 C_2 Q_2 \left[k^4 - \alpha^2 \left(2k^2 - \frac{1}{Q^2} \right) + \alpha^4 \right]^{1/2}}{(1 + Q_2^2 \alpha^2)} \tag{38}$$

$$= \frac{k Q_2}{R_2} \sqrt{\frac{C_1}{C_2}} \frac{1}{|Z_{12}|} \frac{R_2}{(1 + Q_2^2 \alpha^2)^{1/2}}. \tag{39}$$

It may be noticed that G_{11} varies with frequency like the amplitude response of two identical cascaded single-tuned circuits.

The Effects of Mistuning.—In the case where Q_1 is infinite Eq. (20) assumes the form

$$|Z_{12}|$$
$$= \frac{k}{2\pi f_0 \sqrt{C_1 C_2} \left\{ \left(k^2 + \frac{\Delta f^2}{f_0^2} \right)^2 + \left(\frac{\Delta f}{Q_2 f_0} \right)^2 - \frac{2\alpha \, \Delta f}{Q_2 f_0} - 2\alpha^2 \left[k^2 + \left(\frac{\Delta f}{f_0} \right)^2 - \frac{1}{2Q_2^2} \right] + \alpha^4 \right\}^{1/2}} \tag{40}$$

where $\Delta f / f_0$ is the fractional frequency difference between the resonant frequencies of the primary and secondary circuits and $\alpha = \omega/\omega_0 - \omega_0/\omega$.

It would be possible to plot a family of universal resonance curves for various values of k and Δf if full information on the effects of mistuning were required. Here, however, a short discussion of the magnitude of the effect will be given in the case where the coupling is transitional. In that case, the transfer impedance is proportional to

$$\frac{1}{\left[\frac{1}{4} - 2 Q_2^2 \left(\frac{\Delta f}{f_0} \right) \alpha + Q_2^4 \alpha^4 \right]^{1/2}}, \tag{41}$$

where the 3-db bandwidth $\mathcal{B}$ is approximately given by

$$\frac{\mathcal{B}}{f_0} = \frac{1}{\sqrt{2} \, Q_2}. \tag{42}$$

The curve of Eq. (41) no longer has a peak at midband. Instead there is a single peak at

$$\alpha' = \frac{1}{Q_2} \left(\frac{Q_2 \, \Delta f}{2 f_0} \right)^{1/3}, \tag{43}$$

and the peak response is proportional to

$$\frac{1}{\left[\frac{1}{4} - 3 \left(\frac{Q_2 \, \Delta f}{2 f_0} \right)^{4/3} \right]^{1/2}}. \tag{44}$$

The ratio r of the peak to midband response is

$$r = \frac{1}{\left[1 - 3\left(\frac{\Delta f}{B}\right)^{\frac{4}{3}}\right]^{\frac{1}{2}}}. \tag{45}$$

For a fixed Δf it may be seen that the magnitude of the mistuning effect is inversely proportional to $\mathfrak{B}$; hence the wider the bandwidth the less critical does its unequal-Q coupling become.

If the mistuning is produced by a change $-\Delta C$ in the stray capacity C of the tube, then, to the first order,

$$\frac{\Delta C}{C} = \frac{2\,\Delta f}{f_0}$$

and Eq. (45) becomes

$$r = \frac{1}{\left[1 - 3\left(\frac{\Delta C}{2C}\frac{f_0}{B}\right)^{\frac{4}{3}}\right]^{\frac{1}{2}}} \tag{46}$$

From Eq. (46) it appears that the important quantity is the fractional bandwidth, which must be large if the unequal-Q circuit is not to be too critical.

To illustrate the magnitude of the effect numerically, the following assumptions are made:

$$\begin{aligned}
C &= 15 \ \mu\mu\text{f}, \\
\Delta C &= 2.0 \ \mu\mu\text{f}, \\
\mathfrak{B} &= 15 \ \text{Mc/sec}, \\
f_0 &= 45 \ \text{Mc/sec}.
\end{aligned}$$

Then

$$\begin{aligned}
r &= \frac{1}{[1 - 3(\frac{1}{15}\,3)^{\frac{4}{3}}]^{\frac{1}{2}}} \\
&= \frac{1}{(1 - 0.35)^{\frac{1}{2}}} = \frac{1}{(0.65)^{\frac{1}{2}}}.
\end{aligned}$$

Hence the response curve now has a peak of 1.95 db at 51.6 Mc/sec, an appreciable distortion.

Design Procedure When One Q Is Infinite.—The design procedure for a double-tuned circuit with one Q infinite follows essentially the same course as that for an equal-Q double-tuned circuit.

The main results of the last three sections are summarized in Table 5·1.

5·5. The Transitionally Coupled Low-Q Case.—For wide-band receivers the high-Q theory is precise only at exceptionally high midband frequencies. The exact treatment, valid for the low-Q case as well, is

Table 5·1.—Summary of Properties of Inductively Coupled Double-tuned High-Q Circuit

Property	General case, $Q_1/Q_2 = \rho$	Equal-Q case, $Q_1 = Q_2 = Q$	$Q_1 = \infty$
Condition for transitional coupling	$k = \sqrt{\dfrac{1}{2}\left(\dfrac{1}{Q_1^2} + \dfrac{1}{Q_2^2}\right)} = \dfrac{\sqrt{\dfrac{1 + \rho^2}{2}}}{\rho Q_2}$	$k = \dfrac{1}{Q}$	$k = \dfrac{1}{Q_2 \sqrt{2}}$
3-db bandwidth at transitional coupling	$\dfrac{\sqrt{1 + k^2 Q_1 Q_2}\, f_0}{\sqrt{Q_1 Q_2}} = \dfrac{1 + \rho}{\sqrt{2}\,\rho}\dfrac{f_0}{Q}$	$\dfrac{\sqrt{2}\,f_0}{Q} = \dfrac{\sqrt{2}}{2\pi RC}$	$\dfrac{f_0}{\sqrt{2}\,Q} = \dfrac{1}{\sqrt{2}\,2\pi R_2 C_2}$
General expression for midband gain	$\dfrac{g_m k \sqrt{Q_1 Q_2}\,\sqrt{R_1 R_2}}{1 + k^2 Q_1 Q_2}$	$\dfrac{g_m k Q \sqrt{R_1 R_2}}{1 + k^2 Q^2}$	$\dfrac{g_m}{2\pi f_0 \sqrt{C_1 C_2}}\dfrac{1}{k}$
Midband gain with transitional coupling	$g_m \sqrt{R_1 R_2}\,\dfrac{\sqrt{2\rho\,(1 + \rho^2)}}{(1 + \rho)^2}$	$\dfrac{g_m \sqrt{R_1 R_2}}{2}$	$\sqrt{2}\,g_m R_2 \sqrt{\dfrac{C_2}{C_1}}$
Gain-bandwidth factor	$\dfrac{2\sqrt{1 + \rho^2}}{1 + \rho}$	$\sqrt{2}$	2
Gain-bandwidth factor of n-stage amplifier	$\dfrac{2\sqrt{1 + \rho^2}}{1 + \rho}(2^{1/n} - 1)^{1/4}$ $\approx \dfrac{2\sqrt{1 + \rho^2}}{1 + \rho}\dfrac{1}{1.1 n^{1/4}}$	$\sqrt{2}\,(2^{1/n} - 1)^{1/4}$ $\approx \dfrac{\sqrt{2}}{1.1 n^{1/4}}$	$2(2^{1/n} - 1)^{1/4}$ $\approx \dfrac{2}{1.1 n^{1/4}}$
Ratio of response at peak to response at midband	$\dfrac{1 + s^2}{[(1 + \frac{1}{2}b)\,(2s^2 + 1 - \frac{1}{2}b)]^{1/2}}$	$\dfrac{1 + s^2}{2s}$	$\dfrac{k^2 Q_2^2}{(k^2 Q_2^2 - \frac{1}{4})^{1/2}}$

extremely complicated and has been worked out fully only for the case of transitional coupling.[1] Only the main results will be given here. A

[1] A. M. Stone and J. L. Lawson, "Theory and Design of Double-Tuned Circuits," *Electronic Ind.*, p. 62, April 1946; "The Double-tuned Circuit with Transitional Coupling," RL Report No. 784, November, 1945. See also C. P. Gadsden, "Flat-flat Coupling for the Double-tuned Circuit," RL Internal Group Report 61–11/17/44, for an independent derivation.

number of additional symbols that are needed in this section will now be defined.

$$f_1 = \frac{1}{2\pi \sqrt{L_1 C_1}}, \text{ the primary frequency.}$$

$$f_2 = \frac{1}{2\pi \sqrt{L_2 C_2}}, \text{ the secondary frequency.}$$

$f_1' = $ primary frequency with secondary short-circuited.

$f_2' = $ secondary frequency with primary short-circuited.

As in the high-Q case, transitional coupling is defined by the condition that

$$\frac{d|Z_{12}|}{df} = \frac{d^2|Z_{12}|}{df^2} = \frac{d^3|Z_{12}|}{df^3} = 0,$$

where Z_{12} is the transfer impedance and $f = f_0$ is the center frequency. The analysis is presented here for the equal-Q case and the case where $Q_1 = \infty$.

Equal-Q Case.—The design data for the equal-Q transitionally coupled case are shown[1] in Fig. 5·6. If the 3-db fractional bandwidth is given, all the other data can be read off directly. The abscissa is the coefficient of coupling, in terms of which the other quantities are plotted. The circuit Q is obtained from Curve d where $Q/2\pi$ is plotted; note that Q is specified at f_0, not f_1; i.e., $Q/2\pi = f_0 R_1 C_1 = f_0 R_2 C_2$. The resonant frequencies may be found from Curve c. The frequencies f_1', f_2' at which one circuit resonates when the other is short-circuited are of importance in the lining-up procedure.

It may be noted from these curves that the high-Q theory gives results that differ very little from the exact low-Q theory even for values of k as high as 0.40. For example, even if the fractional bandwidth is as high as 0.6, corresponding to a 3-db bandwidth of 18 Mc/sec at 30 Mc/sec, the approximate theory gives a value for k of 0.424, which is only 6 per cent too high. Even more striking is the fact that the gain-bandwidth factor is constant and equal to $\sqrt{2}$ for values of k as high as 0.8, although as k tends to 1, the gain-bandwidth factor tends to 1; the case $k = 0.8$ corresponds to a bandwidth 1.55 times the midband frequency, which is an extreme case. This means that if Q is calculated from the high-Q expression

$$Q = \frac{1}{\sqrt{2}} \frac{\mathfrak{B}}{f_0},$$

then this value of Q will be accurate for values of k as high as 0.8. It will also be noticed from Eq. (2) that $|Z_{12}(f_0)| \sqrt{R_1 R_2}$, the normalized midband value of the transfer impedance, is equal to 0.5 for all k.

[1] Figs. 5·6 and 5·7 are taken from the paper of C. P. Gadsden, *op. cit.*

It may be concluded that for the transitionally coupled double-tuned circuit with equal Q's the approximate high-Q case provides design data accurate to within 6 per cent for values of fractional bandwidth up to 0.6, provided that the primary and secondary circuits are tuned to the fre-
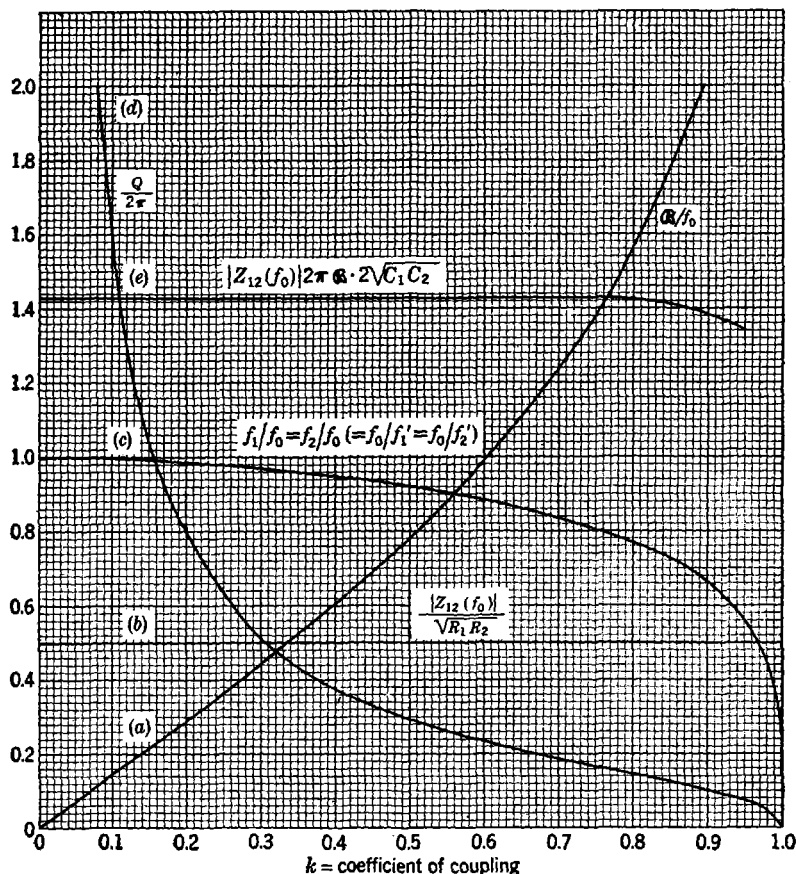
FIG. 5·6.—Design data for low-Q transitionally coupled double-tuned circuit,

$$Q_1 = Q_2 = Q.$$

(a) Fractional bandwidth $\mathcal{B}/f_0$; (b) normalized midband gain; (c) primary and secondary resonant frequencies; (d) $Q/2\pi = f_0 R_1 C_1 = f_0 R_2 C_2$; (e) gain-bandwidth factor

$$|Z_{12}(f_0)|2\pi\mathcal{B}2\sqrt{C_1C_2}.$$

quencies given by Curve c in Fig. 5·6. The approximate theory also provides expressions for the gain-bandwidth factor and the midband gain that are accurate up to fractional bandwidths of 1.55 for the former and even larger fractional bandwidths for the latter.

Even for fractional bandwidths as large as unity, it is still very accurately the case, as in the high-Q approximation, that the bandwidth $\mathbb{B}_r$ at a level $20 \log r$ db down from midband response is related to the 3-db bandwidth $\mathbb{B}$ by

$$\mathbb{B} = \frac{\mathbb{B}_r}{(r^2 - 1)^{1/4}}.$$
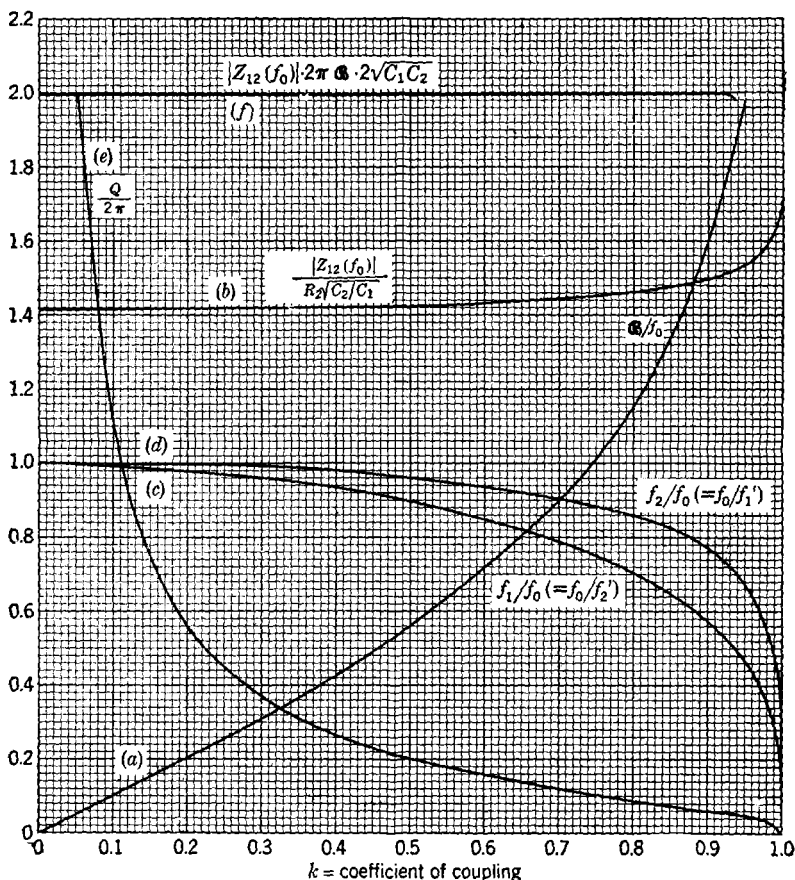
Fig. 5·7.—Design data for low-Q transitionally coupled double-tuned circuit $Q_1 = \infty$, $Q_2 = Q$. (a) Fractional bandwidth $\mathbb{B}/f_0$; (b) normalized midband gain; (c) primary resonant frequency; (d) secondary resonant frequency; (e) $Q/2\pi = f_0 R_2 C_2$; (f) gain-bandwidth factor $|Z_{12}(f_0)| 2\pi \mathbb{B} 2 \sqrt{C_1 C_2}$.

Case When Q_1 Is Infinite.—The design data for the transitionally coupled double-tuned circuit when Q_1 is infinite are shown in Fig. 5·7. The most important difference between this case and the equal Q case is

that the resonant frequencies of the primary and secondary circuits, given by Curves c and d, respectively, are no longer equal. As in the equal-Q case the high-Q theory gives values for k accurate to within 5 per cent for k less than 0.4. The value $k = 0.4$ corresponds to a fractional bandwidth of 0.435 (as opposed to 0.6 in the equal-Q case). The gain-bandwidth factor is constant and equal to 2 for values of k up to 0.9 but drops to $\sqrt{3}$ as k tends to 1. The high-Q formula

$$Q = \sqrt{2}\,\frac{\mathcal{B}}{f_0}$$

is accurate for values of k as high as 0.8. The normalized midband gain, plotted in Curve b, is no longer independent of k, but the high-Q result is accurate to within 5 per cent for fractional bandwidths as large as 1.4.

Apart from the differences mentioned above, the design of a transitionally coupled double-tuned circuit with one Q infinite proceeds along the same lines as for the equal-Q case.

5·6. Stagger-damped Double-tuned Circuits.—If n identical double-tuned transitionally coupled stages are cascaded, the over-all 3-db bandwidth is reduced by the factor $(2^{1/n} - 1)^{1/4}$ as stated in Sec. 5·2. Thus in a conventional eight-stage double-tuned amplifier the over-all bandwidth is only $(2^{1/8} - 1)^{1/4} = 55$ per cent of the bandwidth of each stage. A scheme will now be described[1] that eliminates this shrinking of bandwidth. The idea is to have successive stages over- and undercoupled so that the over-all response curve is maximally flat. For want of a better name this scheme is called "stagger-damping."

Because of the complexity of the calculations for the exact case, the theory is based throughout upon the high-Q approximation. This is a weakness, since stagger-damping is likely to be noncritical enough for practical use only in the low-Q case. However, the wide range of validity of the high-Q theory for the case of transitional coupling, demonstrated in Sec. 5·5, lends support to the belief that results derived from this theory will be sufficiently accurate to yield the high predicted gain-bandwidth factors, at least after some cut and try.

Synthesis of a Maximally Flat Response Curve by Over- and Undercoupled Double-tuned Circuits.—In the first place attention will be confined to the case where one Q is infinite. It was shown in Sec. 5·4 that when $Q_1 = \infty$, the transfer impedance is given by

$$Z_{12} = \frac{1}{2\pi f_0 \sqrt{C_1 C_2}} \frac{jk}{\left[k^2 + \dfrac{j\alpha}{Q_2} + (j\alpha)^2 \right]}. \tag{47}$$

[1] Henry Wallman, "Stagger-damped Double-tuned Circuits," RL Report No. 539, Mar. 23, 1944. Section 5·6 follows this report.

The absolute value of Eq. (47) is

$$|Z_{12}| = \frac{1}{2\pi f_0 \sqrt{C_1 C_2}} \frac{k}{\left[k^4 + \alpha^2 \left(\dfrac{1}{Q_2^2} - 2k^2 \right) + \alpha^4 \right]^{\frac{1}{2}}}. \tag{48}$$

If the individual stages of the n-stage amplifier are over- and under-coupled to yield a maximally flat over-all amplitude response, this response will be of the form

$$\frac{\delta^n}{(\delta^{4n} + \alpha^{4n})^{\frac{1}{2}}},$$

where δ is equal to the fractional over-all bandwidth between half-power points.

Now from Eq. (4·21a) it may be seen that $\delta^n / (\delta^{4n} + \alpha^{4n})^{\frac{1}{2}}$ is the absolute value of

$$\frac{\delta}{\delta^2 + j\alpha 2\delta \sin \dfrac{\pi}{4n} + (j\alpha)^2} \cdot \frac{\delta}{\delta^2 + j\alpha 2\delta \sin \dfrac{3\pi}{4n} + (j\alpha)^2} \cdots$$

$$\frac{\delta}{\delta^2 + j\alpha 2\delta \sin \dfrac{2n-1}{4n}\pi + (j\alpha)^2}. \tag{49}$$

Each factor of Eq. (49) has the same general form as Eq. (47); thus the first factor of Eq. (49), for example, may be realized by a double-tuned circuit loaded on one side only and satisfying

$$k = \delta,$$
$$Q_2 = \frac{1}{2\delta \sin \dfrac{\pi}{4n}}, \tag{50a}$$

the second factor of Expression (49) by a double-tuned circuit loaded on one side only and satisfying

$$k = \delta,$$
$$Q = \frac{1}{2\delta \sin \dfrac{3\pi}{4n}}. \tag{50b}$$

and so on. Hence the selectivity curve $\delta^n / \sqrt{\delta^{4n} + \alpha^{4n}}$ may be realized by an amplifier consisting of n suitably damped double-tuned circuits separated by vacuum tubes. An amplifier of this sort is called a "stagger-damped n-uple."

In the approximate high-Q case this design possesses two very convenient features.

1. All the transformers have the same coefficient of coupling.
2. The coefficient of coupling is the ratio of desired over-all bandwidth to center frequency.

This means that only one transformer type is needed throughout. The only difference between the circuit of one stage and the circuit of another stage is in the value of damping resistor.

It must be pointed out, however, that because this scheme will be used only with wide-band amplifiers, the exact low-Q analysis must be used to calculate the circuit components. Under these circumstances the coefficients of coupling will vary from stage to stage, although the variation will not, in practice, be large.

Gain-bandwidth Factor.—If the transconductance of each stage of the amplifier is g_m, the over-all gain at midband is

$$\frac{g_m^n}{(2\pi f_0 \sqrt{C_1 C_2}\ \delta)^n},$$

so that the mean stage gain is

$$\frac{g_m}{2\pi f_0 \sqrt{C_1 C_2}\ \delta}.$$

The over-all 3-db bandwidth is $f_0\delta$, and hence the stage-gain times over-all bandwidth product is

$$\frac{g_m}{2\pi \sqrt{C_1 C_2}},$$

so that the gain-bandwidth factor is 2. This is the gain-bandwidth factor of a single transitionally coupled double-tuned circuit with one Q infinite and is twice that yielded by staggered single-tuned circuits.

Stagger-damped circuits bear the same relation to double-tuned circuits, all transitionally coupled, as do stagger-tuned circuits to synchronous single-tuned circuits; both schemes eliminate the shrinking of over-all bandwidth that results when identical circuits are cascaded.

Stagger-damped n-uples Using Double-tuned Circuits with Both Q's Finite.—It might be expected, as a corollary of the above analysis, that if a stagger-damped *n*-uple were constructed of equal-Q double-tuned circuits, the over-all gain-bandwidth factor would be equal to $\sqrt{2}$, that is, to the gain-bandwidth factor of a single transitionally coupled equal-Q circuit. Such is not the case, however; the gain-bandwidth factor of an *n*-stage equal-Q stagger-damped *n*-uple is $2^{1/2n}$, which tends monotonically to unity as $n \to \infty$.

It may be seen that if all the stages must be made of equal-Q double-tuned circuits, the stagger-damped scheme is of very little use.

It has been shown in Sec. 5·2 that for a given bandwidth and shape of response, the ratio of midband gain in the equal-Q case to midband gain in the unequal-Q case tends to unity as k tends to unity and to zero as k tends to zero.   Accordingly it would be possible to use equal-Q circuits for the overcoupled stages and unequal-Q circuits for the undercoupled stages and preserve an over-all gain-bandwidth factor close to 2.   Such a design might prove a useful compromise between good gain-bandwidth factor and noncritical circuits.

For the general case where the Q's are unequal the circuit components can be determined quite simply when it is remembered that the shape of the response curve of a circuit is determined once there is given the ratio of peak to midband response and the bandwidth at any level.
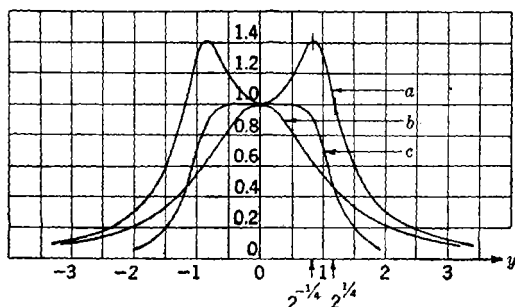
FIG. 5·8.—Stagger-damped pair (a) $1/\sqrt{1 - y^2 \sqrt{2} + y^4}$; (b) $1/\sqrt{1 + y^2 \sqrt{2} + y^4}$; (c) $1/\sqrt{1 + y^8}$.   (a) and (b) are the component selectivity curves of a stagger-damped pair; (c) is the over-all result.

Let $\mathfrak{B}$ be the over-all bandwidth, to the half-power points, of a stagger-damped n-uple; then the circuit components of the rth stage must satisfy the two equations

$$\left.\begin{aligned}
\frac{\mathfrak{B}}{f_0} &= \frac{(1 + s^2)^{\frac{1}{2}}}{\sqrt{Q_1 Q_2}}, \\
2 \sin (2r - 1) \frac{\pi}{4n} &= \frac{\sqrt{b + 2}}{\sqrt{1 + s^2}}.
\end{aligned}\right\} \tag{51}$$

One more relation is needed to determine the system.   The Q-ratio may be given or the magnitude of one of the Q's or the coupling coefficient k or a functional relation of all three, depending upon the particular problem under discussion.   The important thing to emphasize is that there is no need for the third relation to be the same from stage to stage; it is necessary only that the two relations of Eq. (51) be satisfied.

Stagger-damped Pairs and Triples.—The particular cases of stagger-damped pairs and triples will be considered in slightly more detail.

1. *The Resonance Curves.*—In Fig. 5·8 are shown the two resonance curves that make up a stagger-damped pair, and in Fig. 5·9 the three selectivity curves that make up a stagger-damped triple. The frequency variable in each figure is $y = \alpha/\delta$.

The design data for the general case can be found from Eq. (50) by putting $n = 2$ in the stagger-damped pair case and $n = 3$ in the stagger-damped triple case.
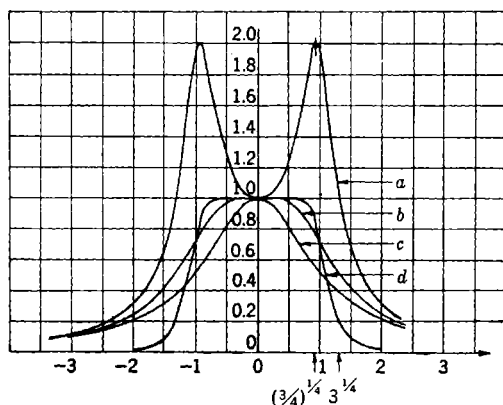
Fig. 5·9.—Stagger-damped triple. (a) $1/\sqrt{1 - y^2}\sqrt{3 + y^4}$; (b) $1/\sqrt{1 + y^4}$, (c) $1/\sqrt{1 + y^2}\sqrt{3 + y^4}$; (d) $1/\sqrt{1 + y^{12}}$. (a), (b), and (c) are the component selectivity curves of a stagger-damped triple; (d) is the over-all result.

2. *Cascaded Pairs and Triples.*—When identical stagger-damped pairs are cascaded, the over-all bandwidth is reduced by the factor $(2^{1/n} - 1)^{1/6}$. when n is the number of pairs, as in Table 5·2.

Table 5·2.—Reduction of Over-all Bandwidth with n Stagger-damped Pairs

n	2	3	4	5	6	7
$(2^{1/n} - 1)^{1/6}$	0.896	0.845	0.812	0.782	0.768	0.754

Consider an eight-stage amplifier, for example, in the form of four stagger-damped pairs. The over-all bandwidth is 81 per cent of the bandwidth of one pair, so that the use of four pairs instead of one octuple entails a loss of only 19 per cent in over-all bandwidth.

When identical stagger-damped triples are cascaded, the over-all bandwidth is reduced by the factor $(2^{1/n} - 1)^{1/12}$, where n is the number of triples, as in Table 5·3.

TABLE 5·3.—REDUCTION OF OVER-ALL BANDWIDTH WITH n STAGGER-DAMPED TRIPLES

n	2	3	4	5
$(2^{1/n} - 1)^{1/2}$	0.929	0.894	0.871	0.848

3. *Transient Response.*—The step-function response of a stagger-damped pair is graphed under $n = 4$ (Fig. 7·7) and shows a 10.9 per cent overshoot. The step-function response of a stagger-damped triple is graphed under $n = 6$ in the same Fig. 7·7 and shows a 14.3 per cent overshoot.

As pairs or triples are cascaded, the percentage overshoot is multiplied by a factor about equal to the square root of the number of pairs or triples.

5·7. Construction and Examples.[1]—It is usually difficult to achieve a coefficient of coupling larger than about $\frac{1}{3}$ with a mutual-inductance-coupled air core transformer or $\frac{1}{2}$ with a powdered-iron core transformer[2] unless the windings are made to overlap. For larger coefficients of coupling the use of the π- or T-equivalent is then necessary. It must be kept in mind that there is a maximum coefficient of coupling possible in a π or T, which occurs when the π or T degenerates to an "inverted-L"; this maximum coefficient of coupling may be noticeably less than unity if the primary and secondary inductances are very different.

The π- or T-equivalent is sometimes employed even when the mutual-inductance-coupled transformer is possible, for the reason that it is easier to make and adjust three separate coils than a transformer with accurately made spacer. Many of the Radiation Laboratory double-tuned circuits employed to couple a crystal converter to the first amplifier grid were of this sort. The π- or T-circuit is somewhat bulkier, however, and requires either a condenser or bifilar winding to block direct current.

Adjustment and measurement of a double-tuned transformer is best carried out with a Q-meter. The primary and secondary are each adjusted until they resonate with the appropriate primary and secondary capacities C_1 and C_2 at the design frequencies f_1 and f_2, (see Fig. 5·6 or 5·7), with the other winding open-circuited. The coupling is then adjusted until the primary winding resonates with C_1 at f_1' (see Fig. 5·6 or Fig. 5·7) with the secondary short-circuited; similarly the secondary

---

[1] This section was written by Henry Wallman.

[2] A coefficient of coupling of $\frac{1}{3}$ corresponds to a fractional bandwidth of about $\frac{1}{3}$ in the case of loading on one side only and a fractional bandwidth of about 0.47 in the case of equal primary and secondary Q's; a coefficient of coupling of $\frac{1}{2}$ corresponds to fractional bandwidths of $\frac{1}{2}$ and 0.7 in the two cases. It is rare that these fractional bandwidths do not suffice.

should resonate with C_2 at f_2' with the primary short-circuited.   The procedure converges rapidly.

In manufacture, close tolerances must be maintained on the thickness of the spacer between primary and secondary as well as on the primary and secondary inductances.

The coefficient of coupling may be measured as follows: Denote by C_{oc} the capacity required to resonate the primary at a convenient frequency with the secondary open-circuited, and denote by C_{sc} the capacity required to resonate the primary at the same frequency with the secondary short-circuited.   Then

$$k = \sqrt{1 - \frac{C_{oc}}{C_{sc}}}. \tag{52}$$

The result should be the same when the primary is replaced by the secondary.

Equation (52) is entirely exact, even for the low-Q case, if only inductance-coupling exists.   In transformers intended for large fractional bandwidth, however, there is usually a certain unavoidable stray-capacity coupling; the result is that a certain amount of cut-and-try adjustment is required.   In order to keep capacity coupling as small as possible, the terminals of the primary and secondary windings that are closest to each other should be those grounded for signal (B^+ and ground, respectively).

The alignment procedure, once the transformers have been wired into the amplifier, follows very much the same method as that outlined above for adjusting the coils and is described in Sec. 8·4.   Because double-tuned amplifiers are almost always fixed-tuned, the alignment procedure is usually employed only in the adjustment of the production prototype.

Example 1.   *Twelve-Stage* 6AK5 100-*db Amplifier at* 60 *Mc/Sec with* 12.5 *Mc/Sec Over-all Bandwidth.*—The circuit diagram of two typical stages of this amplifier, designed by C. E. Ingalls, is shown in Fig. 5·10, and a photograph in Fig. 5·11.   One stage has provision for applying gain control voltage to the grid, and the other has not.

The bandwidth per stage is about 25 Mc/sec, and the stages are adjusted for a coupling just a very little more than transitional.

The primary capacitance is 4.8 $\mu\mu$f and the secondary capacitance is 7.4 $\mu\mu$f.   The primary Q is 2.2 times the secondary Q; this Q-ratio was experimentally determined in order that change of secondary capacitance with gain-control voltage might result in the least tilt in the pass band.

The coefficient of coupling is about 38 per cent.   Powdered-iron cores are used.

The stage gain is 9 db even when the transconductance of the type 6AK5 tubes is only 4000 μmhos.   A conservative rating of this sort

ensures adequate gain even with tubes all of which are close to the lower limit of the JAN-1A specifications on transconductance.

In the interest of low noise figure the first two stages of this amplifier employ the grounded-cathode triode grounded-grid triode circuit

Fig. 5·10.—Circuits of two typical amplifier stages of Example 1.

Fig. 5·11.—Layout of typical amplifier stages of Example 1.

described in Sec. 13·10; the noise figure over the 12.5 Mc/sec bandwidth is 3.5 db. This was 2.0 db better than had been obtained with a type 6AK5 pentode first stage.

Example 2.—The amplifier shown in Fig. 5·12 has a purpose similar to that of the amplifier of Fig. 4·14. The interstage circuits are double-

tuned, with loading on one side only, and transitionally coupled.   The primary capacity is 7.0 $\mu\mu f$, and the secondary capacity is 4.5 $\mu\mu f$.
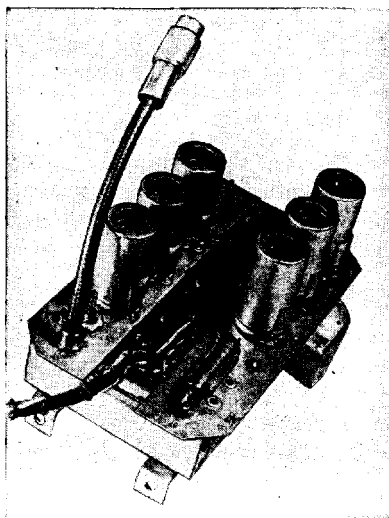
FIG. 5·12.—Four-stage 6AK5 80-db double-tuned amplifier at 30 Mc/sec with 10 Mc/sec over-all bandwidth.   (a) Bottom view; (b) top view.

For an over-all bandwidth of 10 Mc/sec the bandwidth of each of the five double-tuned circuits (including the input circuit) has to be

$$\frac{10}{(2^{1/5} - 1)^{1/4}} = 16 \text{ Mc/sec.}$$

Hence in Fig. 5·7 the value of $\mathscr{B}/f_0$ is 0.53; the value of k is therefore 0.48; this is achieved by use of a powdered-iron core. If the secondary is unloaded, then the primary $Q/2\pi$ at f_0 is 0.21 and the open-circuit primary and secondary frequencies are $30 \times 0.97 = 29.1$ and $30 \times 0.91 = 27.3$ Mc/sec, and the primary and secondary frequencies with the other winding short-circuited are $30/0.91 = 33$ and $30/0.097 = 31$ Mc/sec. The damping resistor R_1 is determined from $Q/2\pi = f_0 R_1 C_1$. If the primary is unloaded, it is the secondary whose $Q/2\pi$ at f_0 is 0.21, and the primary and secondary frequencies above are interchanged. The damping resistor R_2 is determined from $Q/2\pi = f_0 R_2 C_2$.
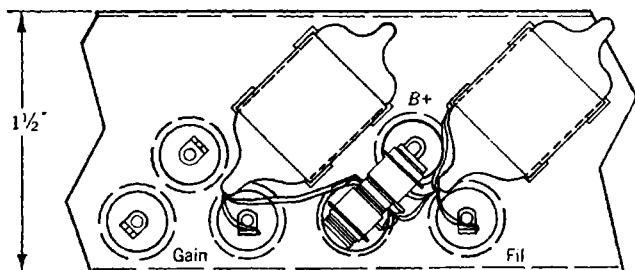
Fig. 5·13.—Layout of two stages of four-stage 80-db stagger-damped amplifier at 30 Mc/sec with 10 Mc/sec over-all bandwidth. The tubes are the Raytheon baseless type CK-604.

In the amplifier of Fig. 5·12 the circuits are loaded alternately in the primary and secondary. Reduction of secondary capacity with increased control-grid bias tilts the response of one stage one way but tilts the response of the next stage the other; a fairly good over-all response is thereby preserved.

The chassis shown in Fig. 5·12 also contains a type 6AL5 detector, type 6AK5 first pulse-amplifier stage, and two pulse output stages employing the Sylvania type SD-834 baseless triodes, which can be seen in the lower left corner of Fig. 5·12a.

Example 3.—The amplifier shown in Fig. 5·13 is interesting in two respects: It employs stagger-damped pairs (Sec. 5·6), and the tubes are baseless.

The tube type is the Raytheon CK-604, which has almost exactly the same internal structure as the type 6AK5 and differs from it only in having a flat press, flexible leads, slightly higher heater temperature, slightly higher interelectrode capacitances, and a slightly smaller envelope, coated with metallic paint as shield. The transconductance is the same, but

because socket capacitances are eliminated, the actual g_m/C ratio is only very slightly lower than for the type 6AK5.

The main virtue of the construction that results from the use of baseless tubes, apart from small size and elimination of parts, is that it permits one to employ circuits that are rather complicated and critical but of high efficiency.   These circuits are matched to the individual tubes, and there is no danger that an unthinking maintenance man will misalign the amplifier by changing tubes.

The entire 80-db amplifier is only $5\frac{1}{2}$ in. long, $1\frac{1}{2}$ in. wide, and 1 in. deep.

The circular components shown in Fig. 5·13 are feed-through button bypass condensers, and the hot-water-bottle-shaped objects are the baseless tubes.   The mechanical design and construction of this amplifier is due to L. A. Harlow.

In a four-stage amplifier made up of two stagger-damped pairs, the bandwidth of each pair has to be only $10/0.896 = 11.2$ Mc/sec (Sec. 5·6) for a 10 Mc/sec over-all bandwidth; compare this with the 16 Mc/sec stage bandwidth required in the preceding example for a 10 Mc/sec over-all bandwidth.   Consequently, for given gain, over-all bandwidth, and interstage capacities, the tubes in a stagger-damped amplifier can be operated with a lower transconductance.   In this stagger-damped amplifier the stage gain of 20 db requires a g_m of only 3800 μmhos; this is only 76 per cent of the nominal g_m, and the tubes should have very long life when operated in that way.

CHAPTER 6

HIGH-FREQUENCY FEEDBACK AMPLIFIERS

By Harry J. Lipkin

6·1. Introduction.—The gain-bandwidth product of an amplifier that uses synchronous single- or double-tuned coupling circuits can be improved by the use of inverse feedback that is a maximum at band center and drops off toward the edges of the band. As shown in Fig. 6·1, a flattening of the response curve of the amplifier with a consequent increase in the bandwidth results. If the increase in bandwidth is greater than the decrease in gain, the gain-bandwidth product is increased.

The simplest feedback circuit of this type is shown in Fig. 6·2. The feedback is produced by means of a feedback resistor connected between the plate and grid of the amplifier stage. The proportion of the output voltage that is fed back depends upon the action of the voltage-divider circuit consisting of
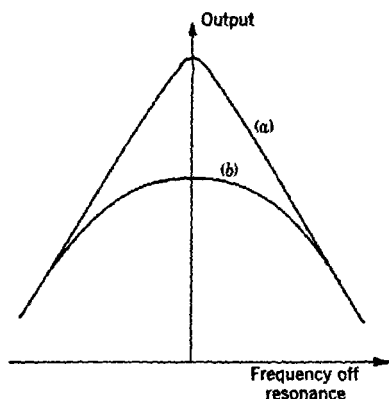
Fig. 6·1.—Amplifier response, (a) without feedback, (b) with inverse feedback.
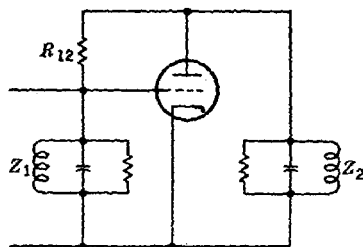
Fig. 6·2.—Simple feedback stage.

the feedback resistor R_{12} and the grid circuit impedance Z_1. Since Z_1 is resistive and a maximum at resonance (band center), there is maximum feedback at this frequency with a phase shift of exactly 180°. As the frequency deviation from band center increases, Z_1 decreases and phase shifts are introduced in the amplifier and in the feedback-voltage divider, thus decreasing the magnitude of the feedback signal and shifting its phase. The negative feedback is therefore of the proper type to increase bandwidth.

232

Multistage amplifiers consisting of feedback stages of this type are illustrated in Fig. 6·3 and are called inverse-feedback chains. Admittances rather than impedances are noted in Fig. 6·3 because of the simpler mathematical analysis possible on this basis. In the general case, any values can be used for the feedback impedances, and the shunt impedances may be two-terminal elements, with variation from stage to stage, as shown in Fig. 6·3*a*, or four-terminal elements as shown in Fig. 6·3*b*.
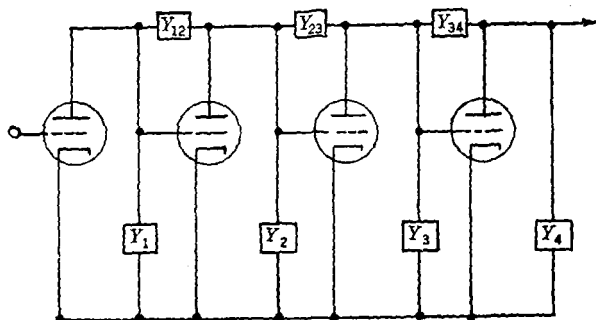
FIG. 6·3*a*.—Feedback chain using two-terminal shunt impedances.
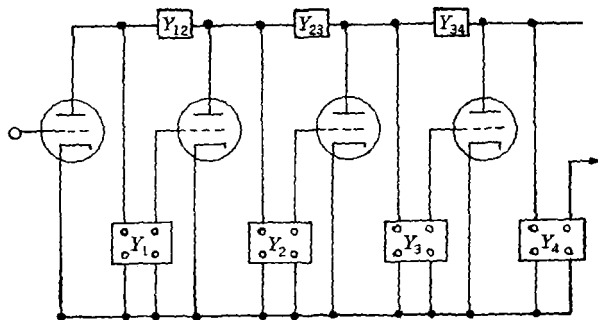
FIG. 6·3*b*.—Feedback chain using four-terminal shunt impedances.

However, the simplest case, where the feedback impedance is a pure resistance and the shunt impedances are single-tuned circuits, is of primary importance and is considered in greatest detail in this chapter. A brief discussion of more complicated feedback chains is given in Sec. 6·7.

Although, in the general feedback chain, the feedback and shunt impedances may be chosen arbitrarily from stage to stage, in practical applications, three special simplified cases of the general feedback chain find frequent use: (1) the uniform chain in which all stages but the first and last are identical, (2) inverse-feedback pairs in which the feedback is

zero in every second stage, and (3) inverse-feedback triples in which the feedback is zero in every third stage.

An important disadvantage of feedback-chain circuits is the removal of the isolation normally existing between vacuum-tube amplifier stages. There is considerable interaction between stages, and a change in the transconductance of any stage affects the feedback and therefore the
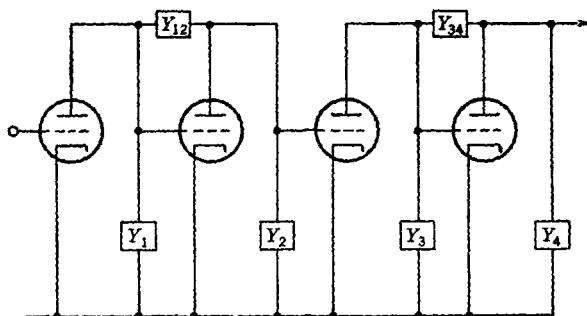
Fig. 6·4.—Feedback pairs.

shape of the amplifier pass band. It is therefore difficult to achieve satisfactory gain control in an amplifier employing a feedback chain. For this reason "feedback pairs" are often used, in which there is feedback only in alternate stages, as shown in Fig. 6·4. Each pair is isolated from the other pairs, and gain control can be achieved by varying the trans-conductance of the first stage of each pair, around which there is no feed-
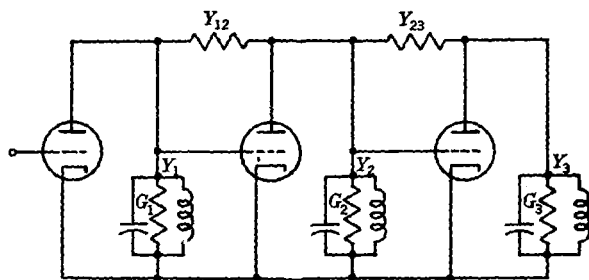
Fig. 6·5.—Feedback triple using single-tuned circuits.

back. "Feedback triples," in which there is no feedback in every third stage, are also often used for the same reason. Feedback pairs and triples are generally used with single-tuned circuits as shunt-loading admittances, as is shown in Fig. 6·5. For this reason, this chapter is mainly devoted to the consideration of pairs and triples using single-tuned circuits. In order to allow the use of the initial portions of the analyses for more complicated cases, the general two-terminal network

is employed up to the point where specialization to the single-tuned circuit is necessary for the analysis.

If all the stages of the general chain are identical, the resulting chain is said to be uniform and lends itself readily to mathematical analysis. Except for the case of feedback pairs and triples, no advantages for the general nonuniform chain have as yet been shown that would justify the added complications of its use over that of the uniform chain.

Although all analyses in this chapter refer to bandpass amplifiers, inverse feedback can be used to good advantage in video amplifiers as well. All the analyses in this chapter can be used in the video case by considering the low-pass bandpass analogue discussed in Sec. 7·1.

6·2. Analysis of the General Chain.—The conventional approach to feedback-amplifier analysis is to determine the gain of the amplifier in the absence of feedback and the amount of feedback and from these quantities to calculate the gain in the presence of feedback from the formula $\alpha/(1 - \alpha\beta)$, where α is the gain of the amplifier without feedback and β is the gain of the feedback circuit. In considering the feedback-chain type of amplifier, this analysis is not practical, because each stage has its own feedback loop, which interacts with adjacent feedback loops. Furthermore, even in the single-loop case (inverse-feedback pair), the exact meaning of α is obscure because the feedback network acts as a load upon the grid and plate circuits, and removing the feedback changes the loading.[1]

The simplest approach to the analysis of the feedback pair or chain is that of writing the nodal equations for the circuit and solving them directly.

Gain of a Single Feedback Stage.—Consider the nth stage of the general feedback-chain amplifier employing two-terminal coupling elements, as in Fig. 6·3a. Let its voltage gain be denoted by $\mathcal{G}_n$. Then if the voltage on the grid of this stage is denoted by e_1, the voltage on the plate is $\mathcal{G}_n e_1$ and the voltage on the plate of the following stage is $\mathcal{G}_n \mathcal{G}_{n+1} e_1$. The currents in the admittances $G_{n-1,n}$, $G_{n,n+1}$, and Y_n are

$$
\begin{aligned}
i_{n-1,n} &= (\mathcal{G}_n - 1)e_n G_{n-1,n}, \\
i_{n,n+1} &= \mathcal{G}_n(1 - \mathcal{G}_{n+1})e_1 G_{n,n+1}, \\
i_n &= \mathcal{G}_n e_1 Y_n.
\end{aligned}
\tag{1}
$$

Since the plate current of the tube is $-\dot{g}_m e_1$, the following Kirchhoff node equation can be written for the plate terminal of the nth stage:

$$
-g_m e_1 = (\mathcal{G}_n - 1)e_n G_{n-1,n} + \mathcal{G}_n(1 - \mathcal{G}_{n+1})e_1 G_{n,n+1} + \mathcal{G}_n e_1 Y_n.
\tag{2}
$$

[1] H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945, pp. 44–45 and 80.

Solving this equation for $\mathcal{G}_n$ yields

$$\mathcal{G}_n = \frac{-g_m + G_{n-1,n}}{G_{n-1,n} + G_{n,n+1} + Y_n - \mathcal{G}_{n+1}G_{n,n+1}}. \tag{3}$$

Thus the gain of any stage of an inverse-feedback chain is given by Eq. (3) in terms of the circuit constants of the stage and the gain of the following stage. It is, therefore, possible to use this equation to determine the response of a chain by starting with the last stage and working backward through the amplifier.

Observe that the gain of an amplifier stage as defined here is the ratio of the output voltage of the stage to its input voltage. The inverse feedback around this stage reduces the input voltage to the stage but does not affect the ratio of output to input voltage. Thus the effect of the feedback around a given stage is not to reduce the gain of that stage, as here defined, but to load the previous stage, thereby reducing the gain of the previous stage. This fact is clearly shown by Eq. (3), where the effect of feedback upon the gain of the nth stage is given by the term $\mathcal{G}_{n+1}G_{n,n+1}$ produced by the feedback in the following stage.

The term $G_{n-1,n}$ in the numerator of Eq. (3) represents the direct transmission of signal around the nth stage through the feedback conductance $G_{n-1,n}$. Since this direct transmission is generally small compared with the amplified signal and is in any case not a function of frequency, the quantity $g_m - G_{n-1,n}$ can be considered as a fictitious transconductance, corrected for the direct transmission and differing from the actual tube transconductance by a small quantity. This reduction in effective transconductance is calculated in Sec. 6·3.

The terms $G_{n-1,n}$, $G_{n,n+1}$, and Y_n represent the parallel combination of all impedances connected to the plate of the nth stage. This then is the total direct loading Y_{nn} (as distinct from feedback loading) on the stage. It is therefore convenient to combine these into a single term. Equation (3) can then be written more conveniently as

$$\mathcal{G}_n = \frac{-g'_{m_n}}{Y_{nn} - \mathcal{G}_{n+1}G_{n,n+1}}, \tag{4}$$

where $g'_{m_n} = g_{m_n} - G_{n-1,n}$ and $Y_{nn} = G_{n-1,n} + G_{n,n+1} + Y_n$.

Note that for the case of no feedback $G_{n-1,n} = G_{n,n+1} = 0$, and Eqs. (3) and (4) reduce to $\mathcal{G}_n = g_m/Y = g_m Z$, the normal equation for the gain of a single stage without feedback.

Gain of General Chain.—Equation (4) can also be written in the form

$$\mathcal{G}_n Y_{nn} - \mathcal{G}_n \mathcal{G}_{n+1} G_{n,n+1} + g'_{m_n} = 0. \tag{5}$$

Let $\mathcal{G}_{1k}$ be the over-all gain of all the stages from the first stage up to an including the kth.   Then, if Eq. (5) is multiplied by $\mathcal{G}_{1,n-1}$, one has

$$\mathcal{G}_{1,n-1}\mathcal{G}_n Y_{nn} - \mathcal{G}_{1,n-1}\mathcal{G}_n\mathcal{G}_{n+1}G_{n,n+1} + g'_{m_n}\mathcal{G}_{1,n-1} = 0, \tag{6a}$$

or

$$g'_{m_n}\mathcal{G}_{1,n-1} + Y_{nn}\mathcal{G}_{1,n} - G_{n,n+1}\mathcal{G}_{1,n+1} = 0, \tag{6b}$$

since

$$\mathcal{G}_{1,n-1}\mathcal{G}_n = \mathcal{G}_{1,n},$$

and

$$\mathcal{G}_{1,n-1}\mathcal{G}_n\mathcal{G}_{n+1} = \mathcal{G}_{1,n+1}.$$

Simultaneous linear equations for the $\mathcal{G}_{1k}$'s can be obtained by applying Eq. (6) to each stage of the amplifier.   Thus

$$\left.\begin{array}{l} Y_{11}\mathcal{G}_1 - G_{12}\mathcal{G}_{12} = -g_{m_1}, \\ g'_{m_2}\mathcal{G}_1 + Y_{22}\mathcal{G}_{12} - G_{23}\mathcal{G}_{13} = 0, \\ g'_{m_3}\mathcal{G}_{12} + Y_{33}\mathcal{G}_{13} - G_{34}\mathcal{G}_{14} = 0, \\ \quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \\ g'_{m_n}\mathcal{G}_{1,n-1} + Y_{nn}\mathcal{G}_{1n} = 0. \end{array}\right\} \tag{7}$$

Solving this set of equations for $\mathcal{G}_{1n}$, which is the over-all gain of the amplifier, yields

$$\mathcal{G} = \mathcal{G}_{1n} = \frac{\begin{vmatrix} Y_{11} & -G_{12} & 0 & 0 & 0 & \cdot & \cdot & -g_{m_1} \\ g'_{m_2} & Y_{22} & -G_{23} & 0 & 0 & \cdot & \cdot & 0 \\ 0 & g'_{m_3} & Y_{33} & -G_{34} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 & g'_{m_{n-1}} & Y_{n-1,n-1} & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & 0 & g'_{m_n} & 0 \end{vmatrix}}{\Delta}$$

$$= \frac{(-1)^n g_{m_1} g'_{m_2} \cdots g'_{m_n}}{\Delta}, \tag{8}$$

where

$$\Delta = \begin{vmatrix} Y_{11} & -G_{12} & 0 & 0 & 0 & \cdot & 0 & 0 \\ g'_{m_2} & Y_{22} & -G_{23} & 0 & 0 & \cdot & 0 & 0 \\ 0 & g'_{m_3} & Y_{33} & -G_{34} & 0 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & g'_{m_{n-1}} & Y_{n-1,n-1} & -G_{n-1,n} \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & g'_{m_n} & Y_{nn} \end{vmatrix} \tag{9}$$

From Eqs. (8) and (9), an important property of inverse-feedback chains becomes evident, namely, their symmetry.   If the stages of an amplifier are arranged in reverse order, with the last stage first, the over-

all response is unchanged. This property is of particular interest in the case of the uniform chain, where all stages are alike except for the first and last load impedances. To reverse the order of the stages of this type of amplifier, it is necessary merely to interchange the first and last load impedances, which are generally called the "terminations." Reversing the terminations of a uniform feedback chain, therefore, does not affect the over-all response.

In the case of the feedback pair, interchanging the terminations is also equivalent to reversing the order of the amplifier and does not change the over-all response. However, in the case of the triple, this fact is no longer true. Reversing the order of the amplifier requires an interchange of the two feedback resistors as well as of the terminations. Thus, for a triple, unless the two feedback resistors are equal, interchanging the terminations changes the response.

Transmission Characteristic and Gain-bandwidth Product.—Inspection of Eq. (9) shows that all terms of the expanded determinant are positive and have the dimensions of conductance to the nth power. The terms can be formed by substituting factors of the form $g'_{m_k} G_{k-1,k}$ in the place of $Y_{k-1,k-1} Y_{kk}$ in the expression $Y_{11} Y_{22} Y_{33} \ldots Y_{nn}$. The complete denominator of Eq. (8) therefore consists of a term $Y_{11} Y_{22} Y_{33} \ldots Y_{nn}$ plus all possible terms that can be formed by one or more substitutions of the type described above.

Thus far the analysis has been rather general, applying to feedback-chain amplifiers using any type of two-terminal network as the shunt-loading admittance. At this point it is convenient to consider the simplest type of feedback chain, namely, that using single-tuned circuits as shunt-loading admittances.

If the amplifier uses single-tuned circuits as shunt admittances, then each Y_{kk} can be expressed as

$$Y_{kk} = G_{kk} + j2\pi C \left(f - \frac{1}{f} \right), \tag{10}$$

where the resonant frequency of the circuits is taken as unity.

The determinant of Eq. (9) therefore becomes

$$\Delta = \begin{vmatrix} G_{11} + j2\pi C \left(f - \frac{1}{f} \right) & -G_{12} & 0 & . & 0 \\ g_{m'_2} & G_{22} + j2\pi C \left(f - \frac{1}{f} \right) & -G_{23} & . & 0 \\ 0 & g_{m'_3} & G_{33} + j2\pi C \left(f - \frac{1}{f} \right) & . & . \\ 0 & 0 & g_{m'_4} & . & 0 \\ 0 & 0 & . & . & -G_{n-1,n} \\ 0 & 0 & . & . & G_{nn} + j2\pi C \left(f - \frac{1}{f} \right) \end{vmatrix} \tag{11}$$

This is clearly a polynomial of the nth degree in $j\left(f - \frac{1}{f}\right)$, in which the coefficient of the term of highest power is $(2\pi C)^n$. Equation (8) can therefore be written as

$$\mathcal{G} = \frac{(-1)^n g'_{m_1} g'_{m_2} \cdots g'_{m_n}}{(2\pi C)^n \left\{ \left[j\left(f - \frac{1}{f}\right)\right]^n + a_{n-1}\left[j\left(f - \frac{1}{f}\right)\right]^{n-1} + \cdots + a_1\left[j\left(f - \frac{1}{f}\right)\right] + a_0 \right\}}, \quad (12)$$

where the a's are constants depending upon the values of the G's and g'_m's.

This expression gives the gain as a complex function of frequency. The amplitude response is given by the absolute value of Eq. (12),

$$|\mathcal{G}| = \frac{g'_{m_1} g'_{m_2} \cdots g'_{m_n}}{\sqrt{(2\pi C)^{2n}\left(f - \frac{1}{f}\right)^{2n} + b_{n-1}\left(f - \frac{1}{f}\right)^{2(n-1)} + \cdots + b_1\left(f - \frac{1}{f}\right)^2 + b_0}}, \quad (13)$$

where the b's are functions of the a's of Eq. (12). Equation (13) can be written

$$|\mathcal{G}| = \left(\frac{g'_{m_1}}{2\pi C}\right)\left(\frac{g'_{m_2}}{2\pi C}\right) \cdots \left(\frac{g'_{m_n}}{2\pi C}\right) \times \frac{1}{\sqrt{\left(f - \frac{1}{f}\right)^{2n} + c_{n-1}\left(f - \frac{1}{f}\right)^{2(n-1)} + \cdots + c_1\left(f - \frac{1}{f}\right)^2 + c_0}}, \quad (14)$$

where $c_k = b_k/(2\pi C)^{2n}$.

For the maximum gain-bandwidth product obtainable in a pass band without dips, all the c_k's except c_0 must be zero. This gives a flat response having a curve of the form $1/\sqrt{(1 + X^{2n})}$. Equation (14) can be written

$$\mathcal{G} = \left(\frac{g'_{m_1}}{2\pi C}\right)\left(\frac{g'_{m_1}}{2\pi C}\right) \cdots \left(\frac{g'_{m_n}}{2\pi C}\right) \frac{1}{\sqrt{\left(f - \frac{1}{f}\right)^{2n} + c_0}}$$

$$= \frac{(\text{mean } g'_m)^n}{(2\pi C)^n \sqrt{\left(f - \frac{1}{f}\right)^{2n} + (c_0^{1/2n})^{2n}}}, \quad (15)$$

where "mean g'_m" is defined as the geometric mean of the g'_m's of the various stages.

The over-all 3-db bandwidth is $c_0^{1/2n}$. The mean gain is the nth root of the over-all gain; hence the gain-bandwidth product is

$$\sqrt[n]{\left(\frac{\text{mean } g'_m}{2\pi C}\right)^n \frac{1}{(c_0^{1/2n})^n}} \, c_0^{1/2n} = \frac{\text{mean } g'_m}{2\pi C}. \tag{16}$$

A comparison of the response of a general feedback chain with that of a stagger-tuned amplifier shows striking similarities. The response, as given by Eq. (12) is $\left(\dfrac{\text{mean } g'_m}{2\pi C}\right)^n$ divided by a polynomial of the nth degree in $j\left(f - \dfrac{1}{f}\right)$, of which the leading coefficient is unity. This expression is identical with that obtained in the stagger-tuned case. Thus, the feedback chain using single-tuned circuits for all shunt-loading admittances turns out to be equivalent to a stagger-tuned amplifier both in the shape of the response curve and in gain-bandwidth product. This is verified by the case of the flat n-uple, as shown by Eq. (16). There are, however, two important cases where this equivalence breaks down.

1. Since g'_m is slightly less than g_m, the gain-bandwidth product is slightly less for inverse-feedback chains than it is for stagger-tuned amplifiers. This difference becomes appreciable at very large bandwidths where the stage gain is low.

2. Certain response curves and bandwidth, although achievable with stagger-tuning, would require negative admittances as shunt-loading elements in the equivalent feedback chain and hence cannot be realized in a practical way by feedback chains.

An important difference between the feedback chain and the stagger-tuned amplifier is in the effect of normalizing the resonant frequency. Note that in Eq. (10) the use of a resonant frequency f_0 merely changes the frequency variable from $f - (1/f)$ to $f - (f_0^2/f)$. The effect of this change in Eq. (12) is also to replace $f - (1/f)$ by $f - (f_0^2/f)$ and does not change the constant $a_0, a_1, \ldots, a_{n-1}$. Thus, the shape of a response curve plotted against $\left(f - \dfrac{f_0^2}{f}\right)$ is independent of the resonant frequency and independent, therefore, of the ratio of the bandwidth to the center frequency. (This is not true in the stagger-tuned case.) If the curve is plotted against the frequency f instead of the variable $\left(f - \dfrac{1}{f}\right)$, the resulting curve is not symmetrical about f_0, but the bandwidth, as well as the curve shape as specified by the flatness or the number and amount of dips,

is unchanged. Thus the effect of retuning, by changing the tuning inductances of all the stages of the feedback chain to a different frequency, does not affect the absolute bandwidth of the amplifier or the flatness or number and amount of dips of the response curve, although the symmetry of the curve may be changed.

To investigate further the nature of the response obtainable from an inverse-feedback chain, i.e., which type of polynomial in $j\left(f - \dfrac{1}{f}\right)$ represents physically realizable feedback amplifiers, it is most convenient to examine first the simple case of the inverse-feedback pair.

6·3. The Inverse-feedback Pair.—The inverse-feedback pair can be considered as the special case of the general chain in which $n = 2$. Thus the frequency response of the pair can be determined directly from Eqs. (8) and (11).

Equivalence to Staggered Pair.—On substituting $n = 2$ into Eqs. (8) and (11), the following expression is obtained for the gain of a feedback pair as a function of frequency:

$$\mathcal{G} = \frac{g'_{m_1} g'_{m_2}}{G_{11} G_{22} + g'_{m_2} G_{12} + j2\pi C \left(f - \dfrac{1}{f}\right)(G_{11} + G_{22}) + \left[j2\pi C \left(f - \dfrac{1}{f}\right)\right]^2}$$

(17)

The denominator can be factored into two linear factors, as shown in Eq. (18).

$$\mathcal{G} = \frac{g'_{m_1} g'_{m_2}}{(2\pi C)^2 \left[d_1 + j\left(f - \dfrac{1}{f}\right)\right]\left[d_2 + j\left(f - \dfrac{1}{f}\right)\right]},$$

(18)

where

$$d_1 = \frac{G_{11} + G_{22} + \sqrt{(G_{11} + G_{12})^2 - 4(G_{11} G_{22} + g'_{m_2} G_{12})}}{4\pi C},$$

$$d_2 = \frac{G_{11} + G_{22} - \sqrt{(G_{11} + G_{22})^2 - 4(G_{11} G_{22} + g'_{m_2} G_{12})}}{4\pi C}.$$

(19)

The quantities d_1 and d_2 are either real or complex conjugates. If they are real, then Eq. (18) is obviously the response of two cascaded synchronous single-tuned stages having dissipation factors d_1 and d_2. If, however, d_1 and d_2 are complex conjugates, of real part d_a and imaginary part d_b, Eq. (18) can be written as

$$\mathcal{G} = \frac{g'_{m_1} g'_{m_2}}{(2\pi C)^2 \left\{d_a + j\left[\left(f - \dfrac{1}{f}\right) + d_b\right]\right\}\left\{d_a + j\left[\left(f - \dfrac{1}{f}\right) - d_b\right]\right\}}.$$

(20)

Comparison of Eq. (20) with Eq. (4·20) shows that the response of a feedback pair is identical with that of a staggered pair, except for the difference between g_{m_2} and g'_{m_2}. For the low-dissipation case, where arithmetic symmetry can be assumed, $f - (1/f)$ is approximately $2(f - 1)$, and Eq. (20) becomes

$$\mathcal{G} = \frac{g'_{m_1} g'_{m_2}}{(2\pi C)^2 \{d_a + j2[f - (1 - \frac{1}{2}d_b)]\}\{d_a + j2[f - (1 + \frac{1}{2}d_b)]\}}. \quad (21)$$

This equation is obviously the response of a staggered pair, staggered by an amount $\pm \frac{1}{2} d_b$. Thus, for a low-dissipation circuit a feedback pair is equivalent to a staggered pair of stages of dissipation factor d_a, staggered by an amount $\pm \frac{1}{2} d_t$, expressed as a fraction of the center frequency.

For the exact calculation of the equivalent staggered pair, rewrite Eq. (17) as

$$\mathcal{G} = \frac{g'_{m_1} g'_{m_2}}{(2\pi C)^2 \left\{ \dfrac{G_{11}G_{22} + g'_{m_2}G_{12}}{(2\pi C)^2} + j\dfrac{G_1 + G_2}{2\pi C}\left(f - \dfrac{1}{f}\right) + \left[j\left(f - \dfrac{1}{f}\right)\right]^2 \right\}}. \quad (22)$$

Comparing Eq. (22) with Eq. (4·20) for a staggered pair (exact case) shows that the two are equivalent if

$$d^2 + \left(\alpha - \frac{1}{\alpha}\right)^2 = \frac{G_{11}G_{22} + g'_{m_2}G_{12}}{(2\pi C)^2}, \quad (23a)$$

$$d\left(\alpha + \frac{1}{\alpha}\right) = \frac{G_{11} + G_{22}}{2\pi C}. \quad (23b)$$

Solving Eq. (23) for d and $\left(\alpha + \dfrac{1}{\alpha}\right)$ yields

$$d = \frac{1}{2}\left[\sqrt{\left(\frac{G_{11}}{2\pi C} + 2\right)\left(\frac{G_{22}}{2\pi C} + 2\right) + \frac{g'_{m_2}G_{12}}{(2\pi C)^2}} \right.$$
$$\left. - \sqrt{\left(\frac{G_{11}}{2\pi C} - 2\right)\left(\frac{G_{22}}{2\pi C} - 2\right) + \frac{g'_{m_2}G_{12}}{2\pi C}} \right] \quad (24a)$$

$$\left(\alpha + \frac{1}{\alpha}\right) = \frac{1}{2}\left[\sqrt{\left(\frac{G_{11}}{2\pi C} + 2\right)\left(\frac{G_{22}}{2\pi C} + 2\right) + \frac{g'_{m_2}G_{12}}{(2\pi C)^2}} \right.$$
$$\left. + \sqrt{\left(\frac{G_{11}}{2\pi C} - 2\right)\left(\frac{G_{22}}{2\pi C} - 2\right) + \frac{g'_{m_2}G_{12}}{2\pi C}} \right]. \quad (24b)$$

Reduction to Normal Form.—If in Eq. (20) the expression

$$G_{11}G_{22} + g'_m G_{12}$$

is set equal to G^2, then

$$
\mathcal{G} = \frac{g'_{m_1} g'_{m_2}}{G^2 + j2\pi C \left(f - \dfrac{1}{f}\right)(G_{11} + G_{22}) + \left[j2\pi C \left(f - \dfrac{1}{f}\right)\right]^2}
$$

$$
= \frac{g'_{m_1} g'_{m_2}}{G^2 \left\{1 + j\dfrac{2\pi C\left(f - \dfrac{1}{f}\right)}{G}\dfrac{G_{11} + G_{22}}{G} + \left[j\dfrac{2\pi C\left(f - \dfrac{1}{f}\right)}{G}\right]^2\right\}}. \quad (25)
$$

The absolute value of Eq. (25) is given by

$$
|\mathcal{G}| = \frac{g'_{m_1} g'_{m_2}}{G^2 \sqrt{1 + \left[\dfrac{2\pi C}{G}\left(f - \dfrac{1}{f}\right)\right]^2\left[\left(\dfrac{G_{11} + G_{22}}{G}\right)^2 - 2\right] + \left[\dfrac{2\pi C}{G}\left(f - \dfrac{1}{f}\right)\right]^4}}
$$

$$
= \frac{g'_{m_1} g'_{m_2}}{G^2 \sqrt{1 + kx^2 + x^4}}, \quad (26)
$$

where

$$
x = \frac{2\pi C}{G}\left(f - \frac{1}{f}\right), \quad (27a)
$$

$$
k = \left(\frac{G_{11} - G_{22}}{G}\right)^2 - 2 = \frac{G_{11}^2 + G_{22}^2 - 2g'_{m_2}G_{12}}{G_{11}G_{22} + g'_{m_2}G_{12}}. \quad (27b)
$$

Equation (26) is the normalized form of the expression for the response of an inverse-feedback pair. The shape of the response curve depends only upon the parameter k, since the quantities G and C merely determine scale factors. Thus any desired response can be achieved if k can be made the proper value.

Inspection of Eq. (27) shows that k can vary between -2 and $+2$. The value $k = -2$, occurring when $G_{12} = 0$ and $G_{11} = G_{22}$, represents two single-tuned circuits in cascade without feedback; the value $k = 2$, occurring when $g'_{m_2} G_{12}$ is large compared with G_{11}^2 and G_{22}^2, gives a response that goes to infinity when $x = 1$, thus producing a double-humped response curve with infinite peaks. Between these two values, the value $k = 0$ results in a response equivalent to that of a flat-staggered pair. All other types of intermediate curves can be obtained, having single or double peaks depending upon whether k is positive or negative.

Since a feedback pair is equivalent to a staggered pair, feedback n-uples consisting of feedback pairs can be built up in the same manner as staggered n-uples are built up from staggered pairs. The values of k are tabulated in Table 6.1 for n-uples from 1 to 4.

Equation (27b) can be solved for $g'_{m_2} G_{12}$:

$$g'_{m_2} G_{12} = \frac{G_{11}^2 + G_{22}^2 - k G_{11} G_{22}}{2 + k}. \tag{28}$$

Table 6·1.—Values of Curve-shape Index for Flat Pairs, Triples, and Quadruples

Curve shape	k	$|\alpha\beta_{min}|$
Two cascaded single-tuned circuits.............	2	0
Flat pair: $(1 + x^4)^{-\frac{1}{2}}$...................	0	1
Pair in flat triple: $(1 + x^6)^{-\frac{1}{2}}$...............	-1	3
Pairs in flat quadruple: $(1 + x^8)^{-\frac{1}{2}}$...........	$\begin{cases} -\sqrt{2} \\ +\sqrt{2} \end{cases}$	$\begin{aligned} 3 + 2\sqrt{2} \\ 3 - 2\sqrt{2} \end{aligned}$

Synthesis of a Feedback Pair.—The problem of designing an inverse-feedback pair to produce a desired response is the problem of determining the three conductances G_{11}, G_{22}, and G_{12}. The response is determined completely by two quantities: k, which determines the shape of the curve, and G, which determines the scale. Thus there are only two conditions to be satisfied by the three variables to be chosen. This extra degree of freedom permits a single infinity of parameter values, each giving the desired response.

If Eq. (28) is substituted into the definition of G one gets

$$G^2 = \frac{(G_{11} + G_{22})^2}{2 + k}. \tag{29}$$

Equations (29) and (28) can be reduced to the following form, which is more convenient for the design of a pair:

$$\frac{G_{11}}{G} + \frac{G_{22}}{G} = \sqrt{2 + k} \tag{30a}$$

$$\frac{g'_{m_2} G_{12}}{G^2} = 1 - \frac{G_{11}}{G} \frac{G_{22}}{G} \tag{30b}$$

Equations (30) are the fundamental equations for the design of an inverse-feedback pair in terms of the desired response. There is an extra degree of freedom present; the disposal of this extra degree of freedom depends upon other considerations, which can be introduced into Eqs. (30) to determine the particular design desired.

Equations (30) show that for a given shape of the response curve, i.e., a given value of k, the quantities G_{11}/G, G_{22}/G, and $g'_{m_2} G_{12}/G^2$ are independent of the value of G and of the center frequency of the amplifier.

Thus, once these quantities have been determined by the shape of the amplifier, it is necessary merely to introduce the value of G as a proportionality factor in order to determine the actual ohmic values of the circuit resistances. From Eq. (26) it is evident that the over-all gain at band center is $g'_{m_1} g'_{m_2} / G^2$. Thus the over-all bandwidth for a flat n-uple containing this pair will be $G/2\pi C$. The value of G for a given amplifier is therefore determined by either the gain or the bandwidth desired.

Note that Eqs. (30) are symmetrical in G_{11} and G_{22}, so that G_{11} and G_{22} can be interchanged in any amplifier without changing the over-all response. The distribution of gain between the individual stages is affected by this interchange, so that considerations other than the shape of the over-all response curve may dictate the use of one or the other scheme.

Although the values of the normalized constants of a feedback pair can be determined for any value of k from -2 to 2, it is not always possible to scale this to a physically realizable amplifier having the desired gain or bandwidth because $g'_{m_2} G_{12}$ is proportional to G^2, whereas g_m is fixed by the type of tube in use. Therefore as G increases, G_{11} and G_{22} increase in proportion to G, while G_{12} increases as G^2. There is, then, a value of G above which G_{12} exceeds G_{11} or G_{22}. This condition is not possible physically, because G_{11} has been defined as the sum of G_{12} and G_1 and must therefore be greater than G_{12}. A similar argument holds for G_{22}. Thus for each value of k, there is a limit on the gain and bandwidth that is obtainable using any specified tube type. Since the limit is an upper limit on G, it is a lower limit on gain and an upper limit on bandwidth.

These limits can easily be calculated from Eqs. (30). Since the gain of the pair is $g_m g'_{m_2} / G^2$, the quantity g'_{m_2}/G is roughly proportional to the gain per stage. It is therefore convenient to solve Eqs. (30) for g'_{m_2}/G.

$$\frac{g'_{m_2}}{G} = \frac{G_{11}^2 - k G_{11} G_{22} + G_{22}^2}{G_{12}(G_{11} + G_{22})\sqrt{2+k}}. \tag{31}$$

Because G_{11} and G_{22} are interchangeable, there is no loss of generality in assuming that G_{22} is the larger of the two. Equation (31) can therefore be written

$$\frac{g'_{m_2}}{G} = \frac{G_{11}}{G_{12}} \frac{1 - k\dfrac{G_{22}}{G_{11}} + \left(\dfrac{G_{22}}{G_{11}}\right)^2}{\left(1 + \dfrac{G_{22}}{G_{12}}\right)\sqrt{2+k}}$$

$$\geqq \frac{1 - k\dfrac{G_{22}}{G_{11}} + \left(\dfrac{G_{22}}{G_{11}}\right)^2}{\left(1 + \dfrac{G_{22}}{G_{11}}\right)\sqrt{2+k}}, \tag{32}$$

since $G_{11} \geqq G_{12}$.

If $g'_{m_2} = g_{m_1}$, Eq. (32) represents the minimum gain per stage. Even if $g_{m_1} \neq g'_{m_2}$, however, the maximum bandwidth obtainable for a flat n-uple containing this pair is

$$\mathcal{B}_{max} = \frac{G}{2\pi C} = \frac{g'_{m_2}}{2\pi C} \frac{\left(1 + \dfrac{G_{22}}{G_{11}}\right)\sqrt{2+k}}{1 - k\dfrac{G_{22}}{G_{11}} + \left(\dfrac{G_{22}}{G_{11}}\right)^2}. \tag{33}$$

In the design of an amplifier, the maximum bandwidth obtainable rather than the minimum gain is the significant quantity. However, Eq. (32) is useful in that it involves no tube characteristics. Therefore, once a given shape of response is specified, the minimum possible gain per stage with $g_{m_1} = g'_{m_2}$ is also specified, regardless of the tube type to be used. The bandwidth obtainable using any particular tube type can then be obtained from the g_m/C ratio of the tube type.

In Eqs. (32) and (33), the extra degree of freedom in the design appears as the ratio G_{22}/G_{11}, which can be chosen arbitrarily while still maintaining the desired response-curve by proper choice of the other parameters. Since G_{22} was assumed to be larger than G_{11}, it becomes obvious from Eqs. (32) and (33) that the value of G_{11}/G_{22} giving the greatest value for $\mathcal{B}_{max}$ is $G_{11}/G_{22} = 1$. For this case Eqs. (30), (32), and (33) become

$$\frac{G_{11}}{G} = \frac{G_{22}}{G} = \frac{\sqrt{2+k}}{2}, \tag{34a}$$

$$\frac{g'_{m_2} G_{12}}{G^2} = \frac{2-k}{4}, \tag{34b}$$

$$G_{min} = \frac{2-k}{2\sqrt{2+k}}. \tag{34c}$$

$$\mathcal{B}_{max} = \frac{g'_{m_2}}{\pi C} \frac{\sqrt{2+k}}{2-k}. \tag{34d}$$

Equations (34) are the fundamental design equations for a symmetrical feedback pair. Since this pair gives the greatest maximum bandwidth in addition to being the simplest case, it is generally used where there are no other considerations requiring unequal loading in the two stages.

Note that for $k = 0$ and $k = -1$, corresponding respectively to the flat pair and the flat triple, the minimum gain, again assuming $g_{m_1} = g'_{m_1}$, comes out respectively 0.707 and 1.5. These values are so low that the restriction upon maximum bandwidth is of little significance, since it is not generally practical to build an amplifier having a gain per stage of less than 1.5. For higher-order n-uples or in cases where a response hav-

ing dips is desired, the restriction upon maximum bandwidth is important, as it is also in the case of the pair and triple where the loading is unsymmetrical.

Another special case of the inverse-feedback pair that is of interest is the limiting case where one of the shunt loading conductances, either $G_1 = G_{11} - G_{12}$ or $G_2 = G_{22} - G_{12}$, is zero. Then either G_{11} or G_{22} is equal to G_{12}. This case has the largest physically realizable ratio of G_{11} to G_{22} for a given over-all response. Because the values of G_{11}/G, G_{22}/G, and $g'_{m_2} G_{12}/G^2$ are not independent of G for this case, the design equations are more complicated and involve the solution of a quadratic equation. The equations for this case can be obtained from Eq. (30) by letting either G_{11} or G_{22} equal G_{12}. The values of G_{12} and G_{ii}, where G_{ii} is either G_{11} or G_{22}, are given by

$$\frac{G_{12}}{G} + \frac{G}{G_{12}} = \frac{g'_{m_2}}{G} + \sqrt{2 + k}, \tag{35a}$$

$$\frac{G_{ii}}{G} = \frac{G}{G_{12}} - \frac{g'_{m_2}}{G}. \tag{35b}$$

Other Formulations for the Response of a Feedback Pair.—Although the exact values of the qualities α and β used in the conventional feedback-amplifier analysis are obscure for an inverse-feedback pair, the product $\alpha\beta$, which represents the transmission around the feedback loop, can easily be calculated. If a voltage e_g is impressed on the grid of the second stage of the pair, the plate signal current is $-g_{m_2} e_g$ and divides between Y_2 and G_{12}. The plate signal voltage is easily calculated to be

$$e_p = -e_g \frac{g_{m_2} - G_{12}}{G_2 + G_{12}} = -e_g \frac{g'_{m_2}}{G_{22}}. \tag{36}$$

The portion of this voltage fed back across the grid is

$$e'_g = e_p \frac{G_{12}}{G_{12} + G_1} = e_p \frac{G_{12}}{G_{11}} = -e_g \frac{g'_{m_2} G_{12}}{G_{11} G_{22}}, \tag{37}$$

and therefore the feedback factor $\alpha\beta$ is

$$\alpha\beta = -\frac{g'_{m_2} G_{12}}{G_{11} G_{22}}. \tag{38}$$

The significance of this feedback factor in the previous analysis can be shown by rewriting the expression for the gain of the feedback pair as follows:

$$\mathcal{G} = \frac{g_{m_1} g'_{m_2}}{Y_{11} Y_{22} + g'_{m_2} G_{12}} = \frac{g_{m_1} g'_{m_2}}{Y_{11} Y_{22}} \frac{1}{1 - \alpha\beta}. \tag{39}$$

Equation (39) has the factor $1 - \alpha\beta$ in the denominator which is characteristic of feedback amplifiers. In this form it becomes clear that the variation of amplifier gain with variations in transconductance of the second stage is reduced by the factor $1 - \alpha\beta$, or $1 + (g'_{m_2} G_{12}/G_{11}G_{22})$, from what it would be in the absence of feedback.

The relation between the feedback factor $\alpha\beta$ and the quantity k determining the shape of the response curve of the pair can be calculated from Eq. (27b).

$$k = \frac{\dfrac{G_{11}}{G_{22}} + \dfrac{G_{22}}{G_{11}} - 2|\alpha\beta|}{1 + |\alpha\beta|}, \tag{40a}$$

$$|\alpha\beta| = \frac{\dfrac{G_{11}}{G_{22}} + \dfrac{G_{22}}{G_{11}} - k}{2 + k}. \tag{40b}$$

Equation (40) shows that the feedback factor required for a given response is not uniquely determined by the curve-shape parameter k, because of the extra degree of freedom present in the design of inverse-feedback pairs, but depends also on the ratio G_{22}/G_{11} of the loading of the two stages. The feedback factor $1 - \alpha\beta$ is a minimum for any given selectivity-curve shape when $G_{11} = G_{22}$. This is the case of the symmetrical feedback pair, treated in Eqs. (34). For this case, Eqs. (40) can be written

$$k = \frac{2(1 - |\alpha\beta_{\min}|)}{(1 + |\alpha\beta_{\min}|)}, \tag{41a}$$

$$|\alpha\beta_{\min}| = \frac{2 - k}{2 + k}. \tag{41b}$$

Values of $|\alpha\beta_{\min}|$ are tabulated in Table 6·1 along with the values of k for the pairs used in the various n-uples.

The maximum value of the feedback factor giving a desired selectivity curve shape is obtained by making $(G_{11}/G_{22}) + (G_{22}/G_{11})$ as large as possible. This therefore corresponds to the limiting case treated in Eqs. (35) where one of G_1 or G_2 is zero. For this case

$$|\alpha\beta_{\min}| = \frac{g'_{m_2}}{G_{11}}, \tag{42a}$$

$$\left(1 + \frac{g'_{m_2}}{G}\sqrt{2 + k}\right)|\alpha\beta_{\min}|^2$$
$$- \left(\frac{g'_{m_2}}{G}\right)\left(\frac{g'_{m_2}}{G} - \sqrt{2 + k}\right)(|\alpha\beta_{\min}| - |\alpha\beta_{\min}|)^2 = 0. \tag{42b}$$

As might be suspected from Eqs. (35), the value of $|\alpha\beta|$ for this case depends upon G and requires the solution of a quadratic equation. The

values are therefore not tabulated. This case is of particular interest, since it represents the pair that has the maximum stability of gain with regard to tube variations (see Sec. 6·5).

Equation (38) can also be written

$$\mathcal{G} = \frac{\dfrac{g_{m_1}}{G_{12}}}{1 + \dfrac{Y_{11}Y_{22}}{g'_{m_2}\,G_{12}}} = \frac{\dfrac{g_{m_1}}{G_{12}}}{1 - \dfrac{1}{\alpha\beta}}. \tag{43}$$

Since the feedback factor $\alpha\beta$ is greater than or equal to unity for all pairs having a flat or double peaked response, the over-all gain $\mathcal{G}$ of a feedback pair is given approximately by

$$\mathcal{G} \approx \frac{g_{m_1}}{G_{12}}. \tag{44}$$

This approximation is too high by a factor of 2 when $\alpha\beta = 1$, i.e., flat symmetrical pair, and is better as $\alpha\beta$ increases.

6·4. Synthesis of a Feedback Chain.—The problem of synthesizing a feedback chain is one of determining the loading and feedback conductances necessary to obtain a desired response. Section 6·2 has shown that the response of an n-stage inverse-feedback chain can be expressed as the reciprocal of a polynomial of the nth degree in $f - (1/f)$, band center being taken as unity. This polynomial can then be factored into n linear factors. The synthesis problem is therefore one of determining n load conductances plus $n - 1$ feedback conductances, or a total of $2n - 1$ quantities, from the n equations that can be written for the n linear factors desired. There is therefore, in general, an $(n - 1)$-fold infinity of possible solutions to the synthesis problem because of the $n - 1$ extra degrees of freedom.

In designing a feedback chain it is consequently possible to impose $n - 1$ auxiliary conditions upon the loading and feedback conductances in addition to the specification of the over-all response. There are various approaches to the design of the feedback chain, each using a different set of auxiliary conditions.

Reduction to Pairs.—The simplest method of feedback-chain design consists in arbitrarily setting every second feedback conductance equal to zero. This method results in an amplifier made up of feedback pairs, either with or without a single stage having no feedback, depending upon whether n is odd or even. Thus, depending upon whether n is odd or even, either $\frac{1}{2}(n - 1)$ or $\frac{1}{2}n - 1$ of the auxiliary conditions are used, and $\frac{1}{2}(n - 1)$ or $\frac{1}{2}n$ remaining degrees of freedom are left.

The problem is now one of designing a number of inverse-feedback pairs to have a given over-all response. The factors of the polynomial

representing the response can be grouped into pairs so that the two factors in each pair are either real or complex conjugates. These pairs of factors can then be normalized to the form of Eq. (26), and each pair can be designed by use of Eqs. (30). Note that there is an extra degree of freedom in the design of each pair, thus making the total number of extra degrees of freedom equal to $n - 1$, as expected. If n is odd, there is one factor remaining alone, after the grouping in pairs; this factor corresponds to a single stage without feedback.

The design of a feedback chain has thus been simplified to the design of a number of pairs. There are various practical advantages resulting from the greater simplicity of this circuit over that of a chain with feedback throughout. There are no known advantages to the use of any other type of feedback chain over that of feedback pairs.

The physical realizability of a chain designed in this manner is limited by the same conditions that limit the design of a feedback pair. Thus, for a given shape of response, the minimum gain per stage is given by Eq. (34c) applied to the pair having the most negative k. Although it may be possible, by using more complicated chain circuits, to obtain a wider range of physically realizable amplifiers, this point has not been investigated because of the involved nature of the algebra encountered.

Synthesis of a Feedback Triple.—The three-stage feedback chain, or feedback triple, is somewhat more involved than the feedback pair, but it is simple enough to allow its use in practical feedback amplifiers. It can be shown that any inverse-feedback triple can be replaced by a combination of a feedback pair plus a single stage, giving the same over-all response, with reduction neither of stability nor of physically realizable bandwidth; consequently there appears to be little specific advantage in the use of the triple. However, the usage of the feedback triple has been sufficient to warrant its treatment here.

Since a triple has three stages, there is a twofold infinity of solutions in the design of a triple for a given response and the design equations have two extra degrees of freedom.

The design equations for a triple can be written in a simple form if the response is specified in its complex form; i.e.,

$$\mathcal{G} \propto \frac{1}{1 + \lambda_1(jx) + \lambda_2(jx)^2 + (jx)^3}. \tag{45}$$

In this case, the two parameters λ_1 and λ_2 must be specified to determine the shape of the response curve. The parameters λ_1 and λ_2 correspond to the single quantity k specifying the response of a pair. Note, however, the important distinction between k and the λ's, in that k is a coefficient of the polynomial formed by taking the absolute value of the response whereas λ_1 and λ_2 are coefficients of the complex expression.

The design equations for the triple can easily be derived by evaluating Eq. (11) for $n = 3$ and equating terms with the corresponding terms of Eq. (45).

Thus

$$G_{11}G_{22}G_{33} + [g'_{m_2}G_{12} + g'_{m_3}G_{23}][\rho G_1 + (1 - \rho)G_{33}] = 1, \qquad (46a)$$

$$(G_{11} + G_{33})G_{22} + G_{11}G_{33} + g'_{m_2}G_{12} + g'_{m_3}G_{23} = \lambda_1, \qquad (46b)$$

$$G_{11} + G_{33} + G_{22} = \lambda_2, \qquad (46c)$$

where G is taken as unity, so that the values obtained for G_{11}, G_{22}, and G_{33} are actually G_{11}/G, G_{22}/G, and G_{33}/G, and $g'_{m_2}G_{12}$ and $g'_{m_3}G_{23}$ are actually these quantities divided by G^2. The symbol ρ is defined as

$$\frac{g'_{m_2}G_{12}}{g'_{m_2}G_{12} + g'_{m_3}G_{23}}.$$

If G_{22} and $g'_{m_2}G_{12} + g'_{m_3}G_{23}$ are eliminated from Eq. (46), a single equation in the three variables G_{11}, G_{33}, and ρ is obtained.

$$\rho(G_{11}^3 - \lambda_2 G_{11}^2 + \lambda_1 G_{11} - 1) \\ + (1 - \rho)(G_{33}^3 - \lambda_2 G_{33}^2 + \lambda_1 G_{33} - 1) = 0, \quad (47a)$$

where

$$G_{22} = \lambda_2 - (G_{11} + G_{33}), \qquad (47b)$$

$$g'_{m_2}G_{12} + g'_{m_3}G_{23} = \lambda_2 - (G_{11} + G_{33})G_{22} - G_{11}G_{33}. \qquad (47c)$$

Equations (47) are the design equations for the feedback triple. There is considerable leeway in the choice of values, since there are two variables in Eq. (47a) that can be chosen arbitrarily within certain limits.

Note that the cases $\rho = 1$ and $\rho = 0$ are the degenerate cases corresponding to the breakdown of the triple into a single stage without a feedback pair. For these cases, either G_{11} or G_{33} is the real root of the cubic equation

$$x^3 - \lambda_2 x^2 + \lambda_1 x - 1 = 0. \qquad (48)$$

Since this means that $jx + G_{11}$ is a factor of the expression for the response, this procedure is clearly the equivalent of factoring the response into a linear factor representing a single stage without feedback and a quadratic factor representing a feedback pair.

Another degenerate case occurs when $G_{11} = G_{33}$. Then each of G_{11} and G_{33} is a real root of Eq. (48), and G_{22} and $g'_{m_2}G_{12} + g'_{m_3}G_{23}$ can be obtained from Eqs. (47). The extra degree of freedom is present in ρ, which can be chosen arbitrarily; thus for this case, the sum of the feedback conductances must be a constant (assuming that $g'_{m_2} = g'_{m_3}$), but this sum can be partitioned between G_{12} and G_{23} in any arbitrary manner. This case is of little practical significance, however, since for most desired

response curves, such as the flat curves, this solution is not realizable physically, as discussed below.

If G_{11} and G_{33} are unequal, it follows at once from Eq. (47a) that one of G_{11} and G_{33} is greater than the real root of Eq. (48) and the other is less. Furthermore, Eq. (47b) shows that G_{11} plus G_{33} must be less than λ_2.

One case of the feedback triple of special interest is that for which a flat response of the form $1/\sqrt{(1 + x^6)}$ is obtained. For this case, $\lambda_1 = \lambda_2 = 2$ and Eqs. (47) become

$$\rho(G_{11}^3 - 2G_{11}^2 + 2G_{11} - 1) + (1 - \rho)(G_{33}^3 - 2G_{33}^2 + 2G_{33} - 1) = 0, \quad (49a)$$

$$G_{22} = 2 - (G_{11} + G_{33}), \quad (49b)$$

$$g'_{m_2}G_{12} + g'_{m_3}G_{23} = 2 - (G_{11} + G_{33})G_{22} - G_{11}G_{33}. \quad (49c)$$

For this case, the real solution of Eq. (48) is unity, and the conditions for physical realizability are

$$G_{11} < 1 \qquad \text{if } G_{33} > 1 \qquad \text{and} \qquad G_{33} < 1 \qquad \text{if } G_{11} > 1, \quad (50a)$$

$$(G_{11} + G_{33}) < 2, \quad (50b)$$

along with the previous conditions for any feedback chain,

$$G_{11} \geqq G_{12}, \quad (50c)$$

$$G_{22} \geqq G_{12} + G_{23}, \quad (50d)$$

$$G_{33} \geqq G_{23}. \quad (50e)$$

Note that the first two of these inequalities place restrictions on the amplifier that hold for all values of G whereas the last three (in actuality the most stringent of the last three for any particular triple) place no restriction on the shape obtainable but set an upper limit on the bandwidth obtainable for the desired curve shape.

The Uniform Chain.—Considerable work has been done on the feedback chain, in which all stages are identical except for the first and last, which are referred to as terminations. The analysis of this type of amplifier can be carried out in a manner analogous to filter theory,[1] and the result obtained that a gain-bandwidth factor (see page 172) of 2 is possible if (1) the stage gain is very high; (2) the termination is perfect, eliminating all reflections; and (3) there is no shunt conductance across the tuned circuits.

[1] H. A. Wheeler, "Wideband Amplifiers for Television," *Proc. I.R.E.*, pp. 433–437, July 1939; E. Feenberg and W. W. Hansen, "Wideband Amplifier Design," Sperry Gyroscope Co. Report, Nov. 11, 1942; A. J. Ferguson, "The Theory of I-f Amplifiers with Negative Feedback," National Research Council (Canada) Report No. PRA-59, October 1942; R. Q. Twiss, "The Theoretical Design and Experimental Response of Single and Coupled Circuit Negative Feedback I-f Amplifiers," Telecommunications Research Establishment (England) Report No. T-1649.

The preceding analysis, however, has shown that for any feedback chain containing single-tuned circuits only, the maximum gain-bandwidth product obtainable for a flat response is unity. If the apparent discrepancy between these two points of view is examined, it will be revealed that if the uniform feedback-chain amplifier is *terminated in a single-tuned circuit* instead of the assumed perfect termination, the effect of the reflections is such as to produce wiggles in the shape of the pass band. If the amplifier design is modified to flatten out these wiggles, the gain-bandwidth factor is brought down to unity. Furthermore, if wiggles are allowable in the pass band, the preceding analysis has shown that it is possible to duplicate the performance of this uniform chain in all respects, including that of gain-bandwidth product, with an amplifier consisting of feedback pairs. It is necessary merely to obtain the mathematical expression for the response desired. A set of functions that can be used for specifying this type of response is the Tschebycheff polynomials.[1]

In order to secure the advantage of the uniform chain when properly terminated, it is necessary to use more complicated terminations than the simple single-tuned circuit. This matter is discussed in more detail in Sec. 6·7.

6·5. Miscellaneous Properties of Inverse-feedback Chains and Pairs. *Gain Stability.*—Because of the presence of feedback, it is to be expected that the gain of an inverse-feedback chain displays less variation with changes in tube characteristics than a bandpass amplifier having no feedback. This fact can be easily verified.

The gain at band center of a feedback chain can be obtained from Eqs. (8) and (9) by substituting Y_{11} for G_{11}. For convenience, assume that all the stages have the same g_m and that $g'_m = g_m$. Then the gain of the amplifier at band center can be expressed as

$$\mathcal{G} = \frac{(g_m)^n}{A_0 + A_1 g_m + A_2 g_m^2 + \cdots + A_{n/2} g_m^{n/2}}, \qquad (51)$$

where the coefficients A_k are functions of the load and feedback conductances.[2] The effect upon the gain of changes in g_m can be obtained by taking the logarithmic derivatives of Eq. (51).

$$\frac{d\mathcal{G}}{\mathcal{G}}$$

$$= \frac{dg_m}{g_m} \left[\frac{nA_0 + (n-1)A_1 g_m + (n-2)A_2 g_m^2 + \cdots + \frac{1}{2} n A_{n/2} g_m^{n/2}}{A_0 + A_1 g_m + A_2 g_m^2 + \cdots + A_{n/2} g^{n/2}} \right]. \quad (52)$$

[1] Courant-Hilbert, *Methoden der Mathematischen Physik*, Springer, Berlin, 1934, p. 75.

[2] For simplicity, an even number of stages is assumed. The results for an odd number of stages are the same, with $\frac{1}{2}(n-1)$ substituted for $\frac{1}{2}n$.

From Eq. (52) it can be seen that if A_0 is large compared with the other coefficients there is very little feedback, and the per cent change in gain is the same as it would be for an n-stage amplifier without feedback, namely, n times the per cent change in the transconductance of a single stage.

If the amount of feedback is very large, then all terms are negligible except the ones involving $A_{n/2}$. The per cent change in the gain is then found to be $n/2$ times the per cent change in the transconductance of a single stage, or half the per cent change in gain that would occur in the absence of feedback. If some intermediate value of feedback is used, the reduction in sensitivity to g_m variation is somewhat less; but in no case does a feedback chain reduce sensitivity to g_m variation by a factor larger than 2.

The apparently anomalous situation where enormous amounts of feedback produce an improvement of only a factor of 2 in gain stability can be better understood if the case of a feedback pair is examined. From Eq. (38) the following equation can be written for the gain of a feedback pair at band center:

$$ \mathcal{G} = \frac{g_{m_1} g'_{m_2}}{G_{11} G_{12} \left(1 + \dfrac{g_{m_2} G_{12}}{G_{11} G_{22}} \right)}. \tag{53} $$

The logarithmic derivative of Eq. (53) gives the following expression for the gain variation:

$$ \frac{d\mathcal{G}}{\mathcal{G}} = \frac{dg_{m_1}}{g_{m_1}} + \frac{dg_{m_2}}{g_{m_2}} \cdot \frac{1}{1 + \dfrac{g_{m_2} G_{12}}{G_{11} G_{22}}} = \frac{dg_{m_1}}{g_{m_1}} + \frac{dg_{m_2}}{g_{m_2}} \cdot \frac{1}{1 - \alpha\beta} \tag{54} $$

Equation (54) shows that the variation in gain due to variations in g_{m_2} is reduced by a factor of $1 - \alpha\beta$ from what it would be in the absence of feedback, as is usually the case in feedback amplifiers. However, there is no change in the gain variation due to changes in g_{m_1} over what would occur in the absence of feedback. This is also to be expected, since g_{m_1} is not included in the feedback loop, there being feedback only around the second stage. Thus, by increasing the amount of feedback, the change in gain due to variations in g_{m_2} can be made negligible, whereas that due to variations in g_{m_1} is entirely unaffected.

Consider now an amplifier consisting of cascaded inverse-feedback pairs. By making the feedback sufficiently large, the variation in gain produced by the variation of g_m in the second stage of each pair can be made small. There remains unchanged, however, the variation due to the first stage in each pair. Thus the per cent variation in over-all gain will be that due to only half of the stages and will be equal to half

the per cent variation that would be produced in the absence of feedback. This result is exactly the same as that obtained for the general chain.

Thus it can be seen that the gain stability of an inverse-feedback chain with feedback around every stage is no better than that of an amplifier consisting of cascaded pairs, i.e., with feedback around alternate stages. The most efficient way, therefore, to obtain gain stability in a feedback chain is to use cascaded pairs, thus stabilizing the gain of half the stages, and to employ some other form of stabilization for the other half of the stages, e.g., large cathode-bias resistors.

Overloading.—The analysis of a feedback pair has shown that there is an extra degree of freedom in designing a pair, which effectively changes the distribution of the gain between stages without affecting the over-all response. This immediately suggests putting as much gain as possible in the second stage and as little as possible in the first, thus reducing the size of the signal appearing on the grid of the second stage. Since the ratio of the signal on the second grid to that on the first has been decreased considerably from its normal value, a larger signal is required on the first grid to overload the second grid. The overload characteristics of the amplifier are therefore improved.

The gains of the two stages can easily be obtained from Eq. (4):

$$\mathcal{G}_1 = \frac{g_{m_1}}{G_{11} + \dfrac{g'_{m_2} G_{12}}{G_{22}}}, \tag{55a}$$

$$\mathcal{G}_2 = \frac{g'_{m_2}}{G_{22}}. \tag{55b}$$

The factor by which the signal on the second grid is reduced by the feedback is the ratio of the mean stage gain of the pair to the gain of the first stage, or

$$\frac{\sqrt{\mathcal{G}_1 \mathcal{G}_2}}{\mathcal{G}_1} = \frac{\sqrt{G_{11} G_{22} + g_m G_{12}}}{G_{22}} = \frac{G}{G_{22}} \tag{56}$$

where

$$g_{m_1} = g'_{m_2} = g_m.$$

This factor is a maximum for the case where G_{12} and G_{22} are equal since the maximum possible gain is then in the second stage. Then, since g_m is generally large in comparison with G_{11}, the gain of the first stage is approximately unity. The factor by which the signal on the second grid is reduced must then be the mean gain per stage.

A word of warning is in order about this improvement in overload characteristics. Because the analysis was made for band center only, the improvement applies to single-frequency signals at the center of the band,

but it does not apply to pulsed signals or to single-frequency signals off band center. The effect of a change in frequency can easily be seen from Eq. (56), where for frequencies off band center, G_{22} must be replaced by Y_{22} and G_{11} must be replaced by Y_{11}, where

$$Y_{22} = G_{22} + j2\pi C\left(f - \frac{1}{f}\right),$$
$$Y_{11} = G_{11} + j2\pi C\left(f - \frac{1}{f}\right).$$

It can be seen that this change causes a rapid increase in $\sqrt{G_1 G_2}/G_1$ as the frequency departs from resonance. In particular, when

$$2\pi C\left(f - \frac{1}{f}\right) = G,$$

which corresponds to the half-power point of a flat n-uple,

$$Y_{22} = G_{22} + jG, \tag{57a}$$
$$Y_{11} = G_{11} + jG, \tag{57b}$$

and

$$\frac{\sqrt{G_1 G_2}}{G_1} = \frac{\sqrt{jG(G_{11} + G_{12})}}{G_{22} + jG} = \frac{\sqrt{j}\sqrt{2 + k}}{jk\dfrac{G_{22}}{G}}, \tag{58a}$$

$$\left|\frac{\sqrt{G_1 G_2}}{G_1}\right| = \frac{\sqrt[4]{2 + k}}{\sqrt{1 + \left(\dfrac{G_{22}}{G}\right)^2}} \approx \sqrt[4]{2 + k}. \tag{58b}$$

From Eq. (58b) it can be seen that the overload improvement factor in Eq. (56) applies only at band center and has dropped off to a factor of at most $\sqrt{2}$ at the frequency corresponding to the half-power point of a flat n-uple.

Because of this effect, a pulse behaves according to the analysis for band center during the flat portion of the pulse, and the signal amplitude required to overload the second grid during this portion of the pulse is much greater than in the absence of feedback. However, overloading occurs during the leading edge of the pulse almost as soon as the pulse amplitude exceeds the amplitude required to overload an amplifier with the gain equally distributed between the two stages.

The exact nature of the overload occurring at the leading edge is somewhat complicated but can be partially understood by considering the waveform of the envelope of the pulse appearing on the second grid. This envelope is shown in Fig. 6·6. The peak is produced by the time required for the feedback to take effect. It is this peaked signal applied

to the grid which makes the output pulse rise more sharply than it would if there were a square pulse applied to the grid.

The overloading at the second grid is of two varieties, one driving the tube into the grid-current region and the other driving the tube to cutoff. The former drives the tube into a region of increasing transconductance, thus increasing the feedback, and the other drives it into a region of decreasing transconductance thus decreasing the feedback. Since an i-f pulse involves both directions, the situation is somewhat complex. However, if the spike of Fig. 6·6 is short compared with the pulse length, the lack of overloading during the flat portion of the pulse may be of considerable advantage.

In general, the overloading during the leading edge of the pulse has the effect of making the leading edge of the output pulse slower than it would be in the absence of feedback. However, the gain-bandwidth improvement due to feedback is not entirely lost, since (1) the trailing edge of the pulse is less affected by the overload and (2) all pulses below the overload level have the fast rise

FIG. 6·6.—Peaked envelope of pulse on second grid of a feedback pair.

corresponding to the increased bandwidth. This is illustrated in Fig. 6·7, which compares the waveforms of signals below and above the overload level for a feedback amplifier with the waveforms of signals of the same

(a)   (b)
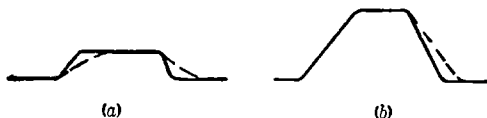
FIG. 6·7.—Response to (a) weak signals and (b) overloading signals for feedback pair (solid line), and narrowband two-stage amplifier with same gain distribution between stages as feedback pair (dashed line).

levels for an amplifier without feedback having the same gain distribution between stages as the feedback amplifier.

High Output Level.—The possibility of putting the bulk of the gain of a feedback pair into the output stage allows a much higher output voltage to be obtained for a given bandwidth than can be obtained without feed-

back. The factor by which the output voltage can be increased by the use of feedback is equal to the overload-improvement factor in Eq. (56). However, this output voltage increase is subject to the same limitations that apply in the use of this circuit for the improvement of overload (see page 104). The improvement in speed effected by feedback is lost on strong signals. Thus the high output voltage is obtained at the expense of speed just as in an amplifier without feedback, but this loss in speed shows up principally on the leading edges of pulses of maximum amplitude. There is no loss of speed on pulses below the overload level.

Bandwidth Switching.—In all the preceding analysis it has been assumed that g'_{m_2} is approximately equal to g_{m_2}. An interesting phenomenon results if g_{m_2} is allowed to decrease to the point where the difference between g_{m_2} and g'_{m_2} becomes appreciable. A point is reached where g_{m_2} is equal to G_{12} and where g'_{m_2} must therefore be zero. At this null condition, the transmission of the signal directly around the second tube through the feedback conductance is exactly equal and opposite to that through the tube. If g_{m_2} is decreased still further, the signal transmitted through the tube is less than that directly transmitted, g'_{m_2} is negative, and the output signal increases, but with a phase reversal. When g_{m_2} is zero, $g'_{m_2} = -G_{12}$, corresponding to direct transmission alone. The gain in this case may be greater than the gain with g_{m_2} at its normal value, because of the absence of feedback when g_{m_2} is zero. The bandwidth is, however, considerably reduced. The sharpened selectivity curve thus obtained when the second tube of the pair is biased to cutoff can be used to simplify tuning the amplifier, since in that condition it is necessary merely to maximize the output signal of a very sharp circuit. This is not quite accurate, however, because of the change in the input capacity of the tube when it is turned off.

Good use can be made of the dependence of bandwidth on g_m in an electrical two-position bandwidth switch. By properly adjusting the values of the circuit constants in a feedback pair it is possible to achieve any desired bandwidth in the narrow position when the second tube is turned off, subject to some restrictions on physical realizability. In addition, the change in gain between the two positions can also be controlled by taking advantage of the extra degree of freedom normally present in the design of a feedback pair and by adding an additional degree of freedom consisting of suitably selecting the value of g_{m_2} in the wide-band position. These two degrees of freedom can be used to determine the gain ratio and the bandwidth ratio of the switch.

The calculation of the design of such a bandwidth switch is somewhat complicated. There are four independent variables: the bandwidths in the two cases, the gain ratio, and the shape of the response curve in the wide-band position, as specified by the factor k. These four variables

TABLE 6·2a.—BANDWIDTH SWITCH DESIGN DATA*

$\dfrac{\mathfrak{B}_n \dagger}{\mathfrak{B}_w}$	0.445				0.386				0.347				0.300				0.253			
$\dfrac{G_w}{G_n}$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$
5.000	8.86	0.681	4.700	2.135	13.39	0.614	4.660	2.350	24.20	0.645	4.730	2.6300	73.20	0.634	4.85	2.9400	....	....	....	....
2.000	19.40	0.759	3.165	1.267	78.40	0.736	3.155	1.414	....	....	....	....	....	....	....	....	....	....	....	....
1.000	64.50	0.844	2.450	0.816	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
0.500	∞	1.000	2.000	0.500	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
0.300	111.90	1.140	1.790	0.335	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
0.200	43.40	1.282	1.670	0.239	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
0.100	15.22	1.550	1.550	0.129	266.00	1.545	1.490		....	....	....	....	....	....	....	....	....	....	....	....
0.050	....	....	....	....	17.81	1.866	1.421	0.0782	97.10	1.896	1.369	0.0192	....	....	....	....	....	....	....	....
0.025	....	....	....	....	9.31	....	1.384	0.0400	16.60	2.285	1.330	0.0470	900.00	2.335	1.28	0.0555	....	....	....	....
0.010	....	....	....	....	6.15	2.745	1.362	0.0163	8.79	2.805	1.307	0.0191	13.85	2.930	1.26	0.0277	397	3.06	1.213	0.0247

* $k = 2$: Curve shape in wide-band position that of two cascaded single-tuned circuits.

† $\mathfrak{B}_w$ = over-all bandwidth of pair in wide case.

TABLE 6·2b.—BANDWIDTH SWITCH DESIGN DATA*

$\dfrac{\mathcal{B}_n\,\dagger}{\mathcal{B}_w}$	0.401				0.374				0.348				0.297			
$\dfrac{\mathcal{G}_w}{\mathcal{G}_n}$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$
5.00	2.345	1.78	4.70	2.37	3.11	1.494	4.72	2.50	3.84	1.368	4.73	2.63	5.95	1.149	4.74	2.94
3.00	. . . .	. . .	. . .	. . . .	2.64	2.165	3.76	1.89	3.72	1.630	3.76	1.99	6.20	1.365	3.85	2.24
2.00	. . . .	. . .	. . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	6.04	1.590	3.22	1.78
1.00	. . . .	. . .	. . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .
0.85	. . . .	. . .	. . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .
0.80	. . . .	. . .	. . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .
0.75	. . . .	. . .	. . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .
0.55	. . . .	. . .	. . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .

$\dfrac{\mathcal{B}_n\,\dagger}{\mathcal{B}_w}$	0.248				0.2245				0.202			
$\dfrac{\mathcal{G}_w}{\mathcal{G}_n}$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$
5.00	9.31	1.100	5.030	3.320	12.52	1.055	5.15	3.540	18.20	1.014	5.290	3.780
3.00	11.21	1.208	3.985	2.523	18.21	1.159	4.06	2.700	35.40	1.100	4.170	2.890
2.00	13.11	1.360	3.320	2.020	49.75	1.256	3.38	2.155	97.50	1.21	3.465	2.31
1.00	9.20	1.970	2.490	1.341	20.34	1.739	2.53	1.440	330.00	1.568	2.580	1.550
0.85	———	———	———	———	———	———	2.38	1.301	———	———	———	———
0.80	5.64	2.720	2.300	1.170	———	———	———	———	———	———	———	———
0.75	. . . .	. . . .	. . . .	. . . .	10.30	2.235	2.30	1.220	40.50	1.900	2.315	1.300
0.55	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	10.13	2.850	2.070	1.067

* $k = 0$: Curve shape in wide-band position that of flat pair.

† $\mathcal{B}_w$ = over-all bandwidth of pair.

TABLE 6·2c.—BANDWIDTH SWITCH DESIGN DATA*

$\dfrac{\mathcal{B}_n}{\mathcal{B}_w}$†	0.200				0.175				0.1505				0.1267			
$\dfrac{\mathcal{G}_w}{\mathcal{G}_n}$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_2}R$	$\dfrac{R_1}{R}$	$\dfrac{R_2}{R}$	$\dfrac{R_{12}}{R}$	$g'_{m_i}R$
5.00	5.39	2.16	5.70	4.39	8.32	2.69	6.00	4.780	14.65	1.670	6.350	5.250	31.4	1.477	6.850	2.85
3.00	3.62	3.62	4.48	3.36	8.76	2.17	4.69	3.655	18.15	1.844	4.975	4.030	77.1	1.612	5.350	4.49
2.00	....	....	....	....	....	....	....	.....	18.40	2.185	4.110	3.250	966.0	1.825	4.405	3.63
1.50	....	....	....	....	....	....	....	.....	12.82	2.680	3.605	2.785	———	———	———	———
1.25	....	....	....	....	....	....	....	.....	7.92	3.960	3.320	2.520	———	———	———	2.50
1.00	....	....	....	....	....	....	....	.....	.....	.....	.....	.....	44.8	2.810	3.215	....
0.90	....	....	....	....	....	....	....	.....	.....	.....	.....	.....	24.6	3.240	3.075	2.36
0.80	....	....	....	....	....	....	....	.....	.....	.....	.....	.....	.....	.....	.....	....

* $k = -1$: Curve shape in wide-band position that of pair in flat triple.
† $\mathcal{B}_w$ = over-all bandwidth of flat triple including pair.

are used to determine the four circuit parameters G_1, G_{12}, G_2, and g_{m_2}. The algebra of the solution is involved and is not reproduced here. However, the results are summarized in Table 6·2, which gives the design data for feedback pairs in terms of the desired gain ratio and bandwidth ratio for three values of k, namely 2, 0, and -1, corresponding respectively to curve shapes in the wide-band position of (1) two cascaded single-tuned circuits, (2) a flat pair, (3) the pair included in a flat triple. The constants of the pair are normalized, i.e., expressed as ratios to G. Since in the design of an amplifier one usually specifies resistances rather than conductances, the quantities G/G_1, G/G_2, and G/G_{12} are expressed as R_1/R, R_2/R, and R_{12}/R, where $R = 1/G = 1/2\pi C \mathscr{B}_w$.

In Table 6·2 values of $\mathscr{G}_w/\mathscr{G}_n$, the ratios of the gain in the wide position to that in the narrow position, are tabulated vertically; values of $\mathscr{B}_n/\mathscr{B}_w$, the ratio of the narrower to the wider bandwidths, are tabulated horizontally. The constants of the amplifier are tabulated in the square corresponding to the bandwidth ratio and gain ratio.

A considerable number of values of gain and bandwidth ratios cannot be obtained because the amplifiers required would contain physically unrealizable elements. Such values are marked with dots. The range of physically realizable values varies widely with the factor k, as can be seen by comparing Table 6·2a, b, and c. An important region not calculated is for the case of $k = -1$, where there are physically realizable values of bandwidth ratio below 0.1 which do not require unreasonable values of the gain.

Dashed values in the table are physically realizable but uncalculated because they are of little practical interest.

6·6. Practical Considerations in Feedback-amplifier Design.—The approach thus far in this chapter has been a theoretical one in which the circuit components have been assumed to behave ideally. These assumptions make the amplifier suitable for a mathematical analysis that does not become too involved and provide solutions that furnish good first-order approximations to the actual circuit constants required to produce the desired response. In actual amplifier design, however, the second-order effects frequently are very important, and considerable care must be taken to reduce them to a negligible level or to compensate for them.

The mathematical approach used in the preceding section becomes extremely involved when these second-order effects are considered. Instead, it becomes expedient to use an approach based mainly upon experiment and a simplified analysis using vector diagrams, such as has been used very successfully by Beveridge.[1]

The main second-order effects found in feedback amplifiers are those

[1] H. N. Beveridge, "Information on Broad-band Feedback I-f Amplifiers," Combined Research Group, Naval Research Laboratory Report CRG-93, Oct. 22, 1945.

due to grid-plate capacity and grid-plate transit angle. Both of these effects tend to make the response curve unsymmetrical about the center frequency. By use of vector diagrams it is possible to determine the extent to which symmetry is impaired and the proper means of compensa-
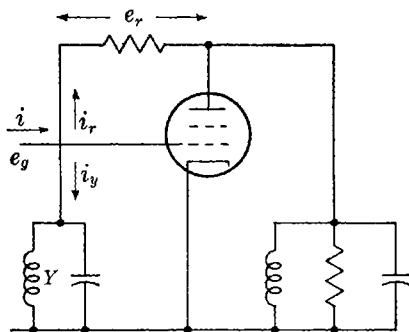
FIG. 6·8.—Feedback stage.

tion. The magnitude of these effects can be discovered by experiment, and their correction determined by the vector diagram.

Vector Diagram for Feedback Stage with No Second-order Effects.— Consider the feedback stage shown in Fig. 6·8. The vector diagram of
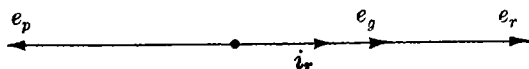
FIG. 6·9.—Vector diagram at resonance.

Fig. 6·9 applies to this stage at resonance. The input current is in phase with the input voltage, since the feedback voltage is also in phase. Figure 6·10 shows the vector relationships for the case of e_g at a higher and at a lower frequency than the resonant frequency. For the higher-
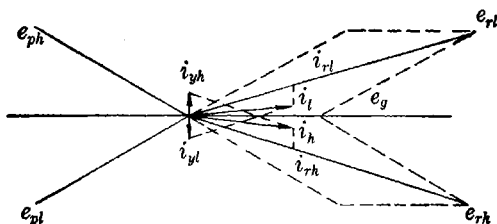
FIG. 6·10.—Vector diagram off resonance.

frequency case e_{ph} is retarded in phase, and for the lower-frequency case e_{pl} is advanced in phase. The current i_{yh} leads e_g, whereas the current i_{rh} lags e_g. The voltage e_r across the feedback resistor is $e_g - e_p$, and the current i_r is in phase with e_r. The input current i is the sum of i_r and i_y.

It can be seen that the input current i is a symmetrical function about the center frequency; that is, i_h and i_l are symmetrically disposed about e_g, if $i_{yl} = -i_{yh}$.
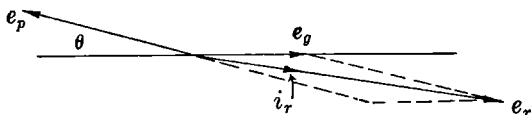
FIG. 6·11.—Vector diagram at resonance showing transit-time effect.

Vector Diagram for Feedback Stage with Transit Angle.—Figure 6·11 shows a vector diagram for the circuit of Fig. 6·8 at resonance when there is a grid-plate transit angle θ.   Here e_p is retarded by the transit angle through the tube.   The current i_r now lags e_g, instead of being in phase
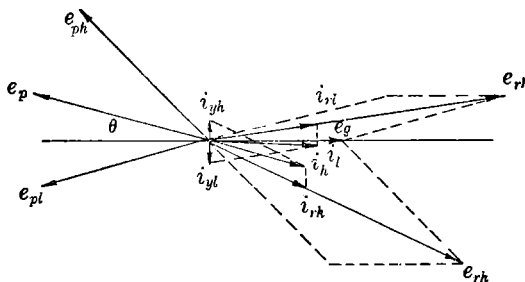
FIG. 6·12.—Vector diagram off resonance showing transit-time effect.

as before, and the input current therefore also lags i_r.   Figure 6·12 shows this effect for frequencies above and below resonance.   Because of the transit angle, e_{ph} and e_{pl} are not symmetrical with respect to e_g. Thus i_{rh} and i_{rl} are not symmetrical; and when the equal currents i_{yh} and i_{yl} are added, the resulting input currents are unequal, i_h being smaller
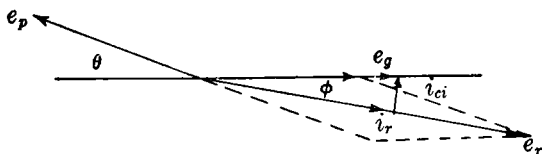
FIG. 6·13.—Correction of transit-time effect at resonance.

than i_l.   Since the output voltages e_p are equal, the gain is therefore higher above resonance than below.

This effect can be corrected by adding a capacity in parallel with the feedback resistor.   The vector diagram for resonance is then as shown in Fig. 6·13, with the addition of the capacitive feedback current bringing the total feedback current back into phase with e_g.   Figure 6·14 shows

that i_{rh}, i_r, and i_{rl}, which were unsymmetrically disposed about e_g, are brought back into a symmetrical disposition about e_g and i_h and i_l are therefore equal.   Although this correction is really only approximate, it is satisfactory.

If the gain is high, the angles θ and ϕ can be assumed to be equal, and the relation between the transit angle and the feedback resistor can be found from Fig. 6·13 to be

$$\tan \theta = \omega C_{12} R_{12}. \tag{59}$$

Effects of Capacity across the Feedback Resistor.—There is always a small amount of capacity across the feedback resistor.   The effect on the response is similar to that of transit angle, except that it is in the opposite direction; it can be used, in fact, to compensate for the effect of transit angle.   If the capacitance across the feedback resistor is larger than necessary to compensate for the transit-time effect, then its presence is serious.   Equation (59) can therefore be considered to define an upper limit on the allowable value of the feedback resistor, where C_{12} is the inherent capacity across the

FIG. 6·14.—Correction of transit-time effect off resonance.

feedback resistor.   If the design considerations demand a feedback resistor higher than this value, as may readily occur at frequencies where the transit-time effect is negligible, other means must be used for preserving the flatness and symmetry of response.

FIG. 6·15.—Feedback resistor tapped on load resistor.

The first thing to be done about feedback-resistor capacity is to reduce it to a minimum.   The capacity across common half-watt resistors is approximately 0.2 $\mu\mu f$ or more.   Certain special low-capacity resistors, such as the International Resistance Company's type MPM,[1] have capacities between $\frac{1}{4}$ and $\frac{1}{10}$ of this capacity.

The obvious method for treating this capacity is to tune it out with a parallel inductance.   This method slightly upsets the feedback and may require a small decrease in the feedback resistor or an increase in the load resistor.   Another possi-

[1] Vol. 17, Chap. 2, Radiation Laboratory Technical Series.

bility is to tap the feedback resistor down on the load resistor, as is shown in Fig. 6·15.[1] The combination of the two load resistors and the feedback resistor can be transformed into an equivalent T, as shown in Fig. 6·16, or an equivalent H, as shown in Fig. 6·15. These use lower resistances in the feedback network than the original circuit and are therefore less troubled by stray capacitance.

If either G_1 or G_2 is zero, this method does not work, but the feedback resistor can still be tapped down on the coil, as shown in Fig. 6·17.
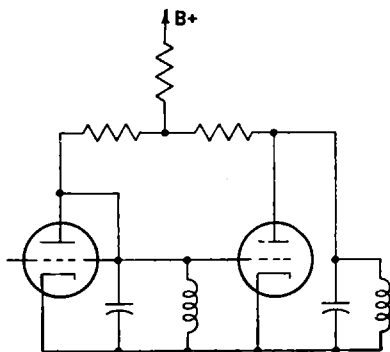
FIG. 6·16.—T-network feedback.   FIG. 6·17.—Feedback resistor tapped on coil.
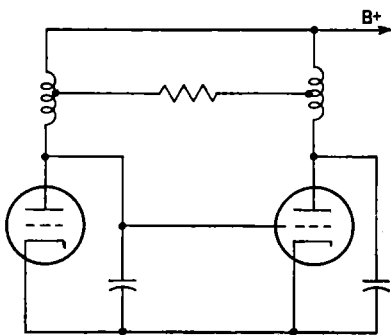
6·7. More Complicated Feedback Amplifiers.—Since it has been shown that the gain-bandwidth product of an amplifier using single-tuned circuits can be improved by the use of inverse feedback, it is natural to expect that amplifiers using more complicated interstage-coupling circuits can similarly be improved.

The mathematical analysis of a general feedback chain using circuits more complicated than the simple single-tuned circuit becomes so involved that the solution is extraordinarily difficult. However, if all the stages but the last and first are made identical, the mathematics is sufficiently simplified so that an analysis can be carried through.

In this section a brief analysis is made for a uniform chain using two-terminal networks. The use of four-terminal coupling networks is considered in great detail by Twiss.[2]

The Uniform Feedback Chain.—The properties of an inverse feedback chain in which all the stages are identical can be determined from Eq. (9).

[1] Bartelink, Watters, and Kamke, "Flat Response, Single-tuned I-f Amplifier," General Electric Company internal report (forthcoming); also B. D. H. Tellegen and C. J. van Loon, U. S. Patent 2155025 (1939); this patent contains a very good early discussion of inverse-feedback bandpass circuits.

[2] R. Q. Twiss, "The Theoretical Design and Experimental Response of Single and Coupled Circuit Negative Feedback Amplifiers," T.R.E. Report No. T-1649.

Assuming that all the g_m's are equal, Eq. (9) becomes for the uniform chain

$$\Delta = \begin{vmatrix} Y & -G_f & 0 & & & & & 0 & 0 & 0 \\ g_m - G_f & Y & -G_f & 0 & \cdot & \cdot & \cdot & 0 & 0 & 0 \\ 0 & g_m - G_f & Y & -G_f & 0 & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & 0 & g_m - G_f & Y & -G_f & 0 & \cdot & 0 & 0 & 0 \\ \cdot & & & & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & & & & & \cdot & \cdot & g_m - G_f & Y & -G_f & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdot & 0 & g_m - G_f & Y & -G_f \\ 0 & 0 & 0 & 0 & 0 & \cdot & 0 & 0 & g_m - G_f & Y \end{vmatrix}. \quad (60)$$

where Y is the loading admittance Y_{ii}, and G_f is the feedback conductance $G_{i-1,i}$.

If Δ_n is the value of Δ for an n-stage amplifier, Δ_n can be expressed in terms of Δ_{n-1} and Δ_{n-2} by expanding Δ_n by minors.

$$\Delta_n = Y \Delta_{n-1} + g_m G_f \Delta_{n-2}. \quad (61)$$

This is a difference equation, the solution for which is given by

$$\Delta_n = A(Y_a)^n + B(Y_b)^n, \quad (62a)$$

where

$$Y_a = \frac{Y}{2} + \sqrt{\left(\frac{Y}{2}\right)^2 + (g_m - G_f)G_f}, \quad (62b)$$

$$Y_b = \frac{Y}{2} - \sqrt{\left(\frac{Y}{2}\right)^2 + (g_m - G_f)G_f}, \quad (62c)$$

$$A = \frac{1}{2} + \frac{Y}{\sqrt[4]{\left(\frac{Y}{2}\right)^2 + (g_m - G_f)G_f}} = \frac{Y_a}{Y_a - Y_b}, \quad (62d)$$

$$B = \frac{1}{2} - \frac{Y}{\sqrt[4]{\left(\frac{Y}{2}\right)^2 + (g_m - G_f)G_f}} = \frac{Y_a}{Y_b - Y_a} \quad (62e)$$

Note that for the case $G_f = 0$, or no feedback, $Y_a = Y$, $Y_b = 0$, $A = 1$, $B = 0$, and Δ_n therefore is Y^n, which is the familiar value for the uniform multistage amplifier without feedback.

If the values for A, B, Y_a, and Y_b are substituted from Eq. (62) into Eq. (61), the square roots cancel out and the result is the polynomial expected from previous theory, having a gain-bandwidth factor of unity for a flat-response curve when single-tuned circuits are used for shunt admittances. However, Eq. (62), unsimplified, shows certain properties of the uniform chain that can be used to secure a greater gain-bandwidth product with the aid of circuits more complicated than single-tuned circuits. Consider the function Y_a if Y is the admittance of a single-tuned circuit. Then

$$Y_a = \frac{G}{2} + j\pi C\left(f - \frac{1}{f}\right) + \sqrt{\left[\frac{G}{2} + j\pi C\left(f - \frac{1}{f}\right)\right]^2 + (g_m - G_f)G_f}. \quad (63)$$

If $G_f \ll g_m$ (this corresponds to high gain), then Eq. (63) is

$$Y_a \approx j\pi C \left(f - \frac{1}{f}\right) + \sqrt{g_m G_f - \left[\pi C \left(f - \frac{1}{f}\right)\right]^2}. \tag{64}$$

From Eq. (64) it is evident that the magnitude of Y_a is independent of frequency for all values of f such that the quantity under the radical remains positive and increases as f varies beyond this range. Since the value of Y_a at resonance is $g_m G_f$ and the width of the band over which $|Y_a|$ is constant is $\sqrt{g_m G_f}/\pi C$, it is obvious that Y_a corresponds to an impedance having a gain-bandwidth factor of at least 2 for any number of stages. This is analogous to the two-terminal infinite filters yielding maximum gain-bandwidth product, as described in Sec. 2·2.

Although the advantage of the flat-amplitude characteristic of Y_a is lost by the addition to it of the Y_b term in any amplifier employing single-tuned circuits only, the use of a more complex circuit in the first or last stage of the amplifier allows the retention of this improvement in gain-bandwidth factor. Consider an amplifier where the load and feedback impedances of the last (or first) stage are Y_t and Y_{ft}. Substituting these values and also substituting Eq. (62a) in Eq. (61) yields

$$\Delta_n = A(Y_t Y_a + g_m Y_{ft}) Y_a^{n-2} + B(Y_t Y_b + g_m Y_{ft}) Y_b^{n-2}. \tag{65}$$

The Y_b term can be eliminated if

$$Y_t Y_b + g_m Y_{ft} = 0. \tag{66}$$

Substituting Eq. (62) in Eq. (66):

$$Y_t = Y_a \frac{Y_{ft}}{G_f}. \tag{67}$$

Then

$$\Delta_n = (Y_a)^n \frac{Y_{ft}}{G_f}. \tag{68}$$

If $Y_{ft} = G_f$, then $\Delta_n = (Y_a)^n$, and $Y_t = Y_a$.

Thus an inverse-feedback chain that is terminated in an impedance Y_a has the response of an amplifier without feedback in which the load impedance in each stage is Y_a. Since Eq. (64) shows that Y_a is the admittance of the infinite two-terminal filter previously discussed, feedback chains, using only a single filter in either the first or last stage instead of one in each stage, can duplicate the performance of amplifiers using filters as coupling networks. This represents an appreciable simplification of the amplifier.

However, in actual practice this is not quite the case. Practical feedback chains do not realize the expected gain-bandwidth factor 2 for the following reasons:

1. The assumption made in arriving at Eq. (64) is true for high gains only. It is equivalent to assuming that the feedback conductance represents a pure transfer admittance, having negligible self-admittance. This assumption is, in general, not true over the range of gains commonly used for wide-band amplifiers. The effect of this difference is to decrease the gain-bandwidth product of a single stage and to round off the corner of the response curve so that the bandwidth decreases more rapidly when several stages are cascaded. This effect is illustrated in Table 6·3.

2. The presence of direct loading, in addition to the loading of the feedback resistor, increases the effect of the latter[1] as explained in (1). This is also illustrated in Table 6·3.

3. Because the numerator of Eq. (8) contains g'_m rather than g_m, there is a small loss in gain-bandwidth product, namely, a factor of

$$\left(1 - \frac{G_f}{g_m}\right) \text{ or } \left(1 - \frac{1}{G^2}\right).$$

4. Some performance is lost because the termination is never ideal.

TABLE 6·3.—GAIN-BANDWIDTH FACTOR OF A PERFECTLY TERMINATED FEEDBACK CHAIN AS A FUNCTION OF MEAN STAGE GAIN, DIRECT LOADING, AND NUMBER OF STAGES

$\dfrac{g_m G_f}{G^2}$	Gain-bandwidth factor			
	1 stage	2 stages	3 stages	4 stages
1.425	2.06	1.580	1.34	1.200
2.160	2.10	1.660	1.43	1.290
7.000	2.17	1.870	1.74	1.645
11.650	2.20	1.940	1.81	1.730
17.000	2.19	1.990	1.88	1.810
∞	2.15	2.045	2.02	2.020

6·8. Practical Examples.—As a first example of an amplifier using inverse feedback, consider the four-stage amplifier shown in the circuit diagram of Fig. 6·18. This amplifier was designed for an over-all band-width of 21 Mc/sec and a gain of 30 to 35 db. A brief calculation of the gain-bandwidth product for a 6AC7 shows that the requirements can be achieved with a flat quadruple, consisting of two inverse-feedback pairs.

The pairs to be used for this amplifier are chosen as symmetrical, as this is the simplest method. The amplifier can then be designed from

[1] The loading effect of the feedback resistor and the direct loading can be somewhat compensated for by use of circuits producing a negative resistance at the input of the amplifier, such as a capacitive cathode circuit.

Eq. (34). However, for a more accurate design, the difference between g_{m_2} and g'_{m_2} is taken into account. Thus, substituting $g_{m_2} - G_{12}$ into Eq. (34b) for g'_{m_2}, a quadratic equation is obtained for G_{12}/G.

$$\left(\frac{G_{12}}{G}\right)^2 - \left(\frac{g_{m_2}}{G}\right)\left(\frac{G_{12}}{G}\right) + \frac{2-k}{4} = 0. \tag{69}$$

If Eq. (69) is solved for G_{12}/G and the square root is approximated by taking the first three terms of its binomial expansion, one gets

$$\frac{G_{12}}{G} = \frac{2-k}{4}\left(\frac{G}{g_{m_2}}\right)\left[1 + \frac{2-k}{4}\left(\frac{G}{g_{m_2}}\right)^2\right]. \tag{70}$$

Note that taking only the first two terms of the binomial expansion is equivalent to letting g_{m_2} equal g'_{m_2} in Eq. (34b).
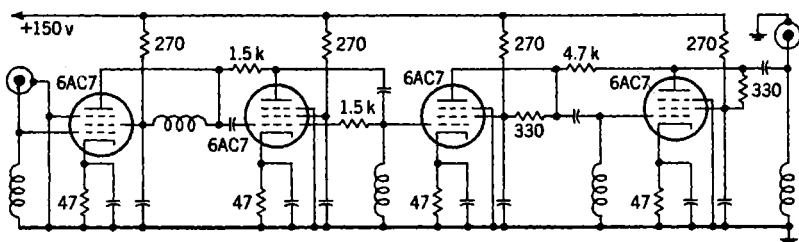
Fig. 6·18.—Flat quadruple feedback amplifier.

From the bandwidth desired and the total interstage shunt capacity, both of which are known, the constant G can be determined since the bandwidth is given by $G/2\pi C$. Assuming a value of 26 $\mu\mu$f for C, G comes out to be 3450 μmhos. Then, assuming $g_m = 8600$, $g_m/G = 2.5$.

Now the constants of the amplifier can be determined from Eqs. (34a) and (70). From Table 6·1, it can be seen that the values of k for the two pairs of a flat quadruple are $\sqrt{2}$ and $-\sqrt{2}$. Thus, using Eq. (34a),

$$\frac{G_{11}}{G} = \frac{G_{22}}{G} = \frac{\sqrt{2 \pm \sqrt{2}}}{2} \begin{cases} 0.925 \text{ for } k = +\sqrt{2},\ G_{11} = 3200 = G_{22}, \\ 0.383 \text{ for } k = -\sqrt{2},\ G_{11} = 1333 = G_{22}. \end{cases}$$

From Eq. (70)

$$\frac{G_{12}}{G} = \frac{2 \mp \sqrt{2}}{4}\frac{1}{2.5}\left[1 + \frac{2 \mp \sqrt{2}}{4}\cdot\frac{1}{(2.5)^2}\right]$$
$$= \begin{cases} 0.060 \text{ for } k = +\sqrt{2},\ G_{12} = 207, \\ 0.388 \text{ for } k = -\sqrt{2},\ G_{12} = 1340. \end{cases}$$

Then

$$G_1 = G_2 = G_{11} - G_{12} = \begin{cases} 3000 \text{ for } k = +\sqrt{2}, \\ 0 \text{ for } k = -\sqrt{2}. \end{cases}$$

Thus, for $k = +\sqrt{2}$,

$$R_1 = R_2 = 330 \text{ ohms}, \qquad R_{12} = 4800 \text{ ohms},$$

and for $k = -\sqrt{2}$,

$$R_1 = R_2 = \infty, \qquad R_{12} = 750 \text{ ohms}.$$

FIG. 6·19.—Circuit diagram of a practical 60 Mc/sec feedback amplifier. The interstage coils are unity-coupled, by being bifilar wound.

A practical difficulty is encountered because of the infinite values obtained for R_1 and R_2, since there must be low-resistance d-c paths from the plate to B and from grid to ground. One method of handling this difficulty would be to split the tuning coil into two coils, each having double the required inductance, and using one in the plate circuit and one in the grid circuit. In the case of this amplifier, another method was used. The 750-ohm feedback resistor was split into parallel resistors of 1500 ohms each, one being connected from grid to grid, the other from plate to plate. In this way a d-c return path is furnished for both plates and both grids if the coil is in the plate circuit of one stage and the grid circuit of the other.

FIG. 6·20.—High-frequency inverse-feedback amplifier.

A second example of a feedback amplifier is due to H. N. Beveridge. Fig. 6·19 is a circuit diagram and Fig. 6·20 a photograph of the last few stages. This amplifier consists of two inverse-feedback triples, having very slight dips in the pass band, with a gain of about 10^5 and a bandwidth of 10 Mc/sec at 60 Mc/sec.

The long thin resistors seen in Fig. 6·20 near the center of the chassis and running along its length are IRC type MPM resistors, used as feedback resistors because of their very low end-to-end capacity, of about 0.02 $\mu\mu$f. The tuning coils are bifilar-wound, i.e., unity-coupled, thereby eliminating the need for blocking condensers; the bifilar construction requires the coils to be very carefully impregnated, however, in order to prevent electrolysis. The coils are at right angles to the length of the chassis; the last one is shown without its impregnation.

The white stand-off bypass condensers visible in Fig. 6·20 provide extremely effective bypassing because of the employment of the "series-resonance" principle; i.e., the capacity of the bypass condensers is series resonant at 60 Mc/sec with the inductance of the condenser leads. A large part of the credit for the extreme simplicity and elegance of the amplifier, as demonstrated by the small number of parts shown in Fig. 6·19, is to be attributed to the effectiveness of the bypassing.

CHAPTER 7

BANDPASS AMPLIFIERS: PULSE RESPONSE AND GENERAL CONSIDERATIONS

By Henry Wallman

PULSE RESPONSE

7·1. Response of Bandpass Amplifier to Carrier-frequency Pulse.
Bandpass Filters That Are Approximately Low-pass Filters Shifted Upward in Frequency.—The response of a bandpass amplifier to a microsecond pulse of a 30 Mc/sec wave can in principle be determined without approximation by applying the methods of Chap. 1; the driving function is

$$f(t) = \sin (2\pi 30 \times 10^6 t), \qquad 0 < t < 10^{-6},$$
$$= 0, \text{ otherwise,}$$

and it is necessary to employ the exact form of the complex bandpass system function. Such calculations are extremely laborious[1] but show just how the carrier-frequency oscillations are built up in the output waveform. The result of such a calculation is as shown in (a) of Fig. 7·1.

In certain applications such detailed information is important; an example is the case in which "phase detection" is employed in conjunction with a coherent carrier-frequency oscillator. Usually, however, only simple rectification follows the bandpass amplifier, and it is sufficient to know the envelope of the carrier-frequency pulse [see (c) of Fig. 7·1].

For amplifiers of small fractional bandwidth it is often possible to simplify the calculations greatly on the basis of the following mathematical argument. Suppose a time function $f(t)$ has the Laplace transform $g(\omega)$. Then $f(t)e^{i\omega_0 t}$ has the transform $g(\omega - \omega_0)$. Now the time function $f(t)e^{i\omega_0 t}$ is used to denote a sine-wave carrier of frequency ω_0 modulated in amplitude by $f(t)$; its spectrum function is seen to be that of $f(t)$ shifted by the carrier-frequency ω_0. Some bandpass filters can be regarded as the displacement in frequency of a lowpass filter, that is, are

---

[1] A. W. Gent, "The Transient Response of R-f and I-f Circuits to a Wave Packet," Standard Telephone and Cables Co., Ltd., Valve Laboratory Report No. G70, 1942. Exact calculations are made for the response of a single-tuned and a double-tuned circuit to a carrier-frequency pulse. It turns out that even for a fractional bandwidth as large as a half, the "envelope" of the carrier-frequency response is well approximated by the response of the low-pass-equivalent filter.

closely approximated by a filter whose mathematical expression can be written in terms of the single variable $j(\omega - \omega_0)$. An example [see Sec. 4·2, in particular Fig. 4·4, and Eq. (4·6)] is a single-tuned circuit of small fractional bandwidth. For such filters it is then true that the envelope of the response (of the bandpass filter to the modulated carrier-frequency wave) is the response (of the corresponding low-pass filter)
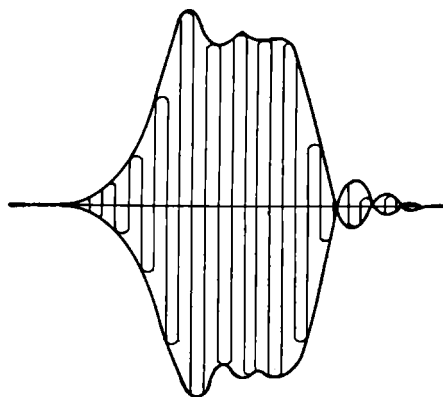
(a)
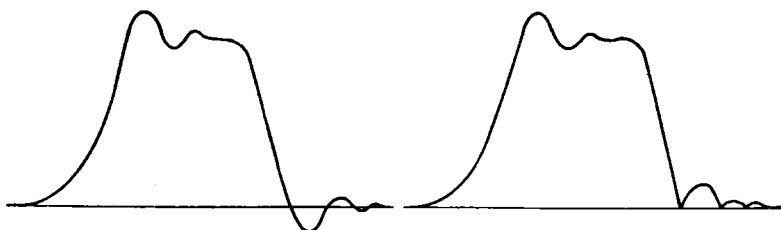
(b)   (c)

Fig. 7·1.—Pulse response: (a) Response of bandpass amplifier to carrier-frequency pulse; (b) pulse response of low-pass equivalent of bandpass amplifier, (c) effect of rectification on (a).

to the envelope. This is shown in (b) of Fig. 7·1. All pulse responses in this chapter are computed on this simplified basis.

Observe that (c) of Fig. 7·1 is the absolute value of (b) of Fig. 7·1; thus the effect of overshoot on the trailing edge of a carrier-frequency pulse is different from its effect on the leading edge.

As an example of the simplified method consider the determination of the response of a single-tuned circuit at 30 Mc/sec, of bandwidth 2 Mc/sec, to a microsecond pulse of a 30 Mc/sec wave. The low-pass

filter corresponding to the 30 Mc/sec single-tuned circuit is a parallel *RC*-circuit having a bandwidth of 1 Mc/sec. Hence the leading edge of the envelope of the carrier-frequency pulse response is closely approximated by the response to a step function of direct current of a parallel *RC*-circuit of 1 Mc/sec bandwidth. The complete pulse response is the sum of the leading-edge response and its 1-μsec-delayed negative.

It is most important to observe that a 2 Mc/sec bandwidth in a oandpass amplifier yields only as much speed as does a 1 Mc/sec bandwidth in a low-pass amplifier. Hence the analogue of Rule 5 of Chap. 2 is this:

If τ denotes the rise time of the envelope of a carrier-frequency pulse and $\mathcal{B}$ the bandwidth between 3-db points of a bandpass amplifier, then[1]

$$\tau\mathcal{B} = 0.7 \text{ to } 0.9.$$

Not every bandpass filter can be regarded as a frequency-shifted low-pass filter. In particular, a filter displaying an amplitude tilt across the band cannot be so regarded, because the amplitude response of a low-pass filter is always an even function of frequency. Hence for such bandpass amplifiers or for detuned amplifiers, the computation of carrier-frequency pulse response is more complicated and is not considered here.

Lowpass-bandpass Analogy.—There is a systematic procedure for generating bandpass filters from low-pass prototypes that is often helpful in simplifying carrier-frequency pulse response calculations. To transform a given low-pass filter to a bandpass analogue centered at a frequency f_0,

1. Leave the resistances unaltered.
2. Replace each capacitance C in the low-pass filter by a capacitance $C/2$, and put in parallel with it an inductance resonating with $C/2$ at f_0.
3. Replace each inductance L in the low-pass filter by an inductance $L/2$, and put in parallel with it a capacitance resonating with $L/2$ at f_0.

The effect is to replace the variable f in the low-pass filter by

$$\frac{1}{2}\left(f - \frac{f_0}{f}\right);$$

this leads to a bandpass filter geometrically symmetrical about f_0 with a bandwidth twice that of the low-pass filter. Figure 4·1, for example, is the bandpass analogue of Fig. 1·26.

[1] For amplifiers with small overshoot the value 0.7 is the one to take.

The variable $\frac{1}{2}[f - (f/f_0)]$ is approximately equal to $f - f_0$ for f close to f_0, and to the extent to which this approximation is valid, that is, to the extent to which the bandpass filter has small fractional bandwidth, a bandpass analogue can be considered as a shift upward in frequency of its low-pass prototype.

Every low-pass filter has its bandpass analogue, but not every bandpass filter has its low-pass analogue. A very common filter that has no exact low-pass analogue is the double-tuned circuit.

If a given bandpass filter is recognized to have a low-pass analogue and if it has small fractional bandwidth, then it is clear that the low-pass analogue is the filter to use in pulse response calculations [see (a) and (c) of Fig. 7·2]. But if no low-pass analogue exists, there may still exist a low-pass circuit whose mathematical expression is closely equivalent to that of the bandpass filter when $f - f_0$ is replaced by f. This is the case in (b) of Fig. 7·2.

FIG. 7·2.—Low-pass-bandpass equivalents: (a) single-tuned circuit, (b) double-tuned circuit, (c) triple-tuned circuit.

Pulse Response and Absolute Value Curve Alone.—For minimum phase-shift networks the absolute value-vs.-frequency characteristic completely determines the phase-vs.-frequency characteristic, and the two characteristics determine the pulse response. All the circuits described in Chaps. 4 to 6 are of the minimum phase-shift type. This fact leads to the important conclusion that if an amplifier has a certain absolute value curve, it has a specific pulse response, without regard to whether it uses stagger-tuned circuits, double-tuned circuits, stagger-damped circuits, combinations of double-tuned circuits and single-tuned circuits, or plate-to-grid inverse feedback, etc.

7·2. One-pole Networks.—As has already been asserted, a single-tuned circuit resonant at frequency f_0 and of small fractional bandwidth is closely approximated by a parallel RC-circuit shifted up in frequency to f_0; hence the envelope of the response to a step-function-modulated sine wave of frequency f_0 is given by the step-function response of a parallel RC-network. The latter response is an exponential curve, of the form shown in Curve 1 of Fig. 1·25.

Similarly the step-function response of n cascaded identical single-tuned circuits is shown in Fig. 1·25 for $n = 1, \cdots, 10$. The time scale $t/(RC)$ in Fig. 1·25 pertains to the low-pass case, however. The correct time scale for the bandpass case is $t/(2RC)$, reflecting the fact that a given

bandwidth in a bandpass amplifier yields only half as much speed as does the same bandwidth in a low-pass amplifier. This assumes that the carrier frequency in the bandpass case is equal to the midband frequency (double-sideband operation).

The correlation between over-all speed and over-all bandwidth in an amplifier made up of cascaded identical single-tuned circuits is extraordinarily good. If τ denotes rise time between 10 and 90 per cent and $\mathcal{B}$ denotes 3-db over-all bandwidth, then

$$\tau\mathcal{B} = 0.70, \tag{1}$$

whether 1, 2, . . . or infinitely many stages are involved.

7·3. Two-pole Networks.—This case includes staggered pairs, double-tuned circuits, and inverse-feedback pairs. All these involve two single-tuned circuits or, equivalently, a pair of complex-conjugate poles in the complex frequency plane.

A Single Two-pole Network.—The fundamental complex form of the networks above can be normalized, in the case of small fractional bandwidth, as

$$\frac{1}{1 + j(x + a)} \frac{1}{1 + j(x - a)}, \tag{2}$$

where x is proportional to frequency off resonance and a is the ratio of the imaginary part of the poles to their real part. The symbol a has a definite physical significance:

1. For stagger-tuned circuits, a is the ratio of the separation of the individual resonant frequencies to the bandwidth of the two individual stages.
2. For double-tuned circuits of equal primary and secondary Q, $a = kQ$. One can also write $a = k/k_t$, where k_t is the coefficient of coupling at transitional coupling.
3. For double-tuned circuits loaded on one side only, $a = \sqrt{4k^2Q^2 - 1}$.
4. In general for a double-tuned circuit of shape index β (see Sec. 5·1),

$$a = \left(\frac{1 + \beta}{1 - \beta}\right)^{\frac{1}{2}}.$$

The value $a = 1$ corresponds to the maximally flat absolute value curve.

The step-function response of Expression (2) is easily found to be proportional to

$$1 - e^{-t}\left(\frac{\sin at}{a} + \cos at\right). \tag{3}$$

The maximum value of Expression (3) is $1 + e^{-\frac{\pi}{a}}$, that is, the fractional overshoot corresponding to Expression (2) is

$$e^{-\frac{\pi}{a}}. \tag{4}$$

Expression (4) is graphed in Fig. 7·3. Two points follow from Expression (4), of which the second is quite important.

1. No matter how small a is, there is always some overshoot. Hence a pair of staggered circuits, no matter how slightly staggered, or an equal-Q double-tuned circuit, no matter how loosely coupled, always displays some overshoot.[1]
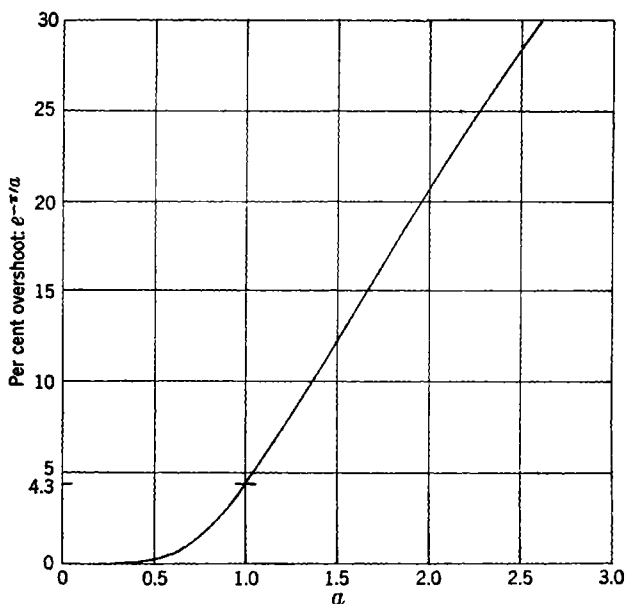
FIG. 7·3.—Overshoot of staggered pair as function of ratio a of separation of the resonant frequencies of the staggered circuits to their bandwidth. The value $a = 1$ corresponds to a flat staggered pair. The curve also applies to double-tuned circuits if a is suitably interpreted:

$a = kQ$ for an equal-Q double-tuned circuit.

$a = \sqrt{4k^2Q^2 - 1}$ for a double-tuned circuit loaded on one side only.

$a = [(1 + \beta)/(1 - \beta)]^{1/2}$ for a general double-tuned circuit having shape index β.

2. From the point of view of transient response there is nothing at all critical about the point $a = 1$, that is, the point representing the maximally flat bandpass. As a passes through unity, which represents critical coupling for an equal-Q double-tuned circuit, the fractional overshoot merely increases smoothly.

In Fig. 7·4 is graphed the per cent overshoot as a function of peak-to-dip ratio; the curve also applies to double-tuned circuits of arbitrary Q-ratio, and inverse feedback pairs.

[1] For $a < 0.5$, the overshoot is negligible, however.
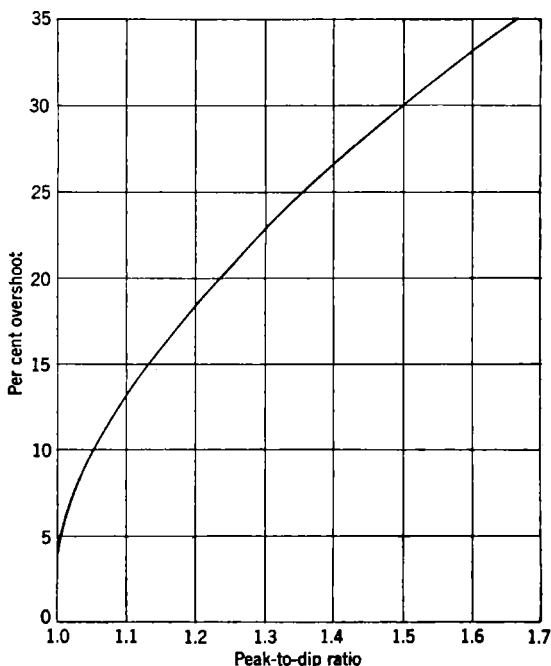
Fig. 7·4.—Overshoot of overstaggered pair, as function of ratio of maximum response to response at midband.   The curve also applies to double-tuned circuits of arbitrary Q-ratio, and inverse-feedback pairs.
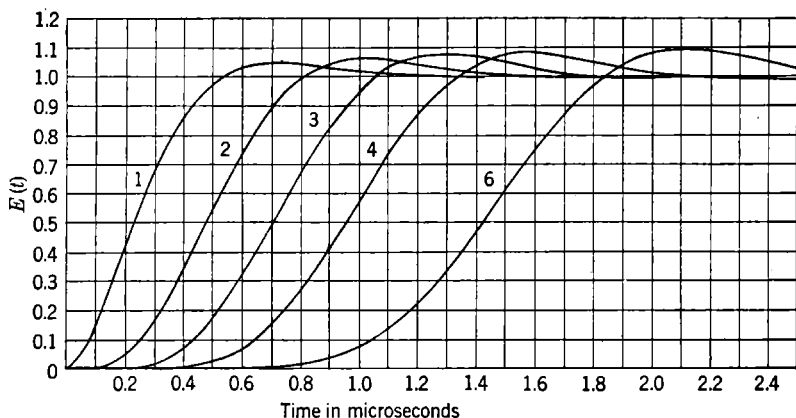
Fig. 7·5.—Step-function response of cascade of *n* flat-staggered pairs, *n* = 1, 2, 3, 4, 6, each pair of bandwidth 2 Mc/sec between 3-db points.   These curves also apply to transitionally coupled double-tuned circuits, flat inverse-feedback pairs, and in general to all minimum-phase circuits having absolute value vs. frequency curves of the form $1/\sqrt{1 + x^4}$.

Cascaded Maximally Flat Two-pole Networks.—In Fig. 7·5 are shown graphs of the step-function responses of 1, 2, 3, 4, and 6 maximally flat two-pole networks, corresponding to $a = 1$ in Expression (2), each individual network having a 2 Mc/sec bandwidth between 3-db points. (The over-all bandwidths are, of course, less than 2 Mc/sec.)

Figs. 7·5, 7·6, and 7·7 can be used for arbitrary bandwidths by remembering that bandwidth and rise-time vary inversely.
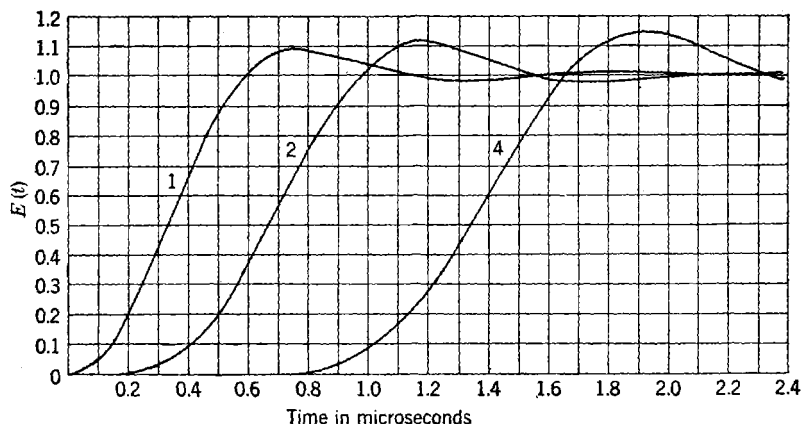
FIG. 7·6.—Step-function response of a cascade of n flat staggered triples, $n = 1, 2, 4$, each triple of bandwidth 2 Mc/sec between 3-db points.   These curves also apply to transitionally coupled triple-tuned circuits, flat inverse-feedback triples, flat combinations of an overcoupled double-tuned circuit and a single-tuned circuit, and in general to all minimum-phase circuits having absolute value vs. frequency curves of the form $1/\sqrt{(1 + x^6)}$.

Some pertinent results are listed in Table 7·1; in the column marked τℬ, the τ refers, as usual, to rise time between 10 and 90 per cent and ℬ to over-all 3-db bandwidth.

TABLE 7·1.—STEP-FUNCTION RESPONSE.   CASCADED MAXIMALLY FLAT TWO-POLE NETWORKS

No. of networks	Overshoot, %	τℬ
1	4.30	0.69
2	6.25	0.72
3	7.70	0.75
4	8.40	0.76
6	10.00	0.79

The use of Table 7·1 is made clear by considering a six-stage amplifier having an over-all bandwidth of 2 Mc/sec.   If the amplifier is made up of flat staggered pairs or flat inverse-feedback pairs, the over-all overshoot is 7.7 per cent and the rise time is 0.37 μsec, because three maximally

flat two-pole networks are involved. If the amplifier consists of transitionally coupled double-tuned circuits, there are six maximally flat two-pole networks; hence the overshoot is 10.0 per cent, and the rise time is 0.39 μsec.

7·4. Maximally Flat Three-pole Networks.—The maximally flat three-pole network can be put in the normalized form

$$\frac{1}{(1 + jx)[1 + jx + (jx)^2]}. \tag{5}$$

In Fig. 7·6 is shown the step-function response of 1, 2, 4 maximally flat three-pole networks, each network having a 2 Mc/sec bandwidth between 3-db points. Data are listed in Table 7·2 on the overshoot and the rise time τ for a given over-all bandwidth $\mathfrak{B}$.

TABLE 7·2.—STEP-FUNCTION RESPONSE OF CASCADED MAXIMALLY FLAT THREE-POLE NETWORKS

No. of networks	Overshoot, %	$\tau\mathfrak{B}$
1	8.15	0.73
2	11.2	0.80
4	14.2	0.87

7·5. Maximally Flat n-pole Networks.—Equation (4·15) gives the complex expression whose absolute value is the maximally flat function $1/\sqrt{1 + x^{2n}}$; the n poles of Eq. (4·15) are equally spaced on the perimeter of the left half of the unit circle, and the step-function response follows from the knowledge of the poles.

As an example consider the case $n = 7$. From Eq. (4·15)

$$\frac{1}{\sqrt{1 + \omega^{14}}} \tag{6}$$

is the absolute value of the minimum phase-shift complex expression

$$\frac{1}{(p - e^{\frac{j4\pi}{7}})(p - e^{-\frac{j4\pi}{7}})(p - e^{\frac{j5\pi}{7}})(p - e^{-\frac{j5\pi}{7}})(p - e^{\frac{j6\pi}{7}})(p - e^{-\frac{j6\pi}{7}})(p + 1)}, \tag{7}$$

$$= \frac{1}{[0.22252 + j(\omega \pm 0.97493)][0.62348 + j(\omega \pm 0.78183)][0.90097}{\qquad\qquad + j(\omega \pm 0.43388)](1 + j\omega),}$$

where $p = j\omega$. The phase lag of Expression (7) is

$$\tan^{-1}\frac{0.44504\omega}{1 - \omega^2} + \tan^{-1}\frac{1.2470\omega}{1 - \omega^2} + \tan^{-1}\frac{1.8019\omega}{1 - \omega^2} + \tan^{-1}\omega. \tag{8}$$

The partial-fraction expansion of Expression (7) is

$$\frac{0.73708p}{(p - e^{\frac{j4\pi}{7}})(p - e^{-\frac{j4\pi}{7}})} + \frac{-2.0649p - 3.3124}{(p - e^{\frac{j5\pi}{7}})(p - e^{-\frac{j5\pi}{7}})}$$
$$+ \frac{-2.9838p}{(p - e^{\frac{j6\pi}{7}})(p - e^{-\frac{j6\pi}{7}})} + \frac{4.3118}{p + 1}. \quad (9)$$
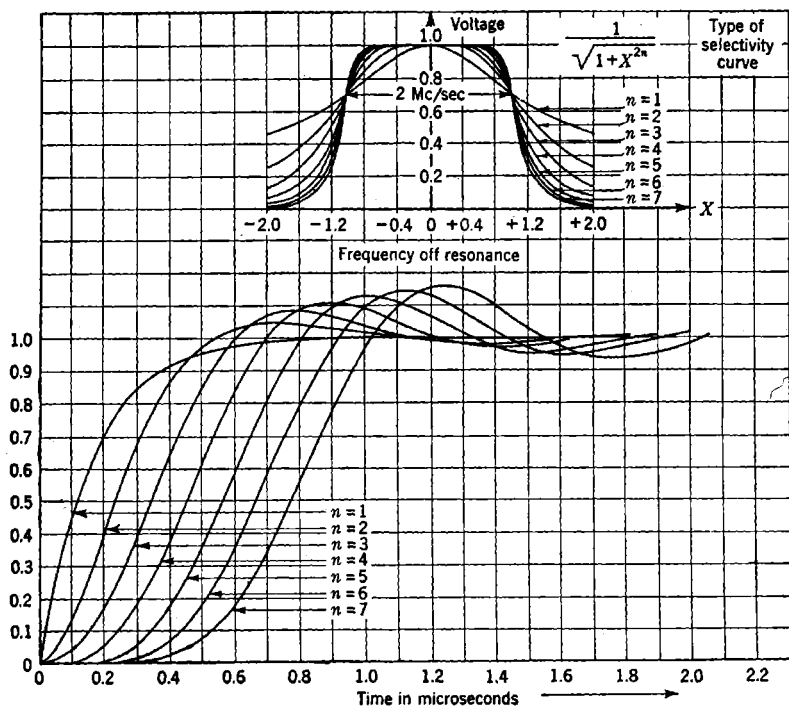
FIG. 7·7.—Step-function response of the (minimum phase shift) networks having the absolute value curves shown above.

The impulse response corresponding to Expression (7) is

$$\left.\begin{array}{l} e^{-0.22252t}(0.73708 \cos 0.97493t - 0.16823 \sin 0.97493t) \\ + e^{-0.62348t}(-2.0649 \cos 0.78183t - 2.9501 \sin 0.78183t) \\ + e^{-0.90097t}(-2.9838 \cos 0.43388t + 6.1964 \sin 0.43388t) \\ + 4.3118e^{-t}, \end{array}\right\} \quad (10)$$

and the step-function response is

$$\left.\begin{array}{l} 1 + 0.75602e^{-0.22252t} \sin 0.97493t \\ + e^{-0.62348t}(3.3124 \cos 0.78183t + 0.00048 \sin 0.78183t) \\ - 6.8774e^{-0.90097t} \sin 0.43388t - 4.3118e^{-t}. \end{array}\right\} \quad (11)$$

Expression (11) is shown as Curve 7 of Fig. 7·7, after its time scale has been changed to correspond to a bandwidth for the maximally flat seven-pole network of 2 Mc/sec.

Figure 7·7 displays the step-function response of maximally flat *n*-pole networks of 2 Mc/sec bandwidth for $n = 1, \cdots, 7$; Table 7·3 lists overshoot and rise time.

TABLE 7·3.—STEP-FUNCTION RESPONSE OF MAXIMALLY FLAT *n*-POLE NETWORKS

n	Overshoot, %	$\tau \mathbb{G}$
1	0	0.70
2	4.3	0.69
3	8.15	0.73
4	10.9	0.78
5	12.8	0.82
6	14.3	0.85
7	15.4	0.89

7·6. Overstaggered Circuits.—In Sec. 4·6 it was pointed out that the only advantages of the maximally flat absolute value curve $1/\sqrt{1 + x^{2n}}$
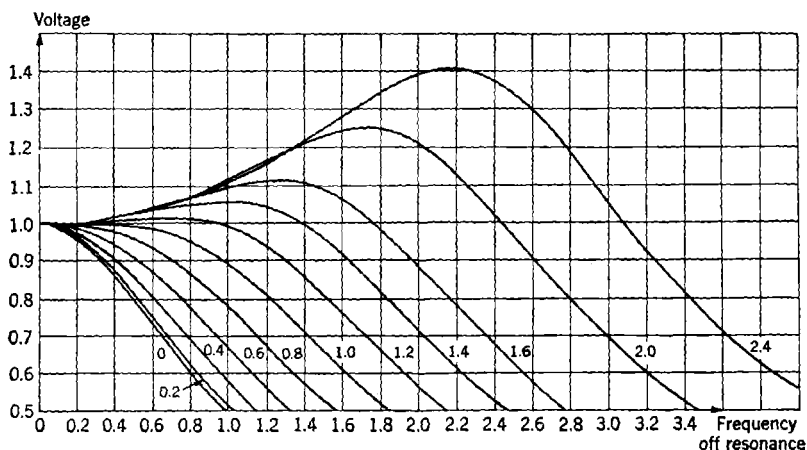
FIG. 7·8.—Overstaggered pairs. The parameter is the ratio a of separation of the resonant frequencies of the staggered circuits to their bandwidth. The value $a = 1$ corresponds to a flat-staggered pair.

are its mathematical simplicity and its easily recognized shape. It was also stated that substantial increases in gain-bandwidth factor could be attained if slight dips in the pass band were allowed; this point will now be demonstrated.

Overstaggered Pairs.—In Fig. 7·8 is shown the right half of the selectivity curves[1] derived from two staggered single-tuned stages of bandwidth 2, the parameter attached to the curves being the variable a of Expression (2), that is, the ratio of separation of the resonant frequencies of the staggered circuits to their bandwidth. The parameter value $a = 0$ corresponds to synchronous stages, $a = 1$ to a flat staggered pair and values of a greater than 1 to overstaggered pairs. Table 7·4 shows the gain factor for these staggered pairs, that is, the mean stage gain at midband, in units of $g_m/2\pi C$. For $a = 1$, for example, which corresponds to a flat-staggered pair of over-all bandwidth $2\sqrt{2}$, the stage gain is $1/(2\sqrt{2})$; this is in accord with the stage gain times over-all bandwidth factor of 1 for a flat-staggered pair. Also listed in Table 7·4 are the per cent overshoot [see expression (4)], the values of the shape index β (see Sec. 5·2) corresponding to each value of a, and the relative gain at band center of a double-tuned circuit of the same shape index and bandwidth as the corresponding staggered pair. It will be seen that a double-tuned circuit loaded on one side only always has an advantage of exactly 2 over the staggered pair whereas for an equal-Q double-tuned circuit the relative advantage lies with the staggered pair for undercoupled circuits, that is, circuits of small overshoot.

TABLE 7·4.—MIDBAND GAIN FACTORS CORRESPONDING TO CURVES OF FIG. 7·8, IN UNITS OF $g_m/2\pi C$

a	Overshoot, %	$\beta = \dfrac{a^2 - 1}{a^2 + 1}$	Stage gain factor for staggered pair	Gain factor for equal-Q double-tuned circuit	Gain factor for double-tuned circuit loaded on one side only
0	0	-1.000	0.500	0	1.000
0.2	0.00	-0.920	0.490	0.192	0.980
0.4	0.04	-0.725	0.464	0.344	0.928
0.6	0.53	-0.470	0.429	0.441	0.858
0.8	2.0	-0.219	0.390	0.487	0.780
1.0	4.3	0	0.353	0.500	0.707
1.2	7.3	$+0.180$	0.320	0.490	0.640
1.4	10.6	$+0.324$	0.290	0.473	0.580
1.6	14.0	$+0.438$	0.265	0.450	0.530
2.0	20.8	$+0.600$	0.223	0.400	0.446
2.4	27.0	$+0.705$	0.192	0.355	0.384

[1] The transition from a flat-staggered *n*-uple, $n = 2, 3, 4, \cdots$, to an overstaggered *n*-uple *of the same over-all bandwidth* turns out to be very simple: The individual resonant frequencies are left unaltered, and the individual bandwidths are all narrowed in the same ratio. This rule is not quite exact but is a close approximation even in the large fractional-bandwidth case, provided the dips are small.

Overstaggered Triples.—In Fig. 7·9 is shown the right half of the selectivity curves yielded by a three-stage amplifier consisting of a centered single-tuned stage of bandwidth 4 and two symmetrically staggered single-tuned stages of bandwidth 2, the parameter being the ratio a of peak separation of the staggered stages to their bandwidth. The value $a = \sqrt{3}$ corresponds to a flat-staggered triple, and values of a greater than $\sqrt{3}$ to overstaggered triples.

The step-function response for such overstaggered triples is proportional to

$$1 - \left(e^{-2t} + \frac{2}{a} e^{-t} \sin at\right). \tag{12}$$

Table 7·5 shows the gain factor for these staggered triples, that is, the mean stage gain at midband, in units of $g_m/2\pi C$. For $a = \sqrt{3}$, for
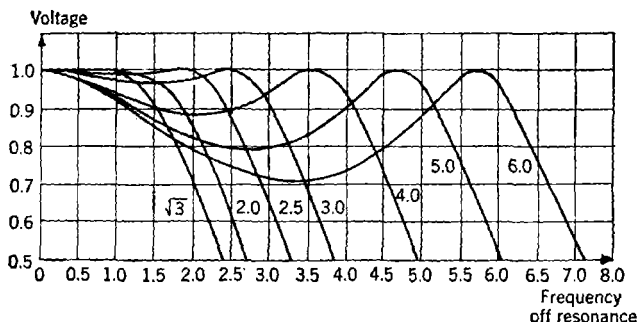
FIG. 7·9.—Overstaggered triples. The parameter is the ratio a of separation of the resonant frequencies of the staggered circuits to their bandwidth. The value $a = \sqrt{3}$ corresponds to a flat triple.

example, corresponding to a flat-staggered triple of bandwidth 4, the gain factor is 0.25, in accord with the gain-bandwidth factor of 1 for a flat-staggered triple. Also listed are the mean gains at midband for the two-stage combinations consisting of a centered single-tuned stage and an overcoupled double-tuned stage that have the same shape and bandwidth as the corresponding overstaggered triple.

Consider the curve corresponding to $a = 2.5$ in Fig. 7·8, for example; the curve shows 1 per cent or 0.1-db dips. The bandwidth between the 84 per cent or 1.5-db points is $5.2 (= 2 \times 2.6)$. Two such overstaggered triples have 0.2-db over-all dips and a 3-db bandwidth of 5.2. Since the mean stage gain factor is 0.205 (see $a = 2.5$ in Table 7·5), the gain-bandwidth factor for the overstaggered triples is 1.07. Two flat-staggered triples, on the other hand, have a gain-bandwidth factor of only 0.86 (see Table 4·6). The two overstaggered triples thus have a

TABLE 7·5.—MIDBAND GAIN FACTORS CORRESPONDING TO CURVES OF FIG. 7·9, IN UNITS OF $g_m/2\pi C$

a	Stage gain factor for staggered triple	Stage gain factor for combination of a centered single-tuned stage and an equal-Q double-tuned stage	Stage gain factor for combination of a centered single-tuned stage and a double-tuned stage loaded on one side only
$\sqrt{3}$	0.250	0.329	0.353
2.0	0.232	0.316	0.334
2.5	0.205	0.296	0.307
3.0	0.184	0.274	0.281
4.0	0.154	0.243	0.248
5.0	0.134	0.220	0.222
6.0	0.119	0.203	0.204

superiority in gain-bandwidth factor of 25 per cent, at the expense of only 0.2-db over-all dips.

When it is considered that the 0.2-db dips are those in the over-all response of an amplifier of six stages, it is clear that the increase of 25 per cent in gain-bandwidth factor is a very large reward for forsaking the maximal flatness of pass band. The increase in gain-bandwidth factor may be employed in several ways: to increase the gain by 25 per cent per stage, or 11.5 db over-all, for a fixed bandwidth; to increase the over-all bandwidth by 25 per cent for a fixed over-all gain; or to leave gain and bandwidth fixed and substantially increase tube life by operating the tubes with 20 per cent reduced transconductance.

The overstaggered triple case just considered is interesting also in its transient-response aspects. The overshoot corresponding to a single overstaggered triple having the constants above ($a = 2.5$) turns out to be 10.2 per cent, and the overshoot for two such triples is about 14.5 per cent; this is very closely the same as the overshoot for a flat-staggered sextuple (see Table 7·3). Moreover, for a given gain the combination of two such overstaggered triples is about 7 per cent faster than the flat sextuple. Hence even from the point of view of pulse response it would be better to build a six-stage amplifier in the form of two slightly over-staggered triples than one flat sextuple. This furnishes another illustration of the lack of inherent advantages in maximally flat pass bands.

GENERAL CONSIDERATIONS

7·7. Gain-bandwidth Factor.—Table 7·6 lists the gain-bandwidth factors for a number of coupling schemes. In each case the gain is the mean stage gain, the bandwidth is the over-all 3-db bandwidth, and the gain-bandwidth factor is the ratio of the gain-bandwidth product to

$g_m/(2\pi 2\sqrt{C_1C_2})$, where C_1 and C_2 are the output and input capacities respectively.[1]

TABLE 7·6.—GAIN-BANDWIDTH FACTOR OF VARIOUS INTERSTAGE COUPLING SCHEMES

Interstage coupling scheme	Normalized selectivity curve	$\mathfrak{GB}$ factor
1. m synchronous single-tuned circuits.......	$\dfrac{1}{(1+x^2)^{m/2}}$	$(2^{1/m}-1)^{1/2}$
2. m flat-staggered n-uples................	$\dfrac{1}{(1+x^{2n})^{m/2}}$	$(2^{1/m}-1)^{1/2n}$
3. m transitionally coupled equal-Q doubled-tuned circuits......................	$\dfrac{1}{(1+x^4)^{m/2}}$	$2^{1/2}(2^{1/m}-1)^{1/4}$
4. m transitionally coupled double-tuned circuits loaded on one side only	$\dfrac{1}{(1+x^4)^{m/2}}$	$2(2^{1/m}-1)^{1/4}$
5. m flat stagger-damped n-uples made up of equal-Q double-tuned circuits.........	$\dfrac{1}{(1+x^{4n})^{m/2}}$	$2^{1/2n}(2^{1/m}-1)^{1/4n}$
6. m flat stagger-damped n-uples made up of double-tuned circuits loaded on one side only...................................	$\dfrac{1}{(1+x^{4n})^{m/2}}$	$2(2^{1/m}-1)^{1/4n}$
7. m flat pairs consisting of a single-tuned circuit and an overcoupled equal-Q double-tuned circuit	$\dfrac{1}{(1+x^6)^{m/2}}$	$3^{1/4}(2^{1/m}-1)^{1/6}$
8. m flat pairs consisting of a single-tuned circuit and an overcoupled double-tuned circuit loaded on one side only........	$\dfrac{1}{(1+x^6)^{m/2}}$	$2^{1/2}(2^{1/m}-1)^{1/6}$
9. m transitionally coupled triple-tuned circuits................................	$\dfrac{1}{(1+x^6)^{m/2}}$	$\alpha\,(2^{1/m}-1)^{1/6}$, where α varies between 1 and 2, depending on the ratio C_1/C_2; $\alpha=1$ for $C_1/C_2=1$, $\alpha=2$ for $C_1/C_2=3$.
10. Plate-grid resistive feedback circuits with simple single-tuned termination are exactly equivalent to stagger tuning except that the effective tube transconductance is slightly reduced (by the shunting affect of the feedback conductance)...........	....	....

[1] For single-tuned circuits the use of $2\sqrt{C_1C_2}$ instead of $C=C_1+C_2$ implies that the tuning coil is employed as an autotransformer tuned with capacity $2C_1$ and tapped down in the impedance ratio C_2/C_1. The improvement in gain-bandwidth factor yielded by assuming the capacity $2\sqrt{C_1C_2}$ instead of C_1+C_2 is almost always negligible; for $C_2=2C_1$, for example, the improvement is 6 per cent.

Table 7·7.—Over-all 3-db Bandwidths Obtainable with Tubes Having a $g_m/(2\pi2\sqrt{C_1C_2})$ Ratio of 62 Mc/Sec.

Type of amplifier	Over-shoot, %	80 db over-all gain. Bandwidth, Mc/sec	100 db over-all gain. Bandwidth, Mc/sec
Three stages:			
Three synchronous single-tuned..........	0	1.5	0.65
Flat-staggered triple....................	8.15	2.9	1.3
Overstaggered triple with 0.1-db dips.....	10.2	3.5	1.55
Three transitionally coupled equal-Q double-tuned...........................	7.7	2.9	1.3
Three transitionally coupled one-side-loaded double-tuned........................	7.7	4.1	1.9
Flat stagger-damped triple made up of one-side-loaded double-tuned circuits.......	14.3	5.8	2.6
Four stages:			
Four synchronous single-tuned............	0	2.7	1.5
Two flat-staggered pairs.................	6.25	5.0	2.8
Flat-staggered quadruples...............	10.9	6.2	3.5
Four transitionally coupled equal-Q double-tuned...............................	8.4	5.8	3.3
Four transitionally coupled one-side-loaded double-tuned........................	8.4	8.1	4.7
Six stages:			
Six synchronous single-tuned.............	0	4.7	3.2
Three flat-staggered pairs...............	7.7	9.5	6.5
Two flat-staggered triples...............	11.2	11.5	7.8
Two overstaggered triples with 0.1-db dips per triple...........................	14.5	14.3	9.8
Flat-staggered sextuple..................	14.3	13.4	9.2
Six transitionally coupled equal-Q double-tuned...............................	10.0.	11.2	7.7
Six transitionally coupled one-side-loaded double-tuned........................	10.0	15.8	10.8
Three flat stagger-damped pairs made up of one-side-loaded double-tuned circuits	∼ 18	20.6	15.6
Nine stages:			
Nine synchronous single-tuned...........	0	6.2	4.8
Three flat-staggered triples..............	∼ 13	17.8	13.8
Three overstaggered triples with 0.1-db dips per triple...........................	∼ 17	23.2	18.0
Nine transitionally coupled equal-Q double-tuned...............................	∼ 12	16.7	13.0
Nine transitionally coupled one-side-loaded double-tuned........................	∼ 12	23.6	18.4
Three flat stagger-damped triples made up of one-side-loaded double-tuned circuits	∼ 25	39.7	30.9

Table 7·7 is derived from Table 7·6 and shows the over-all 3-db bandwidths that can be achieved at over-all gains of 80 and 100 db with various numbers of amplifier stages. The value assumed for $g_m/(2\pi2 \sqrt{C_1 C_2})$ is 62 Mc/sec. This is appropriate to the type 6AK5 tube with capacities $C_1 = 4.4 \ \mu\mu f$ and $C_2 = 6.7 \ \mu\mu f$, and conservatively rated to have a transconductance of 4200 μmhos.

For other tube types, capacities, and transconductances, each of the bandwidths listed in Table 7·7 should be multiplied by the ratio of the other value of $g_m/(2\pi2 \sqrt{C_1 C_2})$ to 62 Mc/sec. For example, for type 6AC7's assumed to have $g_m = 7800 \ \mu$mhos, $C_1 = 8 \ \mu\mu f$, $C_2 = 17 \ \mu\mu f$, the bandwidths in Table 7·7 are to be multiplied by 0.85 ($= \frac{53}{62}$).

Also listed in Table 7·7 is the step-function overshoot accompanying the various coupling schemes.

7·8. Gain Control.—Bandpass amplifiers are usually gain controlled by varying the transconductance of one or more of the tubes by adjustment of suitable electrode potentials. There are two common methods of reducing gain: the first reduces the plate-cathode or screen-cathode potential, and the second increases the cathode-grid bias.

The first method, reducing plate potential, has the advantage of producing less detuning for a given gain reduction, that is, less change in input capacity, and is sometimes employed for this purpose when examining amplifiers in the laboratory for regeneration (see Sec. 8·5). It has the very serious disadvantage, however, that the saturation level of the amplifier is reduced when gain is lowered and hence should never be used when overload capability is of importance.

The second method, increasing the cathode-grid bias, is almost universal. The usual procedure is to return the grid circuits of the gain-controlled tubes to a variable negative supply, although sometimes the cathodes are connected to a variable positive supply.

For quick recovery from overload it is important that the d-c resistance in the grid circuit be small. Hence the d-c resistance in the supply for the gain-control voltage should not exceed about 2000 ohms.

Among tubes of a given type there is, unfortunately, only poor correlation between transconductance and bias voltage for low values of transconductance. For example, among a set of about 200 type 6AK5 tubes a 6-volt bias yielded a 50-db gain reduction with one tube and only 18 db with another. The correlation between transconductance and plate current is very good, however, and it has been suggested that a scheme for accurate gain control should be based on regulating current rather than voltage.

A limit on gain reduction is established by parasitic signal paths through the grid-plate capacitance and decoupling circuits. The maximum possible gain reduction depends on the operating frequency, the

impedance level of the interstage networks, the effectiveness of decoupling, etc., but rarely exceeds about 40 or 50 db per stage.

7·9. Gain Variability.—Sections 7·9, to 7·11 are concerned with aspects of bandpass amplifier design that are important in large production but have only minor significance for laboratory amplifiers.

The problem of gain variability in amplifiers is serious because of the wide variability in transconductance permitted by the JAN-1A specifications. Thus for the type 6AK5 a g_m-variability is permitted between 3500 μmhos and 6500 μmhos under fixed-bias conditions. This is a range of 1.86 to 1, or 5.4 db. For a synchronous single-tuned amplifier, for example, there would therefore be a 5.4-db range per stage between a low-limit tube and a high-limit tube; it is clear that in six-stage amplifiers an over-all gain variability of 32 db between one amplifier and another might occur.

The situation is mitigated by the stabilizing effect of cathode bias. Experience has shown that there is good correlation between transconductance and quiescent plate current; hence if a cathode-bias resistor R_k is employed, the variability in g_m is reduced almost by the factor $1 + g_m R_k$. For a type 6AK5 with $R_k = 200$ ohms, the variability is almost cut in half, to the range between 4000 and 6000 μmhos, say.

For a type 6AC7 tube the transconductance is permitted by the JAN-1A specifications to vary between 7000 and 12,500 μmhos; this is, however, with the stabilizing effect of a 160-ohm cathode-bias resistor already included in the JAN-1A test specifications. The fractional g_m variability is thus seen to be considerably greater for the type 6AC7 than for the type 6AK5.

Returning to the type 6AK5, one may ask what accommodations in design are to be made for the 4000- to 6000-μmho variability in g_m. It is clear that the matter is one of probabilities and that the designer must try to answer this question: "If the design proceeds on the assumption of a certain nominal g_m, say 4200 μmhos, what fraction of the amplifiers will turn out to be deficient in gain?" The answer must then be weighed against the extra complication, in number of stages, components, manhours, etc., of a design based on an assumed g_m of 4,000 μmhos, say.

There are two errors to avoid in this analysis of probabilities: The first is the assumption that the g_m's of the various tubes in an amplifier have independent probability distributions. Unfortunately, there is a tendency for tube characteristics to run in bunches, all the tubes in a given production-run being very closely alike. Hence it can happen that all the sockets of an amplifier are filled with tubes from a batch of relatively low-g_m tubes.

The second error to be avoided, consists in the assumption that all g_m values within the allowed range are equally probable. This would

lead to the pessimistic conclusion that the design g_m must be the lowest possible value. It is, however, very rare that a production run of tubes has an average g_m equal to the low limit.

In the absence of full statistical data there is a somewhat arbitrary decision involving the value to assume for g_m, but it seems fairly conservative to assume in designing amplifiers to be produced in large production that the type 6AK5 has a g_m of about 4200 μmhos. A statistical examination made at the Naval Research Laboratory showed that considerably less than 5 per cent of type 6AK5 tubes, with 200-ohm cathode bias resistors, had g_m's less than 4200 μmhos.

The question remains of what to do about the tubes that have larger g_m, presumably up to about 6000 μmhos even with self-bias; the ratio 6000/4200 is about 1.4 or 3 db. The answer is that provision must be made for additional gain control in the amount of 3 db per stage over and beyond the gain-control range required by the operating requirements of the amplifier. (This extra gain can be controlled, if desired, by a screw-driver adjustment, since it varies only with tube replacement and tube aging.) For example, a certain 12-stage wide-band amplifier of 100-db gain was intended for use as the i-f amplifier of a high-precision inshore-navigation radar system. A gain control of 100 db was needed for the operating radar system; to this had to be added 36 db for tube g_m variability, so that the over-all gain-control range was 136 db in what was nominally a 100-db amplifier. The 136-db gain-control range was achieved by applying gain control to each of stages two to seven.

Plate-to-grid resistive feedback (Chap. 6) reduces the dependence of amplifier gain on tube transconductance by a factor of not quite 2. If larger amounts of gain stabilization are required, other methods must be employed (see Sec. 6·5).

7·10. Capacity Variability.[1]—The JAN-1A specifications permit ± 0.5 $\mu\mu$f variability in the input capacity of type 6AK5 tubes and ± 0.4 $\mu\mu$f variability in output capacity; the total for a single-tuned interstage circuit is ± 0.9 $\mu\mu$f. If the nominal interstage capacity is 10.8 $\mu\mu$f, there is a fractional variability[2] of $\pm \frac{1}{12}$. Hence if random tubes are inserted in an already tuned amplifier, the individual interstage circuits may exhibit fractional frequency variability as large as $\pm \frac{1}{24}$; this is ± 1.25 Mc/sec at 30 Mc/sec and ± 2.5 Mc/sec at 60 Mc/sec. Subject to certain modifying conditions mentioned below, account must be taken of the effect of such detuning.

[1] The relative importance of load resistor variability is usually secondary to that of capacity variability if 5 per cent tolerance resistors are used, except for amplifiers of very large fractional bandwidth, which are noncritical in any case.

[2] For type 6AC7 tubes the JAN-1A capacity variability (output plus input) is ± 3.7 $\mu\mu$f. With a 25 $\mu\mu$f interstage capacity this is a fractional variability of about $\pm \frac{1}{7}$, almost double that for the type 6AK5.

The situation is much like that described in Sec. 7·9 on gain variability, with the important difference that this time the tendency for tube characteristics to run in bunches represents an advantage rather than a disadvantage. The reason is that a set of tubes having uniformly low capacitances produces the result, to the first order, of a uniform shift upward in frequency, without distortion of the pass band. Because the precise frequency of a bandpass amplifier is usually much less important than the shape and width of its pass band, there is little harm done by the frequency shift.

Examination at Naval Research Laboratory of a large sample of type 6AK5 tubes showed that more than 90 per cent of the tubes fell within the inner half of the capacity range permitted by the JAN-1A specifications; that is, tolerances of ± 0.45 $\mu\mu$f in the sum of input and output 6AK5 capacities included 90 per cent of the tubes.

To the tube-capacity variability must be added a certain wiring-capacity variability caused by variation in physical placement of coils, resistors, etc. This variability should be small in a well-designed layout in which the various components are rigidly mounted in prescribed positions.

The next problem is to establish a connection between capacity variability and the performance of amplifiers. For synchronous single-tuned amplifiers the problem is simple and the results are reassuring. Consider, for example, a synchronous single-tuned amplifier at 60 Mc/sec employing type 6AK5 tubes and having a 2 Mc/sec over-all bandwidth. The inner half of the JAN-1A capacity variability leads to a fractional capacity variability of about $\pm \frac{1}{24}$, which corresponds at 60 Mc/sec to a detuning of $\pm \frac{60}{48} = \pm 1.25$ Mc/sec. The individual stage bandwidths are 6 Mc/sec for an over-all bandwidth of 2 Mc/sec. In the least favorable case, in which three of the tubes have the lowest capacities and the other three have the highest capacities, the resonance peaks of each pair are 2.5 Mc/sec apart, for individual 6 Mc/sec bandwidths. The reduction in stage gain turns out to be 0.7 db; a certain additional variability in gain has thus been introduced, but this variability is much like that due to g_m variations and, as a matter of fact, is of smaller extent. The over-all 3-db bandwidth in this "slightly staggered" case is 3.5 Mc/sec; the overshoot is completely negligible, about 0.05 per cent per pair; and usually the increase in bandwidth is not objectionable.

For the more complicated amplifier coupling schemes such as stagger tuning, double tuning, and inverse feedback, the question of assessing amplifier performance in the presence of capacity variability is much more difficult.

The problem can be split into two parts, to neither of which is it possible at this time to give anything like an adequate answer. The

first part is this: From the point of view of pulse response what constitutes adequate preservation of amplifier pass band? Much more work is needed to determine the effect of a tilt or asymmetry in pass band on response to a carrier-frequency pulse. Rough experimental investigations indicate that a tilt across the band of 2 db has extremely little effect on the pulse response, so that the requirements on tilt are apparently not at all critical. A peak on one side of midband does little harm unless the peak is large, 3 db or more; in that case if the carrier-frequency pulse is retuned to the peak frequency, the pulse response tends to have a speed appropriate to the new 3 db bandwidth, but without overshoot. A pass band with a peak on each side of midband seriously increases overshoot, however.

The second part is this: For a given statistical distribution of interstage capacities what is the statistical distribution of amplifier pass bands? The problem is made more difficult by the different effects of capacity variability on the different types of amplifiers. Thus the bandwidth of a double-tuned circuit or an inverse-feedback pair or triple can only be increased by detuning, and the gain can only be reduced; with stagger-tuning, however, the bandwidth can be either increased or decreased and the gain decreased or increased, depending on whether the detuning moves the side circuits farther from midband or closer to midband. Furthermore, with single-tuned circuits (stagger-tuned or inverse-feedback) a batch of tubes with normal output capacities and greater than normal input capacities essentially produces only a reduction in midband frequency, without distortion of the pass band; under the same circumstances the pass band of double-tuned circuits is altered in shape. Indeed, an increase in input capacity together with an equal decrease in output capacity produces no effect at all on single-tuned stages but leads to double-tuned circuits that are double humped.

Analyses have been made in the Radiation Laboratory of the effect on a flat-staggered pair or staggered triple of interstage capacity variability; in the staggered-triple case, for example, there are 27 combinations, corresponding to having each of the three stages tuned to nominal frequency, extreme low frequency, or extreme high frequency. Similar calculations have been made at the Bell Telephone Laboratories for equal-Q double-tuned circuits. Although these calculations are already quite laborious, they do not yield information on the statistical distribution to be expected among the amplifiers.

An experimental statistical investigation was made at the Naval Research Laboratory of three six-stage 6AK5 amplifiers at 60 Mc/sec, each of about 90-db gain and 12 Mc/sec bandwidth. Flat inverse-feedback triples were employed in one amplifier, flat-staggered triples in another, and transitionally coupled equal-Q double-tuned circuits in the

third.   Only one amplifier of each type was used, but 20 sets of six tubes were chosen at random from a stock of 120 6AK5 tubes of three different manufacturers and inserted in the sockets of the various amplifiers. The 60 Mc/sec gains and the 3-db bandwidths were noted.   The results were gain variabilities of 4, 13, and 11 db and bandwidth variabilities of 2, 3, and 3 Mc/sec in the inverse-feedback, stagger-tuned, and double-tuned cases respectively; the gain variability includes the effect of g_m variability as well as that of detuning.   These data indicate a superiority of the inverse-feedback triples over either staggered triples or equal-Q double-tuned circuits of about 3/1 in gain stability and 3/2 in bandwidth stability.   There are several comments that can be made about these results:

1. Because only one amplifier of each type was involved, no information was obtained on the effects of wiring-capacity variability; this variability is likely to be somewhat more serious in the inverse-feedback case, where the influence of grid-plate capacity is especially important (see Sec. 6·6).
2. No account was taken of the tendency of tube capacities to run closely alike in a given production run.   For this reason the results above are probably pessimistic.
3. The relatively poor showing of the equal-Q double-tuned circuits, namely, the conclusion that they are as critical as staggered triples, is hard to understand.   This is especially dubious because staggered pairs are quite surely less critical than staggered triples; the conclusion would then have to be that staggered pairs are less critical than equal-Q double-tuned circuits.
4. For work with short pulses, there remains the important question, already mentioned, of the influence on pulse response of the tilts and asymmetries observed in the various pass bands, and hence of the significance of 3-db bandwidth.

7·11. Pretuned Coils.   *Tuning Adjustments in the Field.*—It is the opinion of the author that a high-gain, high-frequency amplifier should be built in such a way that tuning the amplifier is impossible once it leaves the factory, even if synchronous single-tuned circuits are used; this applies whether the application is military or commercial, and a fortiori to circuits more complicated than single-tuned circuits.

The reasons are that

1. Adequate signal generators are usually not available.
2. Adequate skill is even less common.   There is a great need, as stressed in Sec. 8·5, to make all connections really coaxial if high gain is used; suitable connectors are rare in the field, and

even rarer is the realization of how essential it is to use them. A measurement made with clip leads on a high-gain amplifier is often completely meaningless.[1]

3. Interstage tuning adjustments that are accessible in the field are a great temptation to maintenance men repairing equipment after a breakdown and are often tampered with and thrown completely out of adjustment, even though the likelihood of misalignment was negligibly small.

For these reasons it is felt that field tuning adjustments generally do more harm than good.[2]

Amplifiers in Which the Probability of Tube Replacement Is Small.— Amplifiers can be divided into two types: those in which the probability of tube replacement is small and those in which the probability is large. The first type has as its extreme those amplifiers employing baseless soldered-in tubes, such as Example 3 of Sec. 5·7. It also includes, however, amplifiers of rather small life expectancy employing tubes with sockets.

For this type of amplifier it may be wise to adjust the coils in the factory to the particular tubes in use, and the possibility is presented of using quite complicated circuits, even though these are critical in adjustment. This method (see Example 3 in Sec. 5·7) almost certainly represents the future trend in high-performance amplifiers, of closely integrating the amplifier tubes and the amplifier networks (see p. 84).

Amplifiers in Which the Probability of Tube Replacement Is Large.— For such amplifiers, which make up a very large part of present designs, there are two subcases:

(α) coils tuned in the factory.
(β) pretuned coils.

If, as is being assumed here, the probability of tube replacement is high, the only legitimate purpose of (α) is to accomodate variability in wiring capacities; it is incorrect to use (α) to accomodate for tube-capacity variability as well, for the reason that an amplifier tuned with a set of tubes having capacities off in one direction would be doubly

---

[1] The author recalls an attempt to carry out a simple measurement of cathode-bias voltage in a high-gain 60-Mc/sec amplifier. Attaching clip leads caused the amplifier to regenerate and led to the false conclusion that the tube was drawing excessive current. The precaution of removing the next tube from its socket was not obvious to the man making the measurement, who, although an intelligent and well-trained electrical engineer, was not accustomed to high-gain amplifiers.

[2] The tunable amplifier of Example 3 of Sec. 4·11 was intended for use in well-equipped operating locations of semilaboratory type.

mistuned with a set of replacement tubes having capacities at the other extreme.[1]

But, as already pointed out, in a well-designed mechanical layout the component locations are accurately specified; hence the possibility of being able to accommodate wiring-capacity variability is not very important.

The advantages of (β), i.e., pretuned coils, are substantial:

1. Simpler and less expensive coils.
2. Elimination of the time consumed in tuning (but not checking) the amplifier.
3. Smaller coil size and hence reduced tendency to regeneration (because of the smaller magnetic field around the coils),
4. Reduced interstage capacity.[2]

The conclusion then is that tuning coils should be pretuned in amplifiers in which there is a fairly high probability of tube replacement.   The design of such amplifiers should assure adequate performance with random selection of tubes (see Sec. 7·10).

Most of the amplifiers built at the Radiation Laboratory employed pretuned coils, with a certain concession to wiring capacity variability that proved valuable in production.   A single stage, often the center-tuned stage of a staggered triple, was provided with a tuning adjustment. With a standard set of tubes inserted in the amplifier the tunable stage was adjusted to give a symmetrical pass band, after which the tuning adjustment screw was snipped off and a random set of tubes inserted. The single adjustment proved to be capable of accommodating the normal production variability in wiring capacities.

7·12. Comparison of Amplifier Types.   *Gain-bandwidth Product.*— From Table 7·7, for example the entry under six stages, it is seen that synchronous single-tuned amplifiers have substantially smaller gain-bandwidth products than do any of the other amplifier types.   Flat-staggered pairs have somewhat smaller gain-bandwidth products than do transitionally coupled equal-Q double-tuned circuits, and flat-staggered triples have slightly larger gain-bandwidth products.   Transitionally coupled one-side-loaded double-tuned circuits have an advantage of exactly $\sqrt{2}$ over transitionally coupled equal-Q circuits, and stagger-damped circuits, in turn, surpass the one-side-loaded circuits.

---

[1] Hence if coil tuning is done at all, it should be done with a set of "standard" tubes, that is, tubes having capacities approximately equal to the nominal JAN-1A values.

[2] In an amplifier employing type 6AK5 tubes substitution of pretuned coils for tunable coils reduced the interstage capacity by 10 per cent and hence increased the gain-bandwidth factor by 10 per cent.

Plate-grid resistive feedback amplifiers with single-tuned circuit terminations, i.e., the usual inverse-feedback bandpass amplifiers, are exactly equivalent in gain-bandwidth product to stagger-tuned amplifiers, except that the effective transconductance in the feedback amplifiers is slightly less than in stagger-tuned amplifiers.

In practice the gain-bandwidth advantage of transitionally coupled equal-Q double-tuned circuits over flat-staggered pairs is likely to be less than that indicated in Table 7·7, for the reason that interstage capacities are somewhat larger in the double-tuned case. Between 6AK5 tubes, experience has shown an interstage capacity about 10 per cent larger when the larger coils of double-tuned circuits are used than with single-tuned circuits.

Selectivity.—Selectivity has two aspects: squareness of pass band near midband and attenuation of frequencies far from midband. With regard to the first consideration, amplifiers employing synchronous single-tuned circuits are very poor, amplifiers employing flat-staggered pairs, flat feedback pairs, or transitionally coupled double-tuned circuits are better, staggered or feedback triples are still better, and staggered or feedback quadruples or staggered-damped pairs even better, and so forth. The larger the value of n in selectivity curves of the form

$$\frac{1}{\sqrt{1 + x^{2n}}}$$

the squarer is the top of the pass band, of course, but cascading stages having a given value of n produces negligible squaring-up of the relative pass band.

Remote frequency selectivity can be assessed by the ratio of 40- to 3-db bandwidth, for example. For six-stage amplifiers this ratio is 5.5 for synchronous single-tuned stages, 3.0 for flat pairs, 2.5 for flat triples, 2.35 for transitionally coupled double-tuned circuits, and 1.7 for stagger-damped pairs.

Overshoot.—Table 7·7 shows that synchronous single-tuned stages excel in having small overshoot; this is to be expected as a reflection of the general principle that the squarer the pass band the greater the overshoot.

An amplifier made up of transitionally coupled double-tuned circuits has greater overshoot than an amplifier of the same number of stages employing flat-staggered or inverse-feedback pairs. The reason is that the signal in the stagger-tuned or feedback amplifier encounters only half as many selectivity curves of the form $1/\sqrt{1 + x^4}$ as it does in the double-tuned amplifier.

Consider a two-stage amplifier. For a stage gain G the over-all

bandwidth of a flat-staggered pair is $\mathcal{B}$, where $\mathcal{G}\mathcal{B} = g_m/2\pi C$. The stage bandwidth of a transitionally coupled equal-Q double-tuned circuit of stage gain $\mathcal{G}$ is $\mathcal{B}\sqrt{2}$. Because there are two double-tuned circuits, the over-all bandwidth is less than $\mathcal{B}\sqrt{2}$ and turns out to be 1.12$\mathcal{B}$; that is, the transitionally coupled equal-Q double-tuned amplifier has a 12 per cent larger gain-bandwidth product than the flat-staggered pair. The superiority in gain/rise time ratio turns out, however, to be only 8 per cent. Moreover, the overshoot in the two-stage double-tuned amplifier is 6.25 per cent, as compared with 4.3 per cent for the staggered-pair amplifier. If, to put pulse comparisons on an equal basis, the staggered pair is overstaggered so as also to yield an overshoot of 6.25 per cent, it develops that the superiority in gain/rise time of the equal-Q double-tuned circuits over the staggered pair is only 1 per cent. Indeed if the requirements on an amplifier are that the over-all overshoot be held to less than about 4 per cent (depending on the number of stages), an equal-Q double-tuned amplifier is actually inferior to one using staggered pairs.[1] The conclusion is that when overshoot considerations are important, equal-Q double-tuned circuits should not be regarded as superior in gain rise time ratio to staggered pairs.[2]

Simplicity.—Undoubtedly the simplest amplifier is that employing synchronous single-tuned stages. The coils are simple and are identical in the various stages; tuning is easy; and there are no critical components.

Stagger-tuned and inverse-feedback amplifiers also require only simple tuning coils.

Inverse-feedback amplifiers have the disadvantage of requiring special feedback resistors of extremely small end-to-end capacities, such as the IRC type MPM (see Example 2 in Sec. 6·8). Once these resistors are available, however, inverse-feedback amplifiers have the advantage over stagger-tuned amplifiers that the coils are alike from stage to stage.

Double-tuned transformers are considerably more complicated in construction than single-tuned coils and require accurate control of the spacer thickness that determines mutual inductance, as well as control of both primary and secondary inductances. The greater complexity of the tuning coils constitutes the only important objection to the use of double-tuned circuits in large-production amplifiers. Other objections are the greater size of the coils and the resulting larger capacity. The common objection of greater difficulty of alignment is spurious, inasmuch as the coils ought to be fixed-tuned in any case.

---

[1] This is apart from the consideration mentioned above that interstage capacity is somewhat higher for double-tuned stages than it is for single-tuned stages.

[2] Double-tuned circuits loaded on one side only are, however, superior in gain rise time ratio to staggered pairs, even for the same overshoot.

Gain Control.—Inverse-feedback amplifiers have the disadvantage that it is not possible to control the gain of a stage around which inverse feedback is applied, for the reason that changing the g_m of such a stage would also change its bandwidth. Hence gain control can be applied to only half the stages of feedback pairs, to one stage out of three of feedback triples, and to the first stage only of a general n-stage feedback chain. These restrictions are often bothersome, for example in radar receivers, where very wide-range gain control is needed. The difficulties are not too serious in the case of staggered pairs, however.

Stagger-tuned or double-tuned amplifiers can be gain controlled in any stage or combination of stages.

Gain Stability.—Circuits employing plate-to-grid resistive feedback (whether pairs, triples, or chains) display a certain gain stabilization factor against g_m variations; this factor is always less than 2.

It must be pointed out that although this gain stabilization factor of about 2 is useful for some applications, it is of no great value for radar or television and is inadequate for use in measuring instruments, where gain stabilization factors of about 100 are desired.

Tuning Stability.—This topic has been discussed in Sec. 7·10. In view of the fact that for amplifiers intended for large production the relative criticalness of tuning is one of the most significant points upon which to compare the different amplifier coupling schemes, it is most unfortunate that so few results are known. A few statements can be made, however.

It is known that synchronous single-tuned stages are extremely noncritical in tuning, that staggered pairs are less critical than staggered triples, that equal-Q double-tuned circuits are much less critical than double-tuned circuits loaded on one side only, and that stagger-damped circuits employing one-side-loaded double-tuned circuits are extremely critical.

Nothing at all is known about the relative criticalness of feedback pairs and feedback triples.

The Naval Research Laboratory tests cited in Sec. 7·10 showed a superiority in bandwidth stability of 3/2 of feedback triples over staggered triples.

It is a guess, therefore, that feedback triples and staggered pairs are about equal in tuning stability.

It is a guess that staggered triples are less critical than one-side-loaded double-tuned circuits.

Finally, it is a guess that despite the Naval Research Laboratory tests cited in Sec. 7·10, equal-Q double-tuned circuits are less critical than either staggered pairs or feedback triples, that is, are the least critical circuits next to synchronous single-tuned circuits.

CHAPTER 8

AMPLIFIER MEASUREMENT AND TESTING

By Yardley Beers and Eric Durand

The measurement and testing of the following amplifier characteristics are discussed in this chapter:

1. Gain.
2. Bandwidth and shape of pass band.[1]
3. Pulse response.
4. Overload characteristics.
5. Freedom from regeneration.

In connection with Item 5 this chapter contains a fairly extended treatment of means of eliminating regeneration.

Because of its extreme importance, an entire chapter (Chap. 14) is devoted to the measurement of the signal-to-noise performance of amplifiers

The emphasis in this chapter is on measuring instruments and techniques that are fairly new. Consequently there is no discussion of c-w generators.

8·1. Swept-frequency Signal Generators.—These are used primarily in amplifier alignment and determinations of pass band. In a typical setup, as shown in the block diagram of Fig. 8·1, the generator frequency is periodically swept across a band of frequencies at some a-f rate. The sweep waveform either may be a sawtooth, with a rapid "fly-back" to the starting frequency or may be smooth, sweeping up and then back in frequency. It is important that the amplitude of the signal from the generator be constant over the swept band. Simultaneously, a synchronized voltage sweep is applied to the horizontal plates of a cathode-ray oscilloscope. The sweep waveform must be such that a given horizontal position of the cathode-ray-tube spot corresponds uniquely to a single frequency, regardless of whether the sweep is from low to high or from high to low frequency. The relation between spot position and frequency should be as linear as possible. If the output

[1] There is no discussion of phase-measuring techniques, principally because there was only occasional need in the work of the Radiation Laboratory to employ phase-correcting networks. For minimum-phase-shift networks, pulse response is determined by the absolute value vs. frequency characteristic alone.

voltage of the swept-frequency generator is applied to a circuit and the rectified output of that circuit is applied to the vertical plates of the oscilloscope, the spot traces out the amplitude-vs.-frequency curve of the
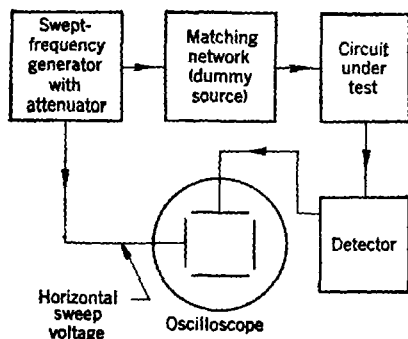

Fig. 8·1.—Block diagram of test setup using swept-frequency generator. The oscilloscope should have a high-gain audio amplifier, and the chain of elements between the detector and the cathode-ray-tube plates, including the audio amplifier, must have low-frequency response extending to well below the sweep rate.
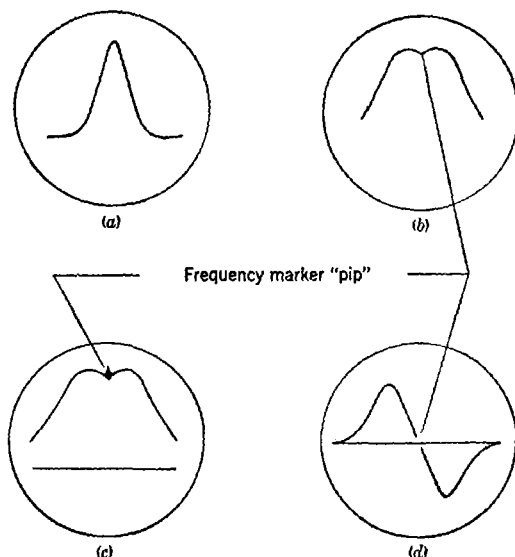

Fig. 8·2.—Typical response curves observed with a swept-frequency signal generator.

circuit, modified by the rectification law.   Response curves so traced for typical amplifiers are shown in Fig. 8·2.

The frequency modulation may be generated electronically using a reactance tube or by varying the reflector voltage of a reflex velocity-

modulated oscillator, or it may be generated mechanically by a device that alters either the inductance or the capacity of the oscillatory circuit.

Reactance-tube generators generally have only a small fractional frequency excursion.

Very wide-band frequency modulation together with great flexibility in choice of band center may be achieved by using a mechanically swept microwave oscillator and by mixing the output signal with that of another microwave oscillator separated in frequency from the first by a suitable amount.   If more power is desired, a carefully designed wide-band amplifier operating at the beat frequency may be added.[1]

A second type of tunable swept-frequency generator that depends on the beat between a fixed and a modulated oscillator has been built commercially by RCA.   Frequency modulation of a fixed-tuned oscillator is accomplished mechanically by using the voice coil of a dynamic speaker to drive a variable condenser; the other oscillator is tunable but unmodulated.   Both oscillators operate in the region between 100 and 200 Mc/sec.

A swept-frequency generator in wide use at the Radiation Laboratory depends on a motor-driven variable condenser, with a single oscillator operating in the desired frequency range.   The frequency excursion is about $\pm \frac{1}{8}$ of the center frequency, and the center frequencies usually employed are 30, 60 and 100 Mc/sec.

Constancy of amplitude is maintained by means of an automatic gain control that adjusts the g_m of the oscillator tube according to the output amplitude.   This type of amplitude limitation results in an output signal that is relatively free of harmonics.

It is not satisfactory to obtain output-voltage constancy by use of an amplitude limiter.   Such means lead to large harmonic content, and what is required of a swept-frequency generator is not merely constancy of the output voltage but constancy of the fundamental-frequency component of the output voltage.

A complete schematic diagram is shown in Fig. 8·3.   The oscillator employs a twin-triode type 6J6, in a conventional push-pull Hartley circuit.   Frequency modulation is accomplished by a special condenser whose rotor plates are driven at 1800 rpm by a synchronous motor. Outlines of the plates and the electrical connections are shown in Fig. 8·3.   Note that the two halves of the rotor are insulated from each other and from ground.   The unit consists essentially of two condensers in series, and the capacity used is that existing between the stator

---

[1] Harold Johnson, "A Wideband Frequency-modulated Alignment Oscillator," RL Report No. 738, May 31, 1945.   The maximum frequency excursion obtained is about 110 Mc/sec with amplitude variation of less than 1 db.   The center frequency may be adjusted anywhere from zero to 400 Mc/sec.

FIG. 8·3.—Swept-frequency signal generator.

sections. Thus, when the stator plates are aligned with the rotor plates, the capacity is a minimum; when they are at 90°, it is a maximum. A complete capacity cycle occurs twice per shaft revolution, so that the period is 60 cps. For each revolution of the shaft, the same value of capacity appears four times. It is important that the four angles involved be accurately complementary and supplementary, so that the four corresponding spots on the cathode-ray tube may be fused. A high degree of mechanical symmetry is therefore required.

For the automatic gain control, a signal from one of the 6J6 grids is applied to a diode rectifier. The resultant negative output is fed back through a cathode follower to determine the bias on the grids of the two sections of the oscillator. A single type 6AQ6 duplex diode-triode serves both functions.

When the circuit under test has appreciable response over the entire swept range, which is often the case when a single stage is being examined, it is hard to tell where the baseline, or position of zero response, lies (see Fig. 8·2b). To overcome this difficulty, a baseline-marker circuit is arranged to turn off the oscillator during one of the four half-sweeps per shaft rotation. For this purpose a multivibrator employing a type 6SN7 tube is used. The plate of the 6AQ6 cathode-follower gain-controlling tube is fed from one of the multivibrator plates, so that when that section is conducting, the oscillator grids are biased beyond cutoff, reducing the output power to zero. When the same section is nonconducting, however, the gain circuit functions normally. The free-running period of the multivibrator is about 30 cps, and the grid-return resistors are unequal so that one phase is about three times as long as the other. It is synchronized to the sweep frequency by a signal taken from the horizontal plate sweep.

Several marker devices may be used to indicate the relation between cathode-ray spot position and frequency. The simplest is called a "passive" or "absorption" marker and consists merely of a tunable high-Q circuit loosely coupled to the oscillator. This circuit absorbs some energy from the circuit at its resonant frequency and produces a small dip at the corresponding point on the trace. The scheme suffers from two drawbacks. First, it is hard to detect the dip when the output amplitude is low, notably in the determination of the "crossover frequency" of a discriminator; and second, if the limiter circuit in an f-m receiver is really effective, the dip will be wiped out. The effects on the oscilloscope of such a marker as well as of the two markers mentioned below are illustrated in Fig. 8·2.

The second device is the "active" marker, which is simply a low-powered tunable oscillator that beats with the swept-frequency oscillator at the point at which the two frequencies are nearly equal. This point is also hard to detect when the output amplitude is low.

The final device, used in the generator of Fig. 8·3, is the "blanking" marker. A small amount of energy is coupled from the oscillator into the high-Q tunable circuit connected to the right-hand grid of the 6SL7. This tube, being connected as an "infinite impedance," or cathode detector, offers little loading, so that a considerable positive voltage is built up at the grid resonant frequency. As the oscillator frequency is swept past this point, therefore, a short positive pulse is generated. This pulse is applied to an amplifier consisting of the left half of the 6SL7 to produce a large negative pulse that may be applied to the intensity grid of the cathode-ray tube to cause a momentary extinguishing of the beam. The sharpness of this blank spot can be increased by biasing the amplifier beyond cutoff so that only the tip of the pulse is effective. This type of marker produces a clear indication even when the response of the circuit is zero and is definitely superior to the other markers.

Since the capacity variation with angle is linear, the frequency variation is crowded at the high-frequency end of the sweep. This disadvantage may be offset by making the horizontal deflecting sweep slow at one end and fast at the other. To obtain such a sweep, sinusoidal power is applied to a type 6SK7 pentode amplifier biased so that it is carried from a region of low g_m to one of high g_m by its grid signal. This distorts the input wave into the desired shape. A phase-shifting network with coarse and fine controls is provided so that the upsweep and downsweep in frequency can be made to coincide on the oscilloscope.

The power supply is self-explanatory. VR-tube regulation is used for the screen of the sweep-amplifier tube and for the negative supply for the automatic-gain-control circuit.

8·2. Direct and Carrier-frequency Pulse Generators.—Although many of the properties of amplifiers can be determined from measurements with c-w generators, their response to transients is best measured with pulsed signals.

A good pulse generator should have as short a rise and fall time as possible, and the regions at the top and after the end of the pulse should be flat and free of "wiggles." Carrier-frequency pulses should be free from frequency modulation during the pulse.

Secondary properties of good pulse generators are (1) independence of pulse output amplitude and shape from line voltage, pulse repetition frequency, and output loading; (2) good shielding; and (3) wide range of available pulse lengths.

If the pulses are initiated by a trigger from an external source, the starting time should be free of "jitter"; that is, the delay between the trigger pulse and the output pulse should be constant. In "self-synchronized" pulse generators, on the other hand, there is normally a

trigger source for initiating operations in external equipment; these triggers should start quickly and rise rapidly and also be free from jitter with respect to the output pulse.

A pulse generator designed and built at the Radiation Laboratory contains not only a pulsed carrier-frequency oscillator, usually at either 30 or 60 Mc/sec, but also two direct-pulse generators that actuate the carrier-frequency section. All three sections contain the latest design principles and are thus illustrative of the current state of the art; the generator is described in detail.

Figure 8·4 is a block diagram of the generator, and Fig. 8·5 is a complete circuit diagram.

The generator is divided into four sections: (1) power supply, (2) short-direct-pulse generator, (3) long-direct-pulse generator, and (4) carrier-frequency pulse generator.

Power Supply.—The supply provides $+400$, $+195$, $+105$, and -150 volts, all but the $+400$ volts being regulated with gas voltage-regulator tubes.

Short-direct-pulse Generator.—Operation of this circuit is as follows: A positive trigger applied to the trigger-shaper gas tetrode V_1 causes it to fire. Because C_1 prevents any sudden change in plate potential, the breakdown results in the appearance of a sharp positive trigger pulse at the cathode. This trigger is coupled, in a way that will be considered later, to the grids of the two gas tetrodes V_4 and V_5, which are the actual pulse generators. Before the appearance of the initiating trigger pulse, these tubes are nonconducting because of the bias voltages on their grids. The cathode of V_4 and the plate of V_5 are held at ground potential by R_{28}, while R_{23} and R_{26} hold the plate of V_4 at about $+250$ volts.

When V_4 fires, its cathode potential rises abruptly, causing the start of the output pulse and simultaneously providing plate voltage for V_5.

Simultaneously, the pulse from the V_1 cathode is applied through R_{17} to the control grid of V_5. The rise of voltage on this grid is exponential, because of R_{17}, the shunt capacity C_{12}, and the grid-to-ground capacity of V_5. The amount of bias on this grid is set by the "pulse length" adjustment R_{19}, so that the time at which the grid voltage reaches the firing point may be varied. There is thus an adjustable delay between the firing of V_1 and that of V_5.

The firing of V_5 marks the end of the pulse, since it results in abruptly lowering the potential at the plate-cathode junction of V_5 and V_4 nearly to ground. An approximately rectangular positive pulse is thus generated at this junction. Some of the details that are responsible for the speed of rise and the flatness of the top of the pulse will now be considered.

The combination of R_{15} and the total stray capacity between the V_4 control grid and ground results in a small delay between the firing of V_1 and that of V_4. This delay is large enough to ensure that when the delay in firing of V_5 is set at its minimum value, V_5 fires at least as soon as V_4. Therefore, the minimum pulse length that can be gen-
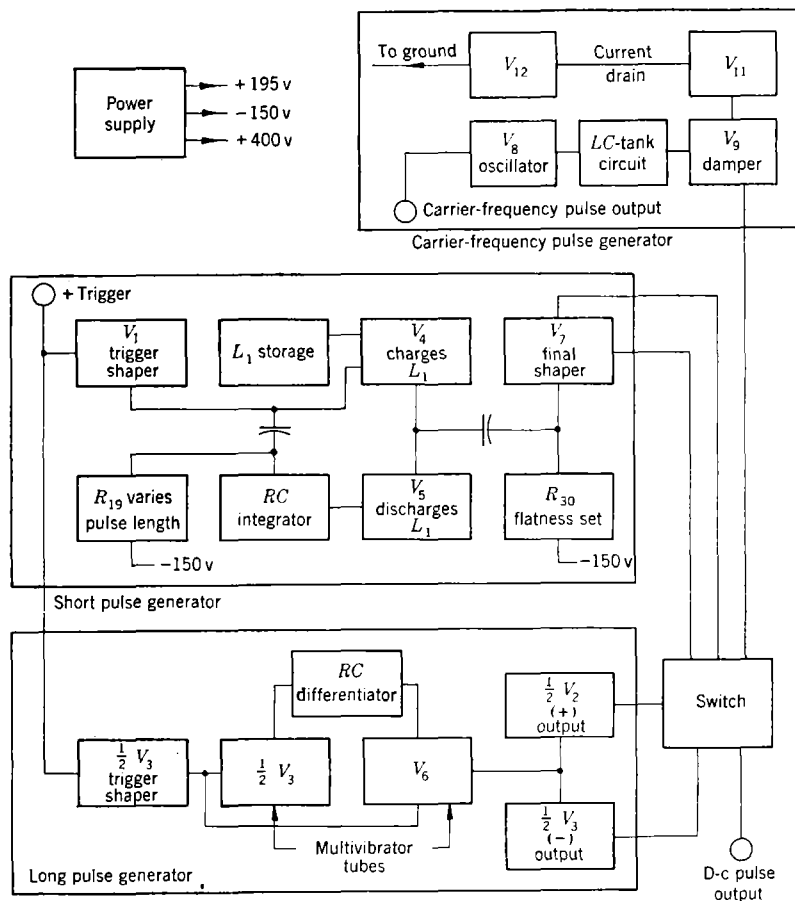

FIG. 8·4.—Block diagram of pulse generator.

erated is determined only by the breakdown time of V_5 and the speed with which the plate voltage on V_5 can rise.

The delay line L_1 provides a flat top for the pulse. Earlier pulse generators used a large capacity between the plate of V_4 and ground. The flatness of the top of the pulse was limited by the ability of this
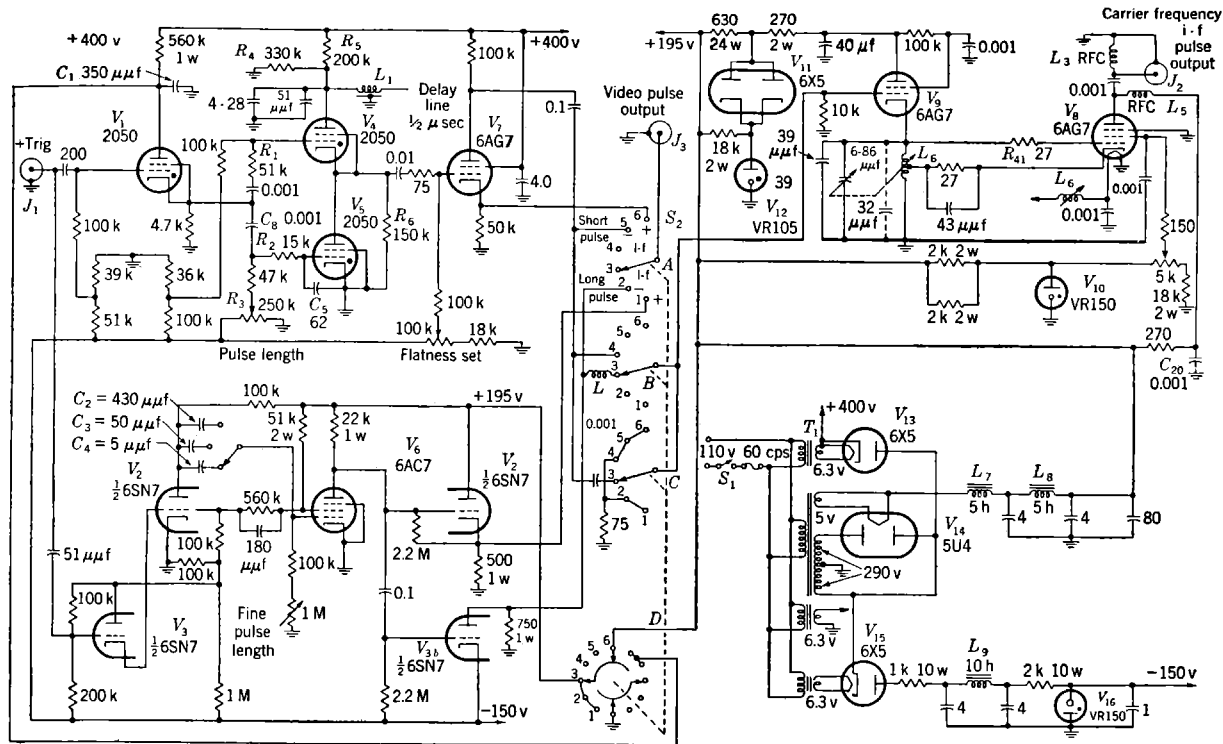
Fig. 8·5.—Circuit diagram of pulse generator. $S_2 - A$, B, and C are nominally 1-gang 11-position switch decks. $S_2 - D$ is nominally a 2-gang 5-position deck, but a sixth position is used as an OFF position. S_2 is shown in position 3. S_2 positions: (a) − long; (b) + long; (c) carrier-frequency long; (d) carrier-frequency short; (e) − short; (f) + short.

capacity to hold its charge. If the capacity was made large, it became difficult to replace the charge during the interval between pulses. The amplitude of the output signal was therefore a function of pulse repetition frequency. The delay line has a small total capacity to ground and can therefore be charged quickly to full potential. On the other hand, it sustains its voltage accurately for the time required for a signal to travel to the end of the line and back. In this pulse generator, the
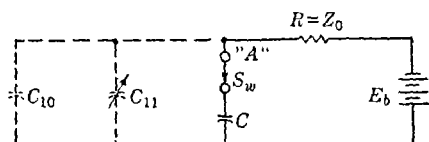

Fig. 8·6.—Equivalent circuit of delay line after firing of V_4. Before the firing of V_4, represented by the switch Sw, the voltage across C is zero and the voltage across C_{10} and C_{11} is the supply voltage E_b. When Sw is closed, the charge on C_{10} and C_{11} is rapidly redistributed and shared by C.

line "length" is either $\frac{1}{2}$ or 1 μsec and therefore pulses as long as 1 or 2 μsec, respectively, can be generated. The flatness of the pulse top is determined mostly by the quality of the line, although the limiting action of the following amplifier V_7 contributes appreciably.

The action of the delay line may be represented by a battery whose voltage is the voltage to which the line is charged in series with a resistance equal to the characteristic impedance of the line for a period equal
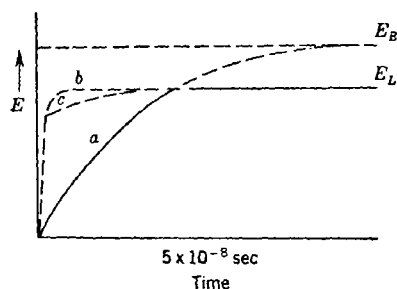

Fig. 8·7.—Shape of voltage rise at the V_4, V_5 junction. (a) Without compensating condensers C_{10} and C_{11}; (b) compensating condensers set correctly; (c) compensating condensers too small. E_L is the voltage at which the grid of V_7 starts to draw current.

to twice the length of the line after the firing of V_4. Figure 8·6 shows the equivalent circuit for the short interval following the firing of V_4. The stray capacity existing between the V_4 cathode and ground, shown as C, may be as much as 50 $\mu\mu$f because it includes interelectrode capacities of V_4, V_5, and V_7. The voltage at the cathode therefore would rise along the curve (a) shown in Fig. 8·7 with a time constant of about 5×10^{-8} sec. When it reaches a value determined by the bias on

the grid of V_7, its rise ceases because of the diode action of the grid-cathode combination. If, however, a capacity C_{10} and C_{11}, shown in dotted lines in Fig. 8·6, is added between the V_4 plate and ground when the switch $Sw(V_4)$ is closed, the junction point A quickly reaches a voltage determined by the ratio of the capacity divider formed by C and C_{10},

C_{11}. By proper adjustment of C_{11}, this voltage may be made equal to the desired final pulse voltage. In this case, the rise time is limited only by the stray inductance in the leads associated with the condensers and may easily be made less than 2×10^{-9} sec. The rise obtained with such compensation is shown in Fig. 8·7b.

The fall time of the pulse is short because it is determined by the very low effective internal resistance of V_5 at the instant of firing or by residual lead inductances.

If the pulse amplitude at the V_4, V_5 junction is so large that heavy current is drawn by the grid of V_7, the current drain on the delay line results in a falling off of its voltage. Therefore, a bias adjustment R_{30} is installed as a "flatness set."

The output tube V_7 is biased far beyond cutoff during the interpulse interval. Thus it provides a sharp termination for the pulse, free from "wiggles." Positive and negative output pulses may be taken from its cathode and plate respectively.

Because of the extremely rapid rise and fall times required of the pulse generator in order that it can give the carrier-frequency pulse a good start (see below), the greatest care must be taken in the layout of the pulse-generator tubes and amplifier.

Long-direct-pulse Generator.—The heart of the long-pulse generator is a one-sided multivibrator, made up of the triode V_2 and the pentode V_6. As far as the multivibrator action is concerned, V_6 acts as a triode, its screen grid g_2 serving as the plate. In the normal condition V_2 is fully cut off by the bias supplied by a resistance network, whereas V_6 is fully conducting. A positive pulse applied to the grid of V_2 through a diode (V_3, cathode and grid) causes V_2 to conduct, and the resulting decrease of its plate potential, transmitted through C_4, C_5, or C_6, cuts off the control grid of V_6. Therefore, the voltage at the plate of V_6 rises; this is the leading edge of the pulse.

Presently, at a time determined by R_{21} and the adjustable R_{22}, the negative voltage at the control grid of V_6 disappears, and V_6 starts to conduct. The resultant decrease of the potential on the screen grid of V_6, coupled through C_9 and later, for direct current, through R_{14}, cuts off V_2 once more, restoring the original condition and terminating the pulse. Pulse length is adjusted by R_{22} and by the choice among C_2, C_3, and C_4.

The diode action provided by V_3 is necessary to prevent the fall of voltage marking the end of the initiating trigger from restoring the multivibrator to its original condition. The connections shown at the plate of V_3 have no significance; it was used as a convenient tie point.

Positive and negative output pulses are obtained from the cathode follower V_{2b} and the amplifier V_{3b} respectively.

Carrier-frequency Pulse Generator.—The oscillator proper V_8 is of the "electron-coupled" type, normally inhibited from oscillation by the heavy damping of the cathode follower V_9 connected across the tuned circuit.   The carrier-frequency pulse is initiated by applying a large negative direct pulse from the short-direct-pulse generator to the control grid of the cathode follower.   This pulse removes the damping and allows the oscillator to start.   Because of the extremely rapid rise of the direct pulse, which occurs in a fraction of a carrier-frequency cycle, the standing current of the cathode follower in the inductance of the oscillatory circuit starts the oscillations.

Suppose for the moment that the oscillator tube were removed. After the direct pulse cuts off the current i_c in the cathode follower, a damped oscillation exists in the LC-circuit, whose amplitude starts at the value $e_0 =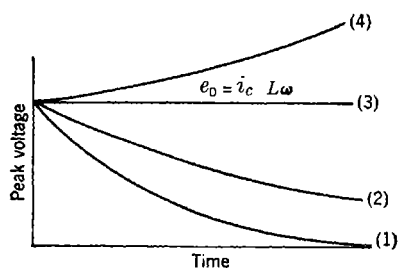 i_cL\omega$ and falls off exponentially at a rate depending on the circuit Q, as shown in Curve 1 of Fig. 8·8. If the tube is reinserted but with the oscillatory circuit adjusted for very weak feedback, the falling off of amplitude is delayed, a "negative resistance" supplied by the tube offsetting part of the circuit losses (Curve 2). Increasing feedback prolongs the delay until a value is reached at which the amplitude remains constant, circuit losses being just offset by the tube (Curve 3).   More feedback results in a rising amplitude of oscillation, the rise continuing until limited by nonlinear conditions in the tube such as grid current or change of g_m (Curve 4).

FIG. 8·8.—Dependence of amplitude of oscillatory voltage on time.

If the condition producing Curve 3 is met, the carrier-frequency pulse rises to its full value in a fraction of a cycle and retains constant amplitude for the duration of the direct pulse.   When the damping is restored, the oscillation dies out rapidly because of the severe loading $(1/g_m)$ of the cathode-follower impedance.

The energy supplied to the LC-circuit by the tube is adjusted by R_{39}, which governs the screen-grid voltage.   In order to make the voltage across the tuned circuit independent of frequency, a constant L/C ratio must be maintained.   Therefore, both inductance and capacity must be suitably ganged and tuned together.   To maintain a flat-topped pulse over the carrier-frequency band, compensation must be made for changes in circuit Q.   A small bypass condenser across the cathode resistor R_{42} increases the amount of regenerative feedback at high frequencies.   Because the oscillator tube always operates over a highly

linear portion of its characteristic, the output signal is exceptionally free of harmonics.

To ensure constant output, the supply voltage for the cathode follower is regulated to $+105$ volts by the gas regulator V_{12}. However, the average cathode-follower current is about 100 ma under low-duty-ratio conditions (short pulses or low repetition rate), whereas, at the maximum duty ratio of 90 per cent, the current is small. Under the heavy-current conditions, the VR-tube might extinguish, since current intended for it is diverted to the cathode follower. A diode V_{11} is therefore provided to allow the plate of the cathode follower to fall below $+105$ volts without extinguishing the VR-tube.

To provide the fast starting discussed above, the short-pulse generator is always used to initiate the oscillator action. The reason is that the rise time of the pulse from the long-pulse generator is about 10^{-7} sec, a time long compared with a quarter-cycle at 60 Mc/sec and hence far too long to give a clean start. By the time the short pulse is over, however, the long-pulse voltage will suffice to keep the cathode follower turned off. However, it is not desirable that the long-pulse rise time be too short lest, because of the grid-to-cathode capacity of V_9, a "wiggle" be imparted to the carrier-frequency oscillation. Also, again because of this capacity, the output impedance of the pulse generator would constitute a complex load on the LC-circuit, causing some frequency modulation and lowering its Q. Both of these effects are overcome by a small r-f choke in the line from the long-pulse generator.

The pulse selector switch S_2 offers a choice of positive and negative, direct- and carrier-frequency pulses, both long and short. In addition to switching the output jacks to the appropriate points in the circuit, the switch interrupts the plate supply for the long-pulse generator in the short-pulse positions and also short-circuits the plate of the trigger shaper V_1 of the short-pulse generator in the long-pulse positions. As was explained above, the short-pulse generator is left in operation to give a quick start when long carrier-frequency pulses are required. A second switch S_1 offers three ranges of pulse lengths in the long-pulse position.

The pulse generator gives satisfactory operation for duty ratios up to 90 per cent and pulse repetition frequencies up to about 2000 per second, the output being independent of repetition frequency up to this value. Operation at 4000 pulses per second results in a slight "sag" in the short-pulse output, which may be eliminated by readjustment of R_{30}.

8·3. Miscellaneous Testing Equipment.—The devices described in this section are modifications of well-known instruments adapted to particular uses.

Vacuum-tube Voltmeters.—Such instruments, when intended for use at high frequencies, usually have the rectifying diode contained in a small probe at the end of a flexible cable. The input capacity of the probe is made up of the anode-cathode capacity of the diode and of stray capacitance. The British diode type CV58, which has extremely small anode-cathode capacity, makes possible a vacuum-tube voltmeter with an input capacity of about 0.6 $\mu\mu$f. Even this capacity is disturbing in bandpass amplifiers, but it can often be allowed for.

At high frequencies the vacuum-tube voltmeter output voltage changes because of series resonance between lead inductance and stray capacity. By mounting a type CV58 tube in a close-fitting brass shell used as one terminal of the probe, the other terminal being the anode lead of the type CV58 diode which projects about an eighth of an inch through the top of the bulb, it is possible to achieve accuracies of a few per cent up to several hundred megacycles per second and usable, although inaccurate, results up to 2000 Mc/sec.

"Synchroscopes."[1]—These are oscilloscopes intended for the observation of short, periodic pulses. Special features are

1. Fast sweeps. The sweeps may occupy only a small part of the interpulse interval. For detailed observation of a microsecond pulse, the pulse may be spread over the entire cathode-ray-tube face.
2. Sweep-starting-time controls, to enable the operator to shift the starting time of the sweep with respect to that of the pulse under examination.
3. Extremely small time jitter.
4. High-fidelity pulse amplifiers. The amplifier of Fig. 2·38, having a rise time of $\frac{1}{50}$ μsec, was designed for use in the model P5 synchroscope.

Attenuators.—For amplifier measurement, attenuators are usually matched to the impedance of a coaxial line, 50, 75, and 100 ohms being common values. Therefore, if a load equal to the attenuator iterative impedance R_0 is attached to one end of the attenuator, the same impedance will be observed looking into the other end. Matched attenuators are, on principle, dissipative, since reflection is always a sign of a mismatch. If a matched attenuator has large attenuation, the impedance looking into the attenuator is approximately R_0, regardless of the impedance of the load at the other end. This fact is utilized to provide sources or loads of known impedance, the attenuator serving to isolate the circuit under test from the variable or unknown impedance of the primary source or the ultimate load.

[1] *Cathode Ray Tube Displays*, Vol. 22, Chap. 7, Radiation Laboratory Series.

The most important type of attenuator for amplifier testing is a calibrated device variable in steps and using small composition resistors in either π- or T-networks.   A typical circuit is shown in Fig. 8·9.   Table 8·1 shows the values that should be used for various attenuations for an

TABLE 8·1.—RESISTANCE VALUES FOR ATTENUATOR SHOWN IN FIG. 8·9

Attenuation, db	R_A	R_B	Test data	
			Resistance P to Q	Resistance P or Q to ground
1	1305	8.7	8.7	657
2	654	17.5	17.2	330
3	438	26.6	25.7	227
5	267	45.6	42.1	144.5
10	144.0	106.7	78.0	91.9
20	91.7	361	123.2	76.8

iterative impedance of 75 ohms.

An attenuator of this type is reliable for frequencies from zero up to a limit determined by stray lead inductance and capacity.   By using small composition resistors carefully mounted, satisfactory operation, with regard to accuracy of calibration and of match, may be obtained
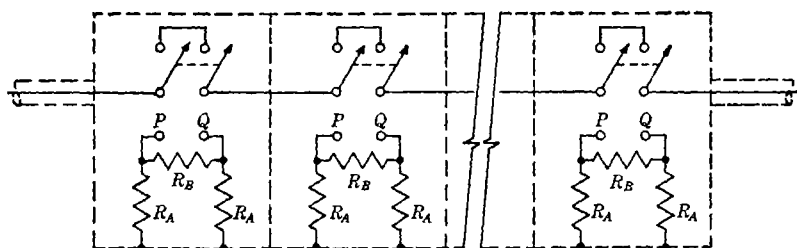
FIG. 8·9.—A 75-ohm attenuator.   The accuracy of any section may be determined with a bridge by measuring the resistances from P to Q and from P and Q to ground with the switch in the out position.   The values that should be obtained are shown in Table 8·1 together with the values of the individual resistors.   For any other iterative impedance, multiply these values by the ratio of the new impedance to 75 ohms.

up to about 100 Mc/sec.   One arrangement that works well in this region is shown in Fig. 8·10.   The over-all attenuation of this unit is over 100 db.   Careful shielding is required if the leakage power due to stray coupling is to be kept well below the desired power.   Note that in the arrangement shown, all leads are very short and each section is fully shielded from the adjoining sections.

Wire-wound resistors cannot be used because of their inductance. Semiprecision composition resistors are invariably bulky and have

excessive capacity to ground. Allen-Bradley $\frac{1}{2}$-watt carbon resistors are therefore used and are selected for 2 per cent accuracy from a large lot, or else resistors measuring less than the desired value are "trimmed" to the correct value by filing out a notch through the insulating case into the composition until the correct value is obtained, after which the notch is sealed with lacquer. In general, because of aging and temperature effects, it is not worth while to try to achieve an accuracy greater than 2 per cent.

In making bandwidth or noise measurements, it is often necessary to be able to change the strength of the input signal by exactly 3 db. For this purpose, a single-element attenuator of the type discussed above may be used. It is generally mounted in a small metal can provided with input and output cable fittings and is installed in the
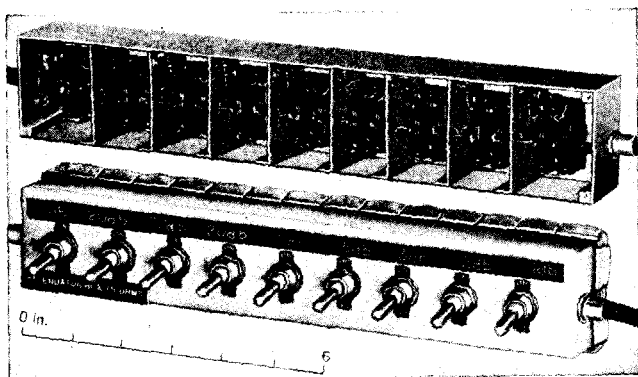
FIG. 8·10.—A 75-ohm attenuator box.

line from the signal generator to the amplifier under test. Some of the swept-frequency signal generators have such attenuators built-in.

Dummy Input Circuits.—The output lead from most signal generators is a coaxial cable. It is necessary to provide the proper termination for this cable to ensure the correct functioning of the attenuator. On the other hand, correct functioning of an amplifier requires that the impedance of the test source be the same as the impedance $Z_i = R_i + jX_i$ of the source used in actual operation.

Therefore, a "dummy input circuit" is used to couple the generator to the amplifier under test. As seen from the generator, the circuit impedance is R_0, the characteristic impedance of the cable; as seen from the amplifier it is $Z_i = R_i + jX_i$.

In microwave radar receivers the i-f amplifier signal source is a crystal mixer whose i-f impedance is made up of a resistance in parallel with a capacitance. The resistance ranges from 200 to 500 ohms, and the

capacitance ranges from 5 to 15 $\mu\mu$f for the mixer proper. If a cable is used to attach the mixer to the amplifier, it is almost always necessary to take the admittance-transforming properties of the cable into account (by means of an admittance circle diagram, for example), rather than to regard the cable as a lumped capacity. This is the case even for a cable as short as 12 in. at 30 Mc/sec or 6 in. at 60 Mc/sec.

Sometimes the mixer is left in place, and the signal is introduced through a "dummy crystal." Figure 8·11 shows the usual arrangement. The impedance seen looking into the input terminals of the amplifier is usually relatively high. The impedance seen from the cable looking into the dummy crystal is therefore only slightly less than the value of the shunt resistance R_2. In practice, R_2 is set equal to R_0, providing the correct termination for the cable, and hence for the attenuator. Looking back into the dummy crystal from the amplifier, one sees R_1 in series with $R_0/2$, since the cable itself, being terminated by R_0 at the generator
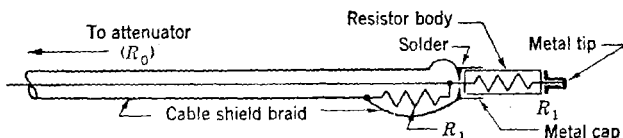

Fig. 8·11.—Dummy-crystal diagram showing approximate physical layout.

or attenuator end, looks like R_0. The value of R_1 is therefore chosen so that $R_1 + (R_0/2) = R_i$.

When a dummy crystal is used, no provision for capacitance need be made, because the mixer itself supplies the right amount. At times, however, it is preferable to remove the mixer and replace it with a complete "dummy mixer." The circuit used is the same as before, but a variable condenser C_i is added from the output end of R_1 to ground to simulate the mixer capacity. The dummy mixer is generally mounted in a small metal can. If the amplifier to be tested includes a cable for the mixer, a suitable cable fitting is mounted directly on the can. If the amplifier is provided with a jack, a short cable may be included as part of the dummy mixer. In either case, the condenser is adjusted to present the same capacity that would be seen at the point of attachment with the regular mixer in place.

Dummy mixers have been built in which various values of R_1 can be selected by means of a switch. In addition to providing values ranging from 200 to 500 ohms, 0 and 5000 ohms are provided to facilitate the study of the effect of variation of crystal impedance on amplifier-bandwidth and noise-figure performance. The former serves to damp out the selective effect of the input circuit in order that effects in later

circuits may be studied; the latter essentially provides a constant-current source useful for special applications.

Attempts have been made to develop signal generators with push-pull output connections for testing the balanced input circuits used with balanced mixers. It is not difficult to obtain equal output voltages at the generator, but it is very difficult to provide matched pairs of attenuators that will give equal output voltages at low level. It has been found, fortunately, that all necessary information can be obtained from tests using a single-ended output signal introduced, by means of a dummy crystal, first into one and then into the other crystal holder, the unused holder being supplied with a second dummy crystal containing a resistor of value R_i, one end of which is grounded. The output voltage that would be obtained with a push-pull input source is 6 db more than the average of the two output voltages obtained in the above
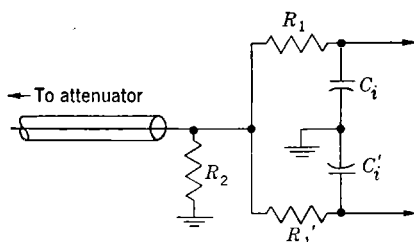

FIG. 8·12.—Double dummy mixer for unbalance tests on balanced input amplifiers.

manner (if it is assumed that the push-pull input circuit is fairly well balanced). Unbalance in push-pull input circuits may be readily measured by feeding equal in-phase signals into both input terminals. A special dummy-crystal pair may be constructed for this purpose, using the connections shown in Fig. 8·12. With such a device, the output response of a perfectly symmetrical circuit is zero, and the ratio of response to in-phase signals to response to push-pull signals is a measure of the unbalance.

Special problems arise in the use of a noise-diode signal generator for the measurement of signal-to-noise ratio. These are considered in Sec. 14·5.

8·4. Measurement and Alignment of Bandpass Amplifiers.—It is possible to make bandpass and alignment tests by using a modulated or unmodulated c-w signal generator, making a series of point-by-point measurements, but the process is tedious. The swept-frequency generator technique discussed in Sec. 8·1 is far superior, in that it enables the operator to see at a glance the entire shape of the pass band. It also permits quick observation, from peaks and dips in the shape of the selectivity curve, of incipient regeneration in the amplifier.

On the other hand, the point-by-point c-w generator method must be used if there is a need, as sometimes arises, for accurate measurements of response more than 20 db down on a selectivity curve.

The basic arrangement of the equipment involved in the swept-frequency-generator technique has already been shown in Fig. 8·1. This section is concerned with the means whereby signals are introduced into and taken out of the amplifier under test and the interpretation of the observations.

Connections to the Amplifier.—The proper way of making connections to an amplifier under test is of very great importance.   It is very common with high-gain amplifiers that poor connections to measuring apparatus introduce more regeneration than there is in the amplifier itself.

The only safe way of dealing with high-gain amplifiers is to have all signal connections, to both input and output terminals, rigorously coaxial.   Furthermore, it is wise whenever possible to incorporate some attenuation between generator and amplifier and between amplifier and measuring device; this has the effect of reducing the gain in the feedback loop from the amplifier output circuits, through the coupling between measuring device and generator, back to the input circuits.

These precautions may be relaxed somewhat in the case of measurements involving but one or two amplifier stages, where it may be permissible to connect a generator by very short leads.   In the long run it is much wiser, however, to spend some time constructing an adequate collection of coaxial fittings, adapters, and attenuators than to carry out the meaningless and variable measurements resulting from hasty connections of input and output leads by means of clips.

Input Connections.—The method of introducing the signal so as to include the effects of the coupling circuit into the first amplifier tube has already been covered in the discussion of dummy input circuits (Sec. 8·3).   To study subsequent circuits the signal is applied to the grid of one of the amplifier tubes.   A 75- or 100-ohm cable from the attenuator is terminated with a suitable resistance; and if only one or two stages are involved, very short leads are run from the center conductor to the grid in question and the outer conductor is soldered to the amplifier chassis.   For measurements involving more than two stages a special probe or dummy tube is built; this probe is plugged into the preceding tube socket and enables measurements to be made with the amplifier chassis cover in place.

The impedance looking toward the signal generator is either 37.5 or 50 ohms, since the terminating resistor is in parallel with the cable impedance.   This low resistance broadens the pass band of the coupling circuit at the point of signal introduction so that the observation refers only to the subsequent circuits.

Figure 8·13 shows the tube and circuit line-up for a typical amplifier. The significant pins on the various tube sockets are designated by the letters *P*, *G*, *K*, for plate, grid, cathode respectively. The tubes and circuits are lettered, and reference is made by these letters in the following discussion.
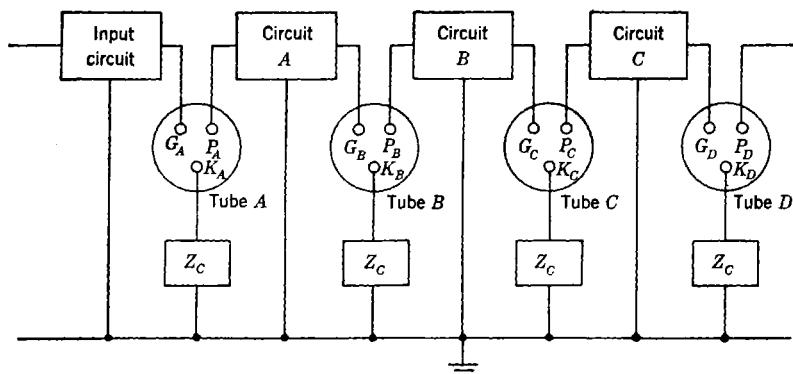

FIG. 8·13.—Tube and circuit line-up in an amplifier. The cathode-circuit impedance Z_c is a short circuit at the intermediate frequency and is equal to the bias resistance (50 to 200 ohms) at pulse and lower frequencies.

Output Connections.—Suppose that Circuit *B* is to be examined. The signal generator is attached either to G_B or P_A, as discussed above. Three types of output connections are possible.

1. *High-impedance Vacuum-tube Voltmeter.*—A high-impedance vacuum-tube detector may be attached to G_C to measure the alternating voltage at this point. Such a detector must be designed to provide negligible loading and detuning. The British type CV58 has been used in probe circuits with an effective capacity of about 0.6 $\mu\mu$f, but even this small value causes considerable detuning because of the small total tuning capacity in the average coupling circuit (about 10 or 11 $\mu\mu$f between type 6AK5 tubes). Most probes show capacities of 4 $\mu\mu$f or more. This method is seldom used.

2. *Low-impedance Voltmeter.*—Vacuum tubes or crystal rectifiers are used, but in either case the rectifier input terminals are shunted with a low resistance. For instance, in the crystal-rectifier type the shunt resistor may be 70 or 100 ohms, which provides correct termination for a coaxial cable. The rectifier may then be located at a point distant from the amplifier and is usually located close to the output indicator, which is generally an oscilloscope with a high-gain audio amplifier.

The point of attachment for the measurement of Circuit B is P_C or G_D (for single-tuned circuits), in contrast with the case of the high-impedance vacuum-tube voltmeter. The low impedance voltmeter so broadens Circuit C that it has negligible effect on the observed selectivity curve.

This method suffers from the presence of "Miller effect" capacity at the input of Tube C. On the average, for type 6AK5 tubes, a total capacity C_{pg} exists from plate to grid, including socket and wiring capacities, of about 0.04 to 0.06 $\mu\mu$f. Because of the Miller effect, there is an apparent input capacity of $(G + 1)C_{pg}$ where G is the stage amplification. In the functioning amplifier, G may be about 5 or 10, so that $(G + 1)C_{pg}$ ranges between 0.24 and 0.66 $\mu\mu$f. But with the low-impedance-voltmeter probe in place, the stage gain is less than unity, causing a proportional reduction in the input capacity of Tube C and a consequent increase in the resonant frequency of the coupling circuit B. The detuning effect may amount to a megacycle per second at 30 Mc/sec or more and makes it difficult to determine interstage resonant frequency accurately.

3. *Self-rectification Connection.*—With moderately large signals applied to the grid of a bandpass amplifier tube, the grid excursion carries over into nonlinear regions and the average cathode current increases slightly. The increase in cathode current is therefore a measure of the signal amplitude; the tube is, in fact, acting as an "infinite impedance" detector. Within limits, the detector response so obtained is square law, the cathode-bias resistor having too small a value to linearize the response by inverse feedback.

To make the observation, a lead is connected from the cathode (K_C in the case in question) of the amplifier tube to the oscilloscope amplifier input terminal, through a low-pass filter consisting of a series 10,000-ohm resistor and a 1000-$\mu\mu$f condenser to ground. The output voltage is developed across Z_C, which at these low frequencies is resistive. Tube C is, of course, left in its socket. This method causes no detuning, either from vacuum-tube-voltmeter probe capacity, as in the first method, or from changing Miller-effect capacity, as in the second. The output voltage is fairly low, however, so that the gain of the audio amplifier must be high. The only serious drawback is that the connection can be made only with the chassis cover removed. This is not a serious drawback, since it is primarily used for accurate tests involving a single stage, when regenerative feedback is likely to be small.

Bandwidth and Bandpass Shape.—In measuring bandwidth it is convenient to observe the value of the output signal with a certain input power and then to double the input power by switching out a 3-db attenuator and to observe the frequencies at which the output signal has the same strength as before. This method of relying on a constant output indication makes it unnecessary to know the output or input law of the indicating device.

Bandpass Amplifier Alignment.—In the next few paragraphs techniques are discussed for accurate alignment of amplifiers intended for laboratory use or as prototypes for production. The adjustments usually consist in selecting suitable values of fixed damping resistors and suitable values of fixed tuning inductances. The production amplifiers are very likely to be fixed-tuned, so that there may be no alignment operation involved in production; however, for both prototype and production amplifiers an over-all check on pass band is recommended, as described below.

Synchronous Single-tuned Amplifiers.—The pass band of individual stages may be examined as described earlier in Sec. 8.4, making use of the "self-rectification connection" of one of the amplifier tubes as the indicating device. Once the individual stages have been given the correct values of damping resistors, the easiest method for aligning a tunable synchronous-tuned amplifier is to use an unmodulated c-w generator set to the desired center frequency and to peak each stage for maximum response, as indicated by the detector output current or voltage.

Stagger-tuned Amplifiers.—The individual stage bandwidths and center frequencies are examined in the same manner employed for synchronous-tuned amplifiers. Once the stages have been given the correct values of damping resistors, the easiest method for aligning a tunable stagger-tuned amplifier is to use an unmodulated c-w generator and peak each stage for maximum response with the generator set at the frequency appropriate to that stage. The sharpness of peaking may be increased by applying the detector output voltage to a high-gain indicating device and "bucking-out" most of the deflection.

Double-tuned Amplifiers.—The adjustment of an individual circuit is best done as follows, with a swept-frequency signal generator connected to the grid of the tube preceding the circuit in question and the output indication taken from the cathode of the tube following the circuit. First the primary is very heavily loaded, with a resistor less than a fifth of that normally in place; this makes the pass band single-tuned, and the secondary inductance is adjusted to the appropriate frequency. The primary loading is then restored to its former value; the secondary heavily loaded; and the primary adjusted to its appropriate

frequency. It should be remembered that if the bandwidth/band center ratio of the double-tuned circuit is high, the primary and secondary frequencies so determined do not coincide with the desired band center (see Sec. 5·5). After the initial loadings are restored on both primary and secondary, the coupling is adjusted to give the desired bandpass shape. If the bandwidth is not correct, the loading resistor or resistors must be changed and the process repeated. When the original double-tuned circuit is designed with care, this process converges rapidly.

Inverse-feedback Amplifiers.—Inverse-feedback chains can be aligned by applying an unmodulated signal to the grid of the last tube in the chain and peaking the output circuit of this tube for maximum detector deflection. The signal generator is moved back to the grid of the next to last stage, and the next to last stage is peaked. This process is continued until all the stages in the chain have been aligned.

The Complete Amplifier.—After alignment in the case of adjustable amplifiers or construction in the case of amplifiers with pretuned coils, the amplifier-bandpass shape should be examined to see that the bandwidth, bandpass shape, and center frequency fall within allowable limits. In this test a swept-frequency generator is used, the signal being introduced through a dummy input circuit (Sec. 8·3) and the output signal taken from the amplifier detector. The observation that is made at this time to detect regeneration will be considered separately in the next section.

8·5. Undesired Feedback Effects (Regeneration) in Bandpass Amplifiers.—A common difficulty with high-gain amplifiers is regeneration, in which energy is transmitted from a point at high gain back to one of low gain. A single stage, a group of consecutive stages, or an entire amplifier may be involved.

Indications of Regeneration.—Very strong regeneration is evident through oscillation. Moderate regeneration is apparent through variations in output voltage with changes in the location of external ground connections or when the external power supply leads are touched by hand or connected to ground through a large condenser.

The most convenient indication of regeneration is the change of pass band with variations in gain-control setting or with variations in plate-supply voltage; this change can be observed very easily by use of a swept-frequency signal generator.

The phase and amplitude of the fed-back voltage depends upon the forward gain and the properties of the feedback parameters. Since all these depend upon frequency, the net gain is increased at some frequencies and decreased at others. Therefore the presence of feedback causes a frequency-dependent distortion of the pass band. Unfortunately, a certain amount of feedback is inevitable, due to Miller effect

and cathode-lead inductance, but in the following discussion these effects will be assumed to be small. Moreover, if the control-grid voltage is varied, the input capacity changes also and to a larger extent than with changes in plate-supply voltage; this change causes a certain amount of detuning. For this last reason, varying gain by changing control-grid voltage is not so reliable a method of testing for change of pass band due to regeneration as varying gain by changing plate-supply voltage. In changing the plate-supply voltage, however, a precaution must be taken against change of pass band because of overload; it must be remembered that the amplifier is more likely to be overloaded with low plate voltage than with high plate voltage.

Just how much change of pass band can reasonably be ascribed to some of these unavoidable effects (Miller-effect detuning, change in input capacity, etc.) and how much must be ascribed to regeneration is difficult to specify. However, it is safe to say that a change in 3-db bandwidth of more than 10 or 20 per cent (which may be either an increase or decrease) or the appearance of subsidiary maxima and minima in the pass band is an indication of regeneration.

Causes of Regeneration.—In wide-band high-gain multistage amplifiers there are several types of feedback that can cause regeneration: improper shielding of input from output leads; improper decoupling of heater, plate-supply, and gain-control circuits; "waveguide" effect; and parasitic oscillations. It is often difficult to determine which of these is the cause of regeneration because all show some of the same symptoms. For example, the presence of "hot" plate-supply leads is not necessarily due to improper decoupling in those leads but may be due to other types of regeneration. Thus the search for the cause and cure of regeneration is largely a trial-and-error process. Nevertheless there are a few helpful remarks that can be made.

The location of components involved in a feedback path can often be determined by exploring the amplifier chassis with a probe connected to a high-gain amplifier of the same frequency; a sharp increase in output deflection from one heater terminal to another, for example, is cause for suspicion.

Coupling between Input and Output Leads.—The cause and cure of regeneration due to insufficient shielding of the input and output leads is more or less obvious. One of the most frequent instances is that in which the input lead is connected through a coaxial cable having its outer conductor improperly grounded. To avoid difficulties the outer conductor should be rolled back in a radially symmetrical way and soldered to the chassis immediately at the point of entry.

In general, regeneration caused by coupling between input and output leads can be prevented by the use of adequate shielding at either the

input or output leads, but it is much more desirable to place the shielding on the input lead because that reduces the pickup of extraneous signals, noise from rotating electrical machinery, and so forth, as well as reducing regeneration.

Heater Circuits.—Regeneration due to improper decoupling in the heater circuit can be detected by observing the pass band with a swept-frequency signal generator and momentarily disconnecting the heaters of one or more tubes from the heater-supply voltage. There should be no discontinuous change in bandpass shape at the instant the disconnection is made—only the continuous change that is due to the cooling down of the tubes. The cure can be effected by improving the heater decoupling.

"Waveguide" Feedback.—Waveguide type of feedback is due to the electromagnetic field set up by one of the interstage-coupling coils in a stage at the high-gain end of the amplifier and propagated down the amplifier box, which acts as a waveguide "beyond cutoff," to a coil located in one of the earlier stages. Its presence can be detected by connecting a wire or placing a post or metal block so that it makes electrical contact with opposite sides of the chassis and noting whether or not the symptoms of regeneration are reduced in magnitude. There are many possible modes, of which one is transmitted through the amplifier box at a loss of about 30 db for a length of the box equal to its width. Hence if the amplifier were 2 in. wide, there would be only 15 db/in. of attenuation, and a 105-db amplifier 7 in. long would be extremely risky. The prevention of this type of feedback is mainly in the design, namely, to mount the amplifier in a chassis as long and as narrow as possible. Further prevention can be effected by placing posts or baffles in the chassis at various points selected by experiment. These should make good electrical contact to both the top and bottom of the chassis.

It should be pointed out that if the frequency and gain per stage are both very large, the amplifier may be unstable because of grid-plate feedback. Also there is a limit to the gain per stage that can be obtained because of waveguide feedback, since the tube-socket diameter sets a practical lower boundary to chassis width. If either of these limitations is being exceeded, then it will be necessary to redesign the amplifier to employ more stages so as to obtain the desired total gain without excessive gain per stage. If a narrow over-all bandwidth is desired under these circumstances, it will be necessary to add capacity to one or more of the interstage-coupling networks.

If none of the previous procedures uncovers the cause of regeneration, the cause can usually be ascribed to improper decoupling of the plate-supply or gain-control leads or to parasitic oscillations.

Decoupling Circuits.—The crucial component in plate-supply, gain-control, and heater leads and cathode bypass circuits is the bypass capacitor.

At low frequencies the inductance in the leads of a bypass capacitor may be neglected, but at 30 or 60 Mc/sec or higher the lead inductance is very important. The inductive reactance of 1 in. of No. 20 wire is 4 ohms at 30 Mc/sec and 8 ohms at 60 Mc/sec.

If at all possible, it is very wise to make use of the inductance of the bypass capacitor leads to achieve series-resonant bypasses. For a 2000-$\mu\mu$f bypass capacitor at 60 Mc/sec (see Fig. 6·20) a $\frac{1}{8}$-in. lead is necessary.

If the capacitor Q is infinite, the impedance to ground at 60 Mc/sec is 0 ohm; if the Q is only 5, the impedance at 60 Mc/sec is only 0.25 ohm. At frequencies off resonance the effect of a finite Q is even less important. Thus at either 50 or 72 Mc/sec the impedance is about 0.5 ohm for $Q = \infty$ and 10 per cent higher for $Q = 5$. Therefore, there is no advantage in high-Q bypass capacitors; as a matter of fact, there is considerable advantage in low-Q bypass capacitors, in order to reduce the likelihood of parasitic oscillations (see below).

There is considerable point, however, in having a high C/L ratio. If a 500-$\mu\mu$f bypass capacitor is series-resonated at 60 Mc/sec, the range over which its impedance is less than about 0.5 ohm extends from 57 to 63 Mc/sec; for a series-resonant 2000-μf capacitor the corresponding range is 50 to 72 Mc/sec. In the latter case, variability of as much as ± 20 per cent in either capacity or lead length still results in extremely effective bypassing.

Parasitics.—Parasitic oscillations are well known to workers with power amplifiers and oscillators. They are oscillations that occur at frequencies lying outside the normal pass band of the amplifier.

Three methods are useful in locating parasitic oscillations. In the first method, use is made of an absorption wavemeter, a grid-dip meter, or other high-Q circuit that may be tuned over the band of frequencies in which the oscillation is expected to be. The amplifier pass band is presented on the face of a cathode-ray tube, using the usual swept-frequency-generator technique, and the wavemeter is tuned over its band while held close to the components of one of the stages of the amplifier. When the oscillation is located both in frequency and in position within the amplifier, a distinct change will be noted in the shape of the pass band or in the grid-dip meter deflection.

Another method is to listen for the parasitic oscillations with a variable-frequency communications receiver.

The third method is to observe the value of the current in the plate-supply lead as the plate voltage is raised from a low value. There will

be observed a discontinuous change in current as the voltage at which the oscillations start is reached.

In some cases, the effects of parasitics show up directly on the cathode-ray tube in the swept-frequency-generator test. The feedback coupling may not be adequate for self-sustained parasitic oscillations in the quiescent amplifier; but when a signal is applied to the input terminals, the grids of the various tubes are swung in the positive direction at the crest of each cycle of signal frequency. This swing causes a momentary increase in tube g_m that may result in the inception of parasitic oscillations, and these oscillations will last until the positive swing has disappeared. This effect shows up as a fuzzy, diffuse region on the cathode-ray-tube face at such frequencies as cause the effect.

A low-frequency parasitic ("motorboating") is generally due to an unfortunate combination of the values of the components in the decoupling networks associated with the input and output circuits of an amplifier and can be cured by changing the values of these components or by introducing damping into these networks. High-frequency parasitic oscillations can be cured by changing the location and length of leads. High-frequency parasitics (500 to 600 Mc/sec) observed in some amplifiers built in the Radiation Laboratory were cured by connecting 10-ohm composition resistors between the screen-grid socket terminals and the plate-supply bypass condensers.

8·6. Pulse Response.—The pulse response of amplifiers is best checked directly by use of pulsed signal generators. The types of input and output connections used are the same as for c-w measurements and have been covered in Sec. 8·4. Suitable pulsed signal generators and the synchroscopes used for observation of the output signal have been discussed in Secs. 8·2 and 8·3. This section is concerned with a discussion of the types of tests that are made and the interpretation of the observations. It is assumed that the shape of the output pulse from the signal generator is essentially perfect and that the speed and flatness of top of the measuring equipment substantially exceeds that of the amplifier under test.

The amplifier under test may be of either the low-pass or bandpass type; direct pulses are used in the first case and carrier-frequency pulses in the second.

In general, it is necessary to reproduce signals of two types: short single pulses separated by long intervals during which no signals appear and long single pulses separated by either long or short intervals. In the latter case, one is concerned with the duty ratio, which is the fraction of the time during which signals are present.

Single Short Pulses.—In making tests with short recurrent pulses, the following points may be observed: (1) the distortion in pulse shape

due to its transmission through the amplifier, (2) the effect of input signal amplitude on the output-pulse shape and amplitude, and (3) the aftereffects of a pulse strong enough to cause amplifier overloading.

The main discussion of pulse response was contained in Chaps. 2 and 7. Pulse distortion due to certain specific causes is illustrated in Fig. 8·14. For carrier-frequency pulses the absolute values of the shapes shown in Fig. 8·14 are to be regarded as those of the carrier-frequency envelopes (see Fig. 7·1). If the pulse shows excessive rise and fall times (Fig. 8·14b), the amplifier over-all bandwidth is inadequate; the narrowing may be either in the bandpass or the low-pass stages or both. If the rise time is short and the fall time long (Fig. 8·14c), the trouble is due to a "polarized time constant." Such an effect is produced, for example, when a capacity is charged from a low-impedance

(a)   (b)   (c)

(d)   (e)   (f)
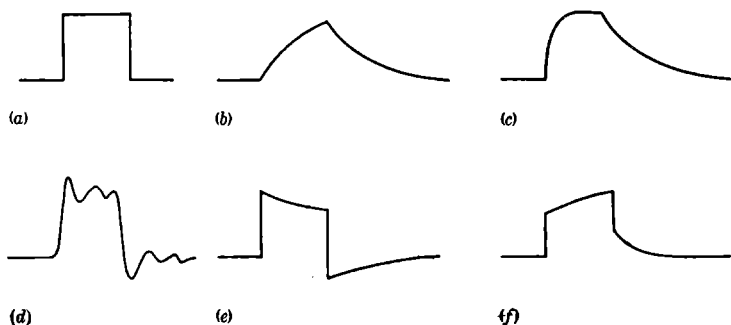
Fig. 8·14.—Effects of amplifier characteristics on output-pulse shape.

source through a diode that conducts during the leading edge of the pulse, whereas the diode no longer conducts during the trailing edge of the pulse; therefore the capacity discharges slowly through a high resistance. Another situation that gives rise to the same effect occurs with large pulses transmitted through a cathode follower, as described in Chap. 2; the positive-going edge is faster than the negative-going edge. In some applications such a "polarized time constant," or "pulse stretcher," is deliberately introduced, but generally it is to be avoided.

Overshoot and ringing (Fig. 8·14d) occur if the pass band of a band-pass amplifier has too square a shape. Similar effects may occur from a dip in the center, due to regeneration (see Sec. 8·5) or to overcoupling the transformers in amplifiers using double-tuned coupling circuits or to overstaggering the off-frequency stages in a stagger-tuned amplifier (see Sec. 7·6).

In a low-pass amplifier, overshoot is generally due to overpeaking. Readjustment of the values of peaking inductances is the general cure.

If both types of amplifier stages are included in the amplifier under test, separate observations should be made on the bandpass and low-pass sections to locate the cause of the trouble.

Another cause of a pulse shape similar to that shown in Fig. 8·14d is an improperly terminated transmission line, the "wiggles" being due to reflections; a place in which this trouble frequently occurs is in the line connecting the synchroscope to the amplifier under test.

The problem of flatness of top does not ordinarily exist for single short pulses.

Single Long Pulses.—A frequently observed effect with long pulses is illustrated in Fig. 8·14e. The downward slope of the top of the pulse and the depression of the baseline after the end of the pulse is generally due to lack of d-c transmission in the pulse amplifier. It can also be produced by poor power-supply regulation when tubes are operated beyond their linear regions with a large duty ratio; this latter effect can easily be detected by connecting the synchroscope to the power-supply leads. A large group of strong short pulses causes a similar depression of the baseline. This effect is very undesirable, because it results in a loss of

FIG. 8·15.—Response of bandpass amplifier to pulse of substantially detuned carrier frequency.

sensitivity to weak signals for a period after the completion of the long pulse or the group of short strong pulses.

Figure 8·14f shows the effects of overcompensating for flat top (see Sec. 2·3).

An important advantage of bandpass amplifiers over video amplifiers is that apart from overload effects leading to a loss in gain, the envelope of a carrier-frequency pulse can never display the lack of flat top shown in Fig. 8·14e. The reason is that bandpass amplifiers do actually transmit the carrier-frequency component, whereas video amplifiers (unless direct coupled) do not transmit the zero frequency component.

Pulses of a carrier frequency substantially different from the center frequency of a bandpass amplifier are reproduced as shown in Fig. 8·15, the flat portion indicating the steady-state transmission of the pulse frequency. Figure 8·15 shows the envelope of the response.

8·7. Overload and "Blackout" Effects.—Amplifiers used with pulses sometimes show a number of undesirable overload effects.

The amplifier may show a loss of sensitivity after a very strong pulse of short duration, frequently indicated by the reduction in the noise output voltage. In some cases the recovery of sensitivity may require a time that is long compared with the interval between pulses, in which case the amplifier is "blocked." The most frequent cause of

this effect is the flow of grid current, which causes a charge to be collected upon a grid coupling condenser. If there is a large time constant associated with the grid circuit, this charge requires a long time to leak away, during which time the gain is reduced (see Chap. 3). The cure is obviously (1) if possible, to supply sufficient negative bias to reduce the grid current and (2) to make the time constant in the grid circuit short compared with the pulse length. This difficulty is frequently encountered in a grid-bias gain-control circuit; to avoid this difficulty, variable bias voltage for gain control of a bandpass amplifier should be supplied from a low-resistance source (2000 ohms at the most) even though the grids normally draw no current.

Another cause of a loss of sensitivity after a strong short pulse is the "blackout" effect; this is the temporary loss in transconductance of an amplifier tube after an overloading pulse. This effect is extremely variable from tube to tube of the same type and even of the same manufacturer and is not fully understood. One explanation that has been proposed is that there are small particles of dielectric on the surface of the control grid, perhaps driven from the cathode emitting surface, which become charged during the pulse and require a finite time to discharge, during which time the tube transconductance is reduced. This effect can usually be distinguished from grid-current effects because it will be found to vary considerably with substitution of tubes, whereas with grid current little change will be noted from one tube to another. Blackout effects are most likely to occur in plate detectors but can be observed in high-level bandpass-amplifier stages as well. The best cure is the substitution of tubes, if permissible. The effect can be reduced by operating the tube with an unbypassed cathode resistor; this has the effect of reducing the cathode-to-grid signal potential but sacrifices gain.

8·8. Measurement of Gain and Determination of Amplifier Law. *Gain.*—The measurement of gain is, in principle, very simple. A signal generator, which may be of any type (c-w, swept-frequency, or pulsed), is connected through an attenuator to the input terminals of an amplifier, and an appropriate detector or indicator (which need not be calibrated) is connected to the output terminals. The signal level s adjusted to give a convenient deflection on the indicating instrument; then the signal generator and attenuator are connected directly to the indicating device and are adjusted a second time to give the same deflection. The gain is given directly by the ratio of attenuator settings.

Unfortunately, however, in practice the measurement of gain involves a much more complicated procedure, and under most circumstances it is difficult to get an accurate value. The calibration of signal generators and attenuators assumes that they are terminated in some definite

impedance.   It is very rare that either the input terminals of the amplifier or the indicating device offer this impedance.   Also it may be necessary to employ matching networks or "dummy-input circuits" to make the signal generator present the correct signal-source impedance to the amplifier.   To allow for all these effects, the signal level measurements must be supplemented by numerous impedance measurements and calculations.

When the impedance levels are high, as is normally the case at the input or output terminals of a vacuum tube, the measurements become simple.   An artificial impedance equal to the correct terminating impedance is connected across the output terminals of the attenuator, both when the signal generator and attenuator are connected to the input terminals of the amplifier and when they are connected to the indicating device.   Almost always this impedance is so low that the effect of the impedances which are present in parallel with it (such as plate-load resistors) can be neglected.   Then the voltage gain is given directly from the ratio of the two attenuator settings, as indicated previously.   The power gain can be found by multiplying the square of the voltage gain by the impedance ratio of the normally present input and output impedances.

Production-line Test for Gain.—There are many applications where an accurate measurement of gain is unnecessary.   In the manufacture of amplifiers it is necessary merely to be sure that the amplifiers coming off the production line have adequate gain for the application for which they are intended.   In such cases a prototype amplifier exists that has been tested in its intended application and is known to have adequate gain.   The test of the other amplifiers consists of determining whether they have more or less gain than the prototype.   This test can be effected by applying a standard signal and noting whether the output is greater or smaller than with the prototype.

In many applications the required amount of gain is such that the weakest detectable signals will have signal strengths comparable to the noise generated within the amplifier.   Therefore it is necessary that the amplifier have sufficient gain to make the noise easily perceptible in the indicating device.   In such cases the noise generated within an amplifier can itself be used as the "standard signal" for testing gain.

The procedure is first to determine that the production amplifiers contain no improper sources of noise by measuring their noise figures (see Chap. 14).   Second, some sort of output indicating device that will give an indication of noise power output, such as an oscilloscope, is connected to the output terminals.   Those amplifiers which give as great or greater deflection on the output device as the prototype have adequate gain.

Amplifier, or Detector, Law.—The "amplifier law" is the name given to the relation between output and input voltage. In the event that the amplifier contains a detector, this relation is more properly called a detector law, because most of whatever nonlinearity exists is usually due to the detector. It is essential to have accurate knowledge of the amplifier or detector law if the amplifier is to be used for the precise measurement of signal amplitudes. A measurement of the law is also important as a design check on certain amplifiers, especially those which are to be part of radio receivers. If the law is nonlinear, undesirable cross modulation will be produced in the presence of interfering signals. Also the "saturation" or "limiting" of the output signal at too low signal-input levels is an indication of poor overload capability.

The amplifier, or detector, law can be determined with the aid of a signal generator, attenuator, and output indicating device, as in the gain measurements, except that the indicating device should be calibrated in terms of voltage, current, or power. The readings of the output device for various positions of the attenuator are noted. It is convenient to make a plot of the logarithm of the output reading against the setting of the attenuator (expressed in decibels), and from this it can easily be seen whether the law is a simple power law or not. At very high gain it is impossible to make an accurate determination of the law because the signal is masked by noise developed by the amplifier.

CHAPTER 9

LOW-FREQUENCY AMPLIFIERS WITH STABILIZED GAIN

By Duncan MacRae, Jr.

9·1. Problems Characteristic of Computer Amplifiers.—The design of many electronic computing devices involves the representation of functions by the magnitudes of alternating voltages. In these computers, amplifiers are needed either to transform the output and input impedances of the different computing elements within the device or as computing elements themselves. They can be used for the latter purpose for example, to multiply a particular function by an arbitrary and sometimes variable factor.

The gain of such an amplifier must be held constant within closely prescribed limits regardless of the variability of vacuum-tube parameters, the manufacturing tolerances of passive components such as resistors and capacitors, possible variations of the ambient temperature, etc. The variation in gain that can be tolerated depends upon the desired accuracy of computation (the figure of ± 0.1 per cent was usually specified by the Radiation Laboratory).

Economy requires that the desired constancy of gain be achieved without using circuit components whose characteristics must themselves be maintained within narrow tolerances. The specified tolerances for resistors and condensers, for instance, are usually not narrower than 5 and 10 per cent respectively, although in some obvious cases resistance values must be more closely specified.

Inverse feedback can be employed to render the gain of computer amplifiers less sensitive to the variations in the values of the circuit parameters. Variations in gain due to changes in tube parameters can be reduced by local feedback methods such as the use of unbypassed cathode resistors or d-c feedback from plate to grid (see Chap. 3). The principal design problems, then, derive from the fact that the feedback circuit requires high gain, which tends to make the amplifier oscillate at frequencies often far removed from the frequency band that it is designed to amplify.[1]

The decision as to the number of stages of amplification necessary to obtain the desired reduction of gain sensitivity to component varia-

[1] H. W. Bode, "Relations between Attenuation and Phase in Feedback Amplifier Design," *BSTJ*, **19**, 421, July 1940.

bility is determined not only by the general requirements of the system but also by the tolerances of the circuit components, the gain attainable from each stage, the bandpass requirements set by the feedback circuit, and the polarity of feedback, which determines for certain feedback circuits whether the number of stages should be even or odd.

The following criteria determine the choice of the particular feedback circuit:

1. The fraction β of the output signal which is to be fed back at the computing frequency must be maintained within the prescribed tolerance.
2. The circuit shall require the minimum number of precision components (e.g., ± 1 per cent resistors).
3. Preferably there should be no need for an amplitude control to compensate for the variability of any of the components used in the amplifier; if such a control is necessary, its required range of adjustment should be kept small in order to minimize effects of maladjustment.
4. The feedback circuit used should cause the amplifier to have the desired input and output impedances.

The choice of feedback methods for computer amplifiers is more restricted than for audio amplifiers or for most filter amplifiers because of the small permissible variability of β; the feedback circuit therefore cannot usually involve tubes. The simplest type of feedback circuit satisfying this requirement is that for which $|\beta| = 1$, that is, a circuit in which the output voltage may either be subtracted from the input voltage and the difference applied to the first grid or be applied directly to the first cathode. Possible subtracting methods include resistive mixing (with an odd number of stages) and subtraction of output voltage from input voltage by means of a transformer. Feedback to the first cathode is most easily accomplished when the load is inductive; if the load is resistive, cathode feedback can be accomplished by the use of an auxiliary inductance to provide a d-c return for the first cathode, or by the use of a transformer. By modifications of these methods, any desired fraction of the output voltage can be fed back to the input terminals: In Circuit a of Fig. 9·1 (resistive mixing) the resistors R_1 and R_2 can be varied; in Circuit b the transformer T can be designed to have any desired ratio, or a potentiometer can be incorporated so as to add a variable fraction of the output voltage to the input voltage; in Circuits c and d an extra tap can be designed into the inductance or transformer. For relatively low-precision applications, each of these circuits can be reduced to a single-stage amplifier: Circuit c will simply become a cathode follower if the amplifier "box" is made an open

circuit; in other cases the amplifier is made a short-circuit or zero-stage amplifier.
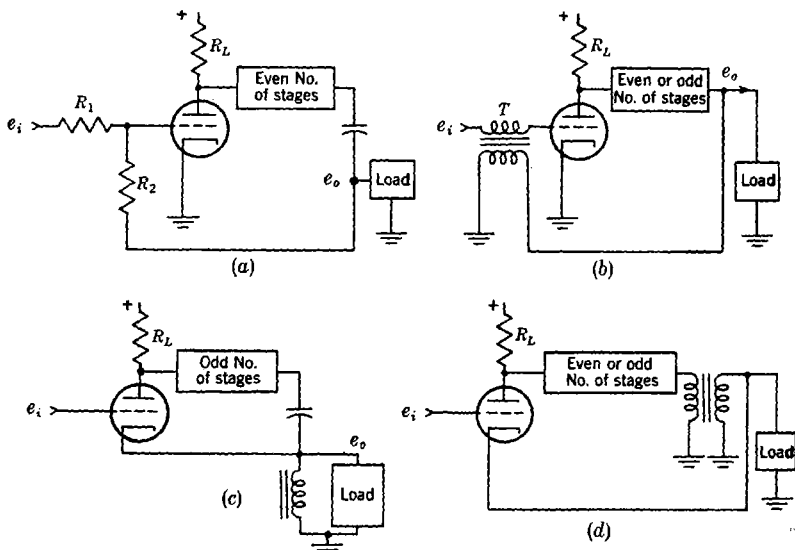
Fig. 9·1.—Feedback circuits. (a) Resistive mixing at grid; (b) transformer subtraction at grid; (c) cathode feedback with auxiliary inductance; (d) cathode feedback with transformer.

9·2. Analysis of Types of Feedback. *Resistive Mixing.*—In the analysis of the resistive-mixing type of feedback circuit (Fig. 9·1a), it is assumed that the source is a generator of voltage e_i having an output impedance Z_s, that the amplifier can be replaced by a generator $-\alpha e_g$ having an output impedance Z_0 in the absence of feedback, and that the first-stage grid is connected to ground through an impedance Z_g (Fig. 9·2). The gain and the output impedance of this circuit can be calculated by writing the circuit equations with

Fig. 9·2.—Schematic circuit for resistive-mixing feedback.

e_i and i as the two independent variables. The gain will then be $\partial e_0 / \partial e_i$ and the output impedance will be $\partial e_0 / \partial i$. Because the equations are linear, these quantities are respectively $(e_0/e_i)_{i=0}$ and $(e_0/i)_{e_i=0}$. Using nodal equations,

$$-\left(\frac{1}{R_2}\right)e_0 + \left(\frac{1}{R_1 + Z_s} + \frac{1}{R_2} + \frac{1}{Z_g}\right)e_g = \frac{e_i}{R_1 + Z_s}, \qquad (1a)$$

and

$$\left(\frac{1}{R_2} + \frac{1}{Z_0}\right) e_0 - \left(\frac{1}{R_2} - \frac{\alpha}{Z_0}\right) e_g = i. \qquad (1b)$$

Then the voltage gain is

$$\mathcal{G} = \left(\frac{e_0}{e_i}\right)_{i=0} = \frac{(-\alpha R_2 + Z_0)}{(1 + \alpha)(R_1 + Z_s) + \left[\dfrac{(R_1 + Z_s)}{Z_g} + 1\right](R_2 + Z_0)}. \qquad (2)$$

Now if

$$Z_g \gtreqqless R_1 + Z_s,$$
$$R_2 + Z_0 \lesseqqgtr R_1 + Z_s,$$

and $\alpha \gg 1$,

$$\mathcal{G} = \frac{-\alpha R_2 + Z_0}{\alpha(R_1 + Z_s)} = -\frac{R_2}{R_1 + Z_s} + \frac{Z_0}{\alpha(R_1 + Z_s)}. \qquad (3)$$

But usually $R_1 + Z_s \gtreqqless Z_0$; and since $\alpha \gg 1$, Eq. (3) becomes, to a very good approximation,

$$\mathcal{G} = -\frac{R_2}{R_1 + Z_s}.$$

Then

$$\frac{d\mathcal{G}}{\mathcal{G}} = \frac{-dR_2}{R_2} + \frac{dR_1}{R_1}\left(\frac{R_1}{R_1 + Z_s}\right) + \frac{dZ_s}{Z_s}\left(\frac{Z_s}{R_1 + Z_s}\right), \qquad (4)$$

which makes it evident that in order for variations in source impedance to have little effect on the gain of the amplifier, Z_s must be very much less than R_1; thereupon fractional variations in R_1 and R_2 affect the gain equally.

The output impedance is given by the following equation:

$$Z_{\text{out}} = \left(\frac{e_0}{i}\right)_{e_i=0} = \frac{\left(\dfrac{1}{R_1 + Z_s} + \dfrac{1}{R_2} + \dfrac{1}{Z_g}\right)}{\left(\dfrac{1}{R_2} + \dfrac{1}{Z_0}\right)\left(\dfrac{1}{R_1 + Z_s} + \dfrac{1}{R_2} + \dfrac{1}{Z_g}\right) - \dfrac{1}{R_2}\left(\dfrac{1}{R_2} - \dfrac{\alpha}{Z_0}\right)}. \qquad (5)$$

As α grows indefinitely, and if

$$Z_s \ll R_1,$$
$$Z_g \gg R_1,$$
$$Z_{\text{out}} \approx \frac{(R_1 + R_2)Z_0}{\alpha R_1}, \qquad (6)$$

which is simply Z_0 divided by the loop gain. A reduction of output impedance is a general property of voltage-feedback amplifiers.

Cathode Feedback.—When cathode feedback is used, the sensitivity of the over-all gain $\mathcal{G}$ to the μ of the first stage cannot be made appreciably

less than $1/\mu$, i.e., $d \ln \mathcal{G}/d \ln \mu \gtreqqless 1/\mu$. To prove this, assume the operating characteristics of the first stage of the amplifier shown in Fig. 9·1c to be given by the usual relation

$$i_p = g_m \left(e_{gk} + \frac{e_{pk}}{\mu} \right) \tag{7}$$

and the remaining stages to have a gain of $- \mathcal{G}_b$.   Then

$$e_{pk} = - R_L i_p - e_k, \tag{8a}$$

where e_{pk} = a-c plate-to-cathode voltage,
   e_k = a-c cathode-to-ground voltage;

$$e_{gk} = e_i - e_k, \tag{8b}$$

where e_{gk} = a-c grid-to-cathode voltage,
   e_i = input signal voltage;

$$e_k = \beta e_0 = \beta R_L i_p \mathcal{G}_b, \tag{8c}$$

where e_0 = output signal voltage,
   β = feedback ratio,
   R_L = first-stage load resistance.
Combining Eqs. (8) with Eq. (7),

$$e_i = i_p \left[\frac{1}{g_m} + \frac{R_L}{\mu} + \beta R_L \mathcal{G}_b \left(1 + \frac{1}{\mu} \right) \right].$$

But the over-all gain is given by

$$\mathcal{G} = \frac{e_0}{e_i} \approx \frac{R_L i_p \mathcal{G}_b}{e_i},$$

and therefore,

$$\frac{1}{\mathcal{G}} = \frac{e_i}{R_L i_p \mathcal{G}_b} = \frac{1}{g_m R_2 \mathcal{G}_b} + \frac{1}{\mu \mathcal{G}_b} + \beta \left(1 + \frac{1}{\mu} \right). \tag{9}$$

Now $\mathcal{G}_b$ will ordinarily be made very large in order to render the effect of component variations on $\mathcal{G}$ negligible.   In this case Eq. (9) may be written as

$$\frac{1}{\mathcal{G}} = +\beta \left(1 + \frac{1}{\mu} \right) \tag{10}$$

Inverting and differentiating, it is found that $d\mathcal{G}/d\mu = 1/\beta(1 + \mu)^2$; hence

$$\frac{d \ln \mathcal{G}}{d \ln \mu} = \frac{\mu}{\mathcal{G}} \frac{d\mathcal{G}}{d\mu} = \frac{\mu}{\beta \mathcal{G}(1 + \mu)^2}. \tag{11}$$

It is evident that this expression becomes smaller as μ becomes larger; now if μ is large, $\mu \approx \mu + 1$.   Then from Eq. (10) $\mathcal{G} \approx 1/\beta$,

where $\mathcal{G}$ is the over-all gain. Substituting these approximations in Eq. (11) and noting the conditions under which they are made,

$$\frac{d \ln \mathcal{G}}{d \ln \mu} \gtreqless \frac{1}{\mu}. \tag{12}$$

Therefore, in order that the gain may be protected against variations of the amplification factor of the first stage, the latter must be as large as possible.

The input impedance of a cathode-feedback amplifier can be calculated with the aid of the preceding analysis. The predominant contribution to the input impedance is that resulting from capacitive coupling of the grid to the plate and cathode. The input impedance is defined as the ratio of grid voltage to total current flowing to the grid node. The total current is the sum of the currents in the capacitances C_{gp} and C_{gk}. These currents are determined by the alternating voltages at the plate and cathode as well as by the input voltage e_i. The voltage differences e_{gp} and e_{gk} have the effect of changing the apparent input capacitance; since the currents in the two branches are in parallel, it is possible to write

$$C_{in} = C_{gp} \frac{e_{gp}}{e_i} + C_{gk} \frac{e_{gk}}{e_i}. \tag{13}$$

Now

$$e_{gp} = e_g - e_p = e_i - e_p,$$

since $e_i = e_g$ by definition. Then

$$\frac{e_{gp}}{e_i} = 1 - \frac{e_p}{e_i}.$$

But

$$e_0 = e_i \mathcal{G} = -e_p \mathcal{G}_b,$$

whence

$$\frac{e_{gp}}{e_i} = 1 + \frac{\mathcal{G}}{\mathcal{G}_b},$$

where $\mathcal{G}$ and $\mathcal{G}_b$ are related by Eq. (9). Similarly

$$e_{gk} = e_g - e_k = e_i - e_k,$$

and

$$\frac{e_{gk}}{e_i} = 1 - \frac{e_k}{e_i} = 1 - \beta\mathcal{G},$$

since $e_k = \beta\mathcal{G}e_i$. Therefore,

$$C_{in} = C_{gp}\left(1 + \frac{\mathcal{G}}{\mathcal{G}_b}\right) + C_{gk}(1 - \beta\mathcal{G}). \tag{14}$$

In deriving Eq. (12) it was assumed that $\mathcal{G}_b \gg 1$, from which it followed that $\beta\mathcal{G} \approx 1$. Because this is an inverse-feedback amplifier, $\mathcal{G}_b > \mathcal{G}$. Moreover, if $\mathcal{G}_b \gg \mathcal{G}$, then

$$C_{in} = C_{gp}. \tag{15}$$

Comparison of Grid and Cathode Feedback.—Certain distinctions may be made between grid and cathode feedback. If the load is resistive, the method shown in Fig. 9·1a of resistive mixing at the grid has the advantage of light weight, for it requires no additional iron-core parts. The input impedance is nearly equal to R_1 because the high gain of the amplifier requires that the alternating voltage applied to the first grid remain small. On the other hand, the input impedance of the circuit in Fig. 9·1b will be approximately equal to the impedance of the inter-winding capacitance of the transformer. The input capacitances of the circuits in Figs. 9·1c and 9·1d are low, approximately equal to C_{gp}. The output impedance of each of these circuits is roughly equal to the output impedance of the output stage divided by the loop gain.

If the load is an inductance, Method c of Fig. 9·1 is probably preferable to Method a because precision resistors are not required. Circuits b and d have an advantage from the standpoint of low power drain on the output stage, because the use of a transformer makes possible push-pull operation, with a corresponding increase in efficiency.

9·3. The Stability Problem.—Certain phase and amplitude requirements must be satisfied by a feedback amplifier in order that it shall not oscillate. These requirements give rise to the problem of synthesizing networks that possess special characteristics, subject to practical limitations imposed by the resistances, condensers, tubes, and wiring. At sufficiently low frequencies a 90° phase lead is introduced by each coupling capacitor and the associated grid resistor; at high frequencies a 90° lag is associated with each stage, determined by its input capacitance and the output resistance of the preceding stage. The design objective is to cause the loop gain $\mathcal{G}\beta$ to decrease to less than unity at frequencies between which the phase shift does not exceed 180°. For a one-stage amplifier the problem is not difficult unless there are elements in the circuit producing a phase shift of 180° or more (such as a transformer plus a coupling capacitor). The two-stage amplifiers that can be made from the feedback circuits of Figs. 9·1b and 9·1d involve transformers that may introduce 180° phase shift at high frequencies; if the circuit of Fig. 9·1c is used, there are also two coupling networks affecting the low-frequency response. Three-stage RC-coupled amplifiers involve at least three coupling networks, each of which will give 90° phase shift at low frequencies, as well as three networks having an analogous effect at high frequencies (the output resistances and the shunt capacitance

of following stages). The limiting transfer characteristic (in both phase and amplitude) determined by these networks is known as the asymptotic characteristic.

The easiest way to satisfy the stability condition at the low-frequency side of the pass band, for a two-stage or three-stage amplifier, is to let one or two coupling networks attenuate at 6 db/octave down to the frequency at which the loop gain is unity or less than unity by the desired amplitude margin.[1] If it is possible to use sufficiently large resistances and capacitances, the other coupling networks may be made to introduce very little phase shift above this frequency and to have almost all their phase-shifting effect at frequencies where the loop gain is less than unity. The analogous procedure for the high-frequency
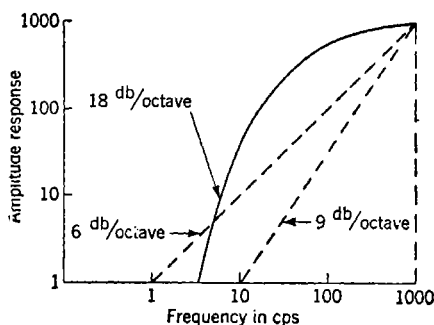
Fig. 9·3.—Low-frequency asymptotic characteristics. Solid line represents possible asymptotic characteristic for a three-stage amplifier.

response is again to make one stage attenuate so that unity gain is reached before the other stages have any effect. This might be achieved, for instance, by putting an additional capacitance in parallel with one of the tube input capacitances. This method cannot be used, however, if the necessary loop gain is too high relative to the ratio of the frequency at which the amplifier is designed to operate and the frequency at which the asymptotic characteristic has unity gain (see Fig. 9·3). In the example shown in Fig. 9·3 the computing frequency is 1000 cps, and the gain at this frequency is 1000 (60 db). If the designer tried to make the amplifier attenuate at 6 db/octave, the effect of the asymptotic characteristic as determined by the inherent properties of the amplifier components would predominate before unity gain was reached and oscillation would probably result. If he could produce a slope of 9 db/octave, however, it might be possible to reach unity gain without having 180° phase shift. Roughly, 6 db/octave corresponds to 90°,

[1] F. E. Terman, *Radio Engineers' Handbook*, McGraw-Hill, New York, 1943, p. 398.

12 db/octave to 180°, etc.; in order to afford a margin of safety, there-
fore, the slope of the attenuation curve must be somewhat less than
12 db/octave.   In order to obtain slopes between 6 and 12 db/octave,
networks more elaborate than single *RC*-coupling networks can be
used.

The characteristics attainable are sharply limited if light weight is
required, for this means that neither condensers larger than some physi-
cally small size such as 0.1 μf nor heavy choke coils are permissible in
the low-frequency networks.
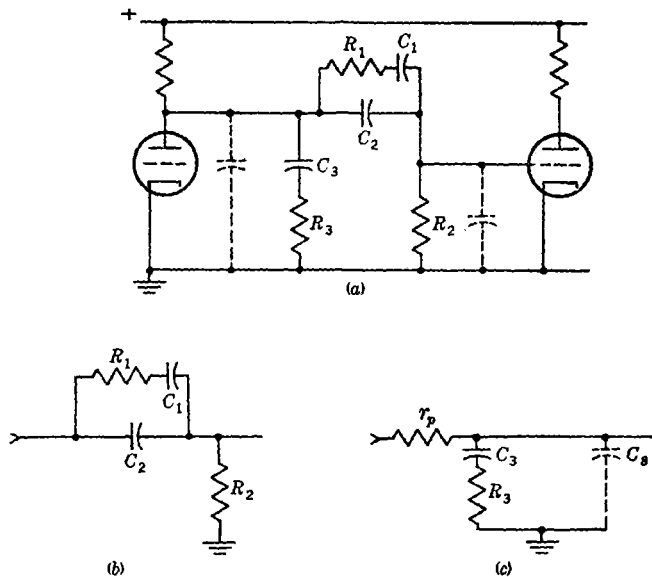
FIG. 9·4.—Typical coupling network.   (*a*) Coupling network; (*b*) low-frequency equivalent
circuit; (*c*) high-frequency equivalent circuit.

The principal topic to be developed here is the stabilization of three-
stage amplifiers by the use of coupling networks similar to those shown
in Fig. 9·4.   A good example of this is given by Bode.[1]   The method
used in this example emphasizes the amplitude response, an easily
measurable function, and the calculation of the phase response from it.
An alternative method, which has advantages in certain special cases,
employs the direct calculation of phase response as a function of fre-
quency, using the amplitude response at only a few points.   The cal-
culation of the phase response of resistance-capacitance networks is
considerably easier than the calculation of the amplitude response.   It

[1] H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand,
New York, 1945, pp. 514–517.

is particularly helpful at low frequencies, where the values of the parameters involved are fairly well determined; at high frequencies the values of stray and input capacitances are less well known before the layout is made, and more experimental work has to be done to complete a design.

The equivalent high-frequency and low-frequency stabilizing networks used (Fig. 9·4) may be considered together. The property of these networks that makes them useful is that each may have a "plateau" in the curve of phase shift vs. frequency. There are other RC-networks having this property, but all those using only two "independent" condensers[1] may be treated mathematically in the same way. If it is assumed that these networks include the effect of the output impedance of the preceding stage and the input impedance of the next stage, each network may be analyzed as a three-terminal network containing R's and C's and operating from zero impedance into infinite impedance. By means of the loop analysis, the ratio e_0/e_i may be expressed as a quotient of determinants. The denominator is a polynomial of the second order in $1/p$; the numerator cannot be of order higher than the second. Thus, the most general representation of e_0/e_i is

$$\frac{e_0}{e_i} = A\,\frac{p^2 + a_1 p + b_1}{p^2 + a_2 p + b_2} = A\,\frac{(p - p_1)(p - p_2)}{(p - p_3)(p - p_4)}, \tag{16}$$

where the roots $p_4 \ldots p_4$ are real quantities of dimension T^{-1} because it has been assumed that the networks contain only resistances and capacitors.

The physical significance of this relationship is that for certain real values of p (corresponding to exponentially varying voltages and currents) the ratio e_0/e_i can be either zero or infinite.[2] However, only the pure imaginary values of p are considered in finding the phase response of the network as a function of frequency. As a further simplification, it may be noted that the equivalent low-frequency network has zero response at zero frequency and unity response at infinite frequency; the equivalent high-frequency network behaves in exactly the opposite way, having zero response at infinite frequency and unity response at zero frequency. Thus, for the low-frequency network,

$$\frac{e_0}{e_i} = \frac{p(p - p_2)}{(p - p_3)(p - p_4)}, \tag{17a}$$

[1] Two condensers are "independent" if current loops can be so drawn that only one passes through each of the condensers. This ensures that the determinant of coefficients in the loop equations is a polynomial of the second order in $1/p$, since $1/p$ occurs twice on the principal diagonal.

[2] Bode, *op. cit.*, pp. 28–30.

and for the high-frequency network,

$$\frac{e_0}{e_i} = \frac{-p_3 p_4}{p_2} \frac{(p - p_2)}{(p - p_3)(p - p_4)}. \tag{17b}$$

Finally, setting $p = j\omega$ and making use of the fact that the phase angle of a product of complex factors is the algebraic sum of their separate phase angles, it is found that for the low-frequency network

$$\phi\left(\frac{e_0}{e_i}\right) = +\frac{\pi}{2} + \tan^{-1}\frac{\omega}{-p_2} - \tan^{-1}\frac{\omega}{-p_3} - \tan^{-1}\frac{\omega}{-p_4}, \tag{18a}$$

and for the high-frequency network

$$\phi\left(\frac{e_0}{e_i}\right) = \tan^{-1}\frac{\omega}{-p_2} - \tan^{-1}\frac{\omega}{-p_3} - \tan^{-1}\frac{\omega}{-p_4}. \tag{18b}$$

In computing the phase response of such a network, it is necessary to find the roots that characterize it and to use them in connection with a single inverse-tangent curve (Fig. 9·5). This curve gives directly the phase shift of a single RC coupling network. The over-all phase response of the network or of several such networks is the algebraic sum of these inverse-tangent curves. Changing the value of p_i is equivalent to adding a constant to $\log_{10} \omega$ and has the effect of moving the curve bodily along the $\log_{10} \omega$ axis without changing its shape. Any partic-

FIG. 9·5.—Inverse-tangent curve.

ular curve may be located by observing that at $\omega = p_i$, $\phi = 45°$. The corresponding frequency, at which the capacitive reactance of a single network is equal to its resistance, will be called the central frequency. The design objective is now to produce a desired total phase response; this is done by successive trials, as will be illustrated in Sec. 9·7.

Consider first the "phase-advance" network, so-called from its use in servomechanisms (see Fig. 9·6).

For this circuit,

$$\frac{e_0}{e_i} = \frac{R_2}{\dfrac{1}{(1/R_1) + pC} + R_2} = \frac{p + \dfrac{1}{CR_1}}{p + \dfrac{R_1 + R_2}{R_1 R_2 C}}. \tag{19}$$

Then the phase response of the circuit will be given by

$$\phi = \tan^{-1}(\omega R_1 C) - \tan^{-1}\left(\frac{\omega R_1 R_2 C}{R_1 + R_2}\right). \tag{20}$$
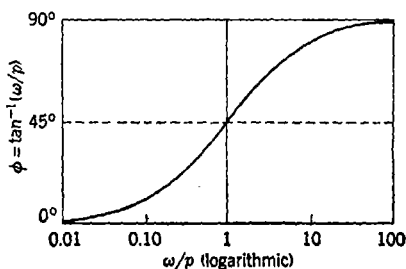
This expression as plotted in Fig. 9·6b has a maximum value at the logarithmic mean of the two roots $-p_1 = +1/CR_1$ and

$$-p_2 = +\frac{(R_1 + R_2)}{R_1 R_2 C}.$$

It may also be seen from Fig. 9·6b that the value of maximum phase shift is

$$2\left[\left(\tan^{-1}\sqrt{\frac{p_2}{p_1}}\right) - 45°\right] = 2\left[\left(\tan^{-1}\sqrt{1 + \frac{R_1}{R_2}}\right) - 45°\right]. \quad (21)$$
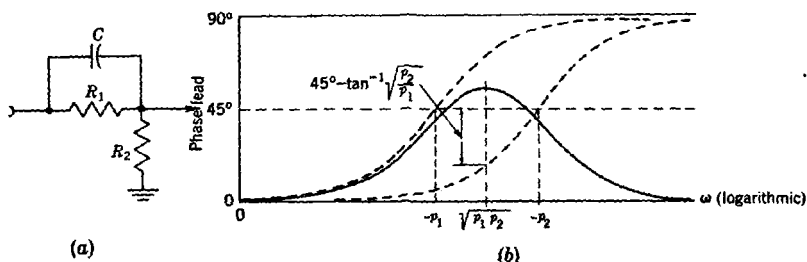
Fig. 9·6.—"Phase-advance" network.   (a) Network; (b) phase response as a function of frequency, $\frac{p_2}{p_1} = 10$; $-p_1 = \frac{1}{CR_1}$; $-p_2 = \frac{(R_1 + R_2)}{R_1 R_2 C}$; $\sqrt{p_1 p_2} = \frac{1}{CR_1}\sqrt{1 + \left(\frac{R_1}{R_2}\right)}$.

The maximum value of phase shift is equal to the difference of the two dotted curves at the center of the horizontal axis and is equal to twice the difference of either dotted curve from 45°. This result is generally useful in connection with linear RC-networks containing only one condenser. It may be noted that

$$\frac{p_2}{p_1} = \frac{\left(\dfrac{e_0}{e_i}\right)_{f=\infty}}{\left(\dfrac{e_0}{e_i}\right)_{f=0}}. \quad (22)$$

This equation may sometimes be used as a convenient way of determining p_2/p_1 in order to find the maximum phase shift; the right-hand side of the equation may be found from experimental gain measurements with and without the condenser (see Sec. 9·13).

A typical low-frequency stabilization network is shown in Fig. 9·7a. The frequency characteristic of this network can be explained qualitatively by considering the effects that occur as the impedance of one capacitance or the other becomes large with decreasing frequency. The usual design procedure is to let R_1 and C_1 be larger than R_2 and C_2 respectively. At the high-frequency end of the characteristic, R_1

has a considerably higher impedance than C_2, and the characteristic, determined largely by R_2 and C_2, starts as an inverse-tangent curve. As the frequency is decreased, the impedance of C_2 increases to the same order of magnitude as that of R_1. To the extent that the impedance of the section consisting of R_1, C_1, and C_2 approaches a real quantity in this region, there is a tendency for the phase shift to return to zero. Finally, as the impedance of C_1 rises to the order of magnitude of R_1, the response of the network approaches that of the part consisting of R_1, C_1, and R_2, this response being another inverse-tangent curve.
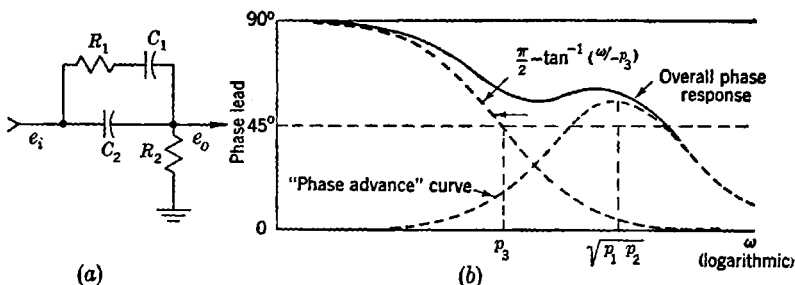
FIG. 9·7.—Low-frequency stabilization network. (a) Network; (b) phase response.
$$\frac{p_2}{p_1} = 10, \quad \sqrt{\frac{p_1 p_2}{p_3}} = 10.$$

The equations for this network can be written

$$\frac{e_0}{e_i} = \frac{R_2}{R_2 + \cfrac{1}{\cfrac{1}{R_1 + (1/pC_1)} + pC_2}}$$

$$= \frac{p\left(p + \dfrac{1}{R_1 C_1} + \dfrac{1}{R_1 C_2}\right)}{p^2 + p\left(\dfrac{1}{R_1 C_1} + \dfrac{1}{R_2 C_2} + \dfrac{1}{R_1 C_2}\right) + \dfrac{1}{R_1 R_2 C_1 C_2}} \quad (23)$$

The roots are, for the numerator,

$$p_1 = 0, \qquad p_2 = -\left(\frac{1}{R_1 C_1} + \frac{1}{R_1 C_2}\right) \quad (24a)$$

and, for the denominator,

$$p_{3,4} = -\frac{\left(\dfrac{1}{R_1 C_1} + \dfrac{1}{R_2 C_2} + \dfrac{1}{R_1 C_2}\right)}{2}$$

$$\pm \sqrt{\frac{\left(\dfrac{1}{R_1 C_1} + \dfrac{1}{R_2 C_2} + \dfrac{1}{R_1 C_2}\right)^2}{4} - \frac{1}{R_1 R_2 C_1 C_2}}. \quad (24b)$$

If, in these expressions, the assumptions are made that $C_1 \gg C_2$ and $R_1 \gg R_2$, the following approximate formulas are obtained:

$$p_2 = -\frac{1}{R_1 C_2}; \quad p_3 = -\frac{1}{R_2 C_2}; \quad p_4 = -\frac{1}{R_1 C_1}. \quad (24c)$$

Here p_2 and p_3 may be said to correspond to a phase-advance network, and p_4 corresponds to the low-frequency limiting response. The phase response is then given by Eq. (18a). The phase-shift characteristic can thus be represented as an algebraic sum of three inverse-tangent curves or as the sum of a curve like that of the phase-advance network and an additional inverse-tangent curve (Fig. 9·7b). In the central region of the characteristic the phase is relatively constant with respect to frequency. In the design of a three-stage amplifier the average
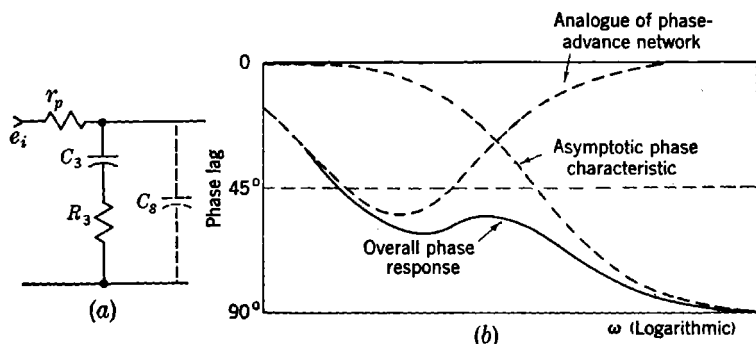
Fig. 9·8.—High-frequency stabilization network. (a) Network; (b) phase response.

phase shift in this region should be kept somewhat less than 60° in order that the response for three such networks will be less than 180°.

Similarly, the network, equivalent at high frequencies to that shown in Fig. 9·4a can be considered to be mathematically equivalent to the series-shunt circuit ($r_p C_s$) plus a circuit analogous to the phase-advance network, which may be called a phase-retard network. This equivalent circuit is shown in Fig. (9·8a). In calculating the response of the equivalent high-frequency network, a similarity to the low-frequency network may be noted. If the components in Fig. 9·7a are renamed by substituting r_p for R_2, C_s for C_2, R_3 for R_1, and C_3 for C_1, the result is found to be the same as for the equivalent high-frequency network, except that the input and ground terminals are interchanged. From this it can immediately be seen that the sum of the responses of the high-frequency network and the renamed low-frequency network is unity; for if at any frequency the response of the high-frequency network is $Z_2/(Z_1 + Z_2)$ (where $Z_1 = r_p$ and Z_2 is the impedance of the network

C_s, R_3, C_3), the response of the renamed low-frequency network is $Z_1/(Z_1 + Z_2)$. Therefore

$$\frac{e_0}{e_i} = 1 - \frac{p(p - p_2)}{(p - p_3)(p - p_4)}$$

where the p_i's may be found from Eq. (24) by renaming the terms. In particular, from Eq. (23),

$$\frac{e_0}{e_i} = 1 - \frac{p\left(p + \dfrac{1}{R_3 C_3} + \dfrac{1}{R_3 C_s}\right)}{p^2 + p\left(\dfrac{1}{R_3 C_3} + \dfrac{1}{r_p C_s} + \dfrac{1}{R_3 C_s}\right) + \dfrac{1}{R_3 C_3 r_p C_s}} \qquad (25a)$$

$$= \frac{1}{r_p C_s} \frac{p + \dfrac{1}{R_3 C_3}}{(p - p_3)(p - p_4)}.$$

Subject to the approximations $R_3 \ll r_p$, $C_s \ll C_3$, the roots are approximately

$$\left. \begin{array}{l} p_2 = -\dfrac{1}{R_3 C_3}, \\[2mm] p_3 = -\dfrac{1}{R_3 C_s}, \\[2mm] p_4 = -\dfrac{1}{r_p C_3}. \end{array} \right\} \qquad (25b)$$

SAMPLE DESIGNS OF COMPUTER AMPLIFIERS

Three computer amplifiers that have been designed for inductive loads and one for a resistive load will be discussed here. The reason for the emphasis on inductive loads is that several types of computer use angle resolvers[1] whose input impedances are chiefly inductive.

TABLE 9·1.—DATA ON RESOLVERS

	Arma Corp.	Bendix Pioneer
Manufacturer	Dwg. No. 213044	XD-759542
Serial No		
Weight, lb	5	1
Stator reactance (400 cps), ohms	j5000	j3000*
Stator-to-rotor voltage ratio (rotor open-circuited)	1.0	0.9
Maximum voltage on stator (400 cps), rms	$\approx$ 100 volts	$\approx$ 60 volts
Corresponding "saturation current," ma	$20\sqrt{2}$	$20\sqrt{2}$
Peak deviation from linearity of rotor voltage with respect to stator voltage (as fraction of max. output), %	Approximately[†] 0.15	Approximately[†] 0.15

* Effective series resistance = 600 ohms.

† These figures are subject to variation, dependent upon the method of measurement. The values given are conservative.

[1] Vol. 17, Chap. 10, "Rotary Inductors," Vol. 21, Chaps. 19 to 21 (uses of resolvers), Radiation Laboratory Series.

Pertinent data on two types of resolver are given in Table 11·1. It was arbitrarily decided to let the resolver stator be the primary winding and its rotor be the secondary or output winding. It can be seen (Table 9·1) that within the specified region of operation, the magnitude of rotor output voltage varies almost linearly with the magnitude of stator input voltage.

In the design of these circuits, electrolytic condensers are used to bypass cathode and screen impedances but are not used for interstage coupling because of the effect that d-c leakage through the condenser might have on the grid bias. The maximum value of coupling capacitance used is limited to 0.1 μf because of the physical size of paper-dielectric capacitors of this value. In a great deal of aircraft equipment, electrolytic capacitors can not be used at all because they do not meet specifications for high-altitude operation.

9·4. Single-stage Drivers. *The Cathode Follower.*—This simplest one-stage feedback amplifier has been used extensively as an impedance-changing device, but the employment of reactive loads requires special designs. A typical a-c cathode-follower circuit is shown in Fig. 9·9.
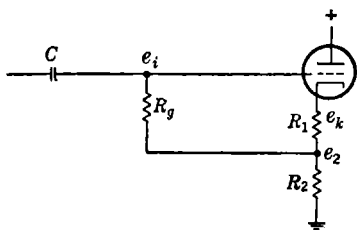
FIG. 9·9.—A-c cathode follower.

The input impedance can be calculated by methods similar to those of Sec. 9·1. The input capacitance is equal to $C_{gp} + [(e_i - e_k)/e_i]C_{gk}$ and is approximately equal to C_{gp}. The resistive component of input impedance (in parallel with this capacitance) is $R_g[e_i/(e_i - e_2)]$. If the current in R_g is neglected relative to the cathode current in computing e_2, the quantity $[e_i/(e_i - e_2)]$ can be easily calculated. The gain of the cathode follower is

$$\frac{e_k}{e_i} = \frac{1}{1 + \dfrac{1}{\mu} + \dfrac{1}{g_m(R_1 + R_2)}}$$

and

$$\frac{e_2}{e_k} = \frac{R_2}{R_1 + R_2}.$$

Therefore,

$$\frac{e_i}{e_i - e_2} = \frac{1}{1 - \dfrac{e_2}{e_i}} = \frac{1}{1 - \dfrac{R_2}{(R_1 + R_2)\left[1 + \dfrac{1}{\mu} + \dfrac{1}{g_m(R_1 + R_2)}\right]}}, \quad (26)$$

and the input resistance is approximately

$$R_g \left(\frac{R_1 + R_2}{R_1} \right)$$

if $\mu \gg 1$, and $g_m(R_1 + R_2) \gg 1$. A nonresistive cathode impedance Z_k could be substituted for $R_1 + R_2$ without invalidating the derivation of e_k/e_i provided that the d-c levels are properly maintained. In this case

$$\frac{e_k}{e_i} = \frac{1}{1 + \dfrac{1}{\mu} + \dfrac{1}{g_m Z_k}},$$

and

$$\frac{e_i - e_k}{e_k} = \frac{1}{\mu} + \frac{1}{g_m Z_k}. \tag{27}$$

The limits of linear operation of a cathode follower are determined by the grid current and cutoff characteristics of the tube. These limits can be calculated by the use of the preceding equations. If it is assumed that the voltage $e_{gk} = e_i - e_k$ has the limits 0 (corresponding to grid current) and e_{co} (at cutoff), and if the region of a-c operation is chosen so that these limits are reached together by the positive and negative peaks of the grid-cathode waveform, then the maximum output is given by

$$\frac{(e_k)_{\max}}{(e_{gk})_{\max}} = \frac{1}{\dfrac{1}{\mu} + \dfrac{1}{g_m Z_k}}. \tag{28}$$

Since the rms value of

$$(e_{gk})_{\max} = \frac{e_{co} - 0}{2\sqrt{2}},$$

and[1]

$$e_{co} \approx \frac{e_p}{\mu},$$

then, finally, the maximum rms output is given by

$$|(e_k)_{\max}| = \frac{e_{co}}{2\sqrt{2}\left| \dfrac{1}{\mu} + \dfrac{1}{g_m Z_k} \right|} = \frac{e_p}{2\sqrt{2}\left| 1 + \dfrac{r_p}{Z_k} \right|}. \tag{29}$$

The value of e_p used for a reactive load will be different from that for a resistive load, for in the former case the load line is an ellipse and the

[1] See, for example, F. E. Terman, *Radio Engineering*, McGraw-Hill, New York, 1937, p. 121.

nearest approach to cutoff will be at a value of e_p less than the plate supply voltage.

It is often undesirable to use an ordinary cathode follower for driving an inductive resolver because the linearity of the stator-to-rotor transfer characteristic is affected by magnetic saturation; moreover, unless resonant tuning is used, the tube must supply a direct current at least as great as the peak alternating current in the inductive load. For cases where the desired range of operating frequencies is too great to permit such tuning, the circuit of Fig. 9·10 which was designed for use with the Arma resolver (Table 9·1), can be used. To make use of the tetrode characteristics of a tube such as the 6V6, it is necessary to bias the screen from a relatively high-impedance source and condenser-couple it to the cathode. The circuit shown in Fig. 9·10 satisfies these conditions because the inductive load itself has the high a-c and low d-c impedance necessary for screen bias. The d-c screen current flowing through the load is less than 5 ma and does not produce serious saturation.

For this circuit the maximum deviation from linearity of peak a-c voltage across the load with respect to peak a-c input voltage was measured as 0.08 per cent of maximum output up to an output voltage of 60 volts rms. However, changing tubes, especially changing from one make of 6V6 to another, changed the gain by ±0.25 per cent.

A single stage with plate-to-grid feedback can also be used to drive an inductive load as is shown in Fig. 9·11. In this circuit the plate current flows through the inductor; and because it may saturate the core, the maximum output voltage is correspondingly limited. The analysis of resistive-mixing feedback in Sec. 9·2 is applicable to this circuit. Compared with the modified cathode follower shown in the previous figure, it has the disadvantages of lower input impedance and increased sensitivity to tube-parameter variability.

FIG. 9·10.—Single-stage driver.
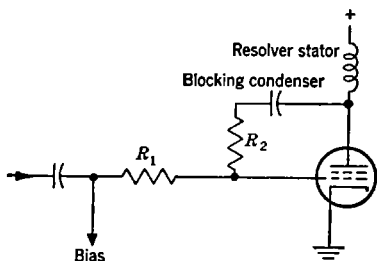
FIG. 9·11.—Single-stage driver with plate-to-grid feedback.

9·5. Driver with Push-pull Output Stage and Regeneration within the Loop.—The circuit in Fig. 9·12 was designed to drive the stator of an

Arma resolver (Table 9·1). Cathode feedback was employed, and a pentode was therefore used in the first stage to secure a high amplification factor. The design procedure for this amplifier was largely experimental, both with regard to antioscillation measures and with respect to maintaining constancy of gain with tube change. The 0.001-μf condenser across the output transformer was used to suppress oscillations; it prevented the high-frequency response from falling off too rapidly with increasing frequency. The 2.2-kilohm resistor and the 0.05-μf condenser provide a low-Q tuning for the load. The feedback from the transformer secondary to the right-hand grid of the output stage is regenerative. The bleeder determining how much signal is fed back was adjusted so that the output stage, disconnected from the pentode, was just on the threshold of oscillation, thus providing an over-all gain very near to unity. The decrease in the linearity of the regenerative stage of the amplifier was offset by the increase of gain; thus the linearity
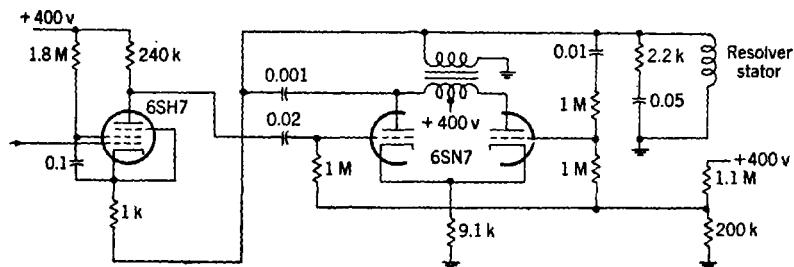
Fig. 9·12.—Driver circuit with push-pull output stage and regeneration.

(and constancy of gain with respect to tube changes, etc.) with the over-all negative feedback was not impaired (see Chap. 10). Since the over-all feedback factor was unity, infinite gain without negative feedback would be needed for unity gain with the negative feedback. If exactly unity gain is not an important consideration, more symmetrical push-pull action can be obtained by reducing the regeneration; this can be done without appreciably sacrificing constancy of over-all gain.

The observed performance of the circuit was as follows: The difference between output and input voltages was less than 0.03 volt up to an output voltage of 70 volts rms. This figure includes the effects due to substituting tubes whose parameters vary over the entire range permitted by the manufacturing tolerances.

TWO-STAGE DRIVER FOR INDUCTIVE LOAD WITHOUT TRANSFORMER
OUTPUT

9·6. General Considerations.—The object of designing this circuit was to reduce the weight in comparison with the amplifier shown in

Sec. 9·5 and to analyze more thoroughly the conditions for preventing oscillations. In order to save weight, miniature tubes were used, and the load, a Bendix resolver (Table 9·1) especially designed for light-weight applications, was condenser-coupled rather than transformer-coupled.

The circuit type of Fig. 9·1c was chosen, and the accuracy requirement set at ± 0.1 per cent probable variation in over-all gain (and ± 0.001 radian in phase) with respect to replacement ot all components. This requirement seemed realizable because a loop gain $\alpha\beta$ of 300 to 500 can be obtained with two stages if $\beta = 1$; the approximate variation of g_m (the largest source of error) is ± 30 per cent, and feedback could reduce the effect of this by a factor of approximately $1/\beta\alpha$.

The circuit was designed to give a maximum output voltage of between 10 and 40 volts rms. The supply voltage, tube type, and load determine this, and the voltage scale of the computer was to be chosen accordingly. It was assumed that the first stage would employ a type 6AK5 and the output stage a type 6C4. The plate supply voltage was to be 250 volts, and the operating frequency was to be 500 cps $\pm$ 2 per cent. The tube-parameter tolerances, as given in the JAN specifications, are shown in Table 9·2. In many cases it would be desirable to supplement these with additional tube data relating more directly to circuit design.[1]

TABLE 9·2.—EXCERPTS FROM JAN SPECIFICATIONS* FOR TYPES 6AK5 AND 6C4 TUBES

Charac-teristic	6AK5 under operating conditions of e_p† $= 120v$, $e_{g_1} = -2v$, and $e_{g_2} = 120v$		6C4‡ under operating conditions of $e_p = 250v$, and $e_g = -8.5v$	
	Min.	Max.	Min.	Max.
i_g	0	$-0.10\mu a$	0	$-1.5\ \mu a$
i_p	3.0 ma	12.0 ma	6.5 ma	14.5 ma
i_{g_1}	0.8 ma	4.0 ma	. . . .	. . . .
g_m	3500 μmhos§	6500 μmhos	1750 μmhos	2650 μmhos

* Issued November–December 1944.

† Maximum allowable value of e_p (according to supplement of Mar. 30, 1945): 180v, design-center value.

‡ The maximum recommended value of grid-leak resistor is 0.25M with fixed bias, 1.0M with cathode bias.

§ Reduction of heater voltage to 5.7 volts causes a reduction of g_m of 15 per cent.

9·7. Design of the Output Stage.—A schematic diagram of the output stage is shown in Fig. 9·13. It incorporates cathode bias, condenser-coupling to the load, and an additional condenser C_4 to increase the

[1] See Chap. 11.

apparent load impedance by parallel-resonant tuning.   Cathode bias was used because it permits a large grid-leak resistor (Table 9·2) and because it stabilizes tube parameters with respect to tube replacement and heater-voltage variation.   In the complete amplifier the tuning condenser C_4 does not appreciably affect the current through the resolver stator for a given input voltage, since the voltage feedback makes the voltage across the load nearly equal to the input voltage throughout the operating range of the circuit.   This condenser extends the operating range of the amplifier.   Parallel-resonant tuning has the disadvantage of increasing the variation in phase shift that is produced by variation in the values of capacitance or of stator impedance, but it will be seen that this is not a serious difficulty.
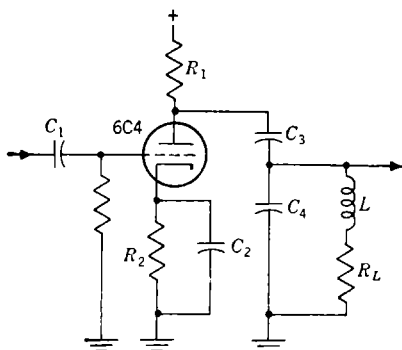
Fig. 9·13.—Output stage of driver.

In order to find the variations in load impedance and in phase shift that might occur, the variations in the load circuit parameters were calculated.   Measurements of a typical resolver-stator impedance showed a variation of 15 per cent over the temperature range −50° to +80°.   The values of the stator impedances at room temperature for all the units measured varied as much as ±12 per cent from their mean; but if condensers were individually selected to tune each resolver, this effect could be almost nullified.   The average value of stator impedance at 500 cps was $600 + j3500$ ohms.   The maximum variation in capacity of condensers of the sort to be used was ±10 per cent apart from temperature-dependent changes.   As a working assumption, the variation to be expected in the LC product will be taken to be ±15 per cent.

It must be determined whether this range of variation of the LC product can lead to an undesirable phase shift through the amplifier or can make the maximum output voltage too small because of the reduced load impedance.   The impedance of a parallel-resonant circuit

in which the resistance R is in series with the inductance is given by

$$|Z| = \sqrt{\frac{R^2 + \omega^2 L^2}{(\omega R C)^2 + (\omega^2 L C - 1)^2}}; \qquad \phi = \tan^{-1} \frac{\omega L}{R} - \tan^{-1} \frac{\omega R C}{1 - \omega^2 L C}.$$

Sample values of $|Z|$ and ϕ can be calculated to show the effect of component variations.   Suppose that

   $\omega = 3140$ radians/sec ($f = 500$ cps),
   $R = 600$ ohms,
   $L = 1.1$ henrys,
   $C = 0.092$ μf (selected so that $\omega^2 L C = 1$).

For the expected variations of L and C the corresponding values of $|Z|$ and ϕ are as follows:

	$L = 1.1\ h$ $C = 0.092$	L and C each increased by 7 %	L and C each decreased by 7 %
$|Z|$	20,300 ohms	15,700 ohms	15,700 ohms
ϕ	$-10°$	$-48°$	$-28°$

Even though the phase angle ϕ corresponding to an increase of 7 per cent in both L and C is 48°, the phase shift from the grid to the output terminal under this condition will be considerably less than 48° because
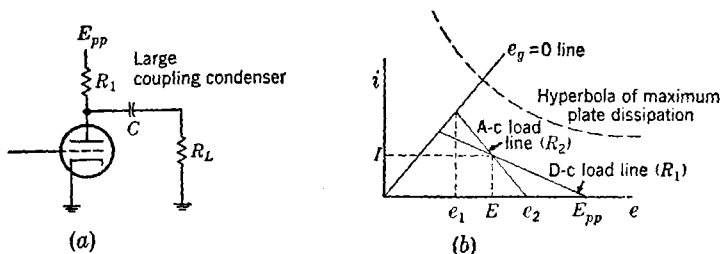
Fig. 9·14.—Triode circuit with condenser coupling to resistive load.   (a) Circuit; (b) operating lines.

the output impedance of the tube (nearly equal to r_p if R_k is sufficiently well bypassed) is smaller than the load impedance.   Since feedback can reduce the phase shift by a factor of approximately $1/\beta\alpha$, a loop gain $\beta\alpha$ of 500 will suffice to keep the phase shift with feedback below 0.1°.

It is necessary to choose the best value of d-c plate-load resistance; for this purpose the following derivation is helpful.   For a triode amplifier whose plate is condenser-coupled to a resistive load of resistance large enough so that the output voltage is not limited by plate dissipation (Fig. 9·14), there is a value of the d-c plate load resistance R_1 for which

maximum output voltage can be obtained without exceeding a given distortion. This value and the corresponding output voltage can be calculated roughly by means of an analysis based on idealized tube characteristics. The procedure will be to state mathematically that the positive and negative voltage swings at the plate, measured from the quiescent voltage, are equal and are limited by cutoff and grid current respectively (Fig. 9·14b). This condition, together with the tube characteristics and component values, determines the operating line and the output voltage for the tube. Maximizing the output voltage with respect to R_1 gives the desired value of R_1.

Let V = d-c plate supply voltage,

E = d-c plate voltage,

I = d-c plate current,

e = instantaneous plate voltage,

e_1 = plate voltage at $e_g = 0$,

e_2 = plate voltage at cutoff,

i = instantaneous plate current,

R_L = load resistance,

$R_2 = \dfrac{R_1 R_L}{(R_1 + R_L)}$.

The value of e at cutoff is

$$e_2 = E + IR_2. \tag{30}$$

The value of e at $e_g = 0$ (assumed to be the grid-current point) is given by the tube characteristic

$$i = \frac{e_1}{r_p} \tag{31}$$

and the equation of the a-c load line

$$e - E = (i - I)(-R_2). \tag{32}$$

Solving for e_1,

$$e_1 = \frac{E + IR_2}{1 + \dfrac{R_2}{r_p}}. \tag{33}$$

If the positive and negative voltage swings from E are equal,

$$e_1 = E - IR_2. \tag{34}$$

By eliminating e_1 from Eqs. (33) and (34), a relation between E and I can be obtained. Combining this with the d-c relation

$$E = E_{pp} - IR_1, \tag{35}$$

it is possible to solve for I and to compute the peak output swing

$$e_{\text{peak}} = IR_2.$$

The rms output is then $e_{\text{peak}}/\sqrt{2}$, or

$$e_{\text{rms}} = \frac{\dfrac{E_{pp}}{\sqrt{2}}}{1 + \dfrac{R_1}{R_2} + \dfrac{2r_p}{R_2}} = \frac{\dfrac{E_{pp}}{2\sqrt{2}}}{1 + \dfrac{r_p}{R_L} + \dfrac{r_p}{R_1} + \dfrac{R_1}{2R_L}}. \tag{36}$$

Maximizing the rms output with respect to R_1 gives

$$R_1 = \sqrt{2r_p R_L} \tag{37}$$

and

$$(e_{\text{rms}})_{\text{max}} = \frac{\dfrac{V}{2\sqrt{2}}}{1 + \dfrac{r_p}{R_L} + \sqrt{\dfrac{2r_p}{R_L}}}. \tag{38}$$

This derivation may be generalized to take into account nonlinear characteristics: If the negative plate swing is a times as great as the positive swing, the equations become

$$R_1 = \sqrt{(1 + a)r_p R_L} \tag{39}$$

and

$$(e_{\text{rms}})_{\text{max}} = \frac{\dfrac{V}{2\sqrt{2}}}{1 + \dfrac{r_p}{R_L} + 2\sqrt{\dfrac{r_p}{(1 + a)R_L}}}. \tag{40}$$

To estimate the maximum output voltage, 18,000 ohms can be used as the average value of $|Z|$, and r_p set equal to 10,000 ohms. According to the treatment for a triode, if the departure of Z from a pure resistance is neglected, R_1 can be selected equal to

$$\sqrt{2 \times 18K \times 10K} = 19K \approx 20{,}000 \text{ ohms}.$$

The output voltage to be expected can be found approximately from Eq. (38): if $B^+ = 250$ volts,

$$(e_{\text{out}})_{\text{rms}} = \frac{\dfrac{250 \text{ volts}}{2\sqrt{2}}}{1 + \dfrac{r_p}{R_1} + \sqrt{2\dfrac{r_p}{R_1}}} = \frac{88 \text{ volts}}{1 + 0.56 + 1.05} = 35 \text{ volts rms}.$$

In Fig. 9·24 this value is compared with experiment.

The d-c and a-c load lines for a 6C4 are shown in Fig. 9·15a. The approximate optimum bias is found to be -7 volts; the corresponding d-c plate voltage is $+170$ volts; and the plate current, 5 ma; hence the

cathode resistor should be 1500 ohms.   The plate dissipation is 0.85 watts (see Fig. 9·15).   A 5.0-μf electrolytic condenser, for which

$$\frac{1}{\omega C} = 64 \text{ ohms at 500 cps,}$$

may be used to bypass the cathode.   The preliminary design of the output stage is shown in Fig. 9·15b.[1]

D-c load line (20 k)
A-c load line (10 k)

(a)

(b)
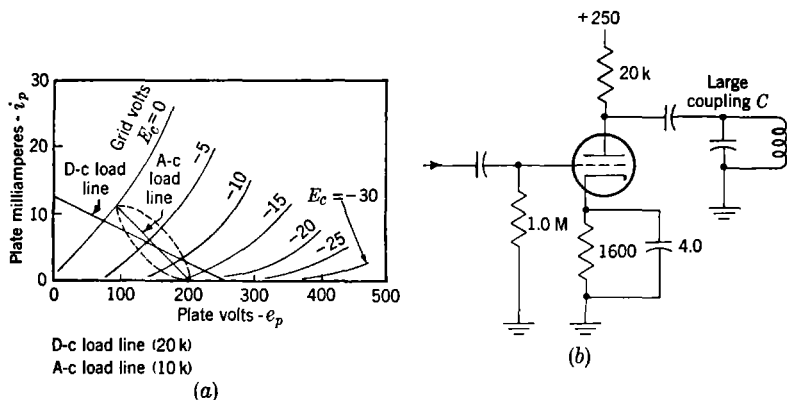
FIG. 9·15.—Operation of 6C4 output stage.
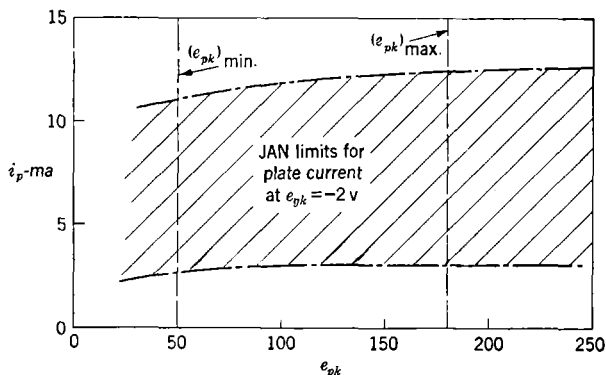
FIG. 9·16.—Type 6AK5 characteristic with tolerances.

9·8. Design of Pentode Stage.—Table 9·2 shows that the expected variations among 6AK5's are somewhat greater than among 6C4's. Moreover, a maximum plate voltage is specified for the 6AK5.   In selecting electrode voltages it is helpful to use characteristic curves that indicate the tolerances of tube performance (Fig. 9·16).   From the plate-current and screen-current specifications on the 6AK5, it can

[1] See Appendix B.

be concluded that the spread of the tube characteristics is roughly equivalent to a change in grid bias of ± 1 volt. This restricts possible designs that use fixed bias, as can be seen from Fig. 9·17a. This graph shows that a plate-load line drawn from a sufficiently large supply voltage will include undesirable operating regions for some tubes if the bias is -2 volts. For example, either the operating point will be near the "knee" of the pentode curves in the case of high-current tubes, or the plate voltage will exceed the absolute maximum rated value in the case of low-current tubes. (The value of 180 volts shown in Fig. 9·17 is actually a design-center value, but it serves to illustrate the point.) Thus, there is a maximum plate-supply voltage at which a particular
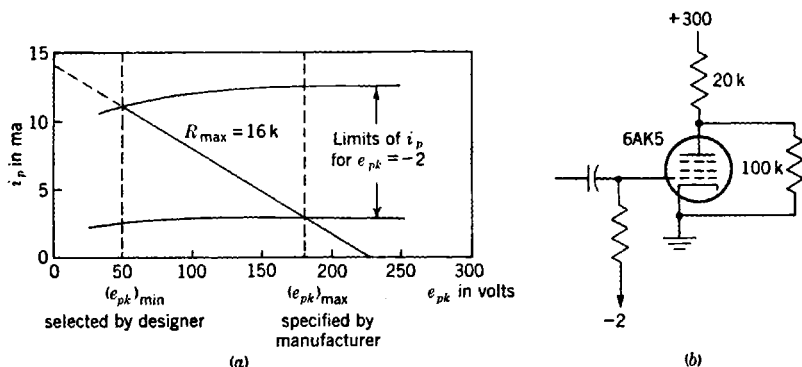
FIG. 9·17.—Tolerances and selection of plate load. (a) Type 6AK5 characteristic; (b) amplifier with bleeder resistor.

fixed bias can be used. If it is desired to operate the tube at fixed bias and to use a higher-voltage plate supply than this, the plate resistor can be replaced by a bleeder from B^+ to ground (Fig. 9·17b).

In the design to be described in this section a cathode resistor will be used to stabilize the tube characteristics. For the stabilization of i_p it is convenient to use the $i_p - e_g$ diagram together with the relation $e_g = -i_p R_k$ graphically expressed as a "load line" (Fig. 9·18a). It is seen that the spread of i_p for different tubes is considerably reduced from the spread at fixed bias by the use of a cathode resistor of the proper value. The optimum value is determined by a compromise between the fact that the plate-current stabilization is greatest for high R_k and the fact that the correlation between a fractional change in i_p and a given fractional change in g_m becomes less at low values of plate current. If linear characteristics are assumed (Fig. 9·18b), the "stabilization factor," the ratio by which the i_p spread is reduced, can be derived. From the geometry, $(A - B)/B = (A/B) - 1 = g_m R_k$

Therefore,

$$\frac{B}{A} = \frac{\Delta i_{p \text{ cathode bias}}}{\Delta i_{p \text{ fixed bias}}} = \frac{1}{1 + g_m R_k}. \tag{41}$$

This is an approximate measure of the reduction of the effect of tube replacement. Similar results are obtained when a combination of fixed bias and cathode bias is used. Analogous considerations also apply to variations of heater voltage (Sec. 10·5, Fig. 10·16). The objects of stabilizing i_p are to keep the plate voltage within rated limits and to reduce the variation of g_m and therefore the variation of gain. Complete
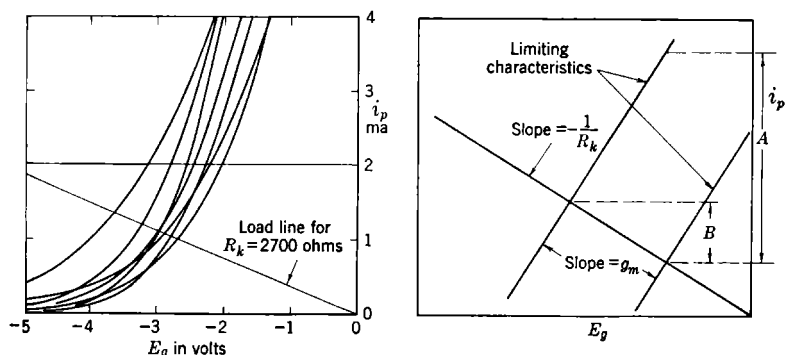
FIG. 9·18.—Stabilization of plate current by cathode resistor. (a) Mutual characteristics of several type 6AK5's, $E_p = E_s = 90'$, $E_f = 6.3'$, four different manufacturers; (b) idealized characteristics, A = spread of i_p, with fixed bias, B = spread of i_p with cathode bias.

data on the g_m-i_p correlation are not available; the curves of Fig. 9·18a indicate the extent of g_m variation. The variation in plate voltage is R_1 times the i_p variation, as is seen from the i_p-e_p diagram.[1]

The screen bias must also be considered. The criteria for the choice of the operating point are the maximization[2] of g_m/i_p and the minimization of plate dissipation. Table 9·3 shows that other things being the same, the smallest values of i_p/g_m (highest gain) are obtained at low screen bias and high (negative) grid bias. It would be desirable to obtain data on 6AK5's beyond the printed characteristics, but in

---

[1] There is a more elegant over-all geometrical representation of this. If the function $i_p(e_p,e_s)$ for fixed screen voltage is represented as a surface in three dimensions, the Ohm's law relations affecting e_p and e_g may be represented by two *load planes*. (A. Preisman, *Graphical Constructions for Vacuum-tube Circuits*, McGraw-Hill, New York, 1943, p. 53.)

[2] For given values of plate supply voltage and d-c plate voltage, the gain of the stage is proportional to g_m/i_p. The reason for this is that the fixed drop across the d-c plate load is $E_{pp} - E_p = i_p R_L$, and the gain is $g_m R_L = g_m(E_{pp} - E_p)/i_p$. This assumes that R_L is the same for alternating current and for direct current.

their absence E_{g_1} is chosen equal to -3 volts, $E_{g_2} = 90$ volts. Since $i_p = 1.1$ ma, the cathode resistor will be 2700 ohms. The average screen current is 0.45 ma; the extent to which it will be stabilized with respect to tube variation by cathode bias is not known; the screen is
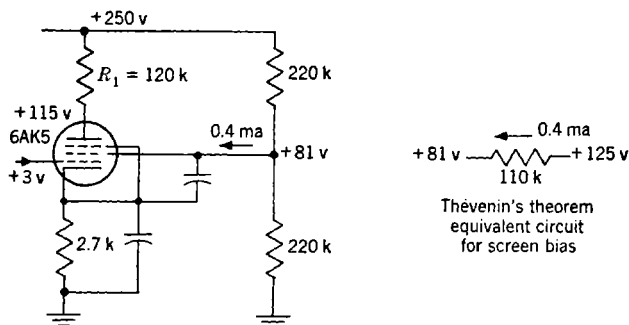
FIG. 9·19.—Preliminary design of pentode stage.

biased from an 0.6-ma bleeder. If the operating e_p is to be 115 volts (the center of the permissible range) and $B+ = 250$ volts, the load resistor R_1 will be $(250 - 115)$ volts$/1.1$ ma $= 123$ kilohms ≈ 120 kilohms. Therefore the circuit assumes the form shown in Fig. 9·19.

TABLE 9·3.—VALUE OF i_p/g_m FOR TYPE 6AK5 AS A FUNCTION OF GRID AND SCREEN VOLTAGES*

e_{c_1}, volts	e_{c_2}, volts	g_m, μmhos	i_p, ma	i_p/g_m, volts
-1	150	9000?	20.00	2.20
-2	150	7000	14.50	2.00
-3	150	5000	8.70	1.70
-4	150	3200	4.70	1.50
-1	120	7200	13.10	1.80
-2	120	4800	7.60	1.60
-3	120	2800	3.70	1.30
-4	120	1400	1.60	1.14
-1	90	4800	7.00	1.46
-2	90	2800	3.00	1.07
-3	90	*1300*	*1.10*	*0.85*
-4	90	400	0.25	0.63
-1	60	2700	2.10	0.78
-2	60	1000	0.40	0.40
-3	60	100?	0	
-4				

* Voltages from data in RCA *Tube Handbook.*

9·9. Constancy of Gain with Respect to Circuit Parameters.—The stabilization factor [Eq. (41)] for i_p and e_p is about 4.5, since

$$g_m R_k = 1300 \times 10^{-6} \times 2700 = 3.5,$$

and the value of 120 kilohms can therefore be used for the load resistor without causing e_p to depart from the desired region.

The sensitivity of over-all gain to component variation will now be calculated. The complete circuit at the present design stage is shown in Fig. 9·20. The d-c resistance of the resolver stator is 170 ohms, which is sufficiently small not to necessitate changing the cathode resistor of the first stage. If perfect bypass at 500 cps is assumed, the gain of the second stage may be expressed as $\mu Z_2/(r_p + Z_2)$ where Z_2 is the effective a-c load.
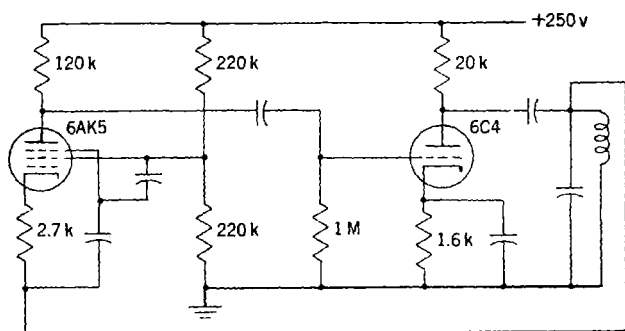
Fig. 9·20.—Driver circuit with tentative resistance values.

The gain of this stage, $\mathcal{G}_b$ in Eq. (9), can be calculated without considering the effect of the pentode stage. If c_i were zero, the cathode input impedance of the pentode would be $1/g_m$; actually the grid is varying more than the cathode, so that when the cathode rises, the current in T_1 increases slightly. This corresponds to a negative conductance in parallel with the load. Its effect is small, however, for the conductance is $g_m(e_k - e_i)/e_k \approx -g_m/300$ using an experimental value of $(e_k - e_i)/e_k$. Since $g_m = 1300$ μmhos, this is a negative resistance of $-300/(1300 \times 10^{-6})$ ohm or -230 kilohms in parallel with the 18-kilohm load. Since the 20-kilohm plate resistor is in parallel with this, the a-c load impedance is the parallel combination of all three of these resistances, or 10 kilohms. Using $r_p = 10$ kilohms, the gain of the second stage is

$$\mathcal{G}_b = \frac{\mu Z_2}{r_p + Z_2} \approx \frac{16 \times 10k}{10k + 10k} \approx 8. \tag{42}$$

The expressions developed for cathode feedback in Sec. 9·2 can now be applied. In this circuit $\beta = 1$. Equation (9) then assumes the form

$$\frac{1}{\mathcal{G}} - 1 = \frac{1}{\mu} + \frac{1}{\mathcal{G}_b}\left(\frac{1}{\mu} + \frac{1}{g_m R_L}\right), \tag{43a}$$

where β = feedback ratio

$\mathcal{G}$ = amplifier gain with feedback,

$\mathcal{G}_b$ = gain of second stage,

R_1 = a-c load of first stage,

μ, g_m = variational characteristics of first stage.

There is an additional network not considered in the previous analysis, namely, the screen-biasing circuit coupled to the first cathode. If the first cathode resistor could be considered as completely bypassed, this would merely be a high impedance in parallel with the load and as such
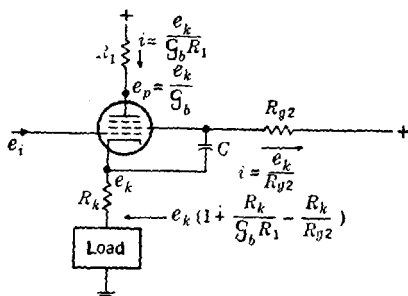
FIG. 9·21.—Pentode stage with unbypassed cathode resistor.   Arrows indicate directions of currents in plate and screen circuits assuming e_i and e_k positive.

could be neglected. If the first cathode is not bypassed, additional terms are introduced into Eq. (43a), which then becomes

$$\frac{1}{\mathcal{G}} - 1 \approx \frac{1}{\mu} + \frac{1}{\mathcal{G}_b}\left(\frac{1}{\mu} + \frac{1}{g_m R_1}\right) + \frac{R_k}{\mathcal{G}_b R_1} - \frac{R_k}{R_{g_2}}. \tag{43b}$$

where R_{g_2} is the Thévenin's-theorem equivalent screen-biasing resistance. The origin of the two new terms may be clarified by reference to Fig. 9·21. The feedback holds e_k at nearly its former value, but the plate current and the current through R_{g_2} now flow through R_k, changing the output voltage accordingly. In this case, by choice of R_{g_2} and R_k, the over-all gain can be made to center at unity rather than at some smaller value, although the sensitivity to $\mathcal{G}_b$, g_m, and μ will not be decreased thereby.

Substituting circuit values in Eq. (43a),

$$\frac{1}{\mathcal{G}} - 1 = \frac{1}{1300} + \frac{1}{8}\left(\frac{1}{1300} + \frac{1}{1.3 \times 160}\right).$$

Now, although both the exact tolerances of μ and g_m and the correlation of g_m with i_p are unknown, it can be seen that the term $1/(8 \times 1300)$ is relatively negligible, and the expression may be written as

$$\frac{1}{\mathcal{G}} - 1 \approx \frac{1}{1300} + \frac{1}{1660}.$$

Even if each of these quantities is subject to a 40 per cent variation, the variation in $1/\mathcal{G}$ will not exceed 0.1 per cent. Therefore the loop gain seems sufficient.

9·10. Stability Against Oscillation.—The loop gain that is important for this problem is found by breaking the loop between the first and second stages, and holding the first grid at a-c ground.[1] In this case the second stage is operating into the low input impedance at the cathode of the first stage. This input impedance is affected by the screen circuit as well as by the plate circuit; it is

$$\frac{1}{g_{m_1} + g_{12}} = \frac{1}{1300 + 500} \times 10^6 \text{ ohms} \approx 550 \text{ ohms}.$$

The gain of the second stage therefore may be expressed roughly as $g_{m_2}/(g_{m_1} + g_{12})$,
where g_m = grid-plate transconductance of the first stage,

   g_{12} = grid-screen transconductance of first stage,

   g_{m_2} = transconductance of second stage.

The loop gain is the product of this quantity and the cathode-to-plate gain of the pentode, the latter being approximately $g_{m_1}R_{L_1}$, where R_{L_1} is the load resistance of first stage. Hence, if both cathodes are bypassed, the loop gain is

$$\mathcal{G} = \frac{g_{m_2}}{1 + \dfrac{g_{12}}{g_{m_1}}} R_{L_1}. \tag{44}$$

It is also this factor which determines the reduction of extraneous voltages (B^+ ripple, for example) by the circuit.

Experimentally, the gain from the second grid to the first cathode with the loop open was found to be approximately one, in fair agreement with the predicted value 2000/1800. The loop gain with the loop open (experimentally) was about 100, and the predicted value was

$$\frac{2000}{1800} \times 120 \times 1.3 = 170.$$

The experiments mentioned were done with a 4-μf bypass capacitor in each cathode circuit.

[1] If the driver is used in another loop, however, the potential of the first grid will vary depending on the effect on it of the output through this loop.

Now that the loop gain at 500 cps is known, the next step is to consider shaping the response at higher and lower frequencies. (It is assumed for the present that the observed gain will suffice to reduce the effects of component variations.) Experimental curves of loop gain (amplitude and phase response) are shown in Fig. 9·22. The con-
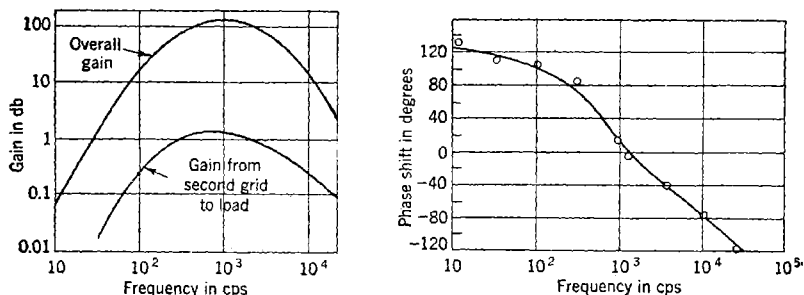
FIG. 9·22.—(a) Amplitude response of resolver driver with loop open; (b) phase response of resolver driver with loop open (experimental points).

denser values used are screen bypass, 4 μf; first-cathode bypass, 0.5 μf; both coupling condensers, 0.1 μf; second-cathode bypass 5.0 μf. It will be shown later that these values are satisfactory. The circuit tested was thus similar to that of Fig. 9·23. At low frequencies the circuit is like an *LC*-network and at high frequencies like an *RC*-network (whose phase shift approaches 90°), hence the different slopes at high

All resistors ½ watt unless indicated
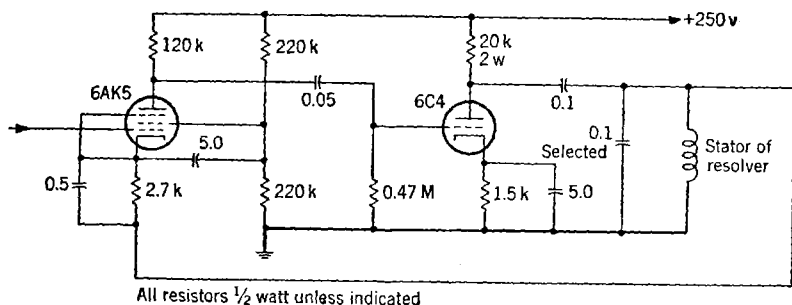
FIG. 9·23.—Final design of resolver-stator driver. All resistors ½ watt unless otherwise indicated.

and low frequencies of the curve in Fig. 9·22. The frequency at which the resolver-stator inductive reactance equals 600 ohms, which is the value of its resistance at 500 cps, is given by $2\pi f = R/L$, $f = 600/7 = 80$ cps. On the low-frequency side, unity gain is reached at about 30 cps (Fig. 9·22a). The frequency at which the *RC*-coupling network has 45° phase shift is $f = 1/2\pi RC$. If $R = 0.51$ megohm, and $C = 0.05$ μf,

$f = 6.4$ cps. Thus the phase shift of the coupling network is effective only at frequencies at which oscillation cannot occur. The rest of the condenser values are also satisfactory in that it appears unlikely that any 180° phase shifts will occur within the region where the loop gain is unity or greater.

A method of determining the approximate values of bypass condensers is to select them so that their maximum phase-shift contributions occur at different frequencies.[1] If a 2.0-μf condenser is used to bypass the screen, the approximate frequency at which this produces maximum phase shift is

$$\frac{1}{2\pi C r_{g_2}} = \frac{1}{2\pi \times 2 \times 10^{-6} \times 30,000} \text{ cps} \approx 3 \text{ cps},$$

where it does not particularly influence the possibility of oscillation. (The screen variational resistance is r_{g_2}.[2]) The approximate effect of the cathode RC-circuits may be seen with the aid of a family of curves showing the falling off and phase shift in output voltages at low frequencies in a resistance-coupled amplifier.[3] The maximum phase shift caused by a phase-advance network (to which each of these is equivalent) is $2 \tan^{-1} \sqrt{p_2/p_1} - 90°$ [Eqs. (21) and (22), Fig. 9·6]. It is found experimentally by measuring the gains with the cathode condenser removed that the ratio of $(p_2/p_1) = (e_2/e_1)_\infty/(e_2/e_1)_0 = 4$ for the triode and $\frac{3}{2}$ for the pentode. The corresponding phase shifts are 37° and 12°. The choice of a 5.0-μf condenser to bypass the second cathode resistor puts the maximum phase shift of that network at about

Fig. 9·24.—Fractional error voltage as a function of output voltage of second stage.

$$\frac{1}{(2\pi \times 1500 \times 5 \times 10^{-6})} \text{ cps} \approx 21 \text{ cps}.$$

The phase-shift effect of the first cathode network may be put at a higher frequency, since the smaller maximum phase shift makes it possible to put the curve nearer to 500 cps without producing much phase shift or attenuation at 500 cps. If, for example, the value of RC is made to correspond to 100 cps,

$$C = \frac{1}{2\pi \times 100 \times 2700} \times 10^6 \ \mu\text{f} \approx 0.6 \ \mu\text{f}.$$

---

[1] The process of staggering networks is treated by Bode, *op. cit.*, pp. 514–517.

[2] F. E. Terman, *Radio Engineers' Handbook*, McGraw-Hill, New York, 1943, p. 358.

[3] *Ibid.*, p. 359.

In these calculations, use has been made of the approximation that the maximum phase shift of the equivalent phase-advance network is at a frequency $1/(2\pi R_k C_k)$; actually this is the central frequency of one of the inverse-tangent curves, but for small phase shifts the error is not serious.

The final circuit design then assumes the form shown in Fig. 9·23. The difference between input voltage and output voltage may be measured by means of an oscilloscope, and the ratio of the peak value of this voltage to the peak value of the output voltage is plotted in Fig. 9·24 as a function of output voltage. Most of this difference voltage, in the region above 20 volts output voltage, corresponds to distortion resulting from cutoff of the second stage.

Up to an output voltage of 20 volts rms this amplifier approximately satisfies the original specifications. The fractional difference between output voltage and input voltage is 0.002; for this quantity to vary by 0.001, variations in component values must combine to produce a fractional variation in gain of at least $\frac{1}{3}$. The phase-shift condition is approximately satisfied, for even a loop phase shift of 48° (calculated for values of L and C at the extreme limits of the tolerances prescribed) would produce an over-all phase shift of only 0.002 sin 48° radian, or about 0.08°, as compared with the requirement of 0.06°.

THREE-STAGE AMPLIFIER FOR RESISTIVE LOAD

9·11. General Considerations.—In the design of this amplifier it is desired to increase a voltage by a factor that will be adjustable about a mean value of 4.5 and that will remain constant to within ± 0.1 per cent with respect to component variations. The input voltage is available from a relatively low-impedance source, and in order to save weight it is desired not to use transformers. A three-stage amplifier with resistive mixing is to be used (Fig. 9·1a).[1] The amplifier is to use miniature tubes. It is required to operate linearly to as high an output voltage as is possible with a given plate supply voltage.

Before selecting the tube types to be used, the loop gain necessary for the desired degree of constancy of over-all gain must be considered. For the purpose of calculating the loop gain $\beta\alpha$ required, it is assumed that the loop gain will vary by ± 40 per cent as a result of all component variations (± 12 per cent peak variation for each stage). This assumption means that half the difference between the extreme values of $1/\beta\alpha$, corresponding to the ± 40 per cent variation of α about a value α_0, must be 0.001 or less. This is expressed mathematically as follows:

$$0.001 \geqq \frac{1}{\beta\alpha_0} \times \frac{1}{2}\left(\frac{1}{0.6} - \frac{1}{1.4}\right), \quad \text{and} \quad \beta\alpha_0 \approx 500. \quad (45)$$

[1] By variation of the mixing resistances, the circuit may also be used for multiplication or division with low output impedance.

The voltage fed back to the first grid is $1/5.5$ times the output voltage of the third stage, this attenuation being characteristic of resistive mixing; for if the ratio $R_B/R_A = 4.5$, the voltage at the first grid will be $4.5/5.5$ times the input voltage plus $1/5.5$ times the output voltage. Therefore the gain of the amplifier from first grid to third plate must be about $5.5 \times 500 = 2750$. If it is assumed that three similar stages are used, the gain of each would be $\sqrt[3]{2750} \approx 14$. This gain can probably be realized if three 6C4's ($\mu \approx 17$) are used, as shown in Fig. 9·25.[1]

If the load resistances of the three amplifier stages are large compared with r_p, the stage gains will be nearly equal to μ and will depend more
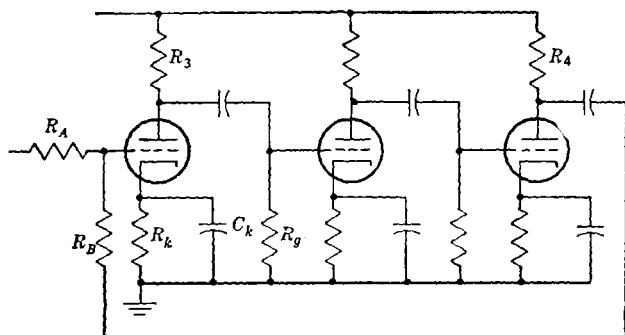
FIG. 9·25.—Schematic circuit of three-stage driver without stabilization networks.

on μ than on other circuit parameters; therefore the variation of μ with tube replacement and aging will be the principal source of error. Data on the variation of μ among 6C4's are given in the JAN specifications: The limits are 15.5 and 18.5 for $E_g = -8.5$ volts and $E_p = +250$ volts. If the equipment is assumed to be calibrated for a mean value of

$$\mu = 17.0 \pm 1.5,$$

the peak variation in μ per stage is then ± 9 per cent.

9·12. Design of Individual Stages.—The first two stages can be identical, since they are both to operate at relatively low level. For a given plate current (limited by the permissible power drain from the plate supply) higher gain can be obtained at low plate voltage. Assume $I_p = 1$ ma, $E_p \approx 30$ volts, $E_{pp} = 250$ volts. The circuit constants may then be determined.

$$R_3 = \frac{(250 - 30) \text{ volts}}{1 \text{ ma}} = 220{,}000 \qquad \text{ohms, an RMA value.}$$

[1] Alternatively, a pentode output stage might be considered in order to increase the maximum output voltage.

If the bias is -2 volts, then $R_k = 2$ volts$/1$ ma $= 2$ kilohms. The operating point on the 6C4 characteristics is shown in Fig. 9·26. The size of the cathode bypass condenser must be determined with some care, because it influences the phase shift at low frequencies and hence the possibility of oscillation. The bypass condensers will be given the smallest values that will provide satisfactory gain for each stage; these may be altered later when the question of oscillation is considered. The gain of a single triode stage is given by
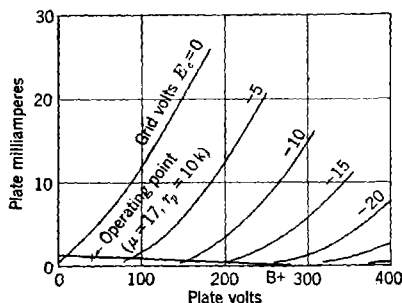
Fig. 9·26.—D-c operation of first and second 6C4 amplifier stages.

$$G_1 = \frac{\mu}{1 + \dfrac{r_p}{R_L} + \dfrac{(\mu + 1)Z_k}{R_L}}, \quad (46)$$

where R_L, the a-c load resistance, is the resistance of the parallel combination of R_3 and R_9 (Fig. 10·25) and is assumed for the present to be 150 kilohms. The equivalent circuit is shown in Fig. 9·1a. The cathode impedance should be sufficiently small not to decrease the gain of a stage more than approximately 5 per cent. The condition that the reactance of $C_k/(\mu + 1)$ be 5 per cent of R_L is satisfied by

$$C_k = \frac{\mu + 1}{\omega R_L \times 0.05} = \frac{18}{2\pi \times 500 \times 150,000 \times 0.05} = 0.8 \quad \mu f. \quad (47)$$

It may not be necessary to use this value of C_k to achieve the desired gain, because the effect of C_k adds in quadrature rather than directly. The gain of each of these stages, with no cathode degeneration, will be approximately

$$\frac{\mu}{1 + \dfrac{r_p}{R_L}} \approx \frac{17}{1 + \dfrac{10}{150}} \approx 16, \quad (48a)$$

where the values $\mu = 17$, $r_p = 10$ kilohms are estimated average values at the operating point (Fig. 9·26).

The output of the third stage is condenser-coupled to the precision resistor R_B (Fig. 9·25). Since the circuit operation keeps the first grid near zero a-c potential, the a-c load line for the third stage has a slope corresponding to the parallel resistance of R_B and the d-c plate load R_4. The value of R_4 for maximum voltage output can be found from the expression[1] $R_4 = \sqrt{2r_p R_B}$, where $r_p \approx 10$ kilohms. This value of r_p

[1] See Sec. 9·7 for derivation.

is the estimated static plate resistance along the zero-bias line, since it is this quantity which enters into the derivation for maximum output. The value of R_B is determined by the availability of high-resistance precision wire-wound resistors. In this design the values $R_A = 100$
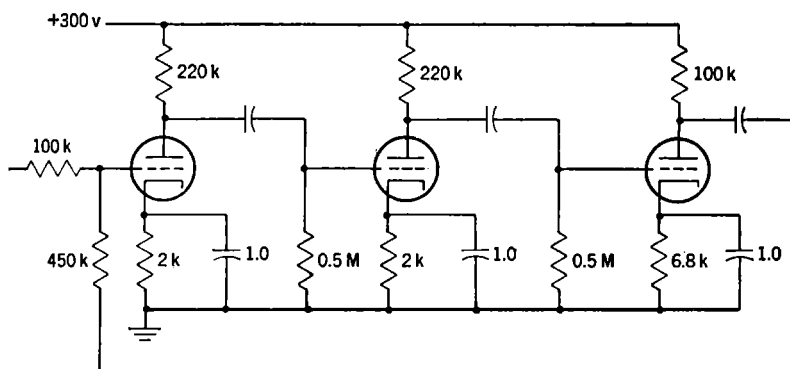
Fig. 9·27.—Preliminary design of three-stage driver exclusive of stabilization networks.

kilohms and $R_B = 450$ kilohms were chosen tentatively, although it might be possible to use higher values and thereby increase the output voltage. The value of R_4 is then $\sqrt{2 \times 10k \times 450k} \approx 100$ kilohms.

The d-c operating conditions are shown in Fig. 9·28. A bias of about -7 volts (arrived at by drawing approximate load lines on the 6C4 characteristics with the object of maximizing the plate voltage swing) seems desirable. This corresponds to a plate current of about 1.0 ma, and $R_k \approx 6.8$ kilohms. A cathode-bypass condenser of about 1 μf will suffice for the third stage. The circuit then assumes the form shown in Fig. 9·27.
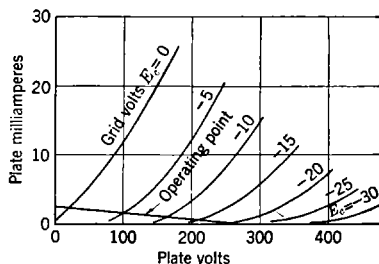
Fig. 9·28.—D-c operation of third amplifier stage.

The values of grid-leak resistors shown are trial values and are likely to be changed subsequently in the design of the stabilization networks. The gain of the third stage will be approximately

$$\frac{\mu}{1 + \dfrac{r_p}{R_L}} \approx \frac{14}{1 + \dfrac{25}{80}} \approx 11, \tag{48b}$$

the values $\mu = 14$, $r_p \approx 25$ kilohms being estimated from the tube characteristics near the operating point as is shown in Fig. 9·28, and the value

$R_L = 80$ kilohms being the parallel resistance of the d-c plate load (100 kilohms) and $R_B = (450$ kilohms). Therefore the expected over-all gain is $(16)^2 \times 11 \approx 2800$, and the loop gain $\mathcal{C}$ is about 500 (54 db).

9·13. Stability against Low-frequency Oscillation.—The frequencies at which the cathode-bypass networks produce maximum phase shift can be calculated approximately from the time constants $R_k C_k$. If the phase shift is small, the frequency of maximum phase shift is approximately that at which one of the RC-networks involved produces 45° phase shift, that is, the frequency at which the reactance of the capacity C_k equals the resistance R_k. For the network $R_k C_k$, this frequency is given by the equation

$$2\pi f R_k C_k = 1.$$

For the first two stages,

$$f \approx \frac{1}{2\pi R_k C_k} = \frac{1}{6.28 \times 2000 \times 1 \times 10^{-6}} = 80 \qquad \text{cps.}$$

For the third stage, the corresponding frequency is about 25 cps. The ratio of the gain with the bypass condenser to that without the condenser determines the maximum phase shift resulting from each cathode-bypass network; this corresponds to the ratio

$$\frac{\left(\dfrac{e_0}{e_i}\right)_{f=\infty}}{\left(\dfrac{e_0}{e_i}\right)_{f=0}}$$

of Eq. (22).

This gain ratio may be found from Eq. (46) for the gain of a triode amplifier with cathode impedance Z_k:

$$\begin{aligned} Z_k &= R_k && \text{at } f = 0, \\ Z_k &= 0 && \text{at } f = \infty ; \end{aligned}$$

therefore,

$$\frac{(\mathcal{G}_1)_{f=\infty}}{(\mathcal{G}_1)_{f=0}} = 1 + \frac{(\mu + 1)R_k}{R_L + r_p}.$$

For the first two stages this has the value

$$1 + \frac{18 \times 2k}{150k + 10k} = 1.22;$$

for the third it is

$$1 + \frac{15 \times 6.8k}{80k + 25k} = 1.97.$$

The corresponding maximum phase shifts, which in Eq. (21) are given by $\phi_m = 2(\tan^{-1} \sqrt{p_1/p_2} - 45°)$, are 5.6° for each of the first two stages

and 16° for the third. The reductions in gain make the remaining stabilization problem somewhat easier, however, for they produce an attenuation of $1.97(1.22)^2 = 2.9$ at very low frequencies without producing an appreciable phase shift. Thus it can be considered that the loop gain that must be reduced to unity by the coupling networks at low frequencies is approximately $500/2.9 \approx 170$. At high frequencies the corresponding figure is 500.

In stabilizing the amplifier at low frequencies it should first be determined whether or not the desired response can be obtained with coupling networks, each consisting of a single resistance and capacitance. To save weight no coupling capacitor greater than 0.1 μf will be used. According to the *RCA Tube Handbook* the grid resistor should not exceed 1.0 megohm. In order to see if the desired response can be obtained in this way, the phase response of the respective *RC* coupling networks can be plotted and the frequency found at which 180° phase shift is obtained. It can be assumed that the low-frequency limit is determined by a time constant of 0.1 sec resulting from one coupling network whose constants have the maximum values given above. It is also assumed that the other coupling networks will not decrease the gain at 500 cps by more than 5 per cent. By a straightforward application of network analysis, the central frequencies f_1 of the other networks can be found.

$$\frac{1}{\sqrt{1 + \left(\frac{f_1}{500}\right)^2}} = 0.95,$$

or

$$\frac{f_1}{500} = \sqrt{0.1} = 0.3, \qquad f_1 \approx 150 \qquad \text{cps},$$

where $f_1 = 1/2\pi RC$. Figure 9·29 shows the phase response of the loop in the case where a phase shift of 45° occurs at 150 cps for two of the coupling networks and at 1.6 cps for the third (corresponding to $R = 1$ megohm and $C = 0.1$ μf). The over-all phase shift other than that of the phase advance of the cathode-bypass networks reaches 180° at $f = 12$ cps; at this frequency the attenuation is approximately $(\frac{12}{150})^2 = 0.0064$. This does not quite suffice to reduce the gain to unity,[1] since

$$170(0.0064) = 1.1.$$

Some small alterations of circuit constants might suffice to reduce this value below unity; however, to provide a larger safety factor additional networks are used. The resulting design will therefore indicate

---

[1] This conclusion may also be reached by use of the formula in Terman, *op. cit.*, p. 398. According to this expression a loop gain of 190 may be stabilized in this way.

how considerably higher gains may be stabilized. The introduction of a network of the type shown in Fig. 9·7 for low-frequency stabilization has the advantage not only of introducing an additional equivalent phase-advance network (which, as was shown, can attenuate without introducing phase shift at low frequencies) but also of moving the low-frequency limiting time constant to a lower frequency. The reason for this is that in the stabilization network of Fig. 9·7, the low-frequency phase response is determined largely by R_1C_1 as shown in Eq. (24c); the limitation on C_1 is still the same as for the single RC-network, but R_1 can be considerably larger, since it is not in the d-c grid-return path.

If the circuit is designed with three such networks, full advantage will be taken of this effect. The values of R_1 and C_1 used are 5.1 megohms
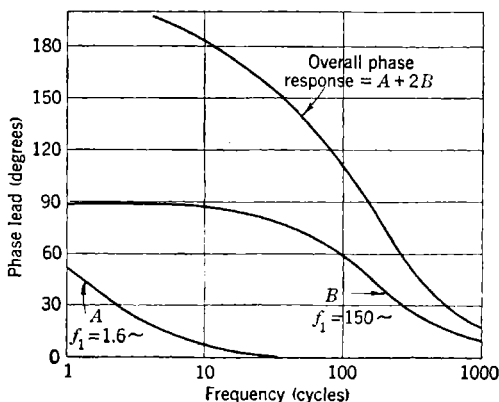
FIG. 9·29.—Phase response of three-stage driver (loop open) with single RC-coupling networks.

and 0.1 μf respectively, the resistance value now being limited only by possible leakage paths. The corresponding central frequency, $1/2\pi R_1C_1$, is 0.3 cps. The use of these networks affords the possibility of adding three phase-advance curves to the limiting curve, which is determined approximately by the R_1C_1's. The design procedure will be to choose these three curves so that they add to give an over-all phase response that rises fairly rapidly to about 150° phase shift (as the frequency is decreased from 500 cps) and remains in this vicinity until the limiting curve takes over and brings the phase shift to 180°. In this the designer may be guided approximately by the phase-area theorem.[1] This states that for networks whose gain changes from one fixed value at zero frequency to another at infinite frequency, the area under the phase-shift curve is proportional to the gain ratio. The theorem is applicable to

[1] Bode, *op. cit.*, p. 286; Terman, *op. cit.*, p. 218.

phase-advance networks. In order to have a fast rise of phase angle with respect to frequency and to increase the phase area between 50 and 500 cps, at least two of the curves should have effect near 500 cps. It seems desirable that the third have effect at lower frequencies. A further restriction is that the two precision resistors at the input grid of the driver constitute part of the coupling network. This restricts the value of R_2 for this network because of the other requirements on R_B (see Fig. 9·25): that it have a high value in order to increase the maximum output but that it cannot be higher than a conveniently obtainable value for precision wire-wound resistors. The general nature of the desired phase response is shown in Fig. 9·30. This diagram
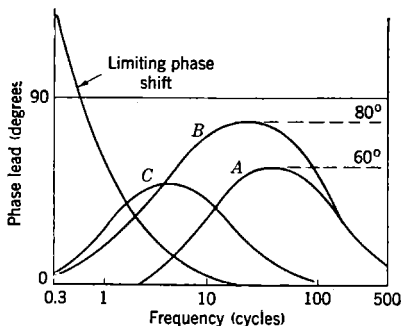
FIG. 9·30.—First approximation of synthesis of phase networks.

was arrived at by successive trials of different phase-advance curves. The following terminology is used: coupling network from first plate to second grid, Network A; from second plate to third grid, Network B; from third plate to first grid, Network C.

The procedure is to find the constants of networks having approximately the response sketched and then to calculate their properties more accurately. It has already been decided that $R_1 = 5.1$ megohms and $C_1 = 0.1$ μf for each of the three networks. Furthermore, for Networks A and B the sharp rise from 500 cps necessitates[1] that $1/2\pi R_2 C_2 \approx 110$, or $R_2 C_2 = 0.0015$ sec. The only other information necessary to the design of the stabilization networks is the value of R_2/R_1 for each network, this ratio being approximately equal to p_2/p_1 for the phase-advance network. The expression $\phi_{max} = 2(\tan^{-1}\sqrt{p_2/p_1} - 45°)$ [Eq. (21)] yields $p_2/p_1 = 14$ for $\phi = 60°$ (Network A) and 120 for $\phi = 80°$ (Net-

[1] It can be shown that if single RC-circuits are used to attenuate at high and low frequencies, the frequencies at which the output is $1/n$ of its maximum value ($n \gg 1$) are closest together if both RC products are equal to $1/\omega_c$, where ω_c is the design or computing frequency. This would mean setting $f_1 = 500$ cps if the amplifier had enough gain to spare a factor of $1/\sqrt{2}$ for each network.

TABLE 9·4.—CONSTANTS OF NETWORKS A AND B

Constant	Network A	Network B
R_1	5.1 megohms	5.1 megohms
C_1	0.1 μf	0.1 μf
R_2 (first approx.)	$\dfrac{5100k}{14} = 360k$	$\dfrac{5100k}{120} = 43k$
C_2 (first approx.)	$\dfrac{0.0015 \text{ sec}}{360k} = 0.0042 \ \mu\text{f}$	$\dfrac{0.0015 \text{ sec}}{43k} = 0.035 \ \mu\text{f}$
$\left.\begin{array}{c} C_2 \\ R_2 \end{array}\right\}$ (final values)	0.005 μf 300k	0.02 μf 75k

work B).   The resulting constants are shown in Table 9·4.   The reason for making a second approximation to R_2 is that the condenser C_2 is available only in values such as 0.01, 0.02, 0.05, and 0.1 μf.   Thus, when the nearest C_2 has been selected, the time constant R_2C_2 is wrong as regards its effect at 500 cps.   The procedure is to readjust R_2 to make R_2C_2 again equal to 0.0015 sec.   The effect on the maximum phase shift of the phase-advance networks is small.   The phase response of these two networks will now be plotted in order to determine more precisely what characteristics the third network should have.   The roots p_i may be found from Eq. (24):

$$p_2 = -\left(\frac{1}{R_1C_1} + \frac{1}{R_1C_2}\right),$$

$$p_{3,4} = \frac{-\left(\dfrac{1}{R_1C_1} + \dfrac{1}{R_2C_2} + \dfrac{1}{R_1C_2}\right)}{2}$$

$$\pm \sqrt{\frac{\left(\dfrac{1}{R_1C_1} + \dfrac{1}{R_2C_2} + \dfrac{1}{R_1C_2}\right)^2}{4} - \frac{1}{R_1R_2C_1C_2}}.$$

For Network A:

Roots	Corresponding frequency $= -\dfrac{p_i}{2\pi}$, cps
$p_2 = -41 \text{ sec}^{-1}$	6.5
$p_3 = -706 \text{ sec}^{-1}$	112
$p_4 = -1.85 \text{ sec}^{-1}$	0.29

Similarly, for Network B:

Roots	Corresponding frequency
$p_2 = -11.8 \text{ sec}^{-1}$	1.9
$p_3 = -676 \text{ sec}^{-1}$	108
$p_4 = -1.93 \text{ sec}^{-1}$	0.31

The corresponding inverse-tangent curves and their algebraic sum, including the curves for p_4 but not for p_2 or p_3 for Network C, are shown in Fig. 9·31.   From this characteristic it appears that the third phase-

advance network should have maximum phase shift at 1.6 cps and that this maximum should be approximately 50°. Therefore, from Fig. 9·6,

$$\sqrt{p_2 p_3} = 2\pi \times 1.6 \text{ cps} = 10 \qquad \text{radians/sec,}$$

and from Eq. (21)

$$50° = 2\left(\tan^{-1}\sqrt{\frac{p_3}{p_2}} - 45°\right) \qquad \text{or} \qquad \sqrt{\frac{p_3}{p_2}} = 2.75.$$
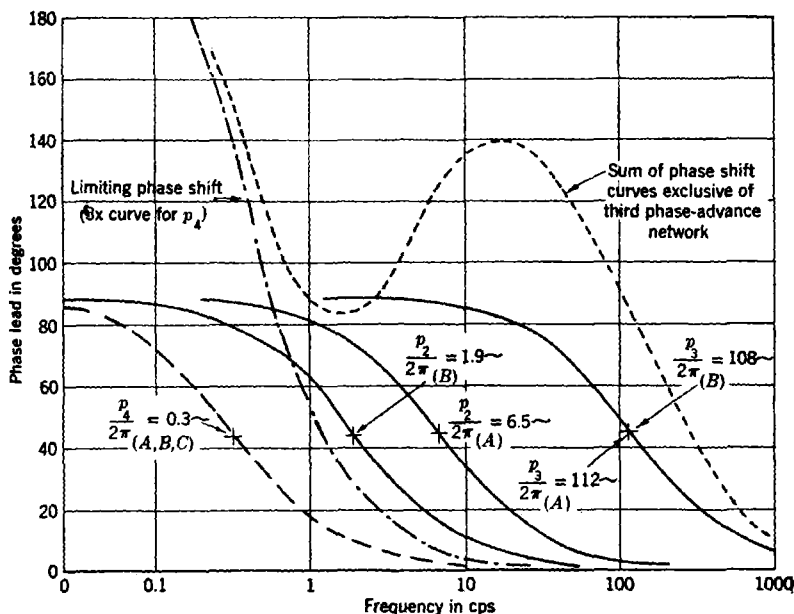
FIG. 9·31.—Phase characteristics for two stabilization networks.

Therefore, for Network C:

Roots	Corresponding frequency, cps
$p_2 = 3.6 \text{ sec}^{-1}$	0.57
$p_3 = 27.5 \text{ sec}^{-1}$	4.4

The constants of Network C may now be selected. It has already been decided that $R_1 = 5.1$ megohms, $C_1 = 0.1$ μf. From the approximate versions of Eq. (24c),

$$p_2 \approx \frac{1}{R_1 C_2}, \qquad p_3 \approx \frac{1}{R_2 C_2}.$$

Therefore, approximately,

$$C_2 = 0.055 \ \mu\text{f (use } C = 0.05 \ \mu\text{f)},$$
$$R_2 = 680 \text{ kilohms.}$$

Finally, for one of these networks, the sum of the resistances of the precision resistors at the input terminals of the driver constitutes R_2. Therefore, the procedure will be to specify values for these resistors consistent with the low-frequency-stabilization design. A satisfactory solution is to use the values of 550 and 125 kilohms, which add to 675 kilohms and have the ratio 4.4. The calculated over-all phase-shift characteristic for the three networks A, B, and C is shown in Fig. 9·32. The phase margin afforded is 20° over a frequency range extending down to 0.5 cps. The attenuation at 0.5 cps due to the three networks can
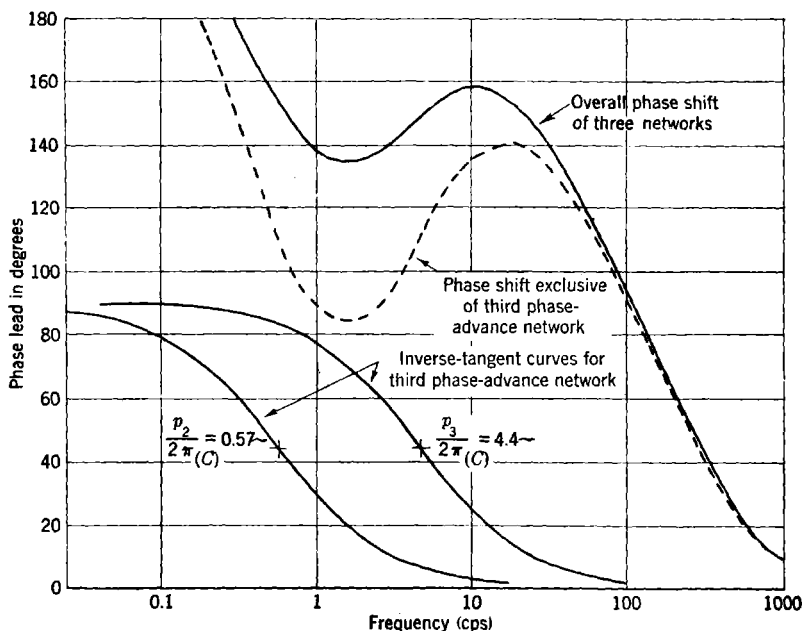
Fig. 9·32.—Calculated phase response of three low-frequency stabilization networks.

now be calculated. The reactance values in the networks at this frequency are shown in Fig. 9·33. It may be seen by inspection that the attenuation of Network A is about $\frac{1}{20}$; that of Network B, about $\frac{1}{50}$; and that of Network C, about $\frac{1}{4}$. Thus the over-all attenuation is $\frac{1}{4000}$ (72 db). This is considerably greater than the factor of 190 (46 db) required. However, the expense in parts is not much greater than if Networks A and B alone had been used (which would probably have sufficed), and the design indicates the order of gain that could be stabilized by this method.

9·14. Stabilization against High-frequency Oscillation.—At high frequency the limiting phase shifts are determined largely by tube char-

acteristics (variational plate resistance and input capacitance). The first step in design of stabilizing networks is to find the limiting (asymptotic) phase-shift characteristic. For the first two 6C4's, the plate resistance is approximately 10 kilohms, and for the third, 25 kilohms. The presence of a series resistor between the third plate and the first grid would, of course, introduce a considerably higher output impedance than r_p, but this can be reduced, if desired, by the use of a bypass condenser. The interelectrode capacitances given[1] are $C_{gk} = 1.8$ μμf, $C_{gk} = 1.6$ μμf, and $C_{gp} = 1.3$ μμf. The input capacitance of a triode amplifier with fixed cathode potential is

$$C_{in} = C_{gk} + (1 - G_1)C_{gp} \tag{49}$$

(cf. Sec. 9·2). Here G_1, the gain of the stage from grid to plate, is a negative quantity. The gains of the three stages as calculated in Eqs.
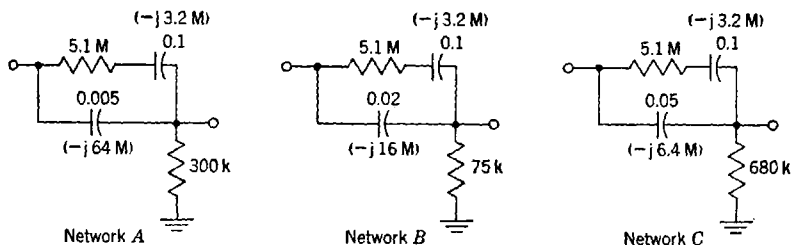
Fig. 9·33.—Low-frequency stabilization networks with reactances at 0.5 cps.

(48a) and (b) are respectively 16, 16, and 11; consequently the corresponding input capacitances are 23.7, 23.7, and 17.2 μμf. Since allowances must be made for stray wiring capacity and the effect of output capacitances of previous stages, safe figures to use are 35, 35, and 25 μμf.[2] The corresponding values of r_p are 10, 10, and 25 kilohms; however, if no condenser is put across the resistor from third plate to first grid, the effective output impedance of the third stage is 600 kilohms. The frequencies at which the phase shifts due to the first and second coupling networks are 45° are 450 and 630 kc/sec respectively. If it is assumed that these phase shifts add to give 90° at about 500 kc/sec, the question arises as to whether or not it is possible to stabilize the amplifier by attenuating with a single RC-network, such as the 600-kilohm resistor combined with a coupling condenser from the third to the first stage. It will be desirable to let the phase shift of the high-frequency networks at 500 cps be a lag of 23° in order to cancel the lead resulting from the low-frequency networks (Fig. 9·32). As was mentioned above, a lag of 45° would be better if the amplifier were capable of com-

---

[1] *RCA Tube Handbook*, Vols. 3 and 4, Radio Corporation of America, Camden, New Jersey, June, 1942.

[2] If pentodes had been used, these figures might have been reduced considerably.

pensating for the corresponding reduction in gain (by a factor of $1/\sqrt{2}$). If the network produces 23° phase shift at 500 cps, its central frequency (of 45° phase shift) must be 1250 cps. This will not quite suffice to attenuate by a factor of $\frac{1}{500}$ before a frequency of 500 kc/sec is reached. If two of the networks are made to attenuate rapidly, they must then have a central frequency such that 500 cps will be the logarithmic mean between it and 110 cps, the corresponding frequency for the low-frequency networks if there is to be zero phase shift at 500 cps. Thus, $500^2/110 = 2270$ cps; and if the third network has a central frequency of 630 kc/sec, stabilization is possible[1] but with a very small safety factor.

The next alternative is to attenuate by means of a single *RC*-circuit and a single phase-retard network. If this is done, it seems desirable to start both phase-shift curves near 500 cps, as was done on the low-
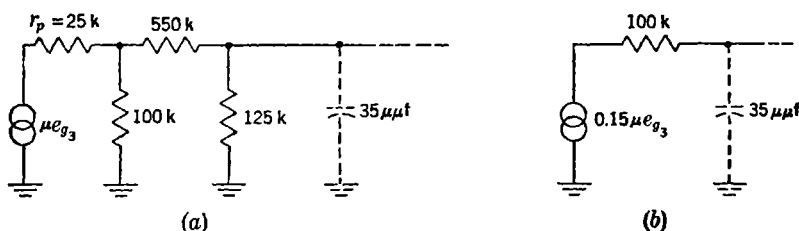
FIG. 9·34.—(*a*) High-frequency network at first grid; (*b*) Thévenin's theorem equivalent circuit.

frequency side. The network from the third plate to the first grid may be transformed by use of Thévenin's theorem, as is shown in Fig. 9·34. Its central frequency, if no additional condensers are used, is 45 kc/sec. The central frequency required for the first two inverse-tangent curves is 2270 cps, as was calculated previously. The procedure for selecting the other inverse-tangent curve in the phase-retard network will be to plot the response exclusive of this curve and then determine its position by a graphical trial-and-error process (Fig. 9·35). This process leads to the conclusion that the central frequency of the final curve should be about 50 kc/sec.

The constants of the networks may now be selected. First, a condenser will be inserted from the first grid to ground in order to produce a central frequency of 2270 cps. The value of this condenser that makes the capacitive reactance equal to the resistance (Sec. 9·3) is

$$C_3 = \frac{1}{2\pi f R} - C_s = \frac{1}{6.28 \times 1.0 \times 10^5 \times 2270} \text{ farad} - 35 \ \mu\mu\text{f}$$
$$= (680 - 35) \qquad \mu\mu\text{f}. \quad (50)$$

[1] Terman, *op. cit.*, p. 398.

A value of 680 $\mu\mu$f will be used.   Next, a phase-retard RC-network as shown in Fig. 9·36 will be introduced into the coupling circuit between the first and second stages.   It has the same form as that discussed in
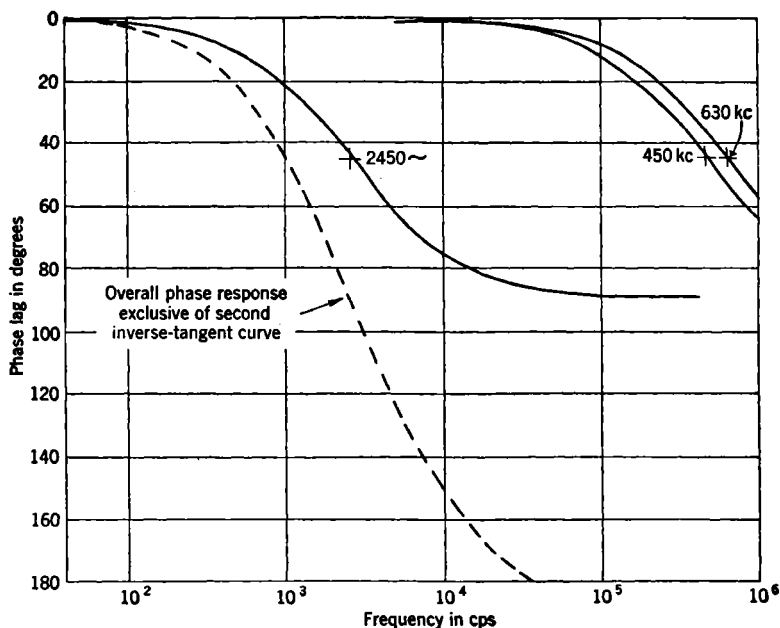

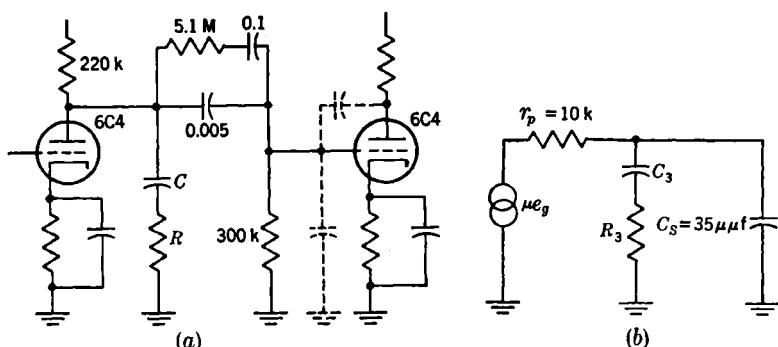FIG. 9·35.—Calculated phase characteristics of high-frequency network.

FIG. 9·36.—Coupling network between first and second stage.   (a) Interstage coupling network; (b) approximate high-frequency equivalent circuit.

Sec. 9·3.   The behavior of the network near 500 cps is determined largely by r_p and C_3 and is characterized by a central frequency of 2270 cps; $r_p = 10$ kilohms; therefore $C_3 = 6800 \ \mu\mu$f, and an approximate value of

0.006 μf is used. The other central frequency is 50 kc/sec and is determined by R_3C_3 of Eq. (25). Therefore

$$R_3 = \frac{1}{6.28 \times 6 \times 10^{-9} \times 5 \times 10^4} = 540 \qquad \text{ohms.}$$

The nearest RMA value, 560 ohms, will be used. The roots p_i and the corresponding central frequencies may now be calculated, using the values $r_{\text{out}} = 10$ kilohms, $C_s = 35$ μμf, $R_3 = 560$ ohms, $C_3 = 0.006$ μf.

$$p_2 = -\frac{1}{R_3C_3}$$

$$p_{3,4} = -\frac{\left(\dfrac{1}{R_3C_3} + \dfrac{1}{r_{\text{out}}C_s} + \dfrac{1}{R_3C_s}\right)}{2}$$

$$\pm \sqrt{\left(\frac{\dfrac{1}{R_3C_3} + \dfrac{1}{r_{\text{out}}C_s} + \dfrac{1}{R_3C_s}}{2}\right)^2 - \frac{1}{R_3r_{\text{out}}C_3C_s}}$$

Therefore

$$-p_2 = 3.02 \times 10^5 \text{ sec}^{-1} \qquad f_2 = 47 \text{ kc/sec}$$
$$-p_3 = 57 \times 10^6 \text{ sec}^{-1} \qquad f_3 = 9.1 \text{ Mc/sec}$$
$$-p_4 = 14{,}900 \text{ sec}^{-1} \qquad f_4 = 2.4 \text{ kc/sec}$$
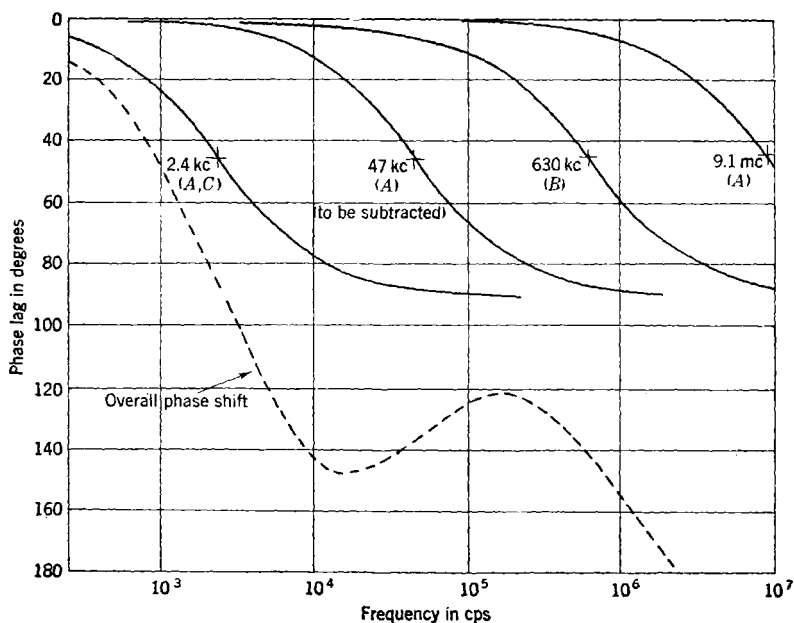
Fig. 9·37.—Over-all phase characteristic of amplifier at high frequencies (calculated).

It may be noted that the limiting high-frequency characteristic for this network has been moved from 450 kc/sec to 9.1 Mc/sec. Whereas the central frequency determined by the plate resistance of the first stage and the input capacitance of the second was 450 kc/sec, the addition of the stabilization network effectively puts a 560-ohm resistance in parallel
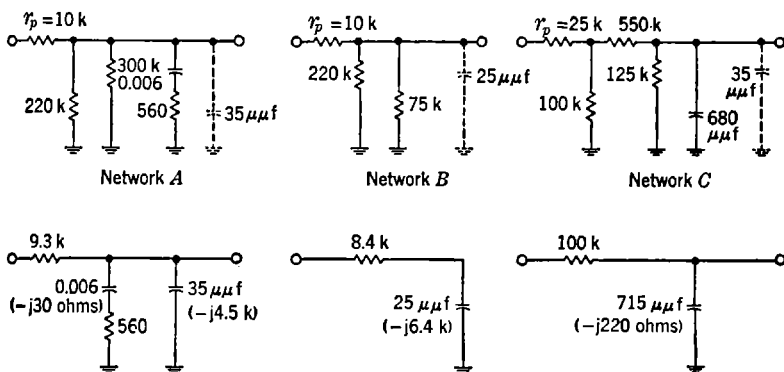
FIG. 9·38.—High-frequency stabilization networks with reactances at 1.0 Mc/sec.

with the input capacitance at high frequencies. The resulting phase response is determined not by the plate resistance of the first stage alone but by the parallel combination of this resistance and 560 ohms. This results from an effect like the one taken into account by means of Thévenin's theorem for the network at the first grid. At sufficiently high
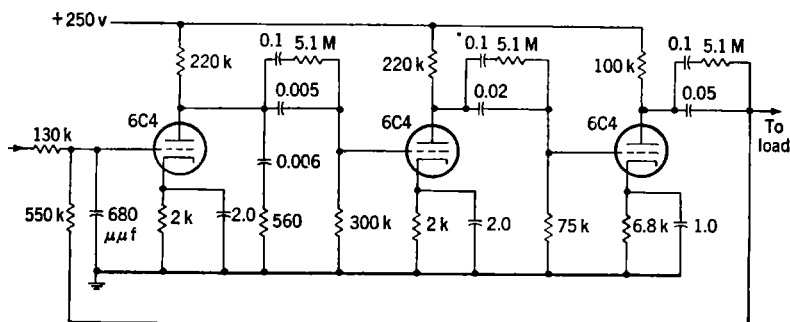
FIG. 9·39.—Design of three-stage feedback amplifier with stabilization networks.

frequencies the impedance of C_3 is small, and the resistances r_p and R_3 form a similar combination. The phase response of the high-frequency networks is plotted in Fig. 9·37. The phase shift is less than 160° up to a frequency of 1 Mc/sec. The attenuation can now be estimated as was done for the low-frequency network. The three interstage networks

are shown in Fig. 9·38, with reactance values calculated at 1 Mc/sec. The plate-load and grid-leak resistances, which have been neglected so far, are included, and it will be seen that their effects are small.   In the case of Network C, the output capacitance of the third triode cannot rigorously be lumped with the input capacitance of the following stage, but this fact has been ignored in getting an approximate result.   The estimated values of attenuation at 1 Mc/sec are $\frac{1}{17}$, $\frac{1}{2}$, and $\frac{1}{150}$ for Networks A, B, and C, respectively.   The corresponding over-all attenuation is 15,000 (84 db).

The final paper design of the amplifier is shown in Fig. 9·39.

9·15. Experimental Checks and Completion of the Design.—A test model of the circuit of Fig. 9·39 was set up.   No oscillations were observed.   The maximum output was about 70 volts rms before distortion was observed at the first grid.   When the 5·1-megohm resistor and 0.1-μf condenser of Network A were disconnected, oscillations occurred at a frequency of 12 cps.   This is about what would be predicted from Fig. 9·31 if the phase-shift curve continued up from 500 cps as would be caused by the two single RC-coupling networks alone.   Disconnecting the 5.1-megohm resistor and the 0.1-μf condenser in both the second and third networks did not produce oscillations.   Disconnecting the 560-ohm resistor and the 0.006-μf condenser in Network A produced oscillations at 180 kc/sec; the curves in Fig. 9·35 labeled 2.4, 450 and 630 kc/sec would add to give 180° phase shift at about 500 kc/sec. When the 680-$\mu\mu$f condenser at the first grid was disconnected, however, no oscillation occurred.   These measurements show that large safety factors were allowed.   They are still no assurance in themselves, however, that if the circuit were produced in quantity, Network A alone could be used for stabilization, because a combination of adverse parameter variations within the allowed tolerances might still cause oscillation.

The measured loop gain was 430, the respective stage gains being 15.5, 13.3, and 11 as against the predicted values of 16, 16, and 11 respectively.   The d-c electrode voltages were roughly as predicted.

The ratio of zero-frequency to infinite-frequency gain as regards the cathode-bypass condensers was found by removing one condenser at a time and measuring gains.   The respective ratios were 1.5, 1.5, and 1.8 as compared with the predicted values 1.22, 1.22, and 1.75, and the measured over-all ratio was 4.0 instead of 2.6.   Some of this discrepancy is probably due to the fact that the R_g's were changed for the stabilization circuits after the calculations of gain ratios were made.   This again means that the precautions taken in low-frequency stabilization were unnecessarily great, for the effective loop gain that had to be reduced by the networks was $\frac{430}{4} = 107$ rather than 190.

The loop gain and loop phase shift were measured over a limited

frequency range (10 cps to 30 kc/sec) and were found to agree with the predicted values.

To complete the circuit diagram, it is necessary to include tolerances and wattages and to specify parts that may be procured. This will not be done in the present case, except to state that every condenser or resistor except the two precision resistors at the first grid could probably be assigned a tolerance of ±20 per cent.

The design of this circuit is by no means complete for production. It should be laid out and built with the proper physical arrangement of parts, and this model should be tested for oscillation, mounting rigidity, life of tubes, etc. The problem of parasitic oscillations that arise from parameters not represented in the circuit diagram (stray wiring capacities, lead inductances) must be attacked separately. If the additional parameters can be represented as lumped constants, an analysis can be made, but the usual procedure is to insert experimentally small shunting capacitors or small series resistors in grid circuits in order to suppress such oscillations.

In designs of circuits to suppress oscillations, it is desirable to have large safety factors. This not only reduces the chance of oscillation but makes it possible to select circuit components that are manufactured to wider tolerances.

The high-frequency network in this circuit, a single phase-retard network, was more effective in reducing gain than was the low-frequency one, which involved three phase-advance networks. This is shown by the fact that over a frequency range of a factor of 2000 (500 cps to 1 Mc/sec) the high-frequency network attenuated by a factor of 15,000 whereas over a frequency factor of 1000 ($\frac{1}{2}$ to 500 cps) the low-frequency network attenuated by a factor of only 4000.

It is of interest to note that the asymptotic characteristic or limiting phase-shift curve is not a fixed limit. The use of low-frequency or high-frequency stabilization networks, as in this circuit, has the effect of pushing both asymptotes farther out (for different reasons in the two cases).

CHAPTER 10

LOW-FREQUENCY FEEDBACK AMPLIFIERS

By Harold Fleisher

10·1. Frequency-selective Networks.—Inverse feedback can be used in the design of low-frequency selective amplifiers to attain a desired over-all characteristic, but this characteristic is not the flat response of the type desired in audio amplifiers. In this case, a network that rejects a certain frequency range can be used in the feedback loop to produce maximum over-all transmission in that frequency range. The circuit, then, is an electronic bandpass filter, which for sufficiently low frequencies (for example, 30 cps) may be of much lighter weight than the corresponding *LC*-filter. Staggered tuning (Chap. 5) can be used to sharpen the bandpass characteristics.

The lattice or bridge network is the most general selective network that can be used in the feedback loop of an amplifier to produce a desired frequency-selective characteristic. Balance of the bridge at a particular frequency is the proper null condition for the rejection of that frequency, and by suitable choice of components this balance can be obtained at any desired frequency. The Wien bridge is an example of such a network.

Bridge networks, however, do not have a common ground between their input and output terminals, and this feature either limits the type of circuit in which they can be used or requires the use of additional coupling devices such as transformers. The latter alternative may be undesirable for very low-frequency applications. Circuits such as the bridged-T and parallel-T (or twin-T) networks, which are equivalent to the Wien-bridge network, are three-terminal networks having a common ground for both input and output terminals and are thus more useful. These networks, however, have several inherent limitations, perhaps the most serious of which is the greater interdependence of the parameters that must be adjusted to set their rejection frequency. This interdependence causes component tolerance specifications to become more critical; and if it is desired to vary the rejection frequency, more interdependent controls are required. For example, to adjust the rejection frequency of a Wien bridge continuously and over a fairly wide range requires that two components (two capacitors or resistors) be simultaneously varied, whereas similar adjustment of a twin-T network requires that three components be simultaneously varied.

384

The networks to be discussed in this section are, specifically, the Wien-bridge network, a bridged-T network involving a coil the Q of which may be used as a parameter to vary its selectivity characteristic, and the twin-T network (see Fig. 10·1).

The mathematical basis for the following discussion will be the transfer function, which will be determined for no-load operation of the network. This function is defined as the ratio of the output voltage to the input voltage and will be denoted by β. Other characteristics, such as input and output impedances, will be of secondary importance, because the networks will be used under no-load conditions.

Such networks may be easily analyzed for β by the use of nodal analysis. Thus, if Terminals 1 and 3 are the input and output terminals of an arbitrary four-terminal network that has a common ground for the
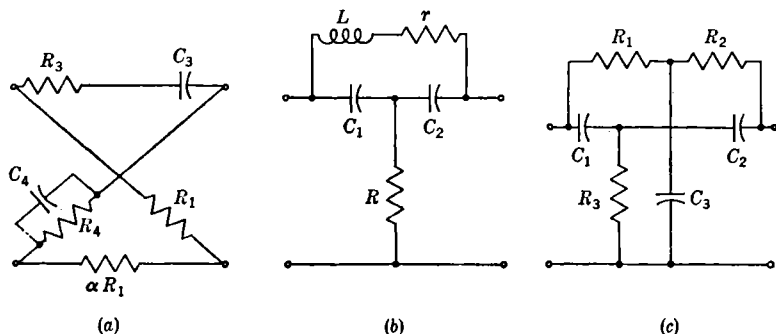
FIG. 10·1.—Frequency-selective networks. (a) Wien-bridge network; (b) bridged-T network; (c) twin-T network.

input and output terminals, and if Δ is the characteristic determinant of the network,[1] then

$$\beta = \frac{e_0}{e_i} = \frac{\Delta_{13}}{\Delta_{11}}, \tag{1}$$

where Δ_{13} and Δ_{11} are the cofactors of Δ, including the proper sign.

Bridged-T Network.—For the bridged-T network in Fig. 10·1b,

$$\Delta = \begin{pmatrix} pC + \dfrac{1}{r + pL} - pC & -pC & -\dfrac{1}{r + pL} \\ -pC & \dfrac{1}{R} + 2pC & -pC \\ -\dfrac{1}{r + pL} & -pC & pC + \dfrac{1}{r + pL} \end{pmatrix} \tag{2}$$

where $p = j\omega$;

[1] See Chap. 1.

and

$$\beta = \frac{e_0}{e_i} = \frac{\begin{vmatrix} -pC & 2pC + \dfrac{1}{R} \\[2mm] -\dfrac{1}{r+pL} & -pC \end{vmatrix}}{\begin{vmatrix} 2pC + \dfrac{1}{R} & -pC \\[2mm] -pC & pC + \dfrac{1}{r+pL} \end{vmatrix}}. \tag{3}$$

The rejection frequency ω_0, at which the null in transmission occurs, is obtained by setting $\Delta_{13} = 0$, but only if $\Delta_{11} \neq 0$. The two conditions imposed on the components to obtain the null (obtained by setting the real part and the imaginary part of Δ_{13} equal to zero) are[1]

$$\omega_0^2 = \frac{1}{rRC^2}, \qquad \text{and} \qquad \omega_0^2 = \frac{2}{LC}. \tag{4}$$

If the Q of the coil at the resonant frequency is used as a parameter $(Q_0 = \omega_0 L/r)$, the two conditions imposed on the network for a null in transmission become

$$Q_0 = \omega_0 2RC, \qquad \omega_0^2 = \frac{2}{LC}. \tag{5}$$

Thus ω_0 is the resonant frequency of the inductance combined with the two series capacitors. If ρ is defined equal to ω/ω_0 and Q_0 is used as a parameter, β for the bridged-T network in Fig. 10·1 may be written

$$\beta_{BT} = \frac{1}{1 - j\dfrac{\rho}{\rho^2 - 1}\left(\dfrac{2}{Q_0}\right)}. \tag{6}$$

Equations (5) and (6) provide most of the network information required to design a frequency-selective amplifier.

Wien-bridge Network.—The null conditions and the value of β for the Wien-bridge network may be obtained in a similar manner.

$$\left.\begin{aligned} \omega_0^2 &= \frac{1}{R_3 R_4 C_3 C_4} = \frac{1}{R^2 C^2} \\[2mm] \frac{1}{\alpha} &= \frac{R_3}{R_4} + \frac{C_4}{C_3} = 2 \end{aligned}\right\}; \tag{7}$$

---

[1] There is no loss in generality in assuming both capacitors equal except in studying the effects of capacitor variability. Equation (4), however, may be written as follows:

$$\omega_0^2 = \frac{1}{rRC_1 C_2}, \qquad \text{and} \qquad \omega_0^2 = \frac{C_1 + C_2}{LC_1 C_2}.$$

$$\beta_{WB} = -\frac{1}{3}\frac{1}{1 - j\frac{3\rho}{\rho^2 - 1}}, \tag{8}$$

where again $\rho = \omega/\omega_0$.[1] The factor $-\frac{1}{3}$ may be neglected in considering the frequency response of the network; but because it is part of the voltage ratio, it must be taken into account when the Wien bridge is to be used in a specific circuit.

Twin-T Network.—The null frequency of the twin-T network is determined by the two equations

and

$$\left.\begin{aligned}\omega_0^2 &= \frac{1}{(R_1 + R_2)R_3C_1C_2},\\[2mm]\omega_0^2 &= \frac{C_1 + C_2}{C_3C_1C_2R_1R_2},\end{aligned}\right\} \tag{9}$$

from which, by eliminating frequency, the null condition as a function of the component parameters may be obtained.

$$\frac{\dfrac{R_1R_2}{R_1 + R_2}}{R_3} = \frac{C_1 + C_2}{C_3} = n. \tag{10a}$$

Because of the separation of variables, n is a real number that may vary from 0 to ∞. It has an optimum value, $n = 1$, although other considerations may require using values of n other than unity, such as $n = 2$ or $n = \frac{1}{2}$.

If $R_1 = R_2 = R$ and $C_1 = C_2 = C$, then

$$R_3 = \frac{R}{2n}, \qquad C_3 = \frac{2C}{n}; \tag{10b}$$

and the transmission of the twin-T network becomes

$$\beta_{TT} = \frac{\rho^2 - 1}{(\rho^2 - 1) - j2\rho\left(\dfrac{n + 1}{\sqrt{n}}\right)}. \tag{11a}$$

If $C_1 = C_2 = C$ is given an arbitrary value, and if $R_1 = R_2$, the circuit may be "balanced" at any prescribed frequency by variation of the resistances alone, because then Eqs. (9) both become

$$\omega_0^2 = \frac{n}{R^2C^2},$$

[1] See Fig. 10·1a for definition of α.

and they are therefore simultaneously satisfied. To achieve balance at an arbitrary frequency, adjustments of two resistors are required; to achieve balance at a prescribed frequency, three resistors must be adjusted.

The optimum value of n may be defined as the value of n for which $|\beta|$ has the steepest slope at the rejection frequency $f_0 = \omega_0/2\pi$. Since

$$\left(\frac{d|\beta|}{d\rho}\right)_{\rho=1} = \frac{\sqrt{n}}{(n+1)},$$

the maximum slope is determined by setting $d/dn \, (d|\beta|/d\rho)_{\rho=1}$ equal to zero, from which it is readily determined that the maximum slope of the attenuation characteristic occurs when $n = 1$. Then

$$\beta_{TT} = \frac{1}{1 - j\dfrac{4\rho}{\rho^2 - 1}}. \tag{11b}$$

Equation (11b) will be used throughout the remainder of this chapter.

It must be noted here that the slope of $|\beta|$ at $\rho = 1$ is a relatively slowly varying function of n and other considerations may require choices of n other than $n = 1$. Thus, if $n = 2$ in Eq. (10b), $R_3 = R/4$, $C_3 = C$, and all three capacitors become equal in value. This makes it possible to couple all the capacitors mechanically and have a continuous variation of rejection frequency. On the other hand, although the maximum slope is -0.5 at $n = 1$, the slope decreases only to -0.472 at $n = 2$. This is approximately a 6 per cent change. Similarly, at $n = \frac{1}{2}$, $R_3 = R$, $C_3 = 4C$, and the maximum slope also decreases to -0.472.

Inspection of Eqs. (6), (8), and (11b) indicates their similarity in form. Equation (11b), the transmission of the twin-T network, may therefore be used as a typical example. In polar form, $\beta = |\beta|e^{j\theta}$, and from Eq. (11b)
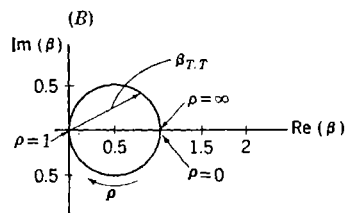
Fig. 10·2.—Phase-amplitude diagram of twin-T network.

$$\left.\begin{aligned}|\beta| &= \frac{1}{\sqrt{1 + \dfrac{16\rho^2}{(\rho^2 - 1)^2}}} \\[2mm] \theta &= \tan^{-1}\frac{4\rho}{\rho^2 - 1}\end{aligned}\right\} \tag{12}$$

Then $|\beta|$ may be expressed as a function of θ: $|\beta| = \cos\theta$, which is seen to be the polar equation of a circle of radius $\frac{1}{2}$, tangent to the imaginary axis and with center at $(\frac{1}{2}, 0)$ (see Fig. 10·2). From Fig. 10·2 it may be seen that as ρ increases from 0 to 1 (ω from 0 to ω_0), θ varies from 0 to $-\pi/2$; and as ρ increases from 1 to ∞ (ω from ω_0 to ∞), θ varies from
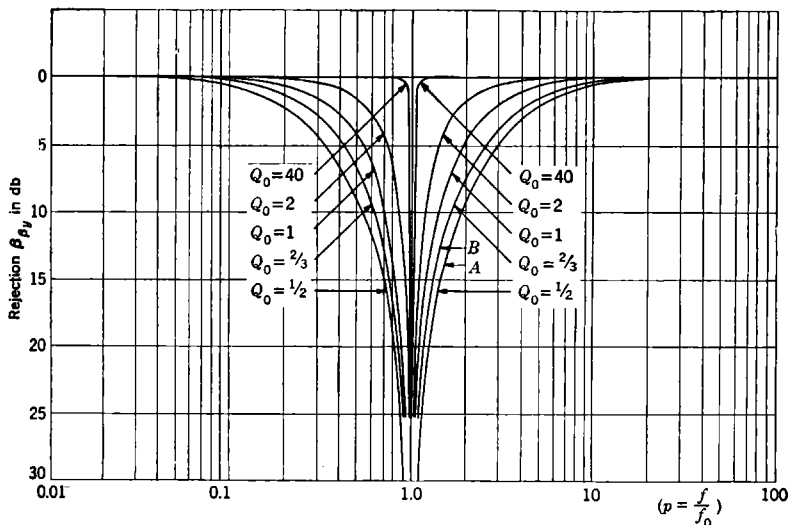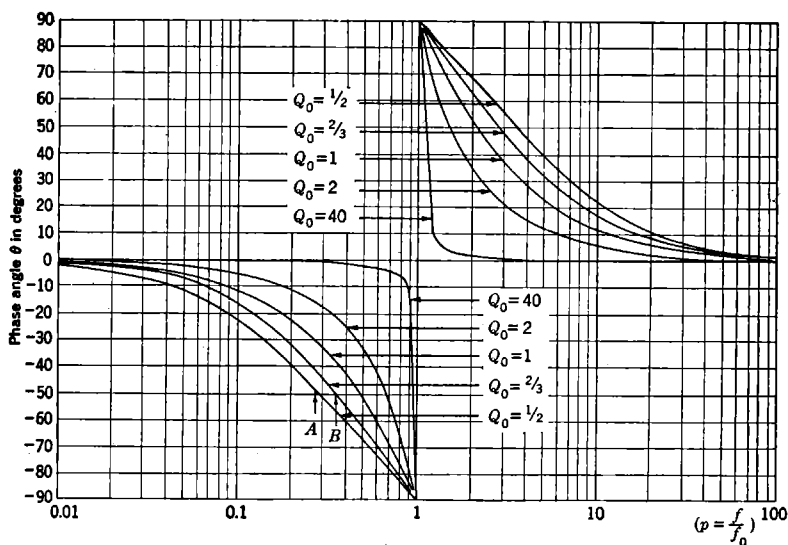
Fig. 10·3.—Characteristics of rejection networks. The values of Q_0 pertain to the bridged-T network; curve A pertains to a twin-T network for which $n = 1$; curve B pertains to a Wien-bridge network. (a) Phase characteristics; (b) amplitude characteristics.

$-3\pi/2$ to -2π (or $+\pi/2$ to 0).   Thus a discontinuity in phase exists at $\rho = 1$.

In Fig. 10·3 are shown the phase and amplitude characteristics of rejection circuits of this type.   The foregoing analysis shows that, whereas in the Wien-bridge and twin-T circuits the exact conditions for maximum sharpness of the rejection band are completely realizable (as, for example, by setting $n = 1$ in the twin-T network), in the case of the bridged-T network the sharpness of the rejection band is determined by the Q of the choke coil employed, and this is set by practical and not by theoretical considerations.   Therefore in Fig. 10·3 are shown characteristics for a bridged-T network employing chokes of several Q values, but only a single curve corresponding to the condition for narrowest rejection band is shown for each of the Wien-bridge and twin-T networks.   It will be seen that the Wien-bridge network corresponds to a bridged-T network whose choke has a Q of $\frac{2}{3}$ and that the twin-T network corresponds to a bridged-T network whose choke has a Q of $\frac{1}{2}$.

The practical possibility of achieving complete rejection at the null frequency can be discussed on the basis of the null conditions determined by the component parameters of the network.   The effects of variability of components due to temperature effects, aging, and manufacturing tolerances must be considered if a given frequency is to be rejected.   If the rejection frequency is to be adjustable over a range of values, the additional problem of simultaneously varying the values of two or more components is introduced.   This second problem may be considerably simplified by the proper choice of component values and of suitable mechanical couplings.

If the twin-T network is used as an example, it is seen from Eqs. (9) that[1] if ω_0 is assumed to be constant and $C_1 \approx C_2$, $R_1 \approx R_2$,

$$\left. \begin{aligned} \frac{\delta R_2}{R_2} &\approx -\left[\frac{\delta R_1}{R_1} + \frac{\delta C_3}{C_3} + \frac{1}{2}\left(\frac{\delta C_1}{C_1} + \frac{\delta C_2}{C_2} \right) \right]. \\ \frac{\delta R_3}{R_3} &\approx \frac{1}{2}\frac{\delta C_3}{C_3} - \frac{3}{4}\left(\frac{\delta C_1}{C_1} + \frac{\delta C_2}{C_2} \right). \end{aligned} \right\} \tag{13}$$

Temperature, tolerance, and aging variations can all be considered by means of Eqs. (13).   By inspection of Eqs. (13), it can be seen that for the ideal case the temperature coefficients of the resistors and capacitors should be equal and opposite, i.e., the capacitors may have a positive temperature coefficient (such as silver mica), and the resistors may have a negative temperature coefficient (such as precision carbon resistors).

---

[1] If R_1 and R_2 are equal within 10 per cent (and the same is true for C_1 and C_2), then Eqs. (13) are accurate within 0.25 per cent.

Commercial components are available, however, that have temperature coefficients of less than ± 50 parts per million per degree centigrade. For many purposes, the probable operating-temperature range is so small that the use of such components will introduce negligible temperature effects.

The choice of the tolerances to which components must be held is determined by the following factors: expense and ease of obtaining the component, facility of adjustment, and required accuracy of adjustment. Components of better than ± 1 per cent accuracy are generally expensive and difficult to obtain. However, as the precision of the component increases, the amount of adjustment required decreases.

Aging is an unknown factor for many components and is determined, among other things, by conditions of use. The effects of aging can be minimized if well-constructed, stabilized components are used.

If sufficiently accurate frequency standards are available, ± 1 per cent or ± 2 per cent components are usable and represent a good compromise between the use of precision components and ease of adjustment. The "ease of adjustment" may be arbitrarily defined as being inversely proportional to the percentage range that the variable components must cover to maintain the null at a given frequency.

In Fig. 10·1c, R_2 may be split into two resistors R_2' and R_2'', and R_3 into R_3' and R_3''. R_2'' and R_3'' are then the trimmer resistors which are determined within the limits prescribed by Eqs. (13). Thus, if ± 1 per cent resistors and ± 2 per cent condensers are used, the conditions on R_2'' and R_3'' are readily determined from Eqs. (13) to be

$$\left.\begin{array}{l} 0 \leqq R_2'' \leqq 0.08R, \\ 0 \leqq R_3'' \leqq 0.06\,\dfrac{R}{2} = 0.03R. \end{array}\right\} \tag{14}$$

The null discussed in this section was assumed to mean zero transmission, but in the next section the actual attenuation required will be discussed in relation to the requirements of frequency-selective amplifiers.

10·2. Frequency-selective Amplifiers.—The frequency-selective amplifiers to be discussed in this section can be qualitatively described as having a frequency characteristic roughly corresponding to the inverse of that of the rejective network and similar to that of a single-tuned RLC-network. A quantitative discussion will be given

Fig. 10·4.—Block diagram of rudimentary feedback amplifier.

under the following assumptions: A single feedback loop is used through the rejective network; the amplifier without feedback is stable at all

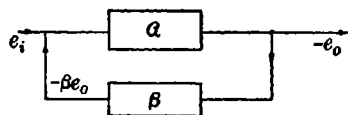frequencies; and the output signal applied to the feedback loop is π radians out of phase with the input signal.[1]

The simplified block diagram of Fig. 10·4, where β may be chosen as the transmission of any of the networks discussed in Sec. 10·1, then applies. If $\mathcal{G}$ is the over-all gain of the amplifier, then

$$\mathcal{G} = \frac{\alpha}{1 + \beta\alpha}, \tag{15}$$

which is the customary feedback formula. In the twin-T network, for example, β may be chosen as

$$\beta_{TT} = \frac{1}{1 - j\dfrac{4}{\rho - \dfrac{1}{\rho}}} = \frac{1}{1 - j\dfrac{4}{u}},$$

where $u = \rho - 1/\rho$. Then,

$$\frac{\mathcal{G}}{\alpha} = \frac{1}{1 + \dfrac{\alpha}{1 - j\dfrac{4}{u}}} = \frac{1 - j\dfrac{4}{u}}{(\alpha + 1) - j\dfrac{4}{u}}. \tag{16}$$

If the bridged-T network is used,

$$\beta_{BT} = \frac{1}{1 - j\dfrac{2}{Q_0 u}}$$

and[2]

$$\frac{\mathcal{G}}{\alpha} = \frac{1}{1 + \dfrac{\alpha}{1 - j(2/Q_0 u)}} = \frac{1 - j\left(\dfrac{2}{Q_0 u}\right)}{(\alpha + 1) - j\left(\dfrac{2}{Q_0 u}\right)}. \tag{17}$$

From either Eq. (16) or (17), it may be shown that $\mathcal{G}/\alpha$ defines a circle in the complex plane with center on the real axis at $\frac{1}{2}(\alpha + 2)/(\alpha + 1)$, having a radius of $\frac{1}{2}\alpha/(\alpha + 1)$ (see Fig. 10·5). The ratio $\mathcal{G}/\alpha$ is unity at $\rho = 1$, ($\omega = \omega_0$) and is a minimum at zero and infinite frequencies. This minimum is readily seen to be $1/(\alpha + 1)$. As α approaches

---

[1] α may then be taken as a positive real number and equal to the gain of the amplifier without feedback.

[2] The over-all gain of an amplifier using either a Wien-bridge or a twin-T network may be obtained from this equation. Thus, if $Q_0 = \frac{1}{2}$, $\mathcal{G}/\alpha$ is given for an amplifier employing a twin-T network; and if $Q_0 = \frac{2}{3}$, $\mathcal{G}/\alpha$, is given for an amplifier employing a Wien-bridge network.

infinity, the radius of the circle approaches $\frac{1}{2}$, and the circle becomes tangent to the imaginary axis.

For comparison, the "selectivity" of the RLC-network shown in Fig. 4·2, may be defined as $Z(u)/R \equiv \gamma_s = 1/(1 + jQu)$, from Eq. (4·1) where $u = (\omega/\omega_0) - (\omega_0/\omega)$. The quantity γ_s defines a circle of unit diameter tangent to the imaginary axis.

The width of the pass band of the series-resonant circuit may be specified in terms of the frequencies at which the power transmitted is half that transmitted at resonance, i.e., at which $\gamma_s^2 = \frac{1}{2}$. It is well known that these frequencies are simply related to the Q of the series-resonant circuit, being given by $1/u = \pm Q_0$ where $Q_0 = \omega_0 L/R$. This makes it convenient to specify the bandwidth directly in terms of Q.

FIG. 10·5.—Phase-amplitude diagram of feedback amplifier employing bridged-T network.

The fact that both γ_s and $\mathcal{G}/\alpha$ define circles in their respective complex planes indicates the similarity in shape of their phase and amplitude characteristics. It is therefore reasonable to define a Q-factor for the feedback amplifier in terms of the frequencies at which $(\mathcal{G}/\alpha)^2 = \frac{1}{2}$. From Eq. (16),

$$\left|\frac{\mathcal{G}}{\alpha}\right|_{TT} = \sqrt{\frac{1 + \dfrac{16}{u^2}}{(\alpha + 1)^2 + \dfrac{16}{u^2}}} \tag{18a}$$

At the half-power frequencies, ω_1 and ω_2, $\left|\dfrac{\mathcal{G}}{\alpha}\right| = \dfrac{1}{\sqrt{2}}$; and from Eq. (18a),

$$u_1^2 = u_2^2 = \frac{16}{(\alpha + 1)^2 - 2}.$$

This can be rewritten as

$$\frac{1}{u_1} = \frac{1}{u_2} = \frac{\sqrt{(\alpha + 1)^2 - 2}}{4}.$$

By analogy with the RLC-network, $Q = 1/u_1 = 1/u_2$. Thus.

$$Q_{TT} = \frac{\sqrt{(\alpha + 1)^2 - 2}}{4}.$$

Since, in general, $(\alpha + 1) > 50$, the numerator can be expanded in a power series, from which it follows that with an error of less than 0.04

per cent,

$$Q_{TT} = \frac{\alpha + 1}{4}. \tag{18b}$$

Similarly, from Eq. (17), the Q of the amplifier using a bridged-T network in the feedback loop is

$$Q_{BT} = \left(\frac{\alpha + 1}{2}\right) Q_0, \tag{19}$$

where Q_0 is the Q of the coil at the resonant frequency ω_0.

The same definition of Q may be applied to the rejective networks themselves; the Q of the twin-T network is found from Eq. (11b) to be $\frac{1}{4}$ and for the bridged-T network to be $Q_0/2$.

It is to be noted that the phase and amplitude characteristics of these amplifiers for the moderate gain (about 40) required to yield a Q of 10 are identical with those of the RLC-network, not only at the resonant frequency ω_0 but also for an appreciable frequency range about the resonant frequency. With a Q of 10 the minimum transmission obtainable (theoretically at zero and infinite frequencies if a perfect amplifier is assumed) is $\frac{1}{40}$ of the maximum transmission at ω_0 or minus 32 db. As α approaches infinity, the characteristics of the frequency-selective amplifier approach those of the RLC-network in both phase and magnitude. Using the Q defined in Eq. (18b) as a parameter, Eq. (18a) may be written as

$$\left|\frac{G}{\alpha}\right| = \sqrt{\frac{1 + \dfrac{u^2}{4^2}}{1 + Q^2 u^2}}, \tag{20}$$

where $u = \rho - 1/\rho$.   The corresponding phase angle is

$$\Theta = \tan^{-1} \frac{1}{Qu} - \tan^{-1} \frac{4}{u}.$$

"Universal" resonance and phase curves for a frequency-selective amplifier using a twin-T network in a degenerative feedback loop may now be plotted from the following equations, using Q as a parameter.

$$\left|\frac{G}{\alpha}\right|_{db} = -10 \log_{10} (1 + Q^2 u^2) + 10 \log_{10}\left(1 + \frac{u^2}{16}\right), \tag{21a}$$

$$\Theta = \tan^{-1} \frac{1}{Qu} - \tan^{-1} \frac{4}{u}. \tag{21b}$$

These curves are illustrated in Fig. 10·6.   Equations (21) apply only to a frequency-selective amplifier using a twin-T network in the feedback loop or to a bridged-T network for which $Q_0 = \frac{1}{2}$. Similar equations, which are more general since they involve the use of two parameters
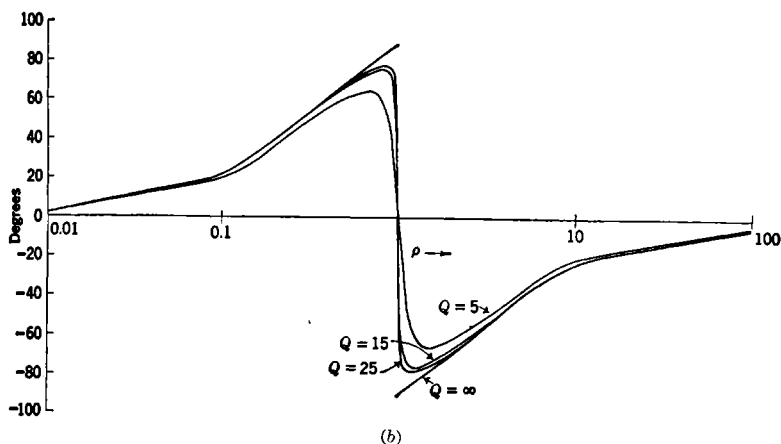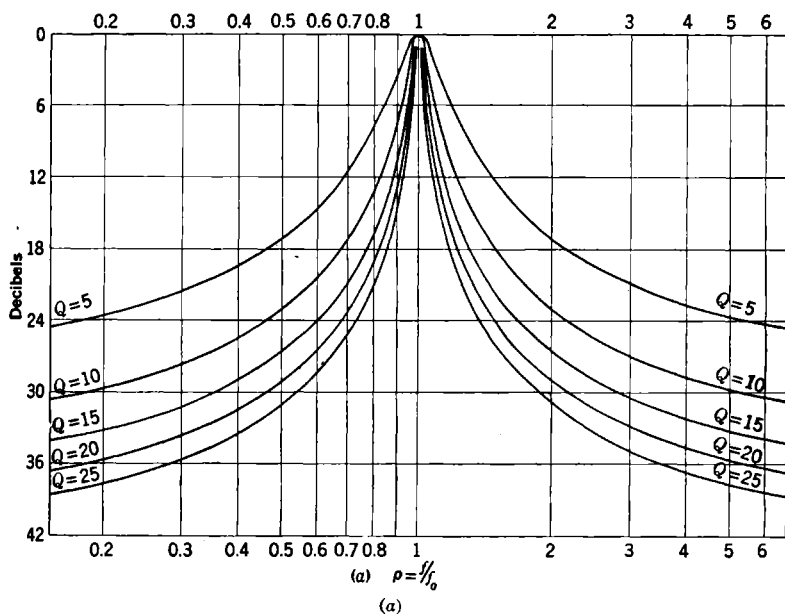
FIG. 10·6.—Resonance and phase curves for a frequency-selective amplifier using a twin-T network in a degenerative feedback loop. (a) Resonance curves; (b) phase curves.

(Q_0 of the coil and Q of the amplifier), may be obtained for amplifiers using a bridged-T network,

$$\left|\frac{\mathcal{G}}{\mathcal{C}}\right|_{db} = -10 \log_{10}(1 + Q^2 u^2) + 10 \log_{10}\left[1 + \left(\frac{Q_0}{2}\right)^2 u^2\right]. \quad (22a)$$

$$\Theta = \tan^{-1}\frac{1}{Qu} - \tan^{-1}\frac{1}{\left(\dfrac{Q_0}{2}\right)u}. \quad (22b)$$

Equation (21a) simplifies to the expression for the transmission of the *RLC*-network,

$$|\gamma_s|_{db} = -10 \log_{10}(1 + Q^2 u^2),$$

with an accuracy of about 1 per cent if $|u| \leqq 0.57$. This condition restricts the frequency range to $0.76 \leqq \rho \leqq 1.32$. Equation (22a) may be similarly simplified to the same order of accuracy if $|u| \leqq 0.28/Q_0$.
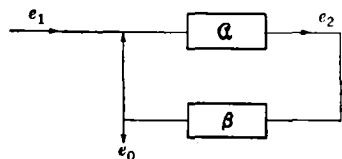
Fig. 10·7.—Block diagram of rejection amplifier.

A special case of the frequency-selective amplifier is the rejection amplifier that effectively "sharpens" the null of the network used in the feedback loop. The transmission of such an amplifier is seen from Fig. 10·7 to be

$$\mathcal{G}_r = \frac{\alpha\beta}{1 + \alpha\beta}. \quad (23)$$

The Q of the rejection amplifier using a twin-T network is approximately $(\alpha - 1)/4$ to the same order of accuracy as was determined for the selective amplifier. Similarly, the Q of the rejection amplifier using a bridged-T network is $Q = Q_0/2(\alpha - 1)$. Thus the frequency discrimination of the network is improved by a factor nearly equal to the gain of the amplifier without feedback.

Network-attenuation Requirements of Frequency-selective Amplifiers.— The actual attenuation of the network at the null frequency as related to the gain of the frequency-selective amplifier without feedback (i.e., at the null frequency) can be discussed by means of the following approximate equation derived for the amplifier with a twin-T network in the feedback loop:

$$\left|\frac{\mathcal{G}}{\mathcal{C}}\right|_{(\rho=1)} \approx 1 + 2\sqrt{2}\,Q|\beta|_{(\rho=1)}, \quad (24)$$

where Q is the Q of the frequency-selective amplifier. A value for $|\beta|_{(\rho=1)}$ slightly different from zero may be obtained by varying a component parameter from the nominal value required for a null. The feedback ratio β may then be expressed as a function of frequency and

of this parameter. The required gain and transmission of the amplifier having been decided upon, the required attenuation of the network at the null frequency is then determined by Eq. (24). This, in turn, determines the amount of adjustment required, as discussed in Sec. 10·1.

The gain and transmission requirements are a consequence of both the Q desired and the over-all stability of the amplifier as determined by its Nyquist diagram.[1] The Nyquist diagram of these amplifiers is essentially that of the β diagram shown in Fig. 10·2. The β of the diagram must be multiplied by α, but at the null frequency $\alpha\beta \approx 0$, and at other frequencies it remains positive. If, for example, the maximum variation of the transmission at the null frequency is allowed to be approximately 10 per cent in order to be safely within stability requirements, and the Q desired is assumed to be 20, then

$$2 \sqrt{2} \, Q|\beta|_{(\rho=1)} = \frac{1}{10}, \qquad \text{and} \qquad |\beta|_{(\rho=1)} = \frac{1}{10} \frac{1}{2 \sqrt{2}Q}.$$

Thus, $|\beta|_{(\rho=1)} \approx 1/600$, which represents an attenuation of about 56 db.

Because an attenuation of approximately 60 db is difficult to measure, the rejection amplifier may be used to measure the actual attenuation of the network at the null frequency. This amplifier not only improves the frequency discrimination of the network but also raises the voltage level of the observed signal by a factor proportional to the gain of the amplifier. Thus, from Eq. (23) the attenuation of the rejection amplifier at the null frequency can be obtained in a manner similar to the derivation of Eq. (24).

$$|\mathcal{G}_r|_{(\rho=1)} \approx 4Q_r|\beta|_{(\rho=1)}, \tag{25}$$

where Q_r is the Q of the rejection amplifier. If, now, the Q of the rejection amplifier is equal to the Q of the frequency-selective amplifier, the following equation is obtained:

$$\left|\frac{\mathcal{G}}{\alpha}\right|_{(\rho=1)} \approx 1 + \frac{|\mathcal{G}_r|_{(\rho=1)}}{\sqrt{2}}. \tag{26}$$

Under the conditions of the previous example, namely, that $Q = 20$ and that the variation in gain be approximately 10 per cent, it is seen that $|\mathcal{G}_r|_{(\rho=1)} = \sqrt{2}/10$, which represents an attenuation of 17 db, and that, from Eq. (25), $|\beta| = |\mathcal{G}_r|/80$ which, again, represents an attenuation of 56 db. Thus the null required of the network, although relatively difficult to measure directly, is easily determined by means of a rejection amplifier of the same Q (i.e., same gain without feedback) as the frequency-selective amplifier in which the network will be used.

[1] See H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945, for a complete discussion of the Nyquist diagram and the design criteria that it imposes.

10·3. The Design of Frequency-selective Amplifiers.—According to the simplified theory discussed in the preceding sections, the amplifier must fulfill the following general conditions: (1) The bandpass characteristic must be essentially flat in the frequency range within which the selected frequency will lie; and (2) the amplifier must be Class A, that is, essentially linear.

Condition 1 means that nowhere within the frequency range should the transmission attenuate at more than a few (≈ 2) decibels per octave and beyond this range it should be sufficiently flat so that the over-all $G\beta$ characteristic falls off at less than 12 db/octave.[1] Because the
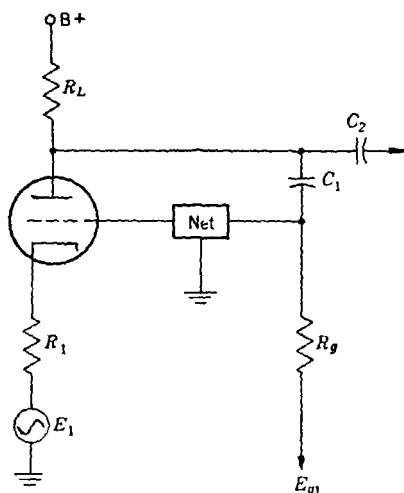
FIG. 10·8.—Single-stage frequency-selective amplifier.

operating-frequency range is limited, the coupling networks are simple and in many cases consist simply of an *RC*-network. Amplifiers of more than three stages require more complicated coupling devices, such as decoupling stages designed to eliminate undesirable feedback loops.

Condition 2 requires that the tubes be operated in the linear part of their characteristics, but for many purposes Class AB operation is allowable.

The "no-load" condition on the network can be obtained by coupling it directly to the grid. Where necessary, mixing networks may be used if they approach the no-load condition by having an impedance of at least 2.5 times that of the output impedance of the network. However, the mixing network not only may introduce an undesirable phase shift

[1] See Bode, *op. cit.* An attenuation of 6 db/octave will result in 90° phase lag for minimum-phase-shift networks.

but will also attenuate the signal to the grid, producing incomplete feedback.

The following discussion of the design of a typical single-stage frequency-selective amplifier will illustrate the requirements imposed on such amplifiers under the given conditions of a specified Q and a definite null frequency. The schematic diagram of the amplifier is shown in Fig. 10·8. As a specific example, let the Q be 15 and the null frequency be 30 cps. Furthermore, let the network used be a twin-T network. This network has the advantage of not requiring a choke, which may be prohibitively large and heavy at 30 cps. It is also assumed that the network has been adjusted to a null as determined in the previous sections. The tube shown is a triode but may be a pentode if necessary. The no-load condition on the twin-T network is obtained by tying it directly to the grid. The proper grid bias is maintained by the direct coupling through the network. The input signal generator is assumed to be of low impedance, and the input signal E_1 may therefore be applied to the cathode.

At the midfrequency of 30 cps, the gain of the amplifier without the selective degenerative feedback is given as

$$\mathbb{G} = \frac{\mu R_L}{r_p + R_L + (\mu + 1)R_1},\tag{27}$$

where r_p is the dynamic plate resistance of the tube and μ is its amplification factor. Because, from Eq. (18b), the gain required is four times the Q, i.e., $\mathbb{G} = 60$, a triode must have a fairly high μ and r_p in order to be used. Many triodes of the receiving type cannot qualify. Many duplex tubes, however, contain such triode components, and many receiving and uhf pentodes, when triode-connected, perform adequately. A typical example might be a tube with a μ of 100 and an r_p of 100,000 ohms. If, in Eq. (27), the impedance R_1 of the generator is neglected, the load resistor R_L is found to be 150,000 ohms. If this value is used to determine a load line, the $B+$ and bias voltages necessary for proper operation of the amplifier can be found.

The high-frequency response of this amplifier can be neglected, but the low-frequency response is extremely important. The constants of the coupling network C_1 and R_g must be so chosen that its 3-db point occurs at a frequency so much less than the null frequency of the network that the additive phase shifts cannot cause the amplifier to oscillate. A conservative design criterion would be to have the 3-db frequency of the coupling network approximately one-tenth the null frequency of the twin-T network. Then the phase lead introduced by the coupling network at the null frequency would be approximately 10°, whereas the phase lag introduced by the twin-T network at the 3-db frequency of

the coupling network would be approximately 22.5°. Thus, for example, the 3-db frequency would be 3 cps.

The grid-leak resistor may be chosen as high as is practicable to insert in the grid circuit of the particular tube, and it must not be overlooked that in the circuit in Fig. 11·8 the two branch arms of the twin-T network are part of the grid leak. If a nominal value of $R_g = 1$ megohm is chosen, from the relation $\omega_{3db}R_gC_1 = 1$, a value for $C_1 = 0.053$ μf is obtained. The choice of value of C_2 is similarly determined for low-frequency response and depends upon the impedance to which it is coupled. For applications at frequencies below 1000 cps, the high-frequency response of the amplifier may safely be neglected because it is determined mainly by the shunt impedances of the amplifier stages. These are negligible for single-stage and two-stage frequency-selective amplifiers but must be taken into account for multistage amplifiers, in common with all feedback amplifiers.

Equation (27) may be used to investigate component tolerances for the amplifier if it is again assumed that $R_1 = 0$. The tolerances for C_1 and R_g are determined by the phase-shift requirements and under the conditions given previously may safely be ± 10 per cent or even ± 20 per cent. From Eq. (27)

$$\frac{\delta \mathfrak{A}}{\mathfrak{A}} = \frac{\delta \mu}{\mu} + \frac{\delta R_L}{R_L}\left(\frac{r_p}{r_p + R_L}\right) - \frac{\delta r_p}{r_p}\left(\frac{r_p}{r_p + R_L}\right). \tag{28}$$

The manufacturing tolerances for many tubes are such that μ and r_p have approximately the same percentage variation, whereas the g_m may have about 1.5 times as much. Thus Eq. (28) may be approximated by

$$\frac{\delta Q}{Q} = \frac{\delta \mathfrak{A}}{\mathfrak{A}} \approx \frac{\delta \mu}{\mu}\left(\frac{R_L}{r_p + R_L}\right) + \frac{\delta R_L}{R_L}\left(\frac{r_p}{r_p + R_L}\right). \tag{29}$$

Using the values given for r_p and R_L and assuming that

$$\frac{\delta \mu}{\mu} \approx \frac{\delta r_p}{r_p} = \pm 10 \text{ per cent,}$$

then

$$\frac{\delta Q}{Q} \approx \pm 6 \text{ per cent} + 0.4 \frac{\delta R_L}{R_L}. \tag{30}$$

Thus, if a manufacturing tolerance of ± 10 per cent is allowable for the Q of the amplifier, the tolerance imposed upon the load resistor also is approximately 10 per cent. On the other hand, because of the tube-tolerance limitation, the manufacturing tolerance of the Q in the example given can never be less than ± 6 per cent and with ± 5 per cent load resistors would be ± 8 per cent.

If the cathode resistor R_1 (in Fig. 11·8), which is assumed to be the generator impedance, is appreciable in value, it must be considered as causing cathode degeneration.   Equation (29) may be then written as

$$\frac{\delta Q}{Q} \approx \frac{\delta \mu}{\mu} \left(\frac{R_L}{r_p + R_L + \mu R_1} \right) + \frac{\delta R_L}{R_L} \left(\frac{r_p}{r_p + R_L + \mu R_1} \right). \tag{31}$$

The additional assumptions in Eq. (31) are that $\mu \gg 1$ and that R_L and R_1 have the same percentage variations.   The effect of R_1 for a high-μ triode is pronounced, even if R_1 is only about $\frac{1}{100}$ of R_L.   For the parameter values previously assumed, if $R_1 = 1500$ ohms, the percentage variation of Q becomes

$$\frac{\delta Q}{Q} \approx \pm 3.7 \text{ per cent} \pm 0.25 \frac{\delta R_L}{R_L}. \tag{32}$$

Thus, the effect of tube-parameter variations (which includes aging) is reduced by nearly one-half, and the allowable resistor tolerances are
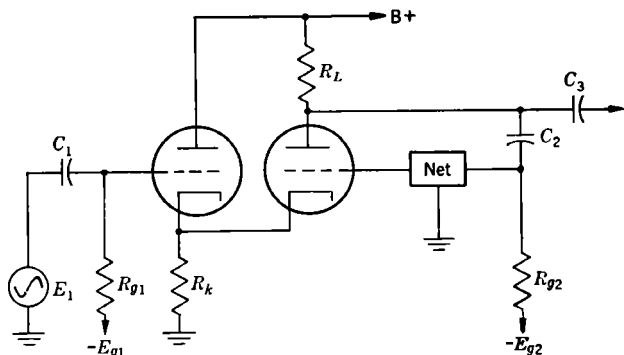
FIG. 10·9.—Frequency-selective amplifier employing cathode follower.

correspondingly widened if the same tolerances for the Q are to be maintained.

A circuit essentially equivalent to that of Fig. 10·8, except that in this case it is possible to use a generator of any internal impedance, is shown in Fig. 10·9.   The cathode of the second tube is directly coupled to the cathode of the first tube, which then acts as a cathode follower, providing a high input impedance for the selective amplifier.   In all other respects the circuits are the same, and the same requirements obtain for the coupling network from the source to the grid of the first tube.

A two-stage triode amplifier, which eliminates the necessity of applying the input signal to the cathode, is the so-called "cascode" amplifier shown in Fig. 10·10.

A qualitative analysis of this circuit shows that a signal e_1 appearing at the grid g_1 of the lower tube is amplified and appears at the upper grid as essentially αe_1, the total amplification then being of the order of α^2 if the two tubes are assumed to be the same. However, the amplification of a signal appearing at the upper grid is reduced by the cathode degeneration caused by the plate resistance of V_1 in the cathode circuit of V_2. In-phase signals at the grids are amplified in phase so that the grids effectively produce the same action but of different magnitude, and they are independent of each other.

From the point of view of the frequency-selective amplifier, this is very desirable, since the upper grid g_2 may be used to insert the signal
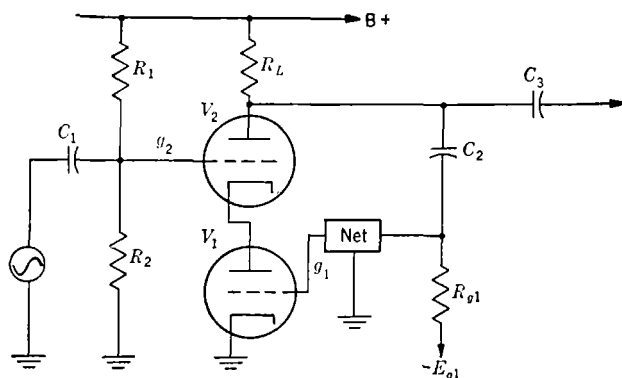
Fig. 10·10.—"Cascode" amplifier.

and the lower grid g_1 may be used as the high-impedance load for the network. Simple midfrequency-gain formulas may be obtained for both upper and lower grids if it is assumed that both tubes are identical. Thus

$$\alpha_{g_1} = \frac{\mu(\mu + 1)R_L}{R_L + (\mu + 2)r_p} \approx g_m R_L, \tag{33a}$$

$$\alpha_{g_2} = \frac{\mu R_L}{R_L + (\mu + 2)r_p}, \tag{33b}$$

where μ and r_p are the amplification factor and dynamic plate resistance of either tube. For the circuit as shown in Fig. 10·11, the gain α_{g_2}, being the loop gain, is the gain to be used in computing the Q of the amplifier. Thus, two triodes of moderate μ can be used to obtain amplifications comparable to those of pentodes and frequency-selective amplifiers of correspondingly high Q. Two triodes in a single envelope are especially useful, and the 6SN7 with $\mu \approx 20$, $r_p \approx 8000$ ohms and the 6SL7 with $\mu \approx 70$, $r_p \approx 50,000$ ohms should be mentioned specifically.

Figure 10·11 illustrates two typical amplifiers, Fig. 10·11a being a low-Q amplifier ($Q \approx 6$) and Fig. 10·11b a high-Q amplifier with a Q of about 15 by virtue of the fact that the cathode follower allows R_L to be increased from 36,000 to 100,000 ohms. The addition of the cathode follower to the feedback loop in Fig. 10·11b isolates the high-impedance plate load from the input terminal of the network and provides
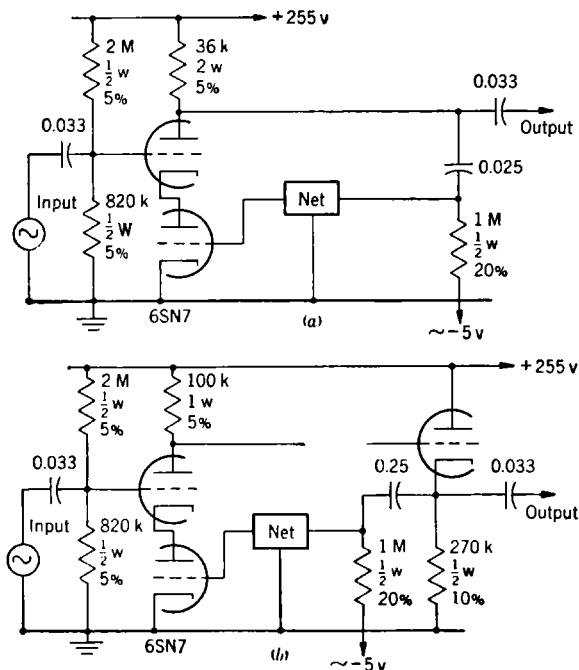

FIG. 10·11.—Two-stage frequency-selective amplifier. (a) Low-Q amplifier; (b) high-Q amplifier.

a source of low input impedance for the network. At the same time it acts as a buffer amplifier between successive stages.

The design criteria represented in these two typical circuits are conservative. All resistor power ratings are overrated, approximately five times, and all coupling networks have their 3-db points at a fraction of a cycle per second.

The photographs in Fig. 10·12 show the general layout and construction features of a lightweight unit that utilizes four frequency-selective amplifiers. A plug-in fixed-frequency twin-T network is also shown in Fig. 10·12a.

Figure 10·13 shows a lightweight selective amplifier and detector

constructed of subminiature tubes (SD834)[1] and components. The amplifier of Fig. 10·13 was designed for a minimum number of components of the widest possible tolerance. Its schematic diagram is shown in Fig. 10·14.
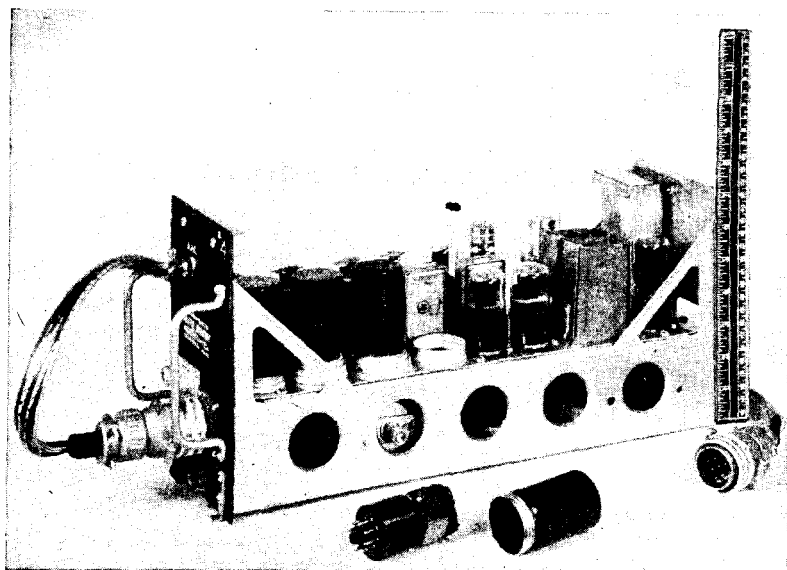
(a)
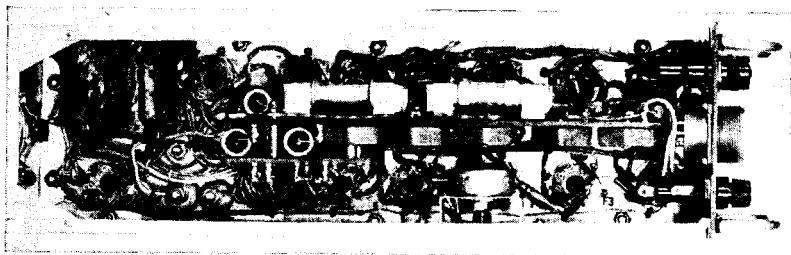
(b)

Fig. 10·12.—Lightweight unit using four frequency-selective amplifiers. (a) Side view, showing plug-in twin-T network; (b) bottom view.

The design of a rejection amplifier is simplified by using a cascode amplifier, since the input and feedback grids are independent. The amplifiers shown in Fig. 10·11 can be readily converted into rejection amplifiers by taking the output voltage from the lower grid g_1 and applying it to an isolating stage such as a cathode follower.

[1] Now known as type 6K4; c.f. "Circuits for Sub-miniature Tubes," *Electronics.* **19,** 5, May 1946.

Although Fig. 10·11b can properly be called a three-stage amplifier, a more typical example is the three-stage direct-coupled amplifier[1]
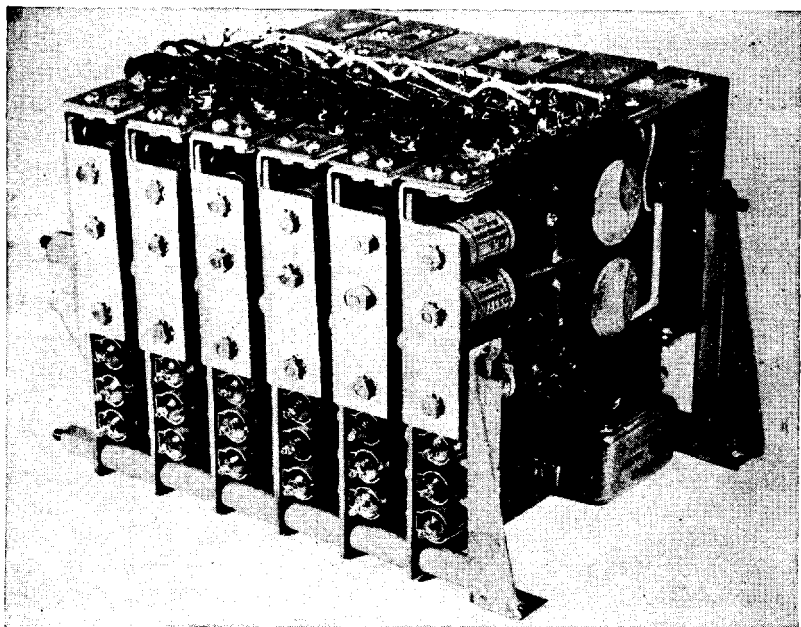
FIG. 10·13.—Photograph showing lightweight selective amplifiers and detectors constructed of subminiature tubes.
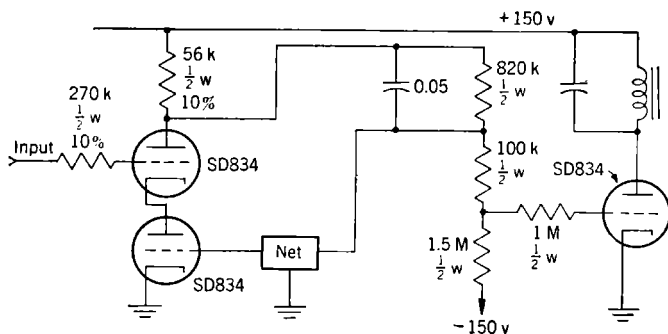
FIG. 10·14.—Circuit diagram, lightweight selective amplifier and detector constructed of subminiature tubes. The detector load is a relay.

shown in Fig. 10·15. This amplifier is designed for battery operation and hence for low power consumption.

[1] H. H. Scott, "A New Type of Selective Circuit and Some Applications," *Proc. I.R.E.*, **26**, 226–235, February 1938. See General Radio Catalog "K" Sound Analyzer type 760A.

The direct coupling eliminates the coupling networks and their attendant phase shifts, thus improving the low-frequency response.
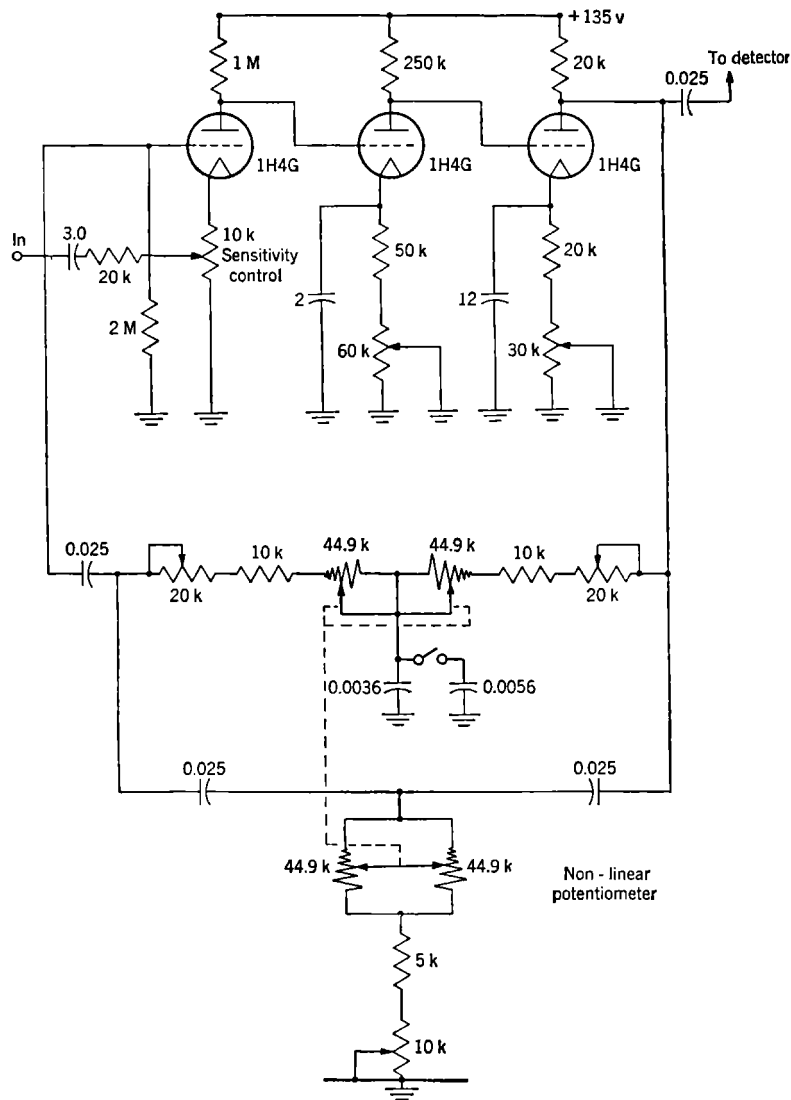
FIG. 10·15.—Three-stage direct-coupled amplifier. *(Courtesy of General Radio.)*

The input voltage is applied to the cathode of the first tube, and the twin-T network ties to the grid through a 0.025-μf condenser. The resistance of the 2-megohm grid-leak resistor is large compared with

the output impedance of the network. The three resistor elements of the twin-T network are coupled mechanically and adjusted so that the null frequency will vary continuously from 25 to 7500 cps. Since the gain of the amplifier without feedback is constant throughout this frequency range, the result is a frequency analyzer with a constant Q throughout the frequency range. This has proved very useful for such types of measurement as low-frequency noise measurement and harmonic analysis.

An increasingly important field of application of these frequency-selective amplifiers is their use as low-frequency (i.e., frequencies below 1000 cps) bandpass and "tone" filters. An important objection to purely passive networks involving inductance and capacitance is the excessive weight and size required to obtain the proper frequency-selection characteristic.

It is known that bandpass filters may be obtained by stagger-tuning single-tuned RLC-networks if there is no interaction among the networks. This isolation may be accomplished by means of buffer amplifiers, and the resulting pass band is a synthesis of the single-tuned stages.[1]

It was demonstrated in Sec. 10·2 that these frequency-selective amplifiers were nearly identical in phase and amplitude response with single-tuned networks, the deviation being inversely proportional to the gain of the amplifier. Because of the single feedback loop involved, two or more such amplifiers can be connected in cascade with no resultant interaction. Thus, each amplifier not only is equivalent to a single-tuned RLC-network but also acts as its own buffer amplifier.[2] Hence these amplifiers may be stagger-tuned to provide adequate bandpass filters in the low-frequency region.

Although any number of stages may be stagger-tuned (see Chap. 4), a filter whose 3-db bandwidth ranges from 40 to 63 cps with a center frequency of 50.2 cps[3] may be taken as a concrete example to which Figs. 10·12a and 12b apply. Then it is found that for an exact staggered quadruple[4] two stages of dissipation factor d_1 staggered at $f_0\alpha_1$ and f_0/α_1 and two stages of dissipation factor d_3 staggered at $f_0\alpha_3$ and f_0/α_3 are needed, where

$$d_1^2 = \frac{4 + \delta^2 - \sqrt{16 + 5.656\delta^2 + \delta^4}}{2} \quad \text{and} \quad \left(\alpha_1 - \frac{1}{\alpha_1}\right)^2 + d_1^2 = \delta^2,$$

$$d_3^2 = \frac{4 + \delta^2 - \sqrt{16 - 6.656\delta^2 + \delta^4}}{2} \quad \text{and} \quad \left(\alpha_3 - \frac{1}{\alpha_3}\right)^2 + d_3^2 = \delta^2,$$

---

[1] See Chap. 4 for a complete discussion on the principles of stagger-tuning.

[2] The use of a cathode follower as in Fig. 10·11b considerably improves the isolation from stage to stage.

[3] The example will be discussed on the basis of geometric symmetry. Hence the center frequency is the geometric mean between the 3-db frequencies.

[4] The dissipation factor for a single-tuned circuit is $1/Q = \Delta f_{3db}/f_{res}$.

f_0 = the center frequency, Δf = the 3-db bandwidth,

$$\delta = \frac{\Delta f}{f_0} = \frac{1}{Q}, \qquad d_1 = \frac{1}{Q_1}, \qquad d_3 = \frac{1}{Q_3}.$$

Using these formulas, it is found that to produce the required bandpass filter there are needed amplifiers tuned to 40.6 and 61.9 cps each with a Q of 15 and two amplifiers tuned to 46.0 and 54.9 cps each with a Q of 6. The amplifier of Fig. 10·11a is representative of the low-Q amplifiers, and that of Fig. 10·11b of high-Q amplifiers.
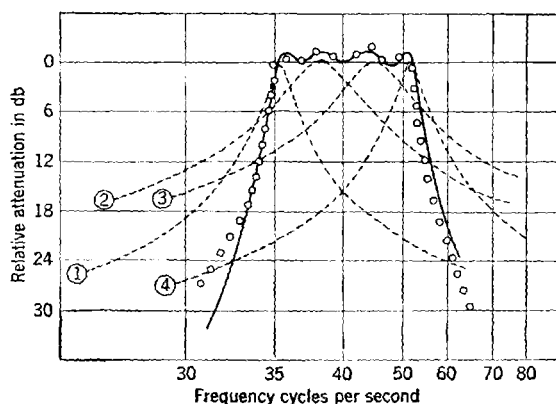

FIG. 10·16.—Pass band, four-section stagger-tuned filter. Calculated values, solid line; experimental values, dotted line. (a) f_1 = 40.6 cps, Q_1 = 15; (b) f_2 = 46.0 cps, Q_2 = 6; (c) f_3 = 54.9 cps, Q_3 = 6; (e) f_4 = 61.9 cps, Q_4 = 15.

Figure 10·16 shows the synthesis of the bandpass characteristic from the four stagger-tuned frequency-selective amplifiers. For comparison, an experimental bandpass characteristic representative of a number of identical units is indicated on this figure by the dotted curve. The attenuation through this filter (i.e., the "insertion loss") was found to be 14 db and was constant within 1 per cent for input signals ranging from 0 to 6 volts rms. Inspection of Fig. 10·16 at the midfrequency (40 cps) shows a loss of about 36 db, and Eq. (33b) shows that the gain of the amplifier with respect to the signal input grid is approximately $1/\mu$ times the gain with respect to the lower grid. Since $\mu \approx 20$, the gains of the amplifiers at their resonant frequencies become, respectively, 3.0, 1.2, 1.2, and 3.0, or a total gain of about 22 db. Hence the mid-frequency loss is 14 db.

The construction features of these "electronic" filters are shown in the photographs of Fig. 10·12. These units also contain a power supply, a-c amplifier, detector, and indicator stages.

CHAPTER 11

DIRECT-COUPLED AMPLIFIERS

By John W. Gray

INTRODUCTION

A direct-coupled amplifier is one in which the input, output, and interstage couplings are conductive, i.e., direct connections or resistive networks. No "blocking" condensers or coupling transformers are employed. Thus, d-c signals may be amplified, whereas the bandpass characteristics of other types of amplifiers drop to zero gain at zero frequency.

Usually zero frequency is the most important part of the pass band, and it has the highest gain, except in certain servoamplifiers where, for damping purposes, the d-c is attenuated in comparison with low audio frequencies. Where rapid fluctuations are to be followed, the pass band must extend correspondingly high in the frequency spectrum, but only in a very special case would a direct-coupled amplifier be expected to pass frequencies much above the audio spectrum.

11·1. Applications of Direct-coupled Amplifiers.—Below are listed several fields of application, with a brief statement of some special features of each. A direct-coupled amplfier is not the only type applicable in each of these cases, since a sequence of modulation, a-c amplification, and demodulation may sometimes be used.

Vacuum-tube Voltmeter.—The purpose of the direct-coupled amplifier in a vacuum-tube voltmeter is to convert, with a linearity and constancy of gain of from 5 to 0.1 per cent, a d-c voltage derived from a high-impedance source into a current in a low resistance. Precision of linearity and gain may be achieved by the use of negative feedback. Except for battery-operated types, the instrument ought to operate independently of supply-voltage variations. The input impedance should be low enough so that the grid current of an ordinary low-power tube will cause negligible additional loading.

Electrometer.—This is a class of vacuum-tube voltmeter in which special precautions are taken to keep the input current very low. Special tubes are used, and low electrode voltages are employed in order to minimize ionization, which would cause grid current to lower the input impedance. Batteries are the most common source of power.

Resistance Measurement.—This application of direct-coupled amplifiers includes ohmmeters for the measurement of high resistance, bolometers, resistance-strain gauges, and photocell amplifiers. Usually a vacuum-tube voltmeter is used in some sort of a bridge circuit.

Mirror Oscillograph Drivers.—A direct-coupled amplifier is needed when, as is often the case, a record of the d-c component of the signal is desired. A higher frequency limit to the pass band and, in general, more current output are required than for a vacuum-tube voltmeter.

Cathode-ray-tube Deflection.—A direct-coupled amplifier is required for some types of measurement with oscilloscopes, such as the measurement of tube characteristics. For electrostatic deflection a large linear voltage swing, in push-pull form but with very little power output, is needed. Magnetic deflection requires large current and power, and the inductance of the coils makes the bandpass problem rather difficult.

Relays and Solenoids.—It may be desired to operate a magnetic device when a d-c voltage, current, or a resistance reach some predetermined value. Linearity is not important here, but sensitivity and stability are.

D-c Servoamplifier.—In this case, a motor must be driven in a direction corresponding to an "error signal," which comprises the amplifier input signal and which the resulting motion of the motor will tend to reduce or reverse. The input signal may be the difference between two variable d-c voltages or between a variable voltage and a fixed potential. The amplifier may furnish all the power to the motor but often only controls its motion by a differential field control or some other device. "Proportional control" is usually, although not always, desired; it implies a rough linearity between error input and motor-control output in the neighborhood of zero error. For large errors the output is expected to be limited in the proper sense. Since the servo loop itself comprises a complete negative-feedback system, no resistive negative feedback is normally employed in the amplifier except, perhaps, to reduce the d-c gain in favor of a-c gain to achieve better damping. D-c stability and sensitivity are the important factors in the amplifier.

Voltage and Current Regulators.—A voltage-regulator amplifier is similar to a servoamplifier, in that an error voltage, the difference between a standard and the regulated output or a fraction thereof, is amplified and used to control the output in order to minimize the error. In a current regulator the output current is passed through a constant resistance, and the resulting voltage is compared with the standard voltage to give the error.

Impedance Changing.—When a variable d-c voltage supplied by a high-impedance source is required to operate a low-impedance load, a d-c impedance changer of unit voltage gain is needed. This, depending

on the requirements, may be a simple cathode follower or a higher-gain direct-coupled amplifier with negative feedback.

Phase-splitting; Subtracting; Level Changing; Summing.—These and similar operations that are sometimes performed in d-c computing circuits can be performed by certain types of direct-coupled amplifiers, usually with large negative feedback.

Integration and Differentiation.—High-gain direct-coupled amplifiers are used in the "amplified time constant" methods of voltage integration and differentiation. In the first case, a series resistance is connected to the input terminals, and a condenser gives over-all negative feedback. In the second case, the resistance and condenser functions are reversed. Thus, the differentiator is not direct-coupled to its input terminal and is independent of the input level, but interstage and output coupling are direct.

11·2. Problems Peculiar to Direct-coupled Amplifiers.—The fact that conductive coupling must be used throughout results in several special problems of design and stability. One obvious design problem arises because plate potentials are considerably higher than grid potentials. It is therefore necessary either to have the cathode of a following stage at this higher potential from ground or to use a device to lower this potential the correct amount. Since batteries are usually impractical here, a negative supply voltage and a resistance divider are generally required, with resulting decrease in gain.

The "zero adjustment" problem is the most difficult one resulting from conductive coupling. If the effects of noise and pickup are neglected, a nonconductively coupled amplifier gives zero output for zero input regardless of circuit parameters such as resistor values, supply voltages, and tube characteristics. This is not true of a direct-coupled amplifier; the d-c output at a given d-c input (e.g., zero) depends on all these things. A certain output current or voltage is usually required for a specified input. In order to meet this requirement some kind of "zero" adjustment is needed to compensate for variation of the above parameters (1) between amplifiers of the same design, due to tolerances of like components, and (2) because of effects of aging, temperature variation, vibration, etc. For rational circuit design, the expected extreme limits of voltage, resistor, and tube characteristics must be known, as well as their effects on the circuit. The zero adjustment should provide sufficient latitude to compensate for these extremes, and the incidence of these extremes, together with the appropriate zero adjustment, should not put any part of the circuit out of its good operating range.

Voltage-supply variation will depend on the kind of regulation provided. Its effect on zero shift and some methods of minimizing or canceling this effect will be discussed in connection with specific circuits.

Resistor tolerances, temperature characteristics, and stabilities are covered in another volume.[1] The use of wire-wound or other high-stability resistors will be recommended for certain critical positions in precision amplifiers. The effects of production tolerances, heater-voltage variation, and aging on the static characteristics of vacuum tubes will also be examined.

SPECIAL ASPECTS AND EFFECTS OF VACUUM-TUBE PROPERTIES

11·3. Variability of Vacuum-tube Characteristics.—The Joint Army-Navy Specification, JAN-1A, for Electron Tubes[2] is a standard of what limits of variability may be expected in tubes of any given type.   Pertinent data therefrom, for common tube types, are presented in the *Components Handbook*.   Maximum and minimum plate current are specified for given plate voltage and grid bias or self-biasing cathode resistor.   The limits of transconductance and amplification factor are also given for the same test conditions.

In direct-coupled amplifiers, the operating plate current and voltage are usually established by design, and grid bias is then adjusted to compensate for tube tolerance.   Therefore, it is desirable to interpret the limits in terms of grid bias.   For the conditions of the JAN test, the grid-bias spread may be computed by dividing the plate-current spread by average transconductance.   If, in the test, bias were obtained from a cathode resistor, the change in bias between upper and lower limits would have to be added to the above figure.   This value is obtained by multiplying the biasing resistance by the spread of cathode current.   The three examples following illustrate the computation:

1. *Fixed bias, pentode:*
   Tube:
     6SJ7 Sharp-cutoff pentode.
   Test conditions:
     $E_b = 250$ volts, $E_{c2} = 100$ volts, $E_{c1} = -3$ volts.
   Plate current:
     Minimum $I_p = 2.0$ ma, maximum $I_p = 4.0$ ma, $\Delta I_p = 2.0$ ma.
   Transconductance:
     Minimum $S_m = 1325$ μmhos, maximum $S_m = 1975$ μmhos,
     Average $S_m = 1650$ μmhos.
   Computed grid bias tolerance:
   $$\Delta E_c = \frac{\Delta I_p}{S_m} = \frac{2.0}{1.65} = 1.21 \text{ volts.}$$

[1] *Components Handbook*, Vol. 17, Chaps. 2 and 3, Radiation Laboratory Series.
[2] Issued by the Joint Army-Navy Electron Tube Committee, Army-Navy Electronic and Electrical Standards Agency, 12 Broad St., Red Bank, N. J.

2. *Cathode bias, triode:*

   Tube:

   6J6 (one section) miniature double triode.

   Test conditions:

   $E_b = 100$ volts, $R_k = 50$ ohms, grid grounded.

   Plate current:

   Minimum $I_p = 5.5$ ma, maximum $I_p = 12.5$ ma,

   $\Delta I_p = 7.0$ ma.

   Transconductance:

   Minimum $S_m = 4000$ μmhos, maximum $S_m = 7300$ μmhos,

   Average $S_m = 5650$ μmhos.

   Actual test grid bias $(-R_k I_b)$:

   Minimum $E_c = -0.275$ volts, maximum $E_c = -0.625$ volts.

   Computed grid-bias tolerance:

   $$\Delta E_c = \frac{7.0}{5.65} + 0.35 = 1.6 \text{ volts.}$$

3. *Cathode bias, tetrode or pentode:*

   Tube:

   6AJ5 low-voltage miniature pentode.

   Test conditions:

   $E_b = E_{c2} = 28$ volts, $R_k = 200$ ohms, grid grounded.

   Plate current:

   Minimum $I_p = 1.8$ ma, maximum $I_p = 4.0$ ma, $\Delta I_p = 2.2$ ma.

   Screen current:

   Minimum $I_{c2} = 0.1$ ma, maximum $I_{c2} = 2.0$ ma.

   Transconductance:

   Minimum $S_m = 2000$ μmhos, maximum $S_m = 3500$ μmhos,

   Average $S_m = 2750$ μmhos.

   Cathode current (sum of $I_p + I_{c2}$):

   Minimum $I_k = 1.9$ ma, maximum $I_k = 6.0$ ma.

   Actual test grid bias $(-R_k I_k)$:

   Minimum $E_c = -0.38$ volts, maximum $E_c = -1.2$ volts

   Computed grid-bias tolerance:

   $$\Delta E_c = \frac{2.2}{2.75} + 0.82 = 1.62 \text{ volts.}$$

The tolerances of the first stage of a direct-coupled amplifier have by far the greatest effect on the zero adjustment. This stage is most often a voltage amplifier, operating with plate voltages and currents considerably lower than were employed in the JAN tests. The question therefore arises as to the validity of the tolerance computed as above.

In the case of a triode, if the assumption is made that the differences between samples of a tube type lie only in plate resistance and in ampli-

fication factor and that the characteristics are linear and obey the formula

$$I_p = \frac{E_b + \mu E_c}{r_p},$$

it follows that variations of μ and/or r_p result in the following variation in grid bias required at a given plate voltage and current:

$$\Delta E_c = \frac{I_p}{\mu} \Delta r_p - \frac{E_c}{\mu} \Delta\mu.$$

The same sort of formula applies to a pentode if μ and r_p refer to screen voltage. Actually, about the only generalization that can be made is that the variation between tubes measured at the grid is less for small currents and grid biases.

11·4. Vacuum-tube Characteristics at Low Currents.—In voltage amplification, which is the function of all but the output stage in most d-c amplifiers and of the output stage as well in some, it is generally desirable to employ each vacuum tube in such a way as to obtain the maximum gain, with little regard for the resulting impedance levels. This involves operation at rather low currents, for which the ordinary published tube data are unsatisfactory for use in a critical analysis. This section is largely the result of special measurements of vacuum-tube characteristics in the low-current region, from which certain generalizations are made regarding voltage amplifiers.

In hot-cathode vacuum tubes the best simple formulation of the relationship between anode current and electrode potentials is generally considered to be the three-halves power law. According to this law, the plate current in a diode is proportional to $e_{pk}^{3/2}$, and in a triode the total anode current (including the grid current if the grid bias is so small that this current is appreciable) is proportional to $[e_{gk} + (e_{pk}/\mu)]^{3/2}$. The basis for the derivation of this law is the assumption of a copious supply of electrons at the cathode surface, all of which are emitted with no excess energy so that the potential gradient at the cathode surface can be zero. For large currents, where e_{pk} or $[e_{gk} + (e_{pk}/\mu)]$ is large compared with the initial electron velocities (expressed in electron-volts), the three-halves power law is valid. For low currents an entirely different relationship obtains.

The portion of the total number of electrons emitted from a hot surface having normally directed velocities greater than e electron-volts is $\epsilon^{-\frac{Te}{11,600}}$, where T is the temperature in degrees Kelvin. Thus, if this expression represents the plate current of a diode, there must be a point in the tube e volts below the cathode potential that the rest of the electrons have insufficient energy to attain. This minimum potential

usually results from a space charge of excess electrons; but if the plate potential is itself negative, the minimum potential in the tube may actually be at the plate. Below the plate potential where this becomes true, the plate current is an exponential function of plate voltage. (Contact potential also is important, but it can be considered to be due to a constant voltage added to the plate supply.) The exponent is such that, for oxide-coated cathodes operating at from 1000° to 1100° K, an increment of plate voltage of about 0.21 volt produces a 10-fold increase of plate current. If the $\log_{10}$ of current is plotted against plate voltage, the result should be a straight line with a slope of $1/0.21$.
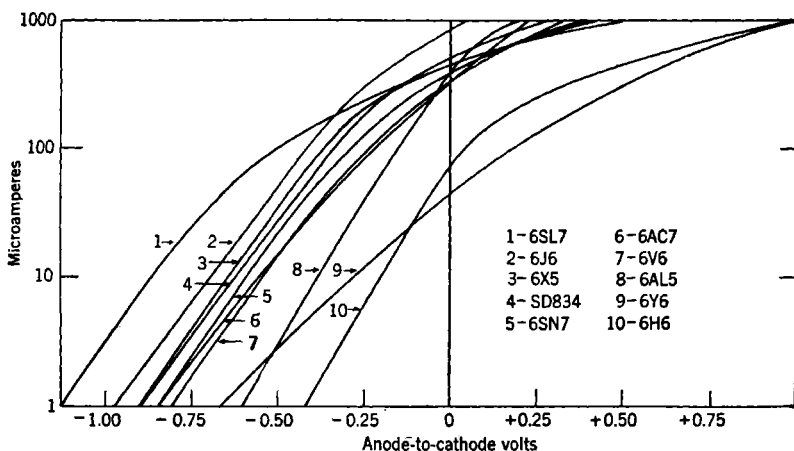
FIG. 11·1.—Low-current characteristics of several diodes and diode-connected tubes, with logarithmic current scale.

The preceding theory is fairly descriptive of all tubes with unipotential cathodes, at low currents. Figure 11·1 shows the logarithmic characteristics of samples of several tube types, either diodes or connected as such (grids connected to plate). The curves are straight for currents below 50 to 500 μa, depending on cathode area and spacing, and most of them have about the same slope: approximately $1/0.25$. (These curves are not necessarily typical of the various tube types. Considerable shift in the voltage scale is encountered from tube to tube. For example, for several 6SL7's, the voltage at 1 μa ranged from -0.7 to -1.3 volts, although the slopes and shapes of the curves were fairly consistent.)

In a triode operated as such, where potential differences exist between plate and grid, the relationship of anode current to grid-cathode potential seems still to be logarithmic, up to the same current level as it is when the plate-to-grid potential is zero. The slope of the curve is decreased by the plate-grid potential, probably because of distortion of the unipotential

surfaces around the grid. This distortion and the decrease of slope of the curve are roughly proportional to grid bias and therefore to $1/\mu$ times the plate potential. Thus, the curve at a given plate potential is, in general, steeper for a high-μ than for a low-μ triode.
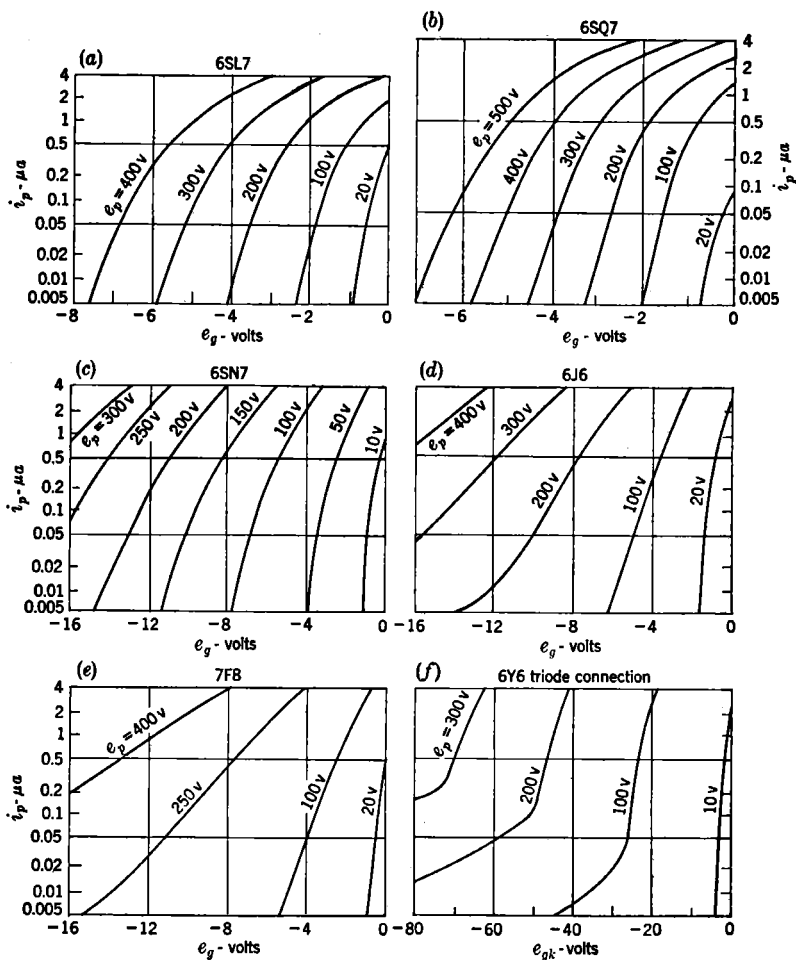

Fig. 11·2.—Logarithmic low-current characteristics of certain triodes.

The curves of Fig. 11·2 are the conventional "mutual" or "transfer" characteristics of certain triodes, but with plate current in a logarithmic scale. The lower portions of the curves are fairly straight and have slopes that approach that of the diode-connected tube for low plate

potentials but that decrease from this value as plate voltage is increased. (An aberration is evident in the case of the triode-connected 6Y6—the loss of control at low currents, because of leakage around the grid.)   As explained above, this decrease of slope with respect to plate voltage is less noticeable for the higher-μ triodes.
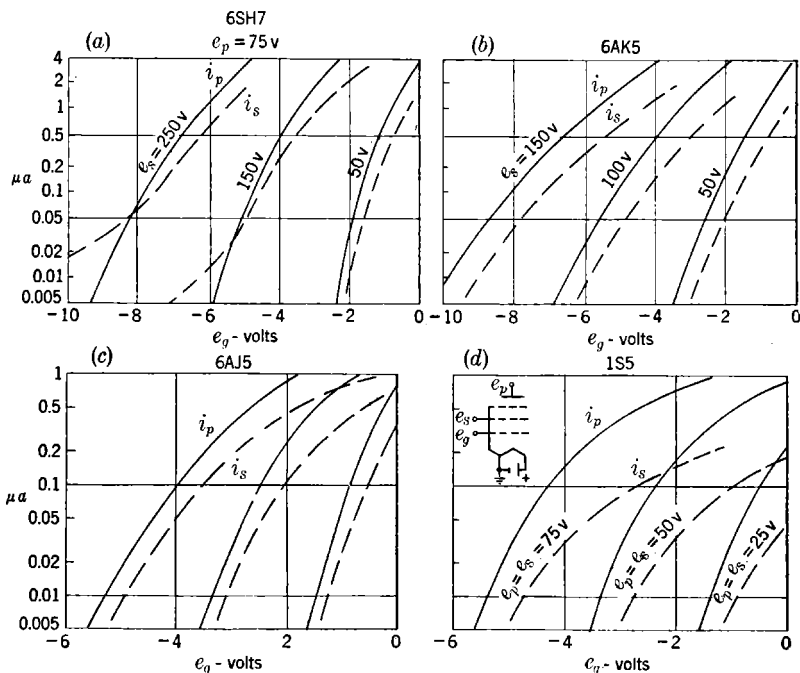
FIG. 11·3.—Logarithmic low-current characteristics of certain pentodes.

If E_0 is the reciprocal of the slope of any curve in Fig. 11·2, i.e., the plate current increases 10-fold for each E_0 increment of e_{gk}, the mutual conductance at this point is

$$g_m = \frac{2.3 i_p}{E_0}$$

and is therefore proportional to the plate current while the curve is straight.   It is evident that for a given plate voltage and plate current, the g_m in the low-current region is greater for high-μ than for low-μ triodes, regardless of rating.   Also, for a given triode at a given plate current, g_m is greater at lower plate voltages.

From the curves and from the formulas developed in a later section, it will be evident that the maximum voltage gain is obtainable in a d-c

amplifier if the tube operates at as low a plate voltage as possible and at a plate current corresponding to the top of the straight part of the curve as in Fig. 11·2.

Similar pentode characteristics are given in Fig. 11·3, from which it is evident that the maximum gain obtains at as low a screen-grid voltage as is permissible without resulting in the flow of control-grid current.

11·5. Grid Current.—A knowledge of the grid-current characteristics of vacuum tubes is essential to an understanding of certain important limitations of direct-coupled amplifiers.
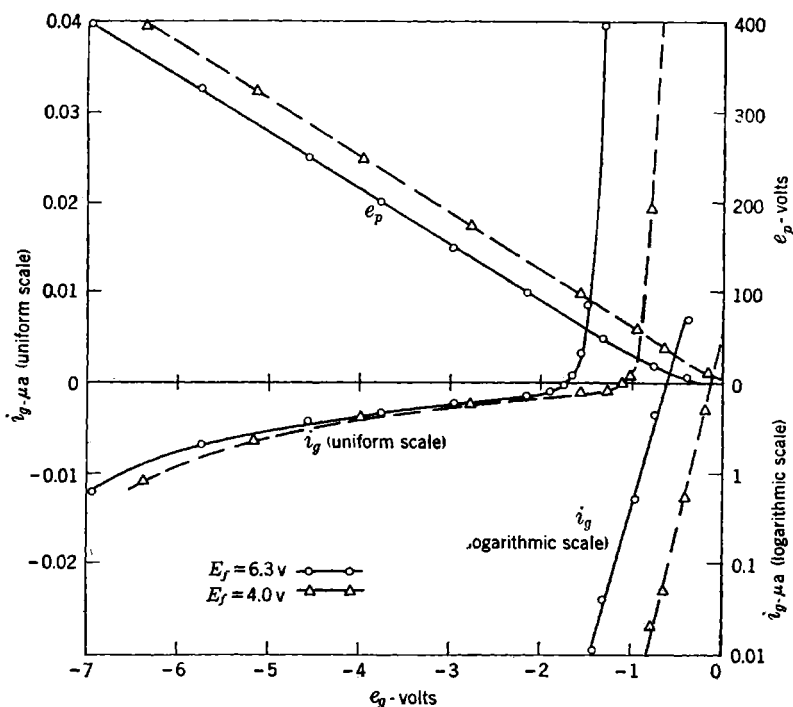
FIG. 11·4.—Negative and positive grid current in a triode operated with constant-plate current (6SL7; $i_p = 0.1$ ma).

Negative Grid Current.—When the grid is negative with respect to the cathode, a current, which is the result of positive ions that are attracted to the grid, flows from the grid terminal. The ionization of the residual gas in the tube depends on voltage gradients, and therefore the negative grid current increases with higher plate voltage as well as with greater negative grid voltage. Since the positive-ion current is

the result of residual gas, its magnitude varies widely from tube to tube. For ordinary receiving tubes under usual operating-voltage conditions it may be anything from 0.00001 to 0.1 μa. The JAN specifications usually limit the grid current to 1 or 2 μa, but generally this value may be considered as being very conservative. The JAN specifications for the 6AK5 grid current is 0.1 and for the 6SU7, 0.02 μa. Thus the latter tube is a good choice where a very high input impedance is needed, although the 6SL7 and many other tube types, when operated at reasonably low plate voltage, may generally be relied on for grid currents as low as this. The effect of plate potential on positive-ion current (negative-grid current) is illustrated for one example of a high-μ triode in Fig. 11·4. (Variation of this part of the curve, from tube to tube, is considerable; the illustration is presented only to show the trend.) Plate current was held constant, so that as the grid voltage was lowered, the plate voltage rose μ times as much. Negative grid current varies about as the square of the plate potential. Reduced cathode temperature has very little effect on this relationship; a given plate potential occurs at a reduced grid bias (see Sec. 11·5), but a given plate potential produces a certain positive-ion current almost independently of cathode temperature.

"Electrometer tubes" are specially designed and constructed not only to have very little gas to be ionized but also in such a way that they are operative at very low electrode potentials in order to minimize the possibility of ionization. In this way and with special insulation, these tubes operate with grid current of the order of 10^{-16} amp. When the positive-ion current is less than about 0.0001 μa, a photoelectric grid current may be large enough to be of importance. Roberts,[1] operating a 606 pentode at plate and screen potentials of 20 volts, observed that a decrease of grid current from 10^{-10} to 10^{-12} amp resulted from shielding the tube from light. A reduced cathode temperature can help in this respect by decreasing the illumination from the cathode.

Another source of negative grid current exists in some tubes, namely, grid emission. This emission may occur if the grid is contaminated and especially if the tube power dissipation is high, even though the grid dissipation itself is negligible. It rarely amounts to more than 1 or 2 μa if the plate dissipation is not above the rating, but even this small amount may be a consideration where a high resistance is employed in the input circuit to a power stage.

Positive Grid Current.—Positive grid current, caused by electrons flowing to the grid from the cathode, is much greater and much less erratic than the negative grid current. The grid receives electrons as if it were the plate of a diode. Because of the initial velocities of the

[1] Shepard Roberts, "A Feedback Microammeter," RSI., **10,** 181, June 1939.

thermoelectrons, many of them have sufficient energy to arrive at the grid even if it is at a considerable negative potential with respect to the cathode. Thus, this positive grid current becomes as great as the negative (positive-ion) grid current, making the net current zero at some negative grid potential, usually about -1.0 to -1.8 volt for oxide-coated unipotential cathodes at temperatures of $1000°$ to $1100°$ K. This is the "floating-grid" potential, which the grid assumes if disconnected. It is largely independent of plate voltage but tends to be slightly more negative at the lower values.

Because of the distribution of electron initial velocities, the grid current increases exponentially with respect to grid potential, increasing itself by a factor of 10 for every 0.2- or 0.3-volt increase of potential. (The fact that the grid potential moves as the logarithm of the grid current is used in certain computing devices.) This exponential type of increase ends at from 10 to 100 μa; thereafter the nature of the increase depends on the tube type and is also considerably affected by plate voltage—the lower the plate voltage the more the grid current.

The logarithmic portion of the grid-current curve is shown in Fig. 11·4. This part of the curve shows much less variation from tube to tube than the negative portion. A reduction of cathode temperature reduces the grid current at a given bias, but it also increases the plate potential at a given plate current. The result is that if the grid current is not to exceed some predetermined magnitude, for a certain plate current the minimum allowable plate potential is not affected by cathode temperature.

Minimum allowable plate potential is of great importance in d-c amplifier design. It depends primarily on the maximum allowable positive grid current, which depends on input resistance, and on the permissible voltage error resulting from their product. For example, a grid current of 0.02 μa in a series input resistance of 1 megohm causes the actual grid voltage to be 20 mv lower than the input potential. In the case of Fig. 11·4, with a plate current of 0.1 ma, this would occur at $e_p =$ about 50 volts, regardless of cathode temperature. In this same example, a reduction of plate current by a factor of 10 (to 0.01 ma) would lower the plate potential by only 15 or 20 volts.

The higher the μ of a triode the higher will be the minimum allowable plate potential (assuming a given allowable grid current and a given plate current), the latter being roughly proportional to the former.

In a pentode, the plate can operate at much lower potentials than in a triode, provided that the screen is at a sufficiently high potential to permit ample negative bias on the control grid. The only requirement on the pentode-plate potential is that it be high enough to keep the screen from taking all of the current; most low-power pentodes will

operate satisfactorily as voltage amplifiers with plate potentials in the neighborhood of 10 volts.

11·6. The Effect of Heater-voltage Variation.—Variation of the cathode temperature of a vacuum tube in which the current is limited by electrode potentials and space charge rather than by cathode emission has the effect of varying the average initial electron velocity and therefore also the electrode voltages required to obtain a given electron flow.[1] When the plate current is very small compared with cathode emission, as is often the case in the first stage of a d-c amplifier, this effect is fairly
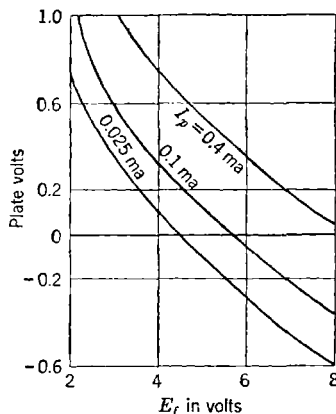
FIG. 11·5.—Diode heater-voltage characteristics (6SL7 as diode).
FIG. 11·6.—Triode heater-voltage characteristics (6SL7, $E_p = 150$ volts).
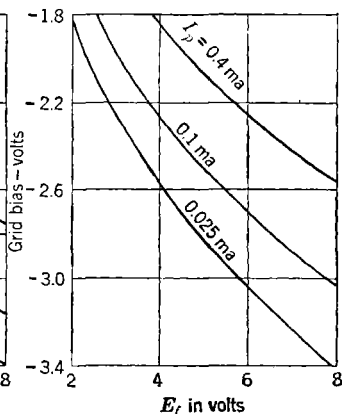
independent of plate current. It may be expressed in terms of the amount by which the cathode voltage must be changed relative to the other electrode voltages in order to hold the current constant as the cathode temperature is varied.

For oxide-coated unipotential cathodes, this amount is approximately 0.2 volt for a 20 per cent change of heater voltage about the normal value, whether the tube is a diode, a triode, or a multigrid tube. Figure 11·5 illustrates the relationship for a diode. (A 6SL7 triode with grid connected to plate was used as a diode in this test, for comparison with Fig. 11·6. An ordinary diode gives the same result, however.) It is seen that the slope $\partial E_p / \partial E_h$ at a given value of heater voltage is not greatly affected by current at the low-current values employed.

[1] L. R. Koller, *The Physics of Electron Tubes*, 2d ed., McGraw-Hill, New York, 1937; W. G. Dow, *Fundamentals of Engineering Electronics*, Wiley, New York, 1937; J. Millman and S. Seely, *Electronics*, McGraw-Hill, New York, 1941. See also Appendix C for a discussion of the drift of vacuum-tube characteristics under constant applied potentials.

Figure 11·6 gives the same sort of information for the same tube used as a triode. Here the plate-to-cathode voltage was fixed; and as the cathode temperature varied, the current was held constant by adjustment of the grid potential. To have been completely comparable with the diode characteristics, the plate-to-grid potential should have been held constant; however, the results differ only by $1/\mu$. It is seen that the slope $\partial E_c/\partial E_h$ is about the same as was the slope $\partial E_p/\partial E_h$ for the diode and that in this case, too, it remains fairly constant with changes in plate current.

If no compensation is employed, a direct-coupled amplifier will suffer a drift in the zero with any variation in the heater voltage. As will be shown later, it is not possible to cancel this drift with inverse feedback. As measured at the input, the drift caused by the first stage is simply $(\partial E_c/\partial E_h)\Delta E_h$; i.e., about 200 mv for a 20 per cent heater-voltage change (assuming an oxide-coated cathode and low plate current). Drift in subsequent stages is less important, depending on the preceding gain.

Various practical methods of counteracting the effect of change in heater voltage will be detailed in later sections. The most important of these rely, in some manner, on the canceling effect of an auxiliary tube that has the same heater supply and therefore undergoes the same sort of drift as the amplifying tube. Usually, equivalent tube types are used for the compensation, but some methods employ a diode to compensate a triode or pentode. The use of double tubes, such as the 6SL7, is preferable, since the cathode characteristics of a pair of triodes in the same envelope, made on the same day by the same manufacturer, are likely to be more alike than the characteristics of a random pair. To check this, the characteristics of 18 6SL7 double triodes (36 triodes) were measured. For each triode was recorded the change in grid bias that was needed to hold constant a plate current of 0.2 ma at a plate voltage of 150 volts while the heater voltage was changed from 10 per cent below normal to 10 per cent above normal. The tubes were of various ages and makes. The average value of ΔE_h for the 20 per cent change was 210 mv, and 90 per cent of the values lay between 191 and 210 mv. Pairing the triodes at random without regard to envelopes, the average difference in ΔE_h between pairs[1] would be 14 mv, whereas the average difference between the two triodes in each envelope was only 8 mv.

A set of mutual characteristics for a 6SL7 is given in Fig. 11·7 for two different heater voltages. For a given current, $\partial E_c/\partial E_h$ is not affected by the value of E_p; nor is $\partial E_c/\partial E_h$ appreciably affected by the value of I_p for currents less than 1 ma, such as are employed in most

[1] In a gaussian distribution of values, the average difference between pairs of values is $2\sigma/\sqrt{\pi}$, where σ is the standard deviation of the distribution.

voltage amplifiers. For higher currents, no longer negligible compared with cathode emission, $\partial E_c / \partial E_h$ changes somewhat more with change in current and is also subject to much more variation from tube to tube.

Table 11·1 gives the average characteristic (in terms of ΔE_c required by a 20 per cent change of E_h about normal) and dispersion of the characteristic for rather limited samples of certain tube types. Both high and low values of current were applied, and the results show that for
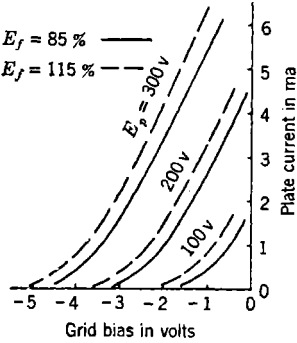
FIG. 11·7.—Effect of heater voltage on 6SL7 characteristics.
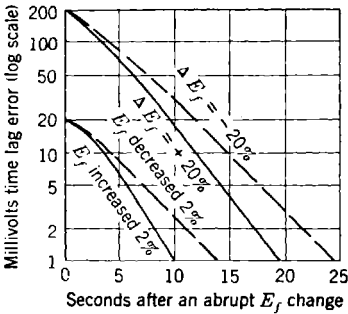
FIG. 11·8.—Heater-cathode time lag of a 6SL7.

high currents the effect of heater-voltage variation is greater and much more erratic. Pentodes were connected as triodes for these tests; this type of operation is legitimate, because it is found that the ratio between plate and screen currents is not affected by this connection.

TABLE 11·1.—EFFECT OF ELECTRODE VOLTAGES AND CURRENTS ON THE AVERAGE DRIFT RESULTING FROM HEATER-VOLTAGE VARIATION AND ON THE DEVIATION FROM THIS AVERAGE, AND THE GRID BIAS AND DEVIATION THEREFROM

Tube type	No. of tubes in sample	Plate (and screen) potential, volts	Plate (and screen) current, ma	Average grid bias (E_h = 63v), volts	Probable deviation of grid bias, volts	Average bias shift required for ΔE_h = 20 %, mv	Probable deviation of bias shift, mv
6SN7	10(20 sec.)	60	2.0	−1.65	±0.15	210	± 15
		240	10.9	−6.70	±0.40	425	± 90
6SJ7	10	60	2.5	−1.60	±0.10	215	± 15
		240	10.0	−7.20	±0.35	400	±110
6AC7	10	60	2.0	−1.30	±0.25	210	± 15
		240	8.0	−5.10	±0.60	290	± 40
6AG7	10	60	10.0	−3.20	±0.20	220	± 20
		240	40.0	−7.90	±0.25	325	± 50

There is a certain amount of thermal lag between a change of heater voltage and the resulting change of emission. Its magnitude, for a sudden increase or decrease of heater voltage on a 6SL7, is shown in Fig. 11·8 in term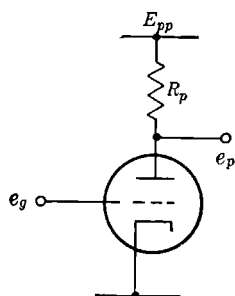s of the lag in bias for a given current. This lag is of particular interest where the method of compensation involves the direct application of heater voltage as a circuit parameter. Where the compensation employs another tube as mentioned above, it is of interest only in so far as the lag varies from tube to tube.

FIG. 11·9.—Simple triode amplifier.

DESIGN PRINCIPLES

11·7. Single-ended Triode Amplifiers.—The amplifier of Fig. 11·9, with fixed cathode potential, may be analyzed by drawing a load line on the graph of plate characteristics in the conventional manner (see Fig. 11·10). The plate current and plate-to-cathode voltage may then be read directly from the figure for any value of grid voltage.

In so far as the characteristics may be considered to be straight and parallel and to start from zero plate current at zero plate and grid voltage,
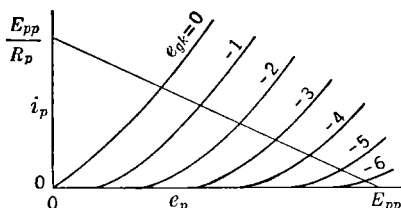
FIG. 11·10.—Plate characteristics with load line.

the operation of the amplifier may be analyzed mathematically from the formula for plate current:

$$i_p = \frac{e_{pk} + \mu e_{gk}}{r_p} \tag{1}$$

where μ is $\left.\frac{\partial e_{pk}}{\partial e_{gk}}\right|_{i_p}$ and r_p is $\left.\frac{\partial e_{pk}}{\partial i_p}\right|_{e_{gk}}$. Substituting $E_{pp} - R_p i_p$ for e_{pk}, the plate current is found to be

$$i_p = \frac{E_{pp} + \mu e_{gk}}{r_p + R_p}, \tag{2}$$

and the plate voltage is

$$e_p = E_{pp} - R_p i_p = \frac{\dfrac{r_p}{R_p} E_{pp} - \mu e_{gk}}{1 + \dfrac{r_p}{R_p}}. \tag{3}$$

From Eq. (2) the current gain (considering R_p as the load) is

$$\frac{\partial i_p}{\partial e_{gk}} = \frac{\mu}{r_p + R_p}, \tag{4}$$

and from Eq. (3) the voltage gain is

$$\frac{\partial e_p}{\partial e_{gk}} = \frac{-\mu}{1 + \dfrac{r_p}{R_p}}. \tag{5}$$

Thus the current gain approaches g_m if R_p is small compared with r_p; and if R_p is large compared with r_p, the voltage gain approaches μ.
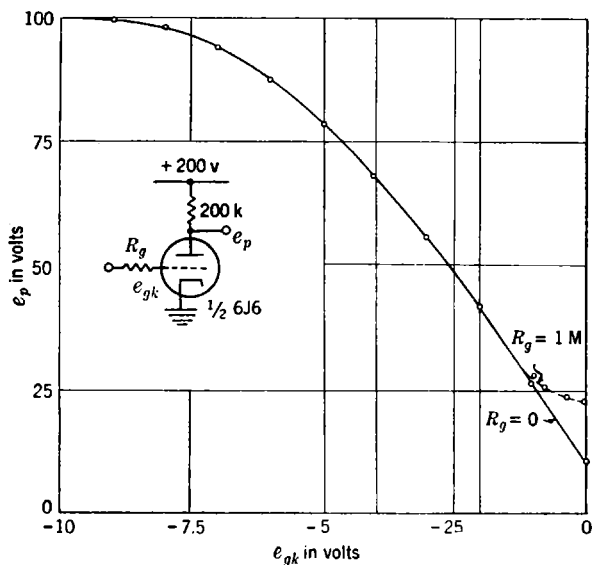
FIG. 11·11.—Typical triode amplifier characteristics.

This relation would indicate that a very high resistance should be employed for R_p if a voltage amplifier is to realize the maximum gain. Actually, r_p is approximately inversely proportional to plate current for small values of i_p, and μ drops slightly as plate current is decreased. Therefore, for a given E_{pp}, the gain ceases to increase as R_p is increased beyond a finite multiple of r_p. For instance, in a 6J6 triode, by measuring the increment of plate voltage obtained between $e_g = -2.5$ and -5

volts, it can be shown that if $E_{pp} = 200$ volts, increasing R_p above about 75,000 ohms gives no increase in the gain. On the other hand, an increase of E_{pp} gives an appreciable increase of gain.

Except where tube rating is exceeded, the factors limiting the range of voltage or current output are the plate-supply voltage (or zero-load current) and the point where grid current becomes excessive as determined by the input impedance. The latter, except for a low input impedance, generally occurs at a grid voltage of from -1 to -2 volts.

The linearity of a simple triode amplifier, considered over the entire output range, is rather poor because the plate resistance increases and the amplification factor decreases toward the low-current end of the load line. Figure 11·11 shows a typical triode amplifier characteristic and is derived by drawing a suitable load line on a family of i_p vs. e_p curves for a type 6J6 triode. The dashed curve results from grid current flowing in the input resistor.
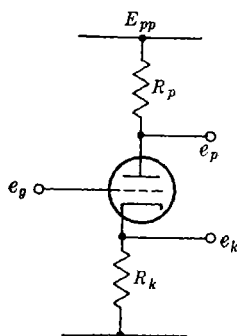
Fig. 11·12.—Triode with cathode and plate resistors.

However, linearity over a given range can be greatly improved by raising the plate-supply voltage to a value considerably above the upper limit required to obtain the desired output voltage. For many triodes, μ is constant with respect to plate voltage if plate current can be maintained constant. Thus, for good linearity as well as for maximum gain, the ideal voltage amplifier would have infinite R_p and infinite E_{pp}. Some constant-current device, such as will be described in connection with pentode amplifiers, may be employed in the plate circuit to approach this ideal. The linearity of output under conditions of constant plate current is illustrated for a 6SL7 in Fig. 11·20.

Triode Amplifier with Cathode Resistor.—When a resistance is inserted in series with the cathode, as in Fig. 11·12, the load line, if it is derived from $R_p + R_k$ instead of R_p, still describes the performance of the amplifier. But since e_{gk} is no longer the same as e_g but is equal to $e_g - i_p R_k$ (assuming no external cathode load), it is not possible to obtain an operating point directly from the load line. For a given e_g, one estimates i_p, computes the corresponding $i_p R_k$ and e_{gk}, refers to the diagram (Fig. 11·10) to see how far the resulting i_p is from the estimate, and repeats the process with some intermediate estimate of i_p. When R_k is large (comparable, for instance, with r_p or greater) and when e_g is positive and sufficiently large so that e_{gk} will be small in comparison, the first estimate of i_p is assumed to be equal to or a little larger than e_g/R_k. This usually proves to be an accurate assumption. When an operating point on the load line has been determined, its interpretation

is as follows: The difference between e_{pk} and E_{pp} is apportioned between R_p and R_k in proportion with their relative values.

When R_k is small, it may be convenient to redraw the plate characteristics to include the effect of this resistor and to consider the new graph as if an equivalent tube (having a different set of characteristics from the original) were being used. The procedure is illustrated in Fig. 11·13, where the plate characteristics of a 7F8 triode are modified by a 1000-ohm cathode resistor. The points are plotted by assuming values of e_g and e_{gk} and computing i_p from $i_p = (e_g - e_{gk})/R_k$. The resulting curves (shown solid) are the plate characteristics of the equivalent tube, upon which load lines may be drawn in the usual manner with no further consideration of R_k.
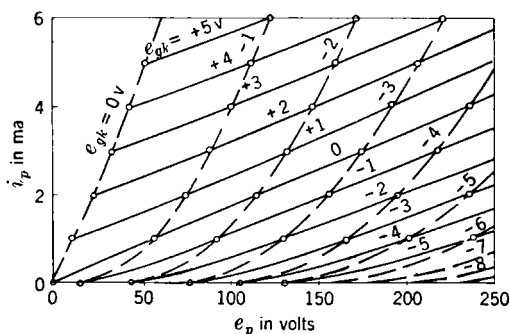
Fig. 11·13.—Modification of triode plate characteristics by a cathode resistor (7F8 triode, $R_k = 1000$ ohms).

It is helpful to analyze the circuit of Fig. 11·12 on the basis of the simple approximation of Eq. (1). Substituting $E_{pp} - R_p i_p - R_k i_p$ for e_{pk} and $e_g - R_k i_p$ for e_{gk}, the plate current is found to be

$$i_p = \frac{E_{pp} + \mu e_g}{r_p + R_p + (\mu + 1)R_k}. \tag{6}$$

A rather useful concept is obtained by comparing Eqs. (6) and (1). It is seen that when a triode has resistors in series with its cathode or plate or both, the triode with its resistors may be considered equivalent to a simple triode having the same μ but having a plate resistance equal to $r_p + R_p + (\mu + 1)R_k$. Figure 11·14 illustrates this equivalence. Figure 11·13 bears this out for the case of the cathode resistor: μ is unchanged in going from the dashed to the solid curves, but r_p is increased by $(\mu + 1)R_k$.

From Eq. (6) it follows that the plate and cathode voltages are (assuming negligible grid current)

$$e_p = \frac{\dfrac{r_p + (\mu + 1)R_k}{R_p} E_{pp} - \mu e_g}{1 + \dfrac{r_p + (\mu + 1)R_k}{R_p}}, \tag{7}$$

and

$$e_k = \frac{E_{pp} + \mu e_g}{\mu + 1 + \dfrac{r_p + R_p}{R_k}}. \tag{8}$$

Equation (7) could have been derived from Eq. (3) by considering R_k as part of an equivalent triode and consequently adding $(\mu + 1)R_k$ to r_p, as suggested in the foregoing paragraph.
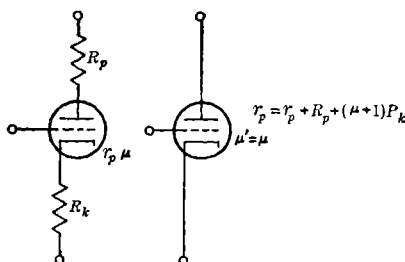

FIG. 11·14.—Triode equivalent of triode with plate and/or cathode resistor.

Equation (7) shows a gain from grid to plate of

$$\mathcal{G} = \frac{-\mu}{1 + \dfrac{r_p + (\mu + 1)R_k}{R_p}}. \tag{9}$$

The cathode resistor affords a negative feedback that, in addition to reducing the gain, reduces the variation in gain caused by variations of r_p and μ. This tends to keep the gain constant when tubes are changed and also to make the output-vs.-input curve more linear. The effect of variations of r_p and μ on the gain may be found by partial differentiation of Eq. (9). The resulting fractional change of gain is

$$\frac{\Delta \mathcal{G}}{\mathcal{G}} = \frac{r_p + R_p + R_k}{r_p + R_p + (\mu + 1)R_k} \frac{\Delta \mu}{\mu} - \frac{r_p}{r_p + R_p + (\mu + 1)R_k} \frac{\Delta r_p}{r_p} \tag{10}$$

Thus if R_k is such that $(\mu + 1)R_k$ is n times $(r_p + R_p)$ the gain is reduced by a factor of $n + 1$ [see Eq. (9)]. The fractional change in gain caused by a given variation of r_p is then reduced by the same factor as that caused by a variation of μ if μ is large compared with n.

If μ is large and μR_k is large compared with $r_p + R_p$, the gain is approximately R_p/R_k. This gain and the deviation from it are shown

by Fig. 11·15a, which is one form of graphical representation of Eq. (9). Figure 11·15b is a different form for use where R_k is low.

The variational plate output impedance of a simple triode amplifier (Fig. 11·9) is equivalent to r_p and R_p in parallel. By applying the equivalent-triode theory developed above, it is found that the plate
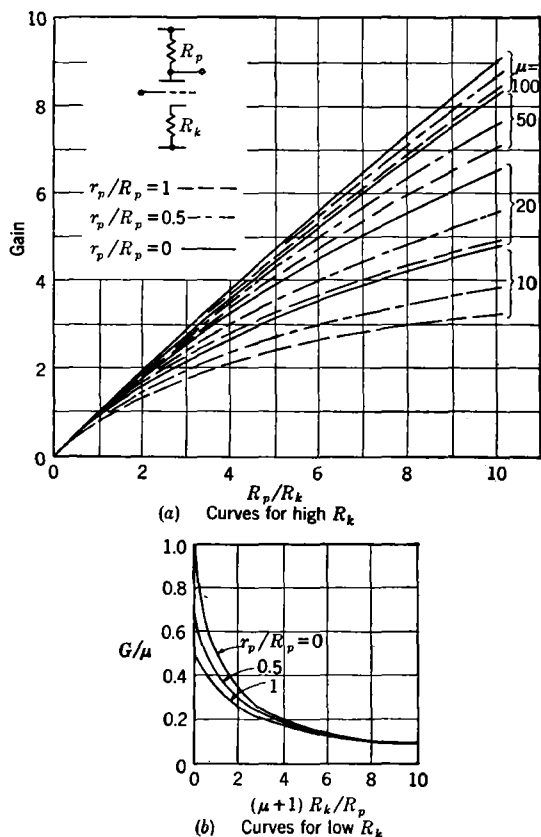

(a)   Curves for high R_k

(b)   Curves for low R_k

FIG. 11·15.—Gain of triode amplifier with cathode resistor.

output impedance for a triode amplifier with cathode resistor (Fig. 11·12) is

$$Z_p = \frac{r_p + (\mu + 1)R_k}{R_p + r_p + (\mu + 1)R_k} R_p. \tag{11}$$

To find the cathode output impedance, consider the change in the current flowing into the cathode terminal in Fig. 11·9 if the cathode is ungrounded

and e_k is varied by external means. This current is $-i_p$, which from Eq. (2) is

$$-i_p = \frac{E_{pp} - e_k + \mu(e_p - e_k)}{r_p + R_p};$$

therefore, if e_k is raised by an increment Δe_k without changing E_{pp} or e_g,

$$\frac{\Delta(-i_p)}{\Delta e_k} = \frac{+1}{r_p + R_p}, \tag{12}$$

which is thus the admittance of the cathode terminal. If this impedance is put in parallel with R_k, the cathode output impedance of a triode with cathode resistor (Fig. 11·12) is

$$Z_k = \frac{r_p + R_p}{\mu + 1 + \dfrac{r_p + R_p}{R_k}}. \tag{13}$$

A similar procedure may also be employed in the derivation of Eq. (11).

D-c Cathode Follower.—If the cathode is to be used as the output terminal, the circuit of Fig. 11·12 is a "cathode follower," although the usual practice is to omit R_p. In the latter case Eqs. (6), (8), and (13) become

$$i_p = \frac{E_{pp} + \mu e_g}{r_p + (\mu + 1)R_k}, \tag{14}$$

$$e_k = \frac{E_{pp} + \mu e_g}{\mu + 1 + \dfrac{r_p}{R_k}}. \tag{15}$$

$$Z_k = \frac{r_p}{\mu + 1 + \dfrac{r_p}{R_k}}. \tag{16}$$

Generally μ and R_k are large enough so that for practical purposes the following approximations can be used:

$$i_p \approx \frac{E_{pp}}{\mu R_k} + \frac{1}{\left(1 + \dfrac{r_p + R_k}{\mu R_k}\right)} \frac{e_g}{R_k}, \tag{17}$$

$$e_k \approx \frac{E_{pp}}{\mu} + \left(1 - \frac{r_p + R_k}{\mu R_k}\right) e_g, \tag{18}$$

$$Z_k \approx \frac{1}{g_m}. \tag{19}$$

The cathode follower is not a voltage amplifier but is a simple and excellent impedance changer or power amplifier, in which a high degree

of stability and linearity is afforded by the inherent negative feedback. The variability of gain may be obtained directly from Eq. (10); and since the same current flows in both R_p and R_k, the fractional gain variation is the same measured across either resistor. Dropping R_p and rearranging Eq. (10) for clarity,

$$\frac{\Delta\mathcal{G}}{\mathcal{G}} = \frac{1 + \dfrac{R_k}{r_p}}{1 + (\mu + 1)\dfrac{R_k}{r_p}}\frac{\Delta\mu}{\mu} - \frac{1}{1 + (\mu + 1)\dfrac{R_k}{r_p}}\frac{\Delta r_p}{r_p}. \qquad (20)$$

For example, if μ is 20 and R_k is twice the average value of r_p, and if μ increases by 20 per cent and r_p decreases by 50 per cent of its average value as the working range is traversed, the percentage change of gain is

$$\frac{\Delta\mathcal{G}}{\mathcal{G}} = \frac{1 + 2}{1 + 21 \times 2} \times 0.2 - \frac{1}{1 + 21 \times 2} \times (-0.5)$$
$$= 0.014 + 0.012 = 2.6 \qquad \text{per cent,}$$

which indicates a 2.6 per cent change in slope of the output-vs.-input curve, or a deviation from linearity of less than ± 0.2 per cent.[1] Equation (15) shows that a given variation of plate-supply voltage results in a variation in the output of a cathode follower of only $1/(\mu + 1 + r_p/R_k)$ times this change of voltage.

The output-vs.-input curve of a cathode follower may be obtained from a load line with a slope of $-1/R_k$ drawn on the plate characteristics (Fig. 11·10). (E_{pp} is, of course, measured from the fixed terminal of R_k in case this is not at ground potential.) The output voltage is simply $E_{pp} - e_{pk}$; and for any given output, the input is this value plus e_{gk} at this point on the load line. The load line reveals immediately, without the necessity of plotting the output curve, the useful range of the cathode follower (as limited by grid current and by low plate current) and the tube power dissipation.

The linearity of a cathode follower may be greatly increased for special applications where precision is paramount by the use of a "constant-current" device in place of R_k. As has been mentioned, the μ of many triodes at constant current is independent of plate voltage. Therefore, good linearity is obtainable if the load line can be made horizontal. The devices of Fig. 11·16 afford good approximations to constant current. The pentode circuit of Fig. 11·16a relies on the fact that the pentode plate current is fairly independent of plate voltage. Current is determined by the self-biasing resistor R and by the screen potential. The latter should be as low as possible, because if the pentode plate falls below the

---

[1] *Cf.* Fig. 11·52, Eq. (102).

level of the screen, the screen current will increase.   This effect, in turn, will decrease the plate current by increasing the biasing current in R_k and by lowering the screen voltage by extra loading of the bleeder.   If a suitable negative voltage source is available, it may be convenient to connect R_k and the grid to this source and to connect the screen to ground.
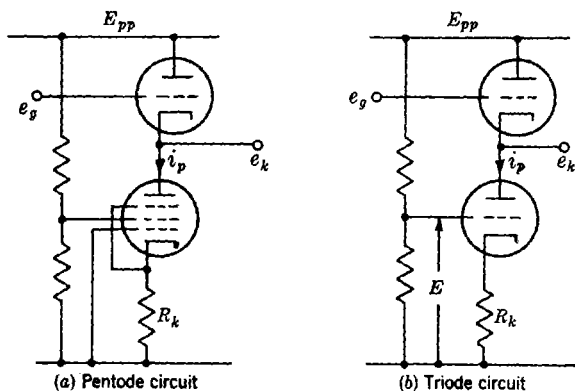

Fig. 11·16.—Constant-current devices replacing R_k of a cathode follower.

The triode circuit of Fig. 11·16b (if it is assumed that e_k remains sufficiently above E so that there is negligible grid current) may be analyzed by application of Eq. (6).

$$i_p = \frac{e_k + \mu E}{r_p + (\mu + 1)R_k}.$$

Thus the equivalent resistance is

$$\frac{\partial e_k}{\partial i_p} = r_p + (\mu + 1)R_k,$$

and it behaves within the limit of useful range as if its nether end were connected to a potential μE below the reference level.   This triode constant-current device has much greater stability with respect to tube drift than has the pentode because a given drift of grid-to-cathode voltage causes less percentage change of the drop across R_k.   Figure 11·20b shows the improvement in linearity of a cathode follower that may be achieved by using such a constant-current device.

If a constant-current device is employed, the cathode-follower output impedance [Eq. (16)] is simply the reciprocal of the transconductance.

11·8. Single-ended Pentode Amplifiers.—If both the cathode and screen grid of a pentode (or tetrode) are at fixed potentials, with no series resistances, the operation of the amplifier may be determined by

means of the load line, as for a triode (see Fig. 11·17).   The plate characteristics employed must apply to the existing screen potential.

A complete mathematical description of the approximate operating characteristics, such as was based on Eq. (1) for the triode, is not feasible here.   However, since μ and r_p are derivatives at an operating point, the gain, which is also a derivative, is correctly expressed by Eq. (5).
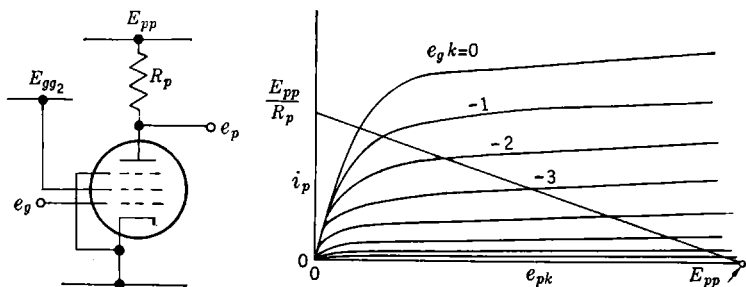

FIG. 11·17.—Pentode amplifier and load line on plate characteristics.

As μ is in practice large and rather indeterminate, the gain is better expressed for computation in the form

$$\frac{\partial e_p}{\partial e_g} = -\frac{g_m R_p}{1 + \dfrac{R_p}{r_p}}. \tag{21}$$

Since, in general, r_p is very great compared with R_p, a good approximation is

$$\frac{\partial e_p}{\partial e_g} \approx -g_m R_p. \tag{22}$$

It would seem that the voltage gain is proportional to the plate-load resistance except for very high values of R_p.   Actually, in the case of any ordinary "sharp cutoff" pentode operating as a voltage amplifier, the transconductance is very nearly proportional to the plate current.   Thus for a given plate-supply voltage and a certain operating range of plate voltage, the gain is independent of R_p over a very wide range of values.   On the other hand, it is roughly proportional to $E_{pp} - e_p$.

If E_{pp}, together with R_p, can be raised without limit, a very high gain may be attained.   The same effect may be achieved by the use of some kind of constant-current device in place of the load resistor.   Such a device is shown in Fig. 11·18.   This is the same as the constant-current element shown in Fig. 11·16b except that here a battery or its equivalent is needed.   The operation of the *triode* (if there is assumed to be sufficient plate-to-cathode voltage to keep its grid from drawing current)

may be determined by substitution of the "equivalent triode" with R_k included.   This substitution changes the plate resistance to

$$r_p + (\mu + 1)R_k,$$

and in applying Eq. (1), $e_{pk} = E_{pp} - e_p$ and $e_{gk} = E$ (Fig. 11·14),

$$i_p = \frac{E_{pp} - e_p + \mu E}{r_p + (\mu + 1)R_k}. \tag{23}$$

Thus, except for an upper working limit somewhat below E_{pp}, this

FIG. 11·18.—Constant-current device as a pentode load resistor.

triode arrangement behaves like a resistor of value $r_p + (\mu + 1)R_k$ (or roughly, μR_k), whose upper terminal is attached to a potential μE above E_{pp}. This device is not often practical because it requires a battery or floating power supply.

In general, with a given plate-supply voltage and without the use of a special circuit such as the foregoing, the maximum voltage gain is obtained when the screen grid is operated at the lowest potential permissible from the standpoint of control-grid current.   The g_m of ordinary sharp cutoff pentodes, at any given plate current, increases slightly with reduced screen potential.

If the screen-to-cathode voltage is subject to variation, the analysis of the operation of an amplifier is more involved than if it is fixed.   For rough calculation, advantage may be taken of the facts that over the usual range of operation, both plate and screen current are fairly independent of plate voltage, and screen current is a substantially constant ratio of plate current (between $\frac{1}{4}$ and $\frac{1}{2}$). An equation similar to Eq. (1) is sometimes useful, although it is not so accurate or complete a description of the characteristics.   The plate current is the same sort of function of screen voltage as it was of plate voltage in the case of the triode.   Therefore, in applying Eq. (1) to a pentode, $\partial e_{g_2}/\partial i_p$ is used instead of $\partial e_p/\partial i_p$, and $\partial e_{g_2}/\partial e_g$ instead of $\partial e_p/\partial e_g$.

$$i_p = \frac{e_{g_2} + \dfrac{\partial e_{g_2}}{\partial e_g} e_{gk}}{\dfrac{\partial e_{g_2}}{\partial i_p}}. \tag{24}$$

Thus, if the screen grid has a series resistor, R_{g_2} (e.g., from a bleeder whose output resistance is appreciable), and if the assumption is made that i_{g_2} is a constant fraction n of i_p, then

$$i_p = \frac{E_{gg_2} - R_{g_2} n i_p + \dfrac{\partial e_{g_2}}{\partial e_g} e_g}{\dfrac{\partial e_{g_2}}{\partial i_p}},$$

from which

$$i_p = \frac{E_{gg_2} + \dfrac{\partial e_{g_2}}{\partial e_g} e_g}{\dfrac{\partial e_{g_2}}{\partial i_p} + n R_{g_2}}, \tag{25}$$

and

$$\begin{aligned} e_p &= E_{pp} - i_p R_p \\ &= E_{pp} - \frac{R_p E_{gg_2} + R_p \dfrac{\partial e_{g_2}}{\partial e_g} e_g}{\dfrac{\partial e_{g_2}}{\partial i_p} + n R_{g_2}}. \end{aligned} \tag{26}$$

The gain is

$$\frac{\partial e_p}{\partial e_g} = - \frac{R_p \dfrac{\partial e_{g_2}}{\partial e_g}}{\dfrac{\partial e_{g_2}}{\partial i_p} + n R_{g_2}} = - \frac{R_p g_m}{1 + n R_{g_2} \dfrac{\partial i_p}{\partial e_{g_2}}}. \tag{27}$$

If the pentode amplifier is further complicated by a cathode resistor, as in Fig. 11·19, the current flowing in this resistor must be considered to include screen current. But if the assumption is made that $i_{g_2} = n i_p$, where n is a constant, then i_k is $(n + 1)i_p$. Equation (24) yields

$$i_p = \frac{E_{gg_2} - n R_{g_2} i_p - (n + 1) i_p R_k + \dfrac{\partial e_{g_2}}{\partial e_g} [e_g - (n + 1) i_p R_k]}{\dfrac{\partial e_{g_2}}{\partial i_p}}$$

from which

$$i_p = \frac{E_{gg_2} + \dfrac{\partial e_{g_2}}{\partial e_g} e_g}{\dfrac{\partial e_{g_2}}{\partial i_p} + n R_{g_2} + \left(\dfrac{\partial e_{g_2}}{\partial e_g} + 1\right)(n + 1) R_k}. \tag{28}$$

The voltage gain is

$$\frac{\partial e_p}{\partial e_g} = - \frac{R_p g_m}{1 + n R_{g_2} \dfrac{\partial i_p}{\partial e_{g_2}} + (1 + n) R_k \left(g_m + \dfrac{\partial i_p}{\partial e_{g_2}}\right)}. \tag{29}$$

Pentode Cathode Follower.—In a-c applications a cathode follower can make capital of the high gain afforded by a pentode, for the purposes of obtaining good linearity and low attenuation, by the use of a condenser to maintain constant screen-to-cathode potential. In d-c circuits, however, where (unless a battery or floating power supply is used to set
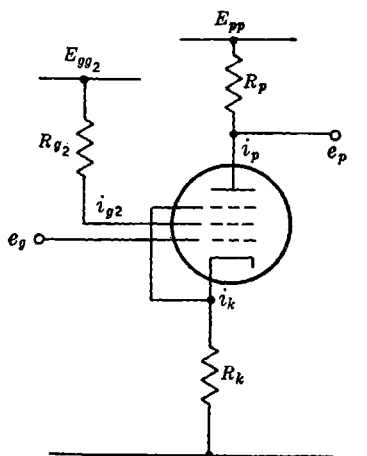

Fig. 11·19.—Pentode amplifier with cathode and screen resistors.

the screen at a fixed potential with respect to the cathode) the screen-to-cathode potential varies along with the plate-to-cathode potential, the pentode operates as a triode. In computing the cathode-follower performance, the triode-connected plate characteristics apply. There are cases, however, where some advantage may be taken of the extra electrode. For example, if a stable output at high power level is desired, and if a regulated, but low-power, voltage source is available, the screen can be tied to this regulated source, and the plate, which furnishes most of the power, can be connected to an unregulated source. The output will then hardly be affected by plate-voltage fluctuation.

11·9. Cascode and Other Series Amplifiers. *Constant-current Plate Load.*—It is evident that the gain of an amplifier is a maximum when both R_p and E_{pp} are infinite and i_p is finite. In this case, also, the output signal is most linear, since the gain, being simply μ, is independent of r_p. For most triodes μ, at any given plate current, tends to be constant over a wide range of plate voltage. The effect of infinite or very great E_{pp} and R_p may be achieved by the use of a "constant-current" device between the plate and the $B+$ voltage supply.

One such device is a diode with such a low filament voltage that its current is limited by cathode emission and is thus independent of its plate-to-cathode potential. A diode with an indirectly heated, oxide-coated cathode does not have really constant current at low temperature because the emission depends partly on plate voltage,[1] but the partial effect is of some value. Figure 11·20a shows the increase in both gain and linearity of a triode amplifier using a diode instead of a fixed plate resistor. A diode having a pure tungsten cathode would be more effective but less convenient. In either instance, the device is subject to con-

[1] See W. S. Dow, *Fundamentals of Engineering Electronics*, Wiley, New York, 1937, p. 204.

siderable drift if the filament supply voltage varies. Figure 11·20b shows analogous data for a cathode follower using the constant-current triode circuit shown in Fig. 11·16b.

A high-vacuum *photoelectric* cell has a current independent of voltage at any given value of illumination and thus may be used as a constant-current device. The current is at most a few microamperes, and the illumination must be constant; however, the scheme may be of value in certain applications.
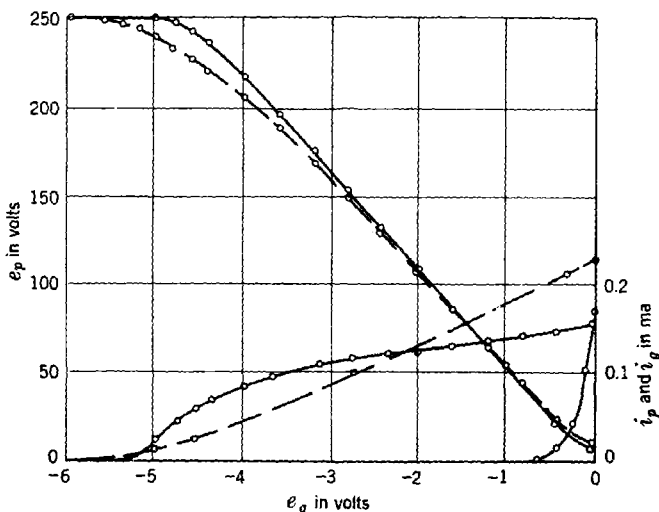

Fig. 11·20a.—Output characteristics of a triode amplifier with and without a "constant-current" diode as plate load (6SL7 triode, $E_{pp} = 250$ volts). Solid curves: plate load is a 6H6 diode, $E_f = 2.4$ volts; dashed curves: plate load is a 1.1-megohm resistor.

The constant-current arrangement of Fig. 11·16b, as used in a cathode-follower circuit, may be adapted as an amplifier plate load as shown in Fig. 11·21. In this adaptation, since the cathode end rather than the plate end of the device varies in potential, a battery or floating supply is needed to determine the grid potential, although no appreciable current is required from this source. A small constant-voltage glow tube, with its current furnished by a resistor from the positive supply, may be used as the voltage standard in place of the battery, but this current is variable and therefore must be small compared with plate current.

The device of Fig. 11·21 is a cathode follower in which the current is approximately E/R. The greater E and μ_2 are the more nearly constant will be the current. From Eq. (6), the current is

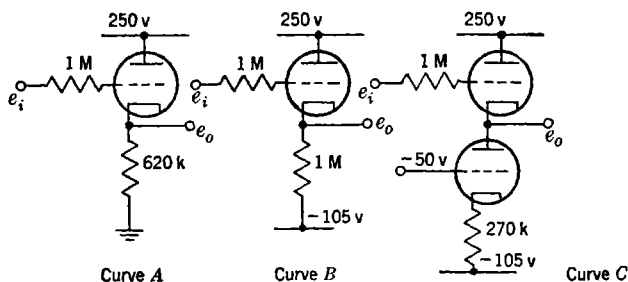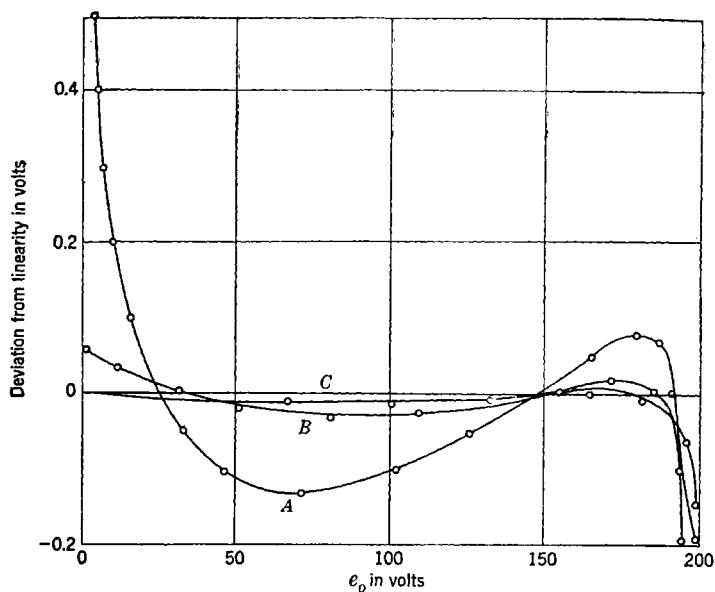$$i_p = \frac{E_{pp} - e_p + \mu_2 E}{r_{p2} + (\mu_2 + 1)R}, \tag{30}$$

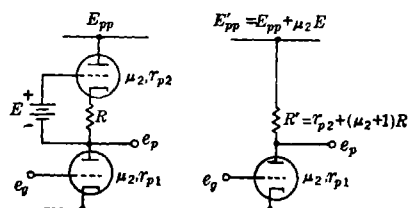Fig. 11·20b.—Cathode-follower linearity.

Fig. 11·21.—Triode amplifier with "constant-current" plate load.   (a) Complete circuit; (b) equivalent circuit.

and its rate of change with respect to amplifier plate voltage is

$$\frac{di_p}{de_p} = -\frac{1}{r_{p2} + (\mu_2 + 1)R}. \tag{31}$$

The device becomes inoperative before e_p reaches E_{pp}, but it is seen from Eq. (30) that over its operating range, the current behaves as if it would be zero when $e_p = E_{pp} + \mu E$. Thus, except for the limited operating range, the device is equivalent to a plain resistance of value

$$R' = r_{p2} + (\mu_2 + 1)R \tag{32}$$

returned to a $B+$ voltage of

$$E'_{pp} = E_{pp} + \mu_2 E,$$

as shown in Fig. 11·21b.
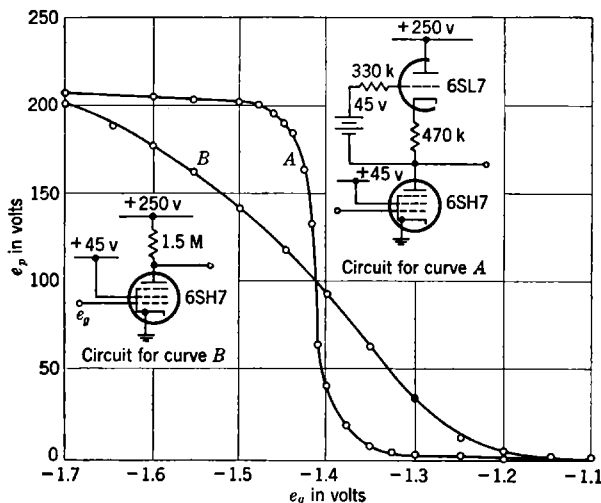
Fig. 11·22.—Output characteristics of a pentode amplifier with and without a constant-current plate load.

The constant-current circuit of Fig. 11·21 is especially effective in increasing the gain of a pentode because the gain with constant current is μ, which, for a pentode at low current, is very high. Figure 11·22 shows the output curve of a 6SH7 pentode with a 6SL7 constant-current plate load, where $E = 45$ volts. The gain is about 7000, whereas with a simple resistor and the same $B+$ voltage it is only 600.

The output impedance of a pentode amplifier with a constant-current load is very high; and if it is followed by a resistance divider, the gain may be seriously decreased. In effect, the divider resistance parallels the equivalent resistance of the constant-current circuit. If the voltage

level must be lowered before being applied to the grid of the next stage, a cathode follower is advisable for the purpose.

If the battery is omitted in Fig. 11·22, the circuit does not, in any sense, comprise a constant-current device. It behaves like a simple resistor, of value $r_{p2} + (\mu_2 + 1)R$, attached to the E_{pp} source. Therefore, no increase of gain obtains. If the two triodes are similar, however, and if the amplifier triode also has a cathode resistor of value R, the arrangement is useful in that the output voltage is more linear than that of a simple amplifier and the effect of cathode temperature variation is canceled (see Sec. 11·11).

The pentode constant-current circuit of Fig. 11·16a may be adapted as a plate load for an amplifier by the use of a battery or its equivalent for the screen supply. This type of supply has the disadvantage that appreciable current is drawn therefrom. A simple divider for the screen supply attached between the cathode and E_{pp} is ineffective, because in this case i_p will tend toward zero as e_p approaches E_{pp}, as is the case with a plain resistance plate load.

Cascode Amplifier.—The preceding constant-current circuits have the aim of making the gain equal to μ by eliminating the $1/g_m R_p$ term from the gain equation

Fig. 11·23.—Cascode amplifier.
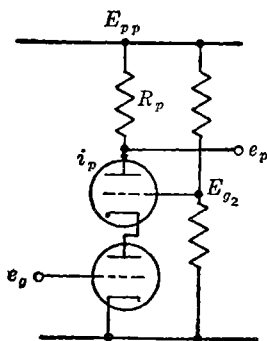
$$\mathcal{G} = \frac{-1}{\frac{1}{\mu} + \frac{1}{g_m R_p}}. \qquad (33)$$

Thus they are most effective for a pentode amplifier where $g_m R_p$ normally is much less than μ. In a triode amplifier, on the other hand, μ is much smaller, and the first term in the denominator of Eq. (33) is usually the more important. It is the result of variation of the plate voltage, which has very little effect in a pentode because of the isolating effect of the screen on the plate. A cascode amplifier tends to hold the triode plate potential fixed but still permits its current to flow in a load resistor. Thus the behavior is similar to that of a pentode, with the advantage that no screen current is required.

A circuit arrangement is shown in Fig. 11·23. The upper tube has a fixed grid, so that the excursions of its cathode are limited. If e_{p1} were really held constant, the current gain of the lower triode would be g_m and the voltage gain from e_g to e_p would be $-g_m R_p$, as in a pentode.

Actually the lower triode has as its plate load the cathode input impedance of the upper tube, which is, from Eq. (12), $(r_p + R_p)/(\mu + 1)$. The current gain is thus

$$\frac{\Delta i_p}{\Delta e_g} = \frac{\mu}{r_p + \dfrac{r_p + R_p}{\mu + 1}}, \tag{34}$$

and the voltage gain is

$$\begin{aligned} \mathcal{G} &= -\frac{\mu R_p}{r_p + \dfrac{r_p + R_p}{\mu + 1}} \\ &= -\frac{1}{\dfrac{1}{g_m R_p} + \dfrac{r_p + R_p}{R_p} \dfrac{1}{\mu(\mu + 1)}}. \end{aligned} \tag{35}$$

Thus, if R_p is large in comparison with r_p, μ^2 takes the place of the μ in Eq. (33).

The value of E_{g2} should be sufficiently high to keep the input grid from drawing appreciable grid current, and the operating range of e_p should be far enough above E_{g2} so that the current in the upper grid will not become comparable with i_p.

11·10. Differential Amplifiers. *Cathode as Input Terminal to an Amplifier.*—The cathode or grid of a tube or both may be employed as input terminals. From Eq. (2), the plate current in Fig. 11·24 is

$$i_p = \frac{E_{pp} + \mu e_g - (\mu + 1)e_k}{r_p + R_p}. \tag{36}$$

The gain of the amplifier with respect to the cathode input terminal is $-(\mu + 1)/\mu$ times that with respect to the grid input terminal. If both input terminals are used, the output voltage is much more sensitive to changes in their difference than to changes in their common level. If their difference is held constant, e_p will change only $R_p/(r_p + R_p)$ as much as

FIG. 11·24.—Triode with two input terminals.

any common change of e_k and e_g, whereas e_p responds by μ times this much to any change of their difference.

The input impedance to the cathode is low and is equal to

$$Z_k = \frac{r_p + R_p}{\mu + 1}. \tag{37}$$

Any resistance in the cathode circuit R_k' has the same effect as adding $(\mu + 1)R_k'$ to the plate resistance. Thus, Eq. (36) becomes

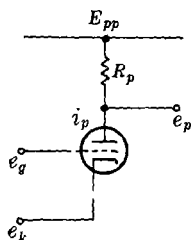$$i_p = \frac{E_{pp} + \mu e_g - (\mu + 1)e_k'}{r_p + R_p + (\mu + 1)R_k'}, \tag{38}$$

where e_k' is the cathode input voltage outside of R_k'.   The output voltage
is still μ times as responsive to input voltage difference as to mean input
voltage level.

In spite of the low impedance of the cathode, it is often convenient
to employ it rather than the grid as the input terminal, because the

amplifier used in this way gives an
output voltage of the same sense as
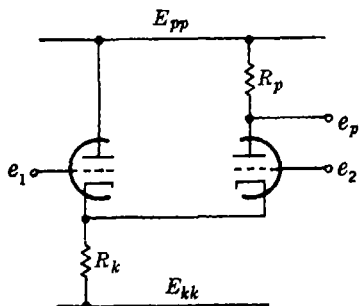that of the input voltage.

*Cathode Input Voltage Applied
through a Cathode Follower.*—A cathode
follower affords a convenient means of
obtaining the necessary low impedance
for the cathode input terminal.   In
addition to high input impedance and
amplification without inversion, the
arrangement illustrated in Fig. 11·25
provides a balancing of the effect of
heater-voltage variation and a con-

Fig. 11·25.—Cathode follower as an
input means.

venient high-impedance terminal (e_2) for zero adjustment or feedback or
both.

The input voltage is attenuated according to the voltage gain of the
cathode follower, which is $\mu/(\mu + 1 + r_p/R_k)$.   Also, the output imped-
ance of the cathode follower, $r_p/(\mu + 1 + r_p/R_k)$, acts as a resistance in
series with the cathode of the amplifier.   Thus, the gain from e_1 to e_p is

$$\mathcal{G}_1 = \left(\frac{\mu_1}{\mu_1 + 1 + \dfrac{r_{p1}}{R_k}}\right)\left(\frac{(\mu_2 + 1)R_p}{r_{p2} + R_p + \dfrac{(\mu_2 + 1)r_{p1}}{\mu_1 + 1 + \dfrac{r_{p1}}{R_k}}}\right), \qquad (39)$$

in which the subscript numbers 1 and 2 refer to the cathode follower and
amplifier triode, respectively.   If these are the same, the gain is

$$\mathcal{G}_1 = \frac{\mu}{1 + \dfrac{2r_p}{R_p} + \dfrac{r_p(r_p + R_p)}{R_p R_k(\mu + 1)}}. \qquad (40)$$

Cathode Follower Used as a Cathode Return.—The circuit of Fig. 11·25
may even be used when voltage inversion is acceptable; in this case
the input voltage may be applied to the amplifier grid.   Possible reasons
for using the circuit in this way are (1) The input voltage may be at some
level far from ground, and a bleeder of sufficiently low resistance for the
cathode return may be undesirable; (2) the cathode follower grid may be
used for negative or positive feedback; (3) the cathode follower may be

used for a zero adjustment for the input level; (4) the effect of heater-voltage variation on zero shift is largely canceled; and (5) the output characteristic may be more nearly linear because of a mutual cancellation of the variation of plate resistance.   The gain from e_2 to e_p is computed from Eq. (9), where the cathode resistor is the cathode-follower output impedance.

$$\mathcal{G}_2 = \frac{-\mu_2 R_p}{r_{p2} + R_p + \dfrac{(\mu_2 + 1)r_{p1}}{\mu_1 + 1 + \dfrac{r_{p1}}{R_k}}}. \tag{41}$$

If the tubes have the same characteristics,

$$\mathcal{G}_2 = -\frac{\mu + \dfrac{\mu}{\mu + 1}\dfrac{r_p}{R_k}}{1 + \dfrac{2r_p}{R_p} + \dfrac{r_p(r_p + R_p)}{R_p R_k(\mu + 1)}}. \tag{42}$$

If the μ's are equal and R_k is large, the last term in the denominator of Eq. (41) is approximately r_{p1}. The sum of the two plate currents, which is equal to the current in R_k, is substantially constant, so that a decrease in the amplifier current causes an almost equal increase in the plate current of the cathode follower.   If the two plate currents are of the same order of magnitude, the variations of plate resistance caused by variation of plate current will be about equal and opposite.   The denominator of Eq. (41) will thus be fairly constant over a considerable output range, and an increase in linearity will result.

Differential Input.—The circuit of Fig. 11·25 is more suitable than that of Fig. 11·24 for the amplification of the difference between two voltages, because both input terminals have a high impedance, the cathodes mutually offset the drift due to heater-voltage variation, and the output voltage may be less affected by the "common-mode variation" (common change of level) of the two input voltages.

This independence of the common mode is a feature of importance when the amplifier is to be used as a comparator,[1] in which the output voltage ideally should be a function only of the difference between the input voltages and not of their average.   The degree of rejection of the common mode may be expressed by the ratio of the gain of the amplifier with respect to the average of the input voltages to that with respect to their difference.   If the two input potentials are e_1 and e_2, an output voltage variation may be expressed as

$$\Delta e_0 = \mathcal{G}_1 \, \Delta e_1 + \mathcal{G}_2 \, \Delta e_2 \tag{43}$$

[1] *Cf.* Vol. 19, Chap. 17.

or

$$\Delta e_0 = \frac{\mathcal{G}_1 - \mathcal{G}_2}{2}(\Delta e_1 - \Delta e_2) + (\mathcal{G}_1 + \mathcal{G}_2)\frac{\Delta e_1 + \Delta e_2}{2}. \qquad (44)$$

The second coefficient in Eq. (44), $\mathcal{G}_1 + \mathcal{G}_2$, should be zero or very small in comparison with the first coefficient $(\mathcal{G}_1 - \mathcal{G}_2)/2$.

In the case of Fig. 11·24 the proportional response to the common mode, expressed as the ratio of $(\mathcal{G}_1 + \mathcal{G}_2)$ to $(\mathcal{G}_1 - \mathcal{G}_2)/2$ is approximately $1/\mu$. For Fig. 11·25, where the triodes have been assumed to be identical, comparison of Eqs. (40) and (43) shows the ratio to be approximately

$$\frac{\mathcal{G}_1 + \mathcal{G}_2}{\dfrac{\mathcal{G}_1 - \mathcal{G}_2}{2}} \approx \frac{-r_p}{(\mu + 1)R_k}. \qquad (45)$$

With given voltage levels, this ratio cannot be reduced indefinitely by increasing R_k because the resulting decrease of current causes r_p to increase
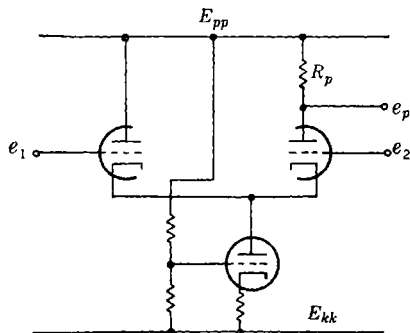

Fig. 11·26.—Use of constant-current device to minimize common-mode effect.

as well. A large negative return potential for B_k is advisable. Figure 11·26 illustrates the use of a constant-current device to simulate a very large R_k without reduction of current.

Even with an effectively infinite R_k, however, the common-mode rejection depends on equality between the μ's of the two triodes. The assumption of $R_k = \infty$ in Eqs. (39) and (41) shows that if the tubes in Fig. 11·22 are dissimilar, even though the constant-current device is perfect, the proportional common-mode variation is

$$\frac{\mathcal{G}_1 + \mathcal{G}_2}{\dfrac{\mathcal{G}_1 - \mathcal{G}_2}{2}} = \frac{\mu_1 - \mu_2}{\mu_1\mu_2 + \dfrac{\mu_1 + \mu_2}{2}}, \qquad (46)$$

$$\approx \frac{\mu_1 - \mu_2}{\mu_1\mu_2}, \qquad (47)$$

or approximately $1/\mu$ times the fractional difference of the μ's. Thus, high-μ tubes are indicated, and double triodes are preferable because they are generally better matched. For 6SL7's, for example, Eq. (47) rarely exceeds 0.1 per cent. The 6SU7 double triode is a specially designed and selected 6SL7 in which the quantity expressed by Eq. (47) is held to a minimum.[1]

From Eqs. (11) and (16), the output impedance of the circuit of Fig. 11·25 is

$$Z_p = \frac{R_p r_p \left(1 + \dfrac{\mu + 1}{\mu + 1 + \dfrac{r_p}{R_k}} \right)}{R_p + r_p \left(1 + \dfrac{\mu + 1}{\mu + 1 + \dfrac{r_p}{R_k}} \right)}. \tag{48}$$

When R_k is large, this is approximately

$$Z_p \approx \frac{2R_p r_p}{R_p + 2r_p}, \tag{49}$$

or R_p in parallel with twice r_p.

The foregoing equations in this section relate mainly to variational quantities. The equations for the d-c levels are cumbersome and are of

---

[1] Care should be exercised not to read into Eq. (45) more than it actually says. For instance, although pentodes have very large μ's, Eq. (45) does not apply to their use, since in its derivation it is assumed that $R_k \gg r_{p1}$; but in view of the large plate resistance characteristic of pentodes it is unlikely that this condition can easily be made to obtain. Thus, when two 6SJ7 pentodes ($\mu > 1000$) are employed in a circuit like that of Fig. 11·22, the common-mode rejection is found to be no better than when two 6SL7 triodes are employed.

Again, if R_k is made large by the use of a large ohmic resistor, the plate currents of the two tubes may be so small that their individual μ's may differ by more than the amount measured under normal operating conditions. Nor can Eq. (45) be expected to apply if, when two tubes of differing μ are employed, some other means is used to balance the amplifier to give zero output for a certain pair of grid potentials.

The above is possibly the explanation of the fact that when highly sensitive amplifiers are used to detect cross talk between the two tubes being tested, a variation of the common-mode bias, for balanced tubes, passes through zones where a minimum, a maximum, or a point of inflection is obtained at the output. In these regions the output is almost independent of small variations of common-mode signals, whereas between these regions the output varies at an appreciable rate. The length of these zones, in which the output is independent—within the noise level of the amplifier—of the common-mode signal, varies with the type of tube. For example, it is found that the region may extend over several hundred millivolts of common-mode signal for type 6SL7 tubes. But for type 6SN7 tubes the region may be as great as 1 or 2 volts. Another series of experiments indicated that this region was between 500 and 600 mv for pentodes (type 1620), when the criterion was the observation of a signal corresponding to a 10 μv difference between the two grid potentials. Editor.

less value in analysis and design. The problem in this regard would be, typically, to determine $e_1 - e_2$ for a certain value of e_p and for a given level of the input voltages. The sum of the two tube currents will be known in the case of the circuit of Fig. 11·26, and in Fig. 11·25 it is
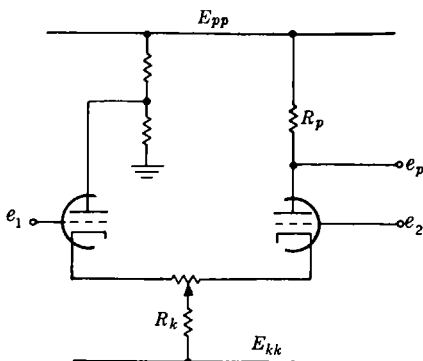

FIG. 11·27.—Method of zero adjustment.

usually approximately e_1/R_k. The current in the amplifier tube is $(E_{pp} - e_p)/R_p$, and that in the cathode follower is the difference between this and e_1/R_k. Thus the approximate e_{pk} and i_p are known for each tube, and the e_{gk}'s may be found by consulting the plate characteristic curves. Since the e_k's are the same, $e_1 - e_2$ is the difference between the e_{gk}'s.
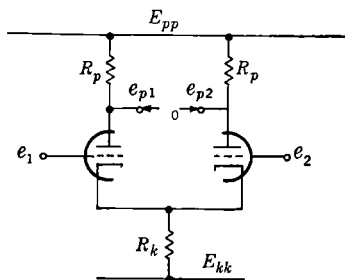
FIG. 11·28.—Symmetrical differential amplifier.

The cathode-follower input voltage e_1 generally will be somewhat lower than e_2 because of the higher plate voltage. The difference may be reduced by employing a lower-potential source for the cathode-follower plate; in fact, a potentiometer to this plate may comprise a zero adjustment. More commonly, however, the zero adjustment, if required, is a potentiometer between the cathodes as shown in Fig. 11·27. This decreases the gain by effectively adding $(\mu + 1)$ times half the resistance of the potentiometer to the r_p of each triode. (The resulting negative feedback may or may not be of value, depending on the application.) It also increases the common-mode susceptibility [Eq. (45)]. The value of this resistance depends on tube tolerances, but it may be held to a minimum by proper choice of the cathode-follower plate potential.

Differential Output.—A differential amplifier with two output terminals, such as the type illustrated by the symmetrical arrangement of Fig. 11·28, finds frequent application. It might be used as the first stage of a multistage amplifier; as a differential voltage output stage (actuating the plates of an oscilloscope, for example); or as a current or power output device, where the load (an ammeter, a relay, a motor field, a pair of magnetic oscilloscope deflecting coils, etc.), appears either between the two plates or, in two parts, as the R_p's.

To find the voltage gain from e_1 to e_{p2}, which is the same as that from e_2 to e_{p1} (assuming equal tubes), Eq. (39) may be applied. Because r_{p1} is increased by R_p, the resulting gain is

$$\mathcal{G}_{12} = \mathcal{G}_{21} = \frac{\mu R_p}{(R_p + r_p)\left[2 + \dfrac{R_p + r_p}{(\mu + 1)R_k}\right]}. \tag{50}$$

Similarly, from Eq. (41), the gain from e_1 to e_{p1} or from e_2 to e_{p2} is

$$\mathcal{G}_{11} = \mathcal{G}_{22} = \frac{-\mu R_p}{(R_p + r_p)\left[2 + \dfrac{R_p + r_p}{(\mu + 1)R_k}\right]}\left[1 + \frac{R_p + r_p}{(\mu + 1)R_k}\right] \tag{51}$$

The gain from either input terminal to the two output terminals across which the voltage e_0 appears is the difference between Eqs. (50) and (51).

$$\mathcal{G} = \frac{\mu R_p}{R_p + r_p} \tag{52}$$

This is identical with the gain of a simple triode amplifier and is not affected by R_k (except in so far as R_k determines the current and therefore affects r_p).

The differential output voltage is unaffected by common-mode variation of the input voltage (provided the μ's and r_p's are equal). Thus, for example, the input voltage might be applied to only one grid, the other remaining fixed, and the output voltage between the plates would be the same as if the input voltage were applied in "push-pull," one half to each grid. In the latter case, however, the output voltage would be truly "push-pull," whereas with one grid fixed the plate of the triode with the moving grid would move $1 + \dfrac{R_p + r_p}{(\mu + 1)R_k}$ times as far as the other plate. Usually this difference is very slight; but if R_k is small and it is important to have balanced output voltage, this may be accomplished by using slightly unequal R_p's.

The differential output voltage will be affected by the input voltage level if there are inequalities between the triode characteristics. This effect may be appreciable where the amplifier is being used to detect the

difference between two potentials, each of which is subject to considerable variation. When the triode characteristics are different, the gain from e_1 to e_0 may be found from Eqs. (39) and (41) in a manner similar to that by which Eq. (52) was obtained. It is

$$\mathcal{G}_1 = \frac{\mu_1 R_p \left[2 + \dfrac{r_{p2} + R_p}{(\mu_2 + 1)R_k} \right]}{\dfrac{\mu_1 + 1}{\mu_2 + 1}(r_{p2} + R_p) + (r_{p1} + R_p) + \dfrac{(r_{p1} + R_p)(r_{p2} + R_p)}{(\mu_2 + 1)R_k}}, \quad (53)$$

where the subscripts refer to the respective triodes. The gain $\mathcal{G}_2$, from e_2 to e_0, is the same as $\mathcal{G}_1$ with opposite sign and subscripts. If the differences between the μ's and r_p's are small, the proportional response to the common-mode input voltage is approximately

$$\frac{\mathcal{G}_1 + \mathcal{G}_2}{\dfrac{(\mathcal{G}_1 - \mathcal{G}_2)}{2}} \approx \frac{1 + \dfrac{(r_p + R_p)}{2R_k}}{\mu + 1}\frac{\mu_1 - \mu_2}{\mu} + \frac{r_p}{2(\mu + 1)R_k}\frac{r_{p2} - r_{p1}}{r_p} \quad (54)$$

Thus, if a given percentage of dissimilarity between either the μ's or the r_p's is assumed, the appearance of the common-mode input voltage in the differential output voltage is lessened by the use of higher-μ triodes and a higher R_k.

In some applications it is desirable that not only the differential output voltage but also the level of the two output voltages be unaffected by the input voltage level. The variation of the output voltage level is approximately $-R_p/2R_k$ times that of the input voltage; this variation may be eliminated by the use of a constant-current circuit in place of R_k. This method also minimizes the effect on the differential output voltage [Eq. (54)], thereby eliminating the effect of difference between the r_p's.

As has been pointed out, unless R_k is very small, the output voltage of a differential amplifier is very nearly "push-pull" even if the input voltage is applied to only one grid. This effect comes about by virtue of the cathode coupling; the cathodes undergo excursions about half as great as those of the moving grid, thus furnishing an input voltage to the cathode of the tube with fixed grid and simulating a true push-pull input. An important result is improved linearity over a large range of output voltages. This result is indicated by Eq. (53), since the sum of r_{p1} and r_{p2} tends to remain more nearly constant than either of them alone. Figure 11·29 illustrates the degree of linearity effected, as well as the balanced output voltage obtained with a single input voltage (*cf.* Fig. 11·11).

Equation (52) indicates the rate of change of differential output voltage with respect to differential input voltage.   By symmetry, the output voltage should be zero at zero input voltage.   Therefore, from Eq. (52), the voltage equation is

$$e_{p2} - e_{p1} = \frac{\mu R_p}{R_p + r_p} (e_1 - e_2). \tag{55}$$

It is evident that the differential output voltage is independent of plate-supply voltage, whereas in the case of the single-triode amplifier [Eq. (3)]
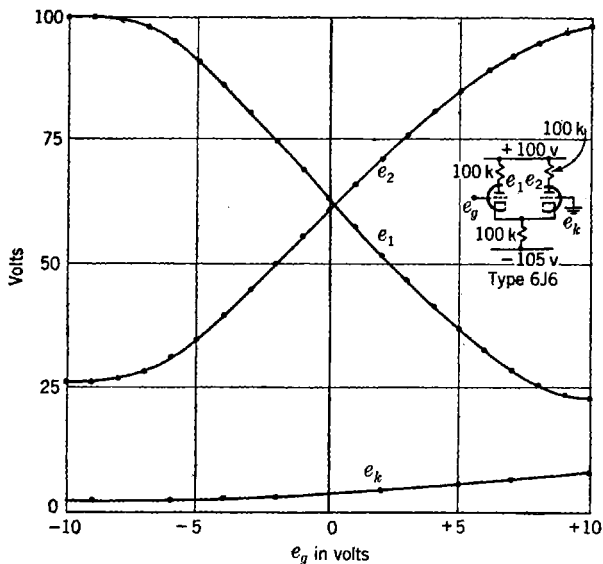

FIG. 11·29.—Example of differential amplifier and characteristic

any variation of E_{pp} appears in the output voltage multiplied by

$$\frac{r_p}{(R_p + r_p)}.$$

Inequalities between the two triodes will result in Eq. (55) being in error by a constant amount.   Often this error is immaterial, as the zero adjustment can be located elsewhere in the circuit.   If zero output voltage at zero input voltage difference is required, a potentiometer between the cathodes, as in Fig. 11·27, usually is satisfactory, although it decreases the gain to some extent.   It is sometimes preferable to employ a potentiometer in the plate circuit to increase one R_p and decrease the other.   In either case, the unbalance required by the zero adjustment tends to spoil the rejection of the common mode and the push-pull

nature of the output voltage; if these are important, a pair of triodes like
the 6SU7, which requires the minimum of zero adjustment, should be
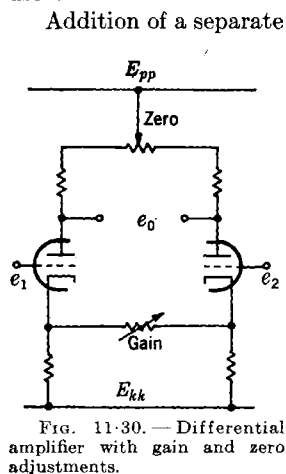used.

Addition of a separate resistance in series with each cathode decreases

FIG. 11·30. — Differential amplifier with gain and zero adjustments.

the gain as given by Eq. (52) by the addition
to the denominator of $\mu + 1$ times this
resistance.   This resistance may be used as a
gain adjustment where desired.   (The in-
crease of linearity accompanying the decrease
of gain is an advantage over simple attenu-
ation.)   To avoid the use of a double poten-
tiometer, the π-connected equivalent circuit
as shown in Fig. 11·30 may be used instead
of the T-connected cathode circuit.   The
gain potentiometer is generally of low resist-
ance as compared with the cathode-return
resistance, and each of these is about twice
the value that R_k would have in the T-circuit.
There is a minimum of interference between
the gain and zero adjustments as shown.

Pentode Differential Amplifier.—Equation (55) may be written in the
form

$$e_{p2} - e_{p1} = \frac{g_m R_p}{1 + \dfrac{R_p}{r_p}} (e_1 - e_2). \tag{56}$$

If pentodes are employed, the plate resistance may be assumed infinite,
and

$$e_{p2} - e_{p1} = g_m R_p (e_1 - e_2). \tag{57}$$

Thus a somewhat higher voltage
gain is obtainable with pentodes
than with triodes.   The special
problem in the case of the pentode
differential amplifier concerns the
screen potential and current.

If a battery or floating supply
can be used to furnish cathode-
screen potential, good common-
mode rejection may be realized.

FIG. 11·31.—Pentode differential amplifier.

Also, since the screen current does not flow in the common cathode
resistor, the output voltage is balanced as in the triode amplifier.

In general, however, the screen grids are supplied from a fixed source
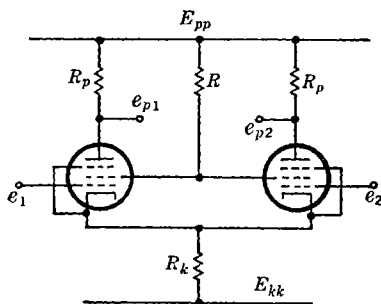or by means of a common resistor as in Fig. 11·31.   The current in R_k

is then the sum of the plate and screen currents. If the total screen current varies, the push-pull aspect of the output voltage will suffer. This effect is likely to be noticeable if either plate potential falls much below the screen potential. Therefore, R_s should be chosen carefully, with this in mind.

11·11. Output Circuits. *Requirements.*—The type of output signal required of a direct-coupled amplifier varies greatly. Therefore, each application must be considered separately in choosing the appropriate output stage. Loads may be classified according to any of the following characteristics; the impedance, voltage, and current range required; whether or not both positive and negative voltage or current are expected; the degree of linearity desired; and the number of terminals needed. An example of a single-terminal load is an amplifier in which a voltage is to be fed back in a d-c computing circuit, wherein the amplifier is a voltage-equating device. A two-terminal load might be the deflecting plates of an oscilloscope, a voltmeter or ammeter, a relay, a motor field, or a magnetic oscillograph. Some loads, such as magnetic oscilloscopes, differential relays, magnetic amplifiers, and servo-motor fields, have three or four terminals for differential excitation.

If there are no preceding stages in the amplifier, the design must take account of the input voltage level and provide an adjustment for variations of components in accordance with the principles of preceding sections. Generally, however, there will be one or more stages of voltage amplification ahead of the output stage, and it will be assumed that the zero adjustment has been taken care of there. The actual output voltage from the preceding stage will be considerably above ground and usually must be dropped down to fit the input voltage level of the output state by a divider that also attenuates.

FIG. 11·32.—Plate as output terminal.

Therefore, a high input voltage level is usually preferable.

Simple Amplifier with Plate as Output Terminal.—If the load is of very high resistance or if the required voltage swing is small, the load may be connected between some fixed point and the plate of the output tube, as in Fig. 11·32. In computing the performance of the amplifier with the load, Thévenin's theorem may be applied, using the unloaded characteristics and the output impedance. Another method involves considering R_L in parallel with R_p as the net plate resistor. In this case a fictitious E_{pp}, which is equal to the value of voltage assumed by e_p with the tube removed, should be employed to determine the output
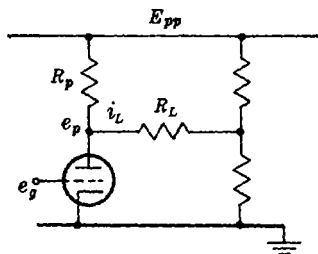
characteristic.   From Eq. (5), with R_p paralleled with R_L, the voltage
gain is

$$\mathcal{S} = \frac{\mu}{1 + \dfrac{r_p}{R_p} + \dfrac{r_p}{R_L}}. \tag{58}$$

For a low-resistance load, such as a meter or a relay, the voltage gain is
of much less interest than the rate of change of load current with input
voltage, which is called the "current gain."   This quantity is equal to
the voltage gain divided by R_L.

$$\frac{\Delta i_L}{\Delta e_g} = -\frac{g_m}{1 + \dfrac{R_L}{r_p} + \dfrac{R_L}{R_p}}. \tag{59}$$

Often the load resistance may be neglected, in which case

$$\frac{\Delta i_L}{\Delta e_g} \approx -g_m. \tag{60}$$

If the output voltage is to be "single-ended" with a large working
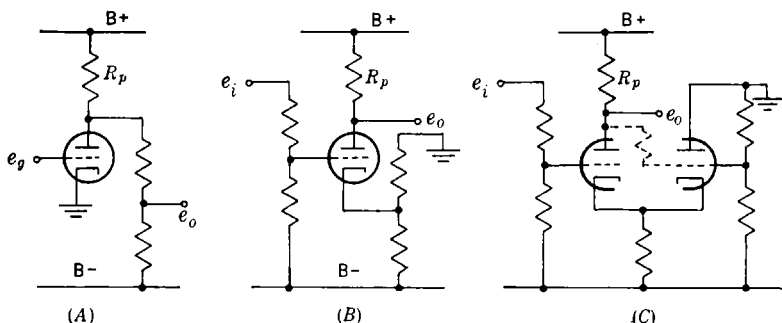range encompassing ground potential, a negative voltage supply will

Fig. 11·33.—Voltage output embracing ground potential.

be needed.   Two general methods of obtaining this voltage are shown in
Fig. 11·33.   In (a) the output voltage is obtained from a divider, and in
(b) the input voltage and the cathode supply are taken from dividers.
If it is assumed that the tube permits a certain minimum plate voltage
without having appreciable grid current and that a certain $B-$ voltage
is available, it is found that the output voltage is somewhat less attenu-
ated in (a) than is the input voltage in (b); also, in (b) the impedance of
the cathode tie point decreases the gain.   On the other hand, the output
impedance in (a) is much higher than in (b), and the upper limit of the
output range is lower.   For these reasons it is sometimes advisable to
employ the depressed cathode method, with a cathode-follower tie point

as in Fig. 11·33c. In addition to providing a low-impedance cathode return, this method permits the application of positive feedback (dashed resistor), which is beneficial if used in conjunction with over-all negative feedback (see Sec. 11·13). In any case a high-resistance plate resistor should be employed because the greater the plate current the higher the minimum allowable plate voltages. A pentode tube is much superior to a high-μ triode in point of low limiting plate voltage and is therefore generally preferable in this application.

Cathode-follower Output Stage.—If the output voltage requirement is the same as in the foregoing paragraph but the current load is appreciable, a cathode follower may be added to any of the circuits of Fig. 11·33. Its cathode resistor should be returned to the $B-$ voltage to permit the output voltage to reach zero without much nonlinearity. Its input voltage will have to drop somewhat below zero, but this amount is minimized by the use of a high-μ tube and by a plate supply that is no higher than needed in view of the maximum output voltage required.
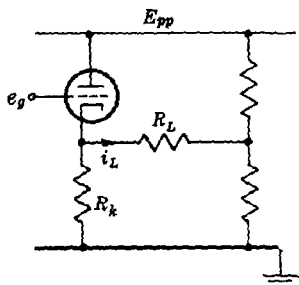

FIG. 11·34.—Cathode-follower output stage.

If one terminal of the load is or can be at some fixed positive potential, the arrangement might be as shown in Fig. 11·34. By Thévenin's theorem, from Eqs. (15) and (16), the voltage "gain" with the load is

$$\mathcal{G} = \frac{\mu}{\mu + 1 + \dfrac{r_p}{R_k} + \dfrac{r_p}{R_L}} \tag{61}$$

(R_L includes the resistance of the load tie point). The current gain is this expression divided by R_L.

$$\frac{\Delta i_L}{\Delta e_g} = \frac{g_m}{1 + (\mu + 1)\dfrac{R_L}{r_p} + \dfrac{R_L}{R_k}}, \tag{62}$$

which, if R_L is negligible, approaches

$$\frac{\Delta i_L}{\Delta e_g} \approx g_m, \tag{63}$$

just as in the case of Fig. 11·32. In fact, if R_L is negligible, the only difference between Figs. 11·32 and 11·34 is the nature of the saturation, and even here the difference is one of sense only. In Fig. 11·34 the maximum reverse current is the current obtained with the tube removed and depends on the magnitude of R_k; the maximum forward load current

depends only on tube capability and plate voltage. A similar, but reverse, situation obtains for Fig. 11·32. Where R_L is appreciable, the current equations (59) and (62) differ by μR_L in the denominator. In this case, the gain of Fig. 11·34 is lower, but its inherent negative feedback gives it greater linearity.

Differential Cathode Follower.—The resistance of the return point for the load resistor may decrease the current gain considerably if a bleeder is used as in Fig. 11·34. It may be advisable to employ a second cathode follower as a tie point as in Fig. 11·35. Similarly to Eq. (62), the current gain is found to be

$$\frac{\Delta i_L}{\Delta e_g} = \frac{g_m}{2 + (\mu + 1)\dfrac{R_L}{r_p} + \dfrac{R_L}{R_k}}, \tag{64}$$

or, for very small load resistance,

$$\frac{\Delta i_L}{\Delta e_g} \approx \frac{g_m}{2}. \tag{65}$$

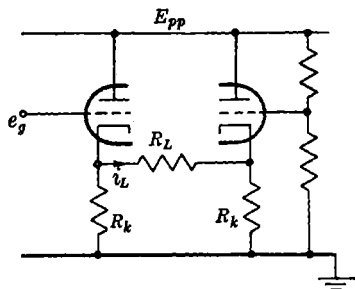Saturation in either direction occurs when one of the tubes cuts off.

Fig. 11·35.—Differential cathode follower.

The load current is then simply the current in R_k. In the forward direction there is still a slight current increase beyond this point, and Eq. (62) then applies with R_L increased by R_k.

This circuit is convenient for driving ammeters and magnetic oscillographs, which have negligible resistance and which need the protection against overloads afforded by the saturation. Good linearity is afforded by mutual cancellation of plate resistance variation. Extra linearity may be obtained at the expense of gain, simply by the addition of resistance in series with the load.

The input terminals may conveniently be at both grids, instead of at only one as shown, from the two plates of a differential amplifier.

Differential Amplifier.—The case of a high-impedance push-pull load, such as oscilloscope plates, is covered by the differential amplifier analysis of Sec. 11·10. The voltage range available depends on the $B+$ voltage supplied, and the push-pull nature of the output voltage, if the input voltage is single-ended, depends on the $B-$ voltage being large in comparison with the input voltage.

To describe the performance of the differential amplifier with a load, as in Fig. 11·36, it is perhaps easiest to begin with the assumption of zero load resistance. In this case, since the cathodes are tied together and the plate voltages are equal, the difference between the plate currents is

$$i_{p1} - i_{p2} = g_m(e_1 - e_2). \tag{66}$$

The currents in the two R_p's are $i_{p1} + i_L$ and $i_{p2} - i_L$ respectively. Since the R_p's are equal and receive the same voltage, these currents are equal, and Eq. (66) becomes (for $R_L = 0$)

$$i_L = \frac{g_m}{2}(e_2 - e_1). \tag{67}$$

Fig. 11·36.—Differential amplifier with load.

The output impedance of the amplifier is the ratio of open-circuit voltage to short-circuit current and thus is found, from Eqs. (67) and (55), to be

$$Z_0 = \frac{2r_p R_p}{R_p + r_p}, \tag{68}$$

or twice the output impedance of a simple amplifier. Application of Thévenin's theorem to Eqs. (55) and (68) gives the voltage gain with load as

$$\mathcal{G} = \frac{-\mu}{1 + \dfrac{r_p}{R_p} + \dfrac{2r_p}{R_L}}, \tag{69}$$

from which the current gain is found to be

$$\frac{\Delta i_L}{\Delta(e_1 - e_2)} = -\frac{g_m}{2 + \dfrac{R_L}{r_p} + \dfrac{R_L}{R_p}}. \tag{70}$$

It is evident that when, as in an ammeter, the load resistance is negligible, the gain of a differential amplifier is identical with that of a differential cathode follower. The linearity of the two types is the same, and the linearity of the differential amplifier may be improved at the expense of gain by insertion of resistance between the cathodes, as in Fig. 11·30. A possible reason for preferring the cathode follower would be the extra damping afforded a meter by the low output impedance. For loads with considerable resistance, like a motor field, a relay, or a solenoid, the differential amplifier is preferable when maximum gain is
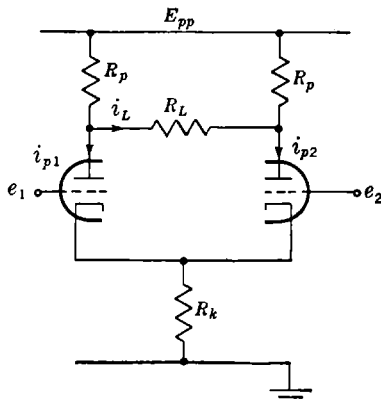
desired. Also, when a quick current response in spite of load inductance is required, it is preferable to use the plate circuit. In this case, the current through the load tends to follow the grid voltage without being influenced by the voltage across the load as much as it would be if the load were in the cathode circuit.

If a pentode differential amplifier, where μ and r_p are almost infinite, is used as an output circuit, Eq. (70) reduces to

$$\frac{\Delta i_L}{\Delta(e_1 - e_2)} = -\frac{g_m}{2 + \dfrac{R_L}{R_p}}. \tag{71}$$

Twice as much current gain can be realized if the load is divided into two parts, as in the case of a differential relay or a magnetic oscilloscope, where each of the two parts is simply one of the R_p's and the output is the difference between these currents. For a pair of triodes, this output current is [adding R_p to r_p in Eq. (61)]

$$i_1 - i_2 = \frac{\mu}{r_p + R_p} (e_1 - e_2). \tag{72}$$

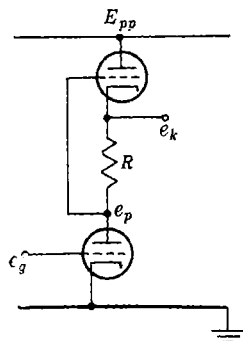In the case of a pair of pentodes, the load resistance does not decrease the gain, and Eq. (66) applies directly.

Two-tube Series Arrangement.—The circuit of Fig. 11·37, where e_k is the output voltage, is sometimes useful as a power amplifier. A similar circuit, having another resistor equal to R in series with the lower cathode and with the output terminal at e_p, affords cancellation of heater-voltage variation and is described in Sec. 11·12. The circuit between E_{pp} and e_p, comprising the plate load for the lower tube, resembles the constant-current circuit of Fig. 11·18, but without the battery it is not actually a constant-current device. It is merely the equivalent of a simple resistance, of value $r_p + (\mu + 1)R$, returned to E_{pp}. Thus in Fig. 11·37 the voltage gain to the e_p terminal is

$$\frac{\Delta e_p}{\Delta e_g} = \frac{-\mu[r_p + (\mu + 1)R]}{2r_p + (\mu + 1)R}. \tag{73}$$

Since all the current flows through R, the cathode is at a point at a distance R from the e_p end of the total equivalent resistance $r_p + (\mu + 1)R$, and therefore e_k moves $\dfrac{r_p + (\mu + 1)R - R}{r_p + (\mu + 1)R}$ times as much as e_p. Thus,

the gain to the cathode (with no load) is

$$\frac{\Delta e_k}{\Delta e_g} \approx \frac{-\mu(r_p + \mu R)}{2r_p + (\mu + 1)R}. \tag{74}$$

To find the gain with a load, it is easiest first to determine the current gain in the case of a zero-resistance load. If R_L is zero in Fig. 11·38, so that e_k is fixed at E, a change Δe_g will produce a plate current increment Δi_{p1} of $\mu \Delta e_g/(r_p + R)$. This increment, in turn, lowers the upper grid by a voltage increment R times this, so that the upper tube current changes by an amount Δi_{p2}, equal to $-\mu^2 R \Delta e_g/r_p(r_p + R)$. Thus, if $R_L = 0$, the net current gain is

$$\frac{\Delta i_L}{\Delta e_g} = -\frac{\mu}{r_p + R} - \frac{\mu^2 R}{r_p(r_p + R)}$$
$$= -g_m \frac{r_p + \mu R}{r_p + R}. \tag{75}$$

The current gain with a load of negligible resistance, where a suitable intermediate voltage source exists for a load tie point, is considerably greater than that for a simple amplifier or a differential amplifier or cathode follower.

Fig. 11·38.—Series amplifier with load.

The output impedance is the ratio of open-circuit voltage gain to short-circuit current gain.

$$Z_0 = r_p \frac{r_p + R}{2r_p + (\mu + 1)R}. \tag{76}$$

From Eq. (74), by means of Thévenin's theorem, the voltage gain with any load resistance is found to be

$$\mathcal{G} = \frac{-\mu(r_p + \mu R)}{2r_p + (\mu + 1)R + (r_p + R)\dfrac{r_p}{R_L}}. \tag{77}$$

The current gain with any R_L is

$$\frac{\Delta i_L}{\Delta e_g} = -g_m \frac{r_p + \mu R}{r_p + R\left[1 + (\mu + 1)\dfrac{R_L}{r_p}\right] + 2R_L}. \tag{78}$$

It is apparent from Eq. (78) that as long as R_L is considerably smaller than r_p, R may be chosen so that the current gain is several times g_m. The maximum range of output current in both positive and negative
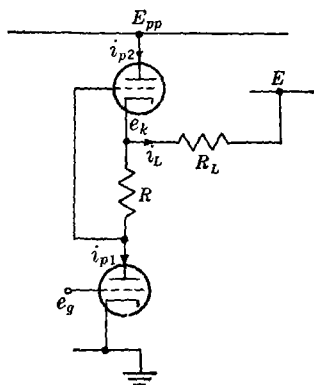
directions is achieved with $E = E_{pp}/2$ and $R = 1/g_m$, but the maximum gain occurs with an R of several times $1/g_m$.
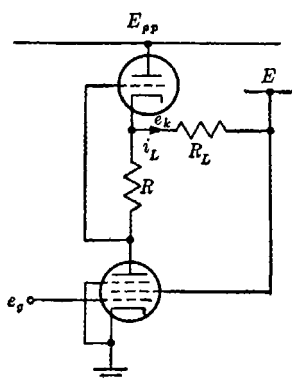
If a pentode is used in the lower position, as in Fig. 11·39, the formulas are simpler. In this case, Eq. (74), the voltage gain with $R_L = \infty$, becomes

$$\frac{\Delta e_k}{\Delta e_g} = -g_m(r_p + \mu R), \qquad (79)$$

where g_m refers to the pentode and μ and r_p to the triode. The output impedance is simply

$$Z_0 = r_p, \qquad (80)$$

because the pentode current is independent of its plate voltage, so that if e_k is moved by external means, the triode bias will remain constant and its current will vary according to plate resistance. From Eqs. (79) and (80) the voltage gain with load is found to be

$$\mathfrak{G} = -g_m \frac{r_p + \mu R}{1 + \dfrac{r_p}{R_L}}, \qquad (81)$$

and the current gain is

$$\frac{\Delta i_L}{\Delta e_g} = -g_m \frac{r_p + \mu R}{r_p + R_L}. \qquad (82)$$

A practical example of this output circuit is given in Fig. 11·42. Both tubes are pentodes, but the upper tube behaves like a triode because its plate and screen both are fixed.

A comparison of this circuit with the differential amplifier shows that for tubes of the same capabilities, the former has at least four times the gain and twice the maximum output current in both directions as the latter. On the other hand, this circuit requires a low-impedance intermediate voltage source, and, for a given available $B+$ voltage, the input voltage level must be considerably lower than that for a differential amplifier.

11·12. Cancellation of Effect of Heater-voltage Variation.—The fundamental effect of heater-voltage variation was explained in Sec. 11·6: A definite change of heater voltage is the equivalent of a definite change of the cathode potential relative to the other electrode potentials. For oxide-coated cathodes, a 10 per cent increase of heater voltage is the same as a cathode-potential decrease of about 100 mv, although this

FIG. 11·39.—Series amplifier using pentode.

figure is subject to some variation from tube to tube and is larger and more erratic if the cathode current is very great. Thus the effect can be canceled either by an equal displacement in the opposite direction of the cathode potential or by approximately the same grid potential displacement in the same direction. If such a cancellation is not applied, the output of the amplifier will shift by its gain times this equivalent cathode-potential decrease. If more than one stage is involved, each will contribute to the heater effect by an amount that depends on the gain of the following stages, but, in general, the gain of the first stage will be great enough so that the heater effect in following stages is negligible in comparison.

Negative feedback cannot reduce the amount of adjustment required at the input to cancel this effect. This fact is obvious in the case of a cathode follower. The effect of an increase of E_f is the same as the insertion of a low-voltage battery of zero resistance in series with the cathode, as shown in Fig. 11·40. If this were done, with no change in e_g, e_k would change by an amount almost equal to the battery voltage ΔE. To cancel this change, e_g must be lowered by just ΔE.

FIG. 11·40.—Heater-voltage variation effect on a cathode follower.

In a majority of applications, some method of canceling the effect of variations of heater voltage is necessary or advisable. (Of course, if the heater supply is well regulated, or if the resulting error is within allowable limits in a specific application, no cancellation will be needed.)

Diode Cancellation.—A comparison of Figs. 11·5 and 11·6 shows that the slopes of these curves for the diode are substantially the same as those for the triode. (At extremely low currents the triode curves become a little steeper, because incomplete grid control over the electrons is more noticeable with increasing temperature.) Thus the variation in a diode, whose heater is connected to the same source as that of the triode, may be used to offset the variation in the amplifying tubes.

One method of diode cancellation is shown by Fig. 11·41. The negative B voltage is large compared with the expected fluctuation at the cathode, so that the current in R may be considered constant (R is required to be large compared with the diode's variational resistance). It is desired that, with no change in e_g, i_p (and thus e_p) will remain constant. To permit this, a certain increase of E_f must be compensated for by a certain rise Δe_k at the cathode of the amplifier. But if the diode has the same cathode characteristic, this same increase Δe_k will occur at the diode, since the diode current is constant. If the two cathodes
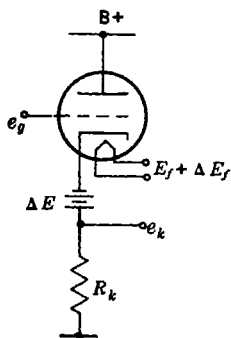
have different heater-voltage characteristics, the compensation will suffer accordingly. But generally the error will be less than one-tenth of that with no correction; e.g., the error due to a 10 per cent change of E_f, measured at e_g, will be less than 10 mv.

The error due to the fact that R is finite is not noticeable in comparison with that due to tube differences unless the voltage across R is less than about 10 volts. It can be shown that this error is only $r_d/(R + r_d)$ of the error with no compensation, where r_d is the diode variational resistance.

The value of R must be such that its current will be greater than the maximum in the amplifying tube; otherwise the diode will cut off, and the amplifier will cease to function as such. The cathode return of the amplifier is through the variational diode resistance r_d (in parallel with
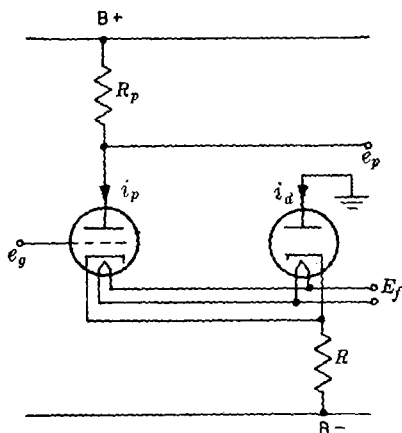

FIG. 11·41.—Balancing E_f variator by means of a diode.

R, which should be large enough to be negligible in comparison). A loss of gain [Eq. (9)] results unless μr_d [or $(\partial e_{g2}/\partial e_g)r_d$ if a pentode is used] is very small compared with R_p. It must be remembered, in selecting R, that r_d is a function of the diode current and may be determined from the slope of the diode current-voltage curve. At very low currents (e.g., below 0.2 or 0.3 ma, for a 6AL5) r_d becomes inversely proportional to the current, with a multiplier that is not a function of the diode type or even of the number of diodes in parallel (since increasing the number would decrease the current and increase the variational resistance of each). Thus, for all diodes with unipotential oxide-coated cathodes, the variational resistance at very low current approximates

$$r_d \approx \frac{0.09}{i_d} \qquad \text{kilohms} \tag{83}$$

if i_d is in milliamperes.[1] This variation of the diode impedance might conceivably be used to linearize a portion of the amplifier output curve by varying the gain in a direction opposite to its natural curvature.

[1] This equation derives from the fact that the current at small values is exponential with respect to voltage: $i_d \approx i_0 e^{11,600 e_{pk}/T}$, where T is the absolute temperature of the cathode and is 1000° to 1100°K in this instance.

Tubes with diodes using the same cathode as the triode or pentode are especially useful in the circuit shown in Fig. 11·41, since the heater-voltage characteristics are bound to be nearly identical. For example, a sample of six 6SQ7 double-diode triodes was tested in the circuit of
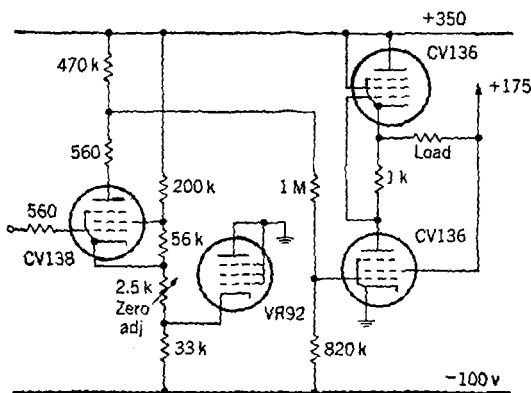

FIG. 11·42.—D-c servo amplifier.

Fig. 11·41, with the diode plates in parallel ($e_p \approx 150$ volts; $i_p \approx 0.1$ ma; $i_d = 0.08$ ma). The adjustments required at the grid by a 20 per cent heater-voltage change were -10, -5, 0, $+2$, $+3$, and $+6$ mv, whereas the adjustments required with the cathode grounded ranged from 195 to 212 mw. The drift due to aging of the heater or cathode may also be fairly well canceled in this way.

Figure 11·42 is the circuit diagram of a direct-coupled velocity servo amplifier,[1,2] which furnishes field excitation for a d-c servo motor. The input stage of this amplifier uses diode cancellation of heater-voltage variation. The zero-adjustment resistance and screen-grid bleeder do

FIG. 11·43.—Diode heater-voltage cancellation at amplifier grid.

not impair this function because the currents here are substantially constant. The diode current is two or three times the cathode current or the pentode.

A method that requires two diodes but needs no negative supply and permits manual adjustment for differences in cathode characteristics is illustrated by Fig. 11·43. The circulating current due to initial electron
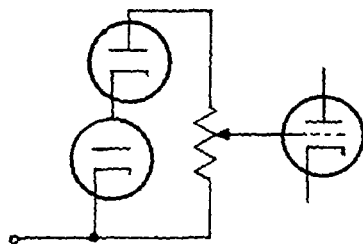
[1] Telecommunications Research Establishment Report.
[2] Cf also Sec. 11·14 for further discussion of this circuit.

velocity (Edison effect) in the diodes inserts a negative voltage in series with the grid. This voltage increases with increasing heater voltage, and its slope factor may be adjusted with the potentiometer. The resistance employed may be anywhere from 10,000 ohms to 1 megohm.

Fig. 11·44.—Balancing E_p variation with a cathode follower.

A single diode is not potent enough because of the load imposed by the resistance.

Cathode Follower and Differential Amplifier Cancellation.—In Fig. 11·44 the diode of Fig. 11·41 is replaced by a cathode follower. The same reasoning that explained the action in the case of the diode circuit is equally appropriate here, as are the remarks concerning the assignment of the value of R_k. For the cancellation to be as effective as possi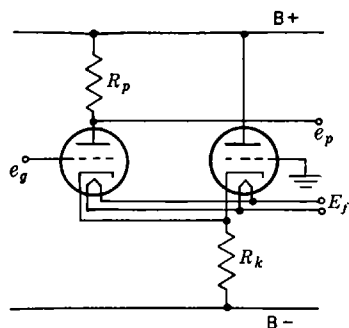ble, the $B-$ voltage must be large enough so that R_k may be very large compared with the reciprocal transconductance of the cathode follower. In the comparison with the diode circuit, this latter quantity corresponds to the diode variational resistance.

The grid of the cathode follower provides a convenient high-impedance point for zero adjustment for the amplifier. This and other features of the circuit have been discussed under differential amplifiers. If the load resistor is transferred to the right-hand triode in Fig. 11·44, to obtain an output of the same sense as the input, or if the circuit is made into a symmetrical differential amplifier, the cancellation of heater-voltage variation still obtains in the same way. Double triodes, such as the 6SL7 or 6SU7, are convenient for this type of application and give reasonable assurance of similar cathode characteristics. (The 6J6 should be very appropriate, since the same cathode is used for both triodes; but it seems subject to considerable drift because of its type of construction.)

Fig. 11·45.—Balancing E_f variation with a self-biased cathode follower.

The circuit of Fig. 11·45, which employs a self-biasing cathode follower as a cathode return for the amplifier tube, permits adjustment for the inequality that may exist between the cathode characteristics of the two tubes. If i_{p1} is to be held constant in spite of any given change
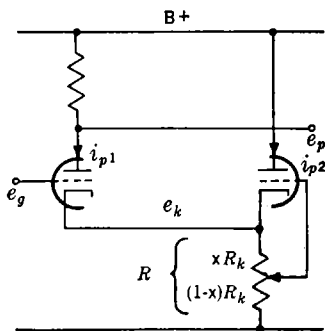
of heater voltage, e_k must be changed by some amount ΔE_1. Since i_{p1} is constant, the resulting change of current in R_k all occurs in the cathode-follower tube.

$$\Delta i_{p2} = \frac{\Delta E_1}{R_k}. \tag{84}$$

If x is the portion of R_k between the cathode and the grid tap, the change of grid bias is

$$\Delta e_{gk} = -x \, \Delta E_1 \tag{85}$$

But this tube behaves as if a small voltage $-\Delta E_2$ (which is approximately equal to $-\Delta E_1$) had been inserted in series with its cathode (as in Fig. 11·40). Thus the effective grid-to-cathode voltage has changed by the amount $\Delta E_2 - x \, \Delta E_1$, and effective plate-cathode voltage has changed by $\Delta E_2 - \Delta E_1$. So from Eq. (1),

$$\begin{aligned}
\Delta i_{p2} &= \frac{\Delta E_2 - \Delta E_1 + \mu_2(\Delta E_2 - x \, \Delta E_1)}{r_{p2}} \\
&= \frac{(\mu_2 + 1) \, \Delta E_2 - (x\mu_2 + 1) \, \Delta E_1}{r_{p2}}.
\end{aligned} \tag{86}$$

Combining (86) and (84),

$$R_k = \frac{r_{p2}}{\mu_2\left(\dfrac{\Delta E_2}{\Delta E_1} - x\right) + \dfrac{\Delta E_2}{\Delta E_1} - 1}. \tag{87}$$

Or, approximately,

$$\left(\frac{\Delta E_2}{\Delta E_1} - x\right) R_k \approx \frac{r_{p2}}{\mu_2}. \tag{88}$$

If the two cathodes have exactly equal characteristics, so that $\Delta E_1 = \Delta E_2$, the portion of R_k below the grid tap is simply

$$(1 - x)R_k = \frac{r_{p2}}{\mu_2}, \tag{89}$$

where r_{p2} is the plate resistance of the cathode-follower tube, and if there is a resistor inserted in series with this plate, it is included in r_{p2}.

The amplifier tube cathode is, in effect, connected to a voltage $E_{pp}/(\mu + 1)$ through a resistance $r_{p2}/(\mu + 1)$, regardless of the value of the portion of R_k above the grid tap (although this value does determine the cathode-follower plate current, which in turn affects r_{p2}). Thus the amplifier tube will be just about cut off at zero grid voltage, and it will have a gain [Eq. (9)] of $\mu R_p/(R_p + 2r_p)$. The value of R_k is not critical but should be between two and eight times r_p/μ. The triodes should be aged for a while and cycled several times through the extremes of heater voltage before the adjustment is made.

Cancellation by a Series Triode.—In the circuit of Fig. 11·46, the plate resistor for the amplifier (lower) tube comprises a circuit resembling the constant-current device of Fig. 11·18 except that the voltage E is omitted. Thus this is not a constant-current device but is simply the equivalent of a resistor whose upper end is attached to E_{pp} and whose value is $r_p + (\mu + 1)R$ [Eq. (23)]. Thus from Eq. (7), if the two triodes are similar,
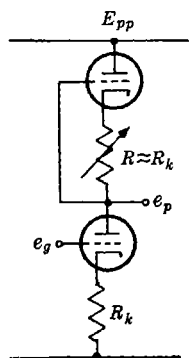

FIG. 11·46.—Cancellation of E_f variation with a series triode.

$$e_p = \frac{\dfrac{r_p + (\mu + 1)R_k}{r_p + (\mu + 1)R} E_{pp} - \mu e_g}{1 + \dfrac{r_p + (\mu + 1)R_k}{r_p + (\mu + 1)R}}. \tag{90}$$

If the two cathodes respond equally to a change of heater voltage, R should equal R_k for cancellation of the effect.[1] With e_g fixed, a given increase of E_h causes a certain small increase of current; but since this increase is the same in both tubes, the grid biases change identically. Therefore, the two plate-to-cathode voltages suffer no change, and e_p remains constant. If $R_k = R$, Eq. (90) becomes

$$e_p = \frac{E_{pp}}{2} - \frac{\mu e_g}{2}. \tag{91}$$

Thus, when e_g is zero, the output voltage is half the plate-supply potential (as is already apparent by symmetry), and the gain is $\mu/2$. The output voltage is linear with respect to e_g because r_p does not appear in Eq. (91). Of course, the two r_p's were assumed equal, but this assumption is not far in error, as the currents in the two tubes are equal.

The output impedance, from Eq. (11), is (if $R_k = R$)

$$Z_p = \frac{r_p + (\mu + 1)R_k}{2}. \tag{92}$$

If the two cathodes have different heater-voltage characteristics, either R or R_k may be adjusted until cancellation is obtained. This adjustment will not usually change the gain from $\mu/2$ by more than 5 per cent.

Cancellation by Means of a D-c Potential Proportional to Heater Voltage. If the d-c load on the power supply is fairly constant, its unregulated output will vary in proportion with the a-c line voltage and therefore with E_h. Thus this output may be employed in some way to offset the

[1] Maurice Artzt, "Survey of D-c Amplifiers," *Electronics*, August, 1945.

effect on the cathode of the variations of E_h. For example, in Fig. 11·47, if E_h increases, e_k will rise a small amount, which, if the adjustment is right, will keep the plate current constant with no change of e_g. If the cathode is on the bleeder at a point 1 volt from ground, then regard-

less of the total voltage on the bleeder, e_k will rise 0.1 volt with a 10 per cent rise in the unregulated source. This rise is just about the required amount. Another resistor, inserted in series with the cathode for degeneration or zero adjustment, will not affect the cancellation.

If the source is not well filtered, filtering may be done at the cathode. But a condenser of a given capacity is more effective if placed as shown, about midway on the bleeder.

Rapid line fluctuations will produce errors whose nature and duration is revealed by Fig. 11·8. Another disadvantage of this method is that the correction is linear with respect to E_h whereas the effect itself is curved (Fig.

FIG. 11·47.—Variation of e_k with e_p.

11·6); therefore, the cancellation is of limited range.

Cancellation by Special Connections of Multigrid Tubes.—The circuit of Fig. 11·48, wherein the second grid of a tetrode or pentode is used as the control grid for the plate current, is actually very much like the circuit of Fig. 11·41. Here the first grid instead of a separate diode

FIG. 11·48.—First grid as diode plate.

plate is grounded (or otherwise fixed) and acts as a diode plate to furnish a large part of the cathode current. If the cathode temperature is increased, its potential will rise; and since i_g is practically constant (e_p is assumed to be held constant, and the variation of e_k is small compared with the drop across R_k), its rise will be just enough to neutralize the effect on the other electrodes in the tube. The advantage over the diode-cancellation method of Fig. 11·41 is that the same cathode area is used in the cancellation as in the amplification. The drift that is caused by change of average initial electron velocity with time is also canceled in the same way as that due to temperature change.

Ordinary pentodes are generally unsatisfactory in this circuit because the screen grid has so much control over the plate current that the latter is completely cut off (except of extremely high plate voltage) when the screen-grid potential is lowered far enough so that its own current is
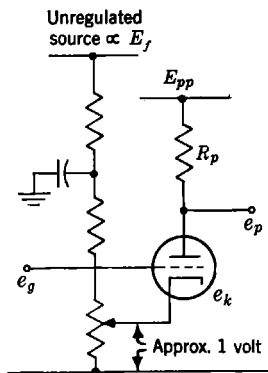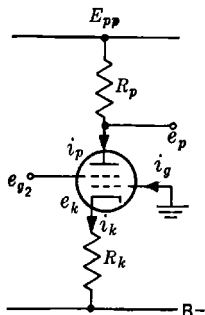
negligible. The suppressor grid, on the other hand, has very little control. In a tetrode like the 6V6, however, the effect of the screen grid is between these two extremes. Figure 11·49 shows the plate characteristics of a 6V6 in the circuit of Fig. 11·48, with a negative supply of 45 volts, and $R_k = 90,000$ ohms, so that i_k is just about $\frac{1}{2}$ ma. The curves are similar to those of an ordinary triode with a μ of about 40,
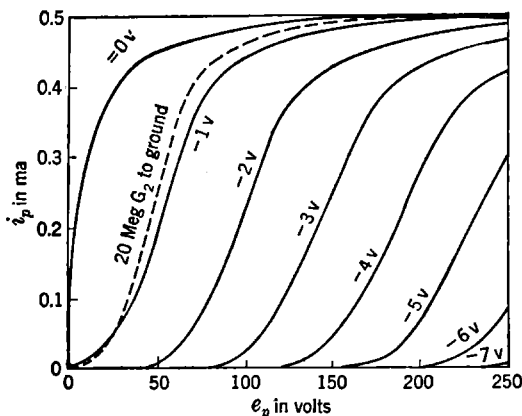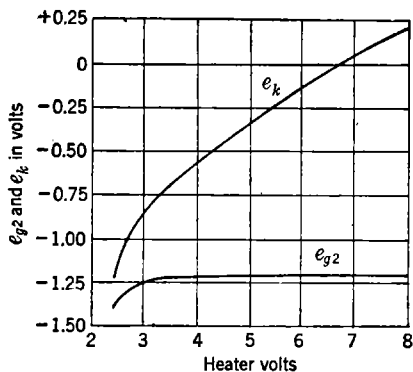

Fig. 11·49.—Plate characteristics of a 6V6 triode.

Fig. 11·50.—Effect of heater voltage on e_{g2} and e_k.

except that the current has an upper limit of $\frac{1}{2}$ ma. Figure 11·50 shows the effect of heater voltage on e_k and e_{g2}, the latter being adjusted so as to maintain constant plate current. When i_p and i_g are about the same, e_{g2} is fairly constant over a large temperature range.

The 6AS6 miniature pentode has a specially designed suppressor grid that has unusually effective control over the plate current. This tube may be used in a manner similar to the one in the circuit of Fig. 11·48,

but with the suppressor as the control grid and the screen grid at some positive potential. This method of operation has the advantage of producing plate characteristics more like those of a pentode, but the cancellation of heater-voltage effect seems to be less effective over a wide operating range than in the case of the tetrode. A pentagrid converter tube might also be applicable in the same way as the 6AS6. More research is indicated in this field, with a possible objective of developing a suitable tube for the purpose.

11·13. The Use of Feedback in D-c Amplifiers. *Conductive Negative Feedback.*—Conductive, or direct-coupled, negative feedback can reduce the dependence of gain upon output d-c voltage or current (i.e., nonlinearity or amplitude distortion) and tube characteristics. It cannot reduce the effect of zero drift or displacement of a given amplifier (as referred to the input terminals) no matter whether it is due to heater-voltage variation, tube aging or replacement, or microphonics. This fact was illustrated in Sec. 11·12 for the case of a cathode follower, in which the zero drift referred to the input terminals was found to equal the tube drift in terms of the shift of grid-to-cathode potential required to maintain constant current. In some instances, as in the arrangement of Fig. 11·51, negative feedback can actually increase the effect of drift

FIG. 11·51.—Direct feedback using resistance adding.

as seen at the input terminals. In Fig. 11·51, if the amplifier drifts so that a given change Δe_g is required to hold e_0 constant, the change required at the input terminals is $\Delta e_1 = \Delta e_g(R_1 + R_2)/R_2$.

However, if a given over-all gain is required, negative feedback can reduce the effect on zero displacement of changes in certain parts of an amplifier by permitting the use of more amplification ahead of these parts. This is often of great value, since a power-output stage is much more susceptible to shifts caused by variations of load, supply voltage, heater voltage (because of greater tube currents), etc., than is a voltage amplifier with very low power output.

The common form of the feedback equation is

$$\mathcal{G} = \frac{\alpha}{1 - \alpha\beta}. \tag{93}$$

where α is the gain without feedback, $\mathcal{G}$ is the gain with feedback, and β is the fraction of output voltage that is added to the input voltage. This equation is based on the assumption of simple addition involving no attenuation of the input voltage and no disturbance of the amplifier parameters by the feedback action. In Fig. 11·51 the resistance addition
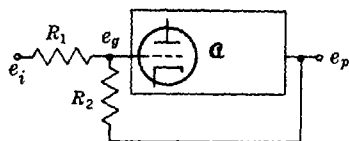
of input voltage to a fraction of the output voltage attenuates the input voltage by the factor $R_2/(R_1 + R_2)$, and Eq. (93) therefore becomes[1]

$$G = \frac{R_2}{R_1 + R_2} \frac{\mathfrak{a}}{1 - \dfrac{\mathfrak{a}R_1}{R_1 + R_2}}. \tag{94}$$

For negative feedback the sense of the amplifier must be such that $\mathfrak{a}$ is negative.

Perhaps the most common form of negative feedback is through the cathode. Some of the aspects of this method, as applied to single-stage voltage amplifiers, were covered in Sec. 11·7. Equation (93) cannot be applied directly because the operation of the amplifier itself is somewhat affected by this type of feedback. From Eq. (9) it is seen that the insertion of a cathode resistor R_k gives the triode amplifier a gain of

$$G = - \frac{\mu}{1 + \dfrac{r_p}{R_p} + (\mu + 1) \dfrac{R_k}{R_p}}, \tag{95}$$

whereas the gain without this degeneration is simply

$$\mathfrak{a} = - \frac{\mu}{1 + \dfrac{r_p}{R_p}}. \tag{96}$$

Combination of these equations to obtain a form like Eq. (93) results in

$$G = \frac{\mathfrak{a}}{1 - \mathfrak{a} \dfrac{\mu + 1}{\mu} \dfrac{R_k}{R_p}}. \tag{97}$$

Thus the feedback factor is, in effect,

$$\beta = \frac{\mu + 1}{\mu} \frac{R_k}{R_p}, \tag{98}$$

---

[1] In the resistance addition of e_0 and e_i to obtain e_g,

$$\Delta e_g = \frac{R_2 \, \Delta e_i + R_1 \, \Delta e_0}{R_1 + R_2}. \tag{i}$$

But the amplifier gain is

$$\mathfrak{a} = \frac{\Delta e_0}{\Delta e_g}, \tag{ii}$$

and over-all gain with the feedback is

$$G = \frac{\Delta e_0}{\Delta e_i}. \tag{iii}$$

Combination of Eqs. (i), (ii), and (iii) gives Eq. (94).

which is rather different from the common form in that it is a function of a characteristic of the triode, although a given percentage change of μ will cause only $1/(\mu + 1)$ times as much percentage change in β.

If β is constant in Eq. (93), differentiation shows that a change $\Delta\mathcal{G}$ of the gain of the amplifier will produce a change with the feedback of

$$\Delta\mathcal{G} = \frac{\Delta\mathcal{G}}{(1 - \mathcal{G}\beta)^2} \tag{99}$$

$$= \frac{\mathcal{G}^2}{\mathcal{G}^2} \Delta\mathcal{G}, \tag{100}$$

or

$$\frac{\Delta\mathcal{G}}{\mathcal{G}} = \frac{\mathcal{G}}{\mathcal{G}} \frac{\Delta\mathcal{G}}{\mathcal{G}}. \tag{101}$$

This equation indicates that the introduction of negative feedback reduces fractional deviations from constancy of gain in the same ratio as it reduces the gain. Of course, if cathode feedback is employed so that Eq. (98) obtains, Eq. (101) is not complete for any change of μ but is complete for a change or r_p. This agrees with the conclusions drawn from Eq. (10) in Sec. 11·7.

Although negative feedback tends to stabilize the gain with respect both to changes of circuit parameters at any given value of output voltage and to changes of the output voltage itself, these two factors may be considered separately to some extent. One or the other of them may be of greater importance in a given application, and this may affect the nature of the feedback employed.

FIG. 11·52.—Relation of nonlinearity to change of gain.

Variation of gain with respect to output voltage results in nonlinearity. The common way of expressing nonlinearity is in terms of maximum deviation of the curve of output voltage plotted against input voltage from the best linear approximation thereto, given as a percentage of the total working range of output voltage. This quantity may be related to the change of gain over the length of the curve if the nature of this change is known. For example, if the gain, which is the slope of the curve, changes uniformly with respect to input, the curve is part of a parabola, as shown in Fig. 11·52. The dashed line is the best linear approximation, and δ is the maximum deviation. Thus the non-linearity is expressed as δ/E_0, where E_0 is the range of output voltage.
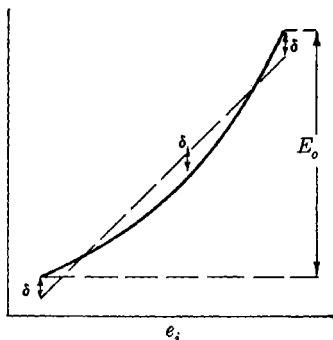
A geometrical calculation reveals that the nonlinearity is

$$\frac{\delta}{E_0} = \frac{1}{16} \frac{\Delta \mathcal{G}}{\mathcal{G}_0}, \tag{102}$$

where $\mathcal{G}_0$ is the slope of the dashed line and $\Delta \mathcal{G}$ is the difference between the slopes at the extremes of the curve. In most cases where the curvature is greater near one end, the fractional nonlinearity is an even smaller fraction of the fractional maximum change of gain. In any event, Eqs. (102) and (101) show that the fractional nonlinearity is decreased by negative feedback in the same ratio as is the gain.

If there is more than one stage in an amplifier, the question may arise as to whether the negative feedback should be over all or over the individual stages. The total amplification is

$$\mathcal{A} = \mathcal{A}_1 \mathcal{A}_2 \mathcal{A}_3, \cdots, \tag{103}$$

where $\mathcal{A}_1$, $\mathcal{A}_2$, etc., are the amplifications of the stages. If each has its own deviation in gain, which is assumed to be small, then differentiation shows that the fractional deviation of $\mathcal{A}$ is simply the algebraic sum of the individual fractional deviations:

$$\frac{\Delta \mathcal{A}}{\mathcal{A}} = \frac{\Delta \mathcal{A}_1}{\mathcal{A}_1} + \frac{\Delta \mathcal{A}_2}{\mathcal{A}_2} + \frac{\Delta \mathcal{A}_3}{\mathcal{A}_3} \cdots. \tag{104}$$

If over-all feedback, which reduces the gain to $\mathcal{G}$, is employed, Eq. (101) gives the deviation of gain as

$$\frac{\Delta \mathcal{G}}{\mathcal{G}} = \frac{\mathcal{G}}{\mathcal{A}} \left(\frac{\Delta \mathcal{A}_1}{\mathcal{A}_1} + \frac{\Delta \mathcal{A}_2}{\mathcal{A}_2} + \frac{\Delta \mathcal{A}_3}{\mathcal{A}_3} + \cdots + \right). \tag{105}$$

On the other hand, if feedback is on an individual basis so that each stage has its gain reduced from $\mathcal{A}_1$ to $\mathcal{G}_1$, etc., then the linearity of each stage is improved according to Eq. (100):

$$\frac{\Delta \mathcal{G}_1}{\mathcal{G}_1} = \frac{\mathcal{G}_1}{\mathcal{A}_1} \frac{\Delta \mathcal{A}_1}{\mathcal{A}_1}, \text{ etc.}$$

Thus, as in Eq. (104), the over-all deviation in this case is

$$\frac{\Delta \mathcal{G}}{\mathcal{G}} = \frac{\mathcal{G}_1}{\mathcal{A}_1} \frac{\Delta \mathcal{A}_1}{\mathcal{A}_1} + \frac{\mathcal{G}_2}{\mathcal{A}_2} \frac{\Delta \mathcal{A}_2}{\mathcal{A}_2} + \frac{\mathcal{G}_3}{\mathcal{A}_3} \frac{\Delta \mathcal{A}_3}{\mathcal{A}_3} + \cdots +. \tag{106}$$

If the over-all gain is to be the same in both cases, i.e., if $\mathcal{G} = \mathcal{G}_1 \mathcal{G}_2 \mathcal{G}_3, \cdots$, then Eq. (106) yields a larger result than Eq. (105), since $\mathcal{G}/\mathcal{A}$ will be much smaller than any of the individual ratios $\mathcal{G}_1/\mathcal{A}_1$, etc. Thus the over-all degeneration is by far the more effective in linearizing and stabilizing the gain.

In many cases, it is more convenient to apply the degeneration to individual stages, especially in view of the simplicity of the cathode-resistor method or the fact that proper phasing of the feedback may not otherwise be possible. The question then arises as to what the apportionment of degeneration should be among the stages. This will depend to some extent on whether linearity or stability of gain with respect to circuit elements is desired. If linearity is of greater importance, most of the degeneration should be placed where there is the greatest range of voltage (or current) to be encountered because the fractional nonlinearity tends to be greatest there. But if it is desirable to have constancy of gain with respect to tube parameters, etc., each stage should be stabilized regardless of the magnitude of its output voltage swing.

FIG. 11·53.—Over-all cathode feedback.

Over-all negative feedback to the cathode of the first stage may be applied as shown in Fig. 11·53, where the rest of the amplifier beyond the first stage is generalized and assigned the gain α'. From Eq. (13) the portion of the output fed back to the cathode is found to be
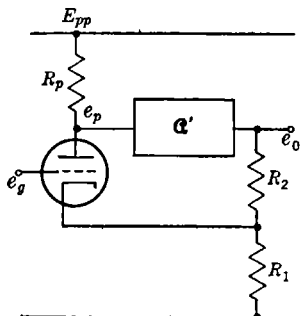
$$k = -\frac{R_1}{R_1 + R_2 + (\mu + 1)\dfrac{R_1 R_2}{r_p + R_p}}$$

This attenuating network also has the effect of inserting in series with the cathode of the first tube a resistance of value

$$R_k = \frac{R_1 R_2}{R_1 + R_2}.$$

As explained in Sec. 11·7 this has the effect of increasing the plate resistance of the triode by $(\mu + 1)R_k$. Thus,

$$\Delta e_p = \frac{(\mu + 1)k\,\Delta e_0 - \mu\,\Delta e_g}{1 + \dfrac{r_p + (\mu + 1)R_k}{R_p}}. \tag{107}$$

But $\Delta e_p = \Delta e_0/\alpha'$, and the over-all gain may therefore be found as

$$\mathcal{G} = \frac{\Delta e_0}{\Delta e_g} = \frac{\mu}{(\mu + 1)k - \dfrac{1}{\alpha'}\left[1 + \dfrac{r_p + (\mu + 1)R_k}{R_p}\right]}. \tag{108}$$

But the gain of the first stage alone (including the effect of R_k) is

$$\alpha'' = -\frac{\mu}{1 + \dfrac{r_p + (\mu + 1)R_k}{R_p}}. \tag{109}$$

Substituting this in Eq. (105),

$$\mathcal{G} = \frac{\mu}{k(\mu + 1) + \dfrac{\mu}{\alpha'\alpha''}}. \tag{110}$$

The total amplification without the feedback (but with R_k) is $\alpha = \alpha'\alpha''$, so that the gain with feedback, in the form of Eq. (93), is

$$\mathcal{G} = \frac{\alpha}{1 + \dfrac{\mu + 1}{\mu}k\alpha}, \tag{111}$$

which is very much like the single-tube case of Eq. (97).   The feedback factor here is

$$\beta = -\frac{\mu + 1}{\mu}k \tag{112}$$

The negative sign shows that for negative feedback, α must be positive, i.e., α' must be negative.

The possible variation of β can cause variation of gain with this type of feedback, no matter how large α is.   Differentiation of Eq. (111) shows that if μ and α are both subject to small changes, the resulting change of gain is

$$\frac{\Delta \mathcal{G}}{\mathcal{G}} = \frac{\mathcal{G}}{\alpha}\frac{\Delta\alpha}{\alpha} + \frac{k\mathcal{G}}{\mu}\frac{\Delta\mu}{\mu}. \tag{113}$$

If α is very large compared with $\mathcal{G}$, Eqs. (111) and (113) become

$$\mathcal{G} \approx \frac{\mu}{\mu + 1}\frac{1}{k} \tag{114}$$

and

$$\frac{\Delta \mathcal{G}}{\mathcal{G}} \approx \frac{1}{\mu + 1}\frac{\Delta\mu}{\mu}. \tag{115}$$

Thus the limiting factor in the degree of linearization and gain stabilization achieved is expressed by Eq. (115).   If a considerable range is covered by the input voltage, the plate-to-cathode voltage will vary, since the cathode voltage follows the input voltage.   Because of the concommitant variation in μ, nonlinearity may result, but it can be kept to a minimum by keeping the plate current in the tube practically constant, i.e., by having α' large enough so that the variation of e_p will be small

compared with the drop across R_p.   Obviously a high-μ tube is preferable in the input position.   As an example of the effect of tube replacement, μ for 6SL7's at a plate current of 0.2 ma varies approximately from 61 to 73 for different tubes.   Therefore, the maximum change of gain to be expected from Eq. (115) would be 0.26 per cent.

If a pentode with its screen at a fixed potential is employed in the input stage, the preceding analyses and formulas still apply, except that μ must be replaced by $\partial e_{sk}/\partial e_{gk}$ and r_p does not appear.   The screen may be kept at a constant potential above the cathode by means of an auxiliary network, as in Fig. 11·54 where $R_4/R_3 = R_2/R_1$.   Thus the expression on the right-hand side of Eq. (115) is very nearly equal to zero.   If the variation in pentode plate current is small, and if the screen potential never swings much above the plate potential, the screen current will be reasonably constant, so that R_3 and R_1 may be rather large.   A simpler arrangement than that of Fig. 11·54 puts the screen on a bleeder from cathode to $B+$.   The operation is then somewhere between the two conditions described.
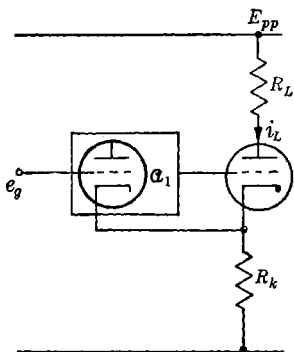

FIG. 11·54.—Cathode and screen feedback.

In all of the previous discussion, the loading effect of the feedback network on the amplifier has been assumed to be included in α.   If the amplifier also delivers power to an external load, which is not constant, α will be caused to vary, and $\mathcal{G}$ will vary in accordance with Eq. (113).

The feedback arrangements of Figs. 11·53 and 11·54 as well as those of Fig. 11·51 are *voltage* negative feedback; they operate to stabilize the output voltage with respect to changes of load; i.e., the effective output impedance is decreased.   In fact it may readily be demonstrated that the output impedance of the amplifier is decreased in the ratio $\mathcal{G}/\alpha$.

This is not the case with the simple cathode-resistor system of Fig. 11·12 where the voltage output terminal is at the plate.   Equation (11) shows that the output impedance is actually increased.   This type of feedback, which operates to stabilize the current in the load resistor by feeding back a voltage developed by this current in a fixed resistor, is termed *current* feedback.   Figure 11·55 shows a more elaborate example of current feedback.   The output is the current in R_L.   As this current also flows in R_k, the fedback voltage is proportional to it.   (The current of the input tube also flows in R_k, but this is usually much smaller and is also nearly constant.)   Figure 11·55 would be the same as Fig. 11·53 if the output voltage were considered to be the voltage at the cathode

of the output tube (and $k = 1$ in Fig. 11·53). Thus the output current may be derived from these considerations by simply dividing the output voltage by R_k. The current in R_L is linearized with respect to e_g and stabilized with respect to changes of R_L, E_{pp}, etc., in a manner identical with that of the output voltage of Fig. 11·53.

The cathode feedback arrangement of Fig. 11·53 leaves the input circuit with no compensation for variation of heater voltage. In the resistance-adding feedback of Fig. 11·51, compensation may be provided at the input stage, but this type of feedback is often less desirable than the other because of the lower input impedance and because of the need for high-resistance precision resistors. Compensation may be provided
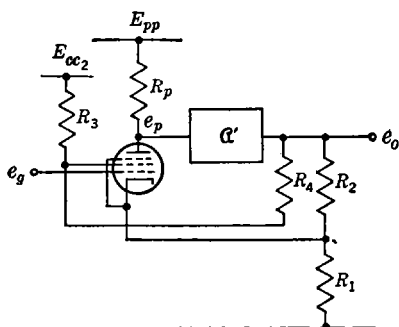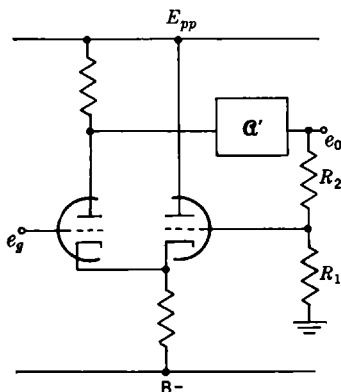

FIG. 11·55.—Example of current feedback.   FIG. 11·56.—Cathode-follower feedback.

in the case of cathode feedback, by feeding back to the plate of a diode arranged as in Sec. 11·12. A more common method employs a cathode follower as in Fig. 11·56, which has the advantage that R_1 and R_2 do not reduce the gain of the input stage and therefore may be large enough to avoid serious loading of the output stage. The effective resistance in series with the cathode of the first triode is, from Eq. (16),

$$\frac{r_p}{\mu + 1 + \dfrac{r_p}{R_k}}$$

where r_p and μ refer to the cathode follower. If the triodes are similar, this effective resistance has the effect on the gain of the triode of slightly more than doubling its plate resistance. From Eq. (15) it is apparent that the effective feedback factor of Eq. (112) should be multiplied by $\mu/(\mu + 1 + r_p/R_k)$ so that if the μ's are the same,
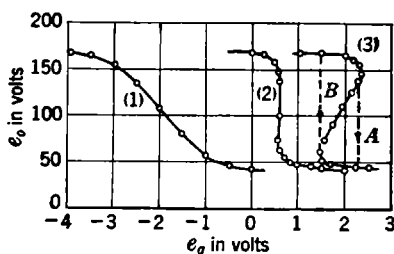
$$\beta = -\frac{\mu + 1}{\mu + 1 + \frac{r_p}{R_k}} k, \tag{116}$$

where $k = R_1/(R_1 + R_2)$. The two μ's, however, are subject to small changes independently of each other. If μ_1 and μ_2 refer respectively to the amplifier and cathode follower and r_p to the latter, Eq. (115) becomes

$$\frac{\Delta\mathcal{G}}{\mathcal{G}} \approx \frac{1}{\mu_1 + 1} \frac{\Delta\mu_1}{\mu_1} - \frac{1 + \frac{r_p}{R_k}}{\mu_2 + 1 + \frac{r_p}{R_k}} \frac{\Delta\mu_2}{\mu_2}. \tag{117}$$
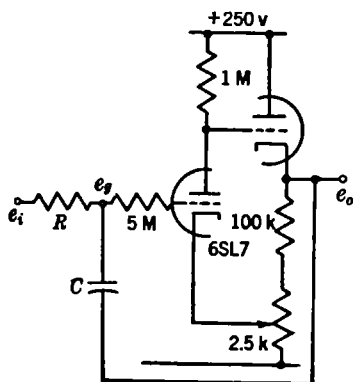
(a) (b)

FIG. 11·57a.—Linear sweep or integrating circuit employing positive feedback.
FIG. 11·57b.—Amplifier characteristics of Fig. 11·57a with different feedback adjustments.

Much the same conditions obtain if the second triode in Fig. 11·56 is used as the amplifier so as to obtain the opposite sense of output voltage or if the two triodes are employed as a differential amplifier. The peculiarities of these arrangements are discussed in Sec. 11·10.

Positive Feedback.—Positive feedback has the effect of increasing the gain. Also, because Eq. (101) applies as well to positive as to negative feedback, it decreases the linearity and stability of gain. In some applications, such as when the output-voltage range is small and when it is feasible to readjust the feedback when tubes are replaced, this limitation is not serious. Amplifiers used in voltage regulators, linear sweep circuits, and servomechanisms may sometimes employ positive feedback to advantage.

In a linear sweep circuit, or voltage integrator, employing a direct-coupled amplifier and a resistor and condenser in a "fed-back time-constant" arrangement, perfect operation requires infinite gain. This

may be obtained by adjusting the feedback factor β so that it is the reciprocal of the amplification, thereby causing the denominator in Eq. (93) to disappear. The circuit will still operate if β is so great that the denominator is actually negative—there will simply be an error of the same nature, but of opposite sign, as if the gain were finite and positive. Figure 11·57 gives an example of a simple linear sweep circuit or voltage integrator employing a high-gain amplifier with over-all capacitive negative feedback through C and R, such that, in so far as e_g is constant, the rate of change of output voltage is proportional to the input voltage. The amplifier between e_g and e_0 is a single-stage amplifier with a cathode-follower output stage and positive feedback to the cathode, as in Fig. 11·53. The curves are input vs. output voltage characteristics for the amplifier. The upper and lower limits are caused by grid current

FIG. 11·58.—Voltage regulator employing positive feedback.

in the cathode follower and in the amplifier triode respectively (the grid resistor being for the purpose of demonstrating the latter). Curve 1 represents the characteristics with no positive feedback and shows a maximum gain of about 50. For Curve 2 the feedback was adjusted so as to afford infinite gain over a limited range. In the case of Curve 3 even more feedback was used, resulting in negative gain over a portion of the output voltage range. This reflex part is unstable; without R and C the output voltage would jump along dashed line A or B (respectively) if e_g were being raised or lowered, but with R and C in operation the whole curve is actually traced out.
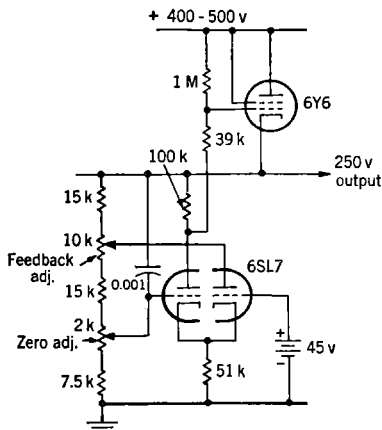
For perfect operation, an electronic voltage regulator, requires an amplifier with infinite gain, which may be approached or achieved conveniently with the aid of positive feedback. The regulator circuit also includes over-all negative feedback, which results in stability even though the positive feedback may be slightly more than is needed for infinite gain. The regulator of Fig. 11·58 may be adjusted to give an almost constant output over a limited range of load-current or supply-voltage variation. Positive feedback is provided by the resistive coupling between the cathode-follower plate and the amplifier grid. If the amplifier grid voltage is lowered, the decrease in plate current results in an almost equal increase in plate current in the cathode follower,

which acts on the bleeder resistance to add to the input change. The cathode follower also provides heater-voltage compensation and loading of the reference source. The small condenser increases the negative feedback at high frequency, thus preventing oscillation. This circuit is given not as an example of a good voltage regulator but only as an illustration of the use of positive feedback as a simple means of increasing amplification.

In a servoamplifier the input signal comprises the error, and the gain should therefore be as high as possible. Because linearity is not of importance, it is permissible to employ positive feedback in many applications. Usually the upper limit of gain is determined by consideration of stability of the servomechanism, but there are instances[1] where infinite gain may be employed to advantage, and the feedback may be even greater, as in Fig. 11·57b, Curve 3, without causing oscillation, although an error of the reverse sense results.

If the feedback is adjusted for infinite gain [$\beta = 1/\alpha$ in Eq. (93)] and thereafter the amplification without feedback changes by a certain percentage (e.g., because of tube replacement), the gain with feedback will become some finite positive or negative quantity whose magnitude will be proportional to the original amplification. If $\beta = 1/\alpha$ in Eq. (93) and if $\alpha + \Delta\alpha$ is substituted for α, the gain with feedback is found to be

$$\mathcal{G} = -\alpha \frac{1 + \Delta\alpha/\alpha}{\Delta\alpha/\alpha}. \tag{118}$$

Thus, for example, if β in Fig. 11·57 were adjusted to give Curve 2 and then the tube were replaced by one with an amplification that was 10 per cent higher, the curve would assume a shape slightly more like that of Curve 3, with a slope of $-50(1 + 0.10)/0.10 = -550$.

When positive feedback is to be applied to a multistage amplifier, the question arises as to whether to apply it over all or to the individual stages. In the case of negative feedback it was determined that from the standpoint of constancy of gain, it is better when arranged overall. The same reasoning gives the opposite answer for positive feedback. When $\mathcal{G}/\alpha$, $\mathcal{G}_1/\alpha_1$, etc., are *greater* than one, the expression for $\Delta\mathcal{G}/\mathcal{G}$ in Eq. (106) is smaller than that in Eq. (105). Thus, it is a general rule that positive feedback, if employed, should be applied only over single stages rather than over the whole amplifier. Of course, it should be applied to the stages that are subject to the least percentage variation of gain.

It follows, then, that a multistage amplifier with over-all negative feedback may be further stabilized as to gain by the use of positive feedback over individual stages. Figure 11·59 gives a computed illustration

[1] For example, in velocity servomechanisms employing RC feedback for stability.

of this principle in the case of an amplifier of two stages, each with a normal amplification of 20 and each being subject to the same variation. For each curve, the gain was computed from Eq. (93); the feedback factors were assumed to remain constant at values giving an over-all gain of five at the normal value of α. The individual positive feedback is seen to give a great improvement in constancy of gain. The improvement is appreciable even when the positive feedback is applied to only one of the stages.

Fig. 11·59.—Variation with respect to α of the over-all gain G of two stages each having amplification α with fixed over-all negative feedback adjusted so that $G = 5$ when $\alpha = 20$. Curve (1), no positive feedback; Curve (2), fixed individual positive feedback adjusted to give each stage a gain of 100 when $\alpha = 20$; Curve (3), fixed individual positive feedback adjusted to give each stage infinite gain when $\alpha = 20$; Curve (4), fixed positive feedback over only one of the stages adjusted to give it infinite gain when $\alpha = 20$.

Figure 11·60 is the circuit of a two-stage voltage amplifier employing SD-834 subminiature triodes ($\mu \approx 15$), with over-all negative feedback as in Fig. 11·51. The input voltage is applied to the cathode of the first amplifier tube by way of a cathode follower, permitting the necessary voltage inversion in two stages. Individual positive feedback is provided for
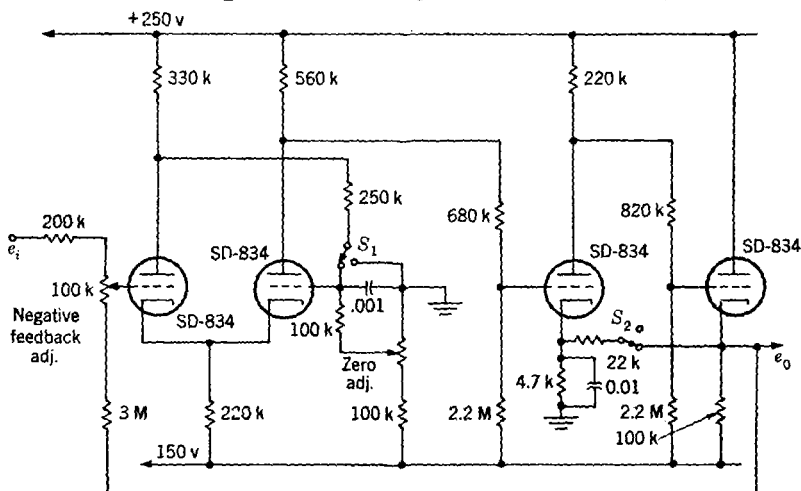

Fig. 11·60.—Two-stage voltage amplifier with gain stabilized at 10 by over-all negative feedback. Positive feedback operative for each stage with switches S_1 and S_2 in positions shown.

each stage: in the first stage by coupling from the cathode-follower plate to the amplifier grid and in the second stage by coupling from the output stage (cathode follower) to the amplifier cathode, as in Fig. 11·57a. In the linearity tests whose results are presented in Fig. 11·61, the nonlinearity of the amplifier was exaggerated by operation up to an output voltage of 100 volts, where the second-stage amplifier tube is near cutoff. Figure 11·61 shows an improvement in the linearity of more than 10-fold obtained by the positive feedback, with the same over-all gain. Substituting a weaker set of tubes, without readjustment of feedback, lowered the gain from 10 to 7.95 when the positive feedback was not employed and from 10 to 9.78 when it was employed. A stronger set of tubes had the effect of changing the gain from 10 to 10.30 without and to 9.91 with positive feedback. In the latter case, the positive feedback was more than enough for infinite gain in each stage, but the
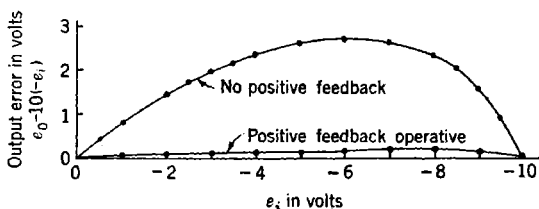

Fig. 11·61.—Deviation from linearity of output voltage, with and without positive feedback for each curve; negative feedback was adjusted so $e_o = 0$ at $e_i = 0$ and $e_o = 100$ at $e_i = -10$.

condensers prevented oscillation by reducing the effect at high frequencies without impairing the negative feedback.

No improvement can be obtained by the preceding method in the gain stability of a single-stage amplifier. Two β's between the same output and input points simply add algebraically.

EXAMPLES OF SPECIAL-PURPOSE AMPLIFIERS

In the next three sections are circuit diagrams and qualitative explanations of a few examples of direct-coupled amplifiers. The circuits are selected with the primary view of illustrating the principles discussed in previous sections, and they are not necessarily the best that have been developed for the purpose.

11·14. Current-output Amplifiers. *Meters.*—The single-stage differential amplifier of Fig. 11·62, which actuates a $\frac{1}{2}$-ma meter, illustrates the independent gain and zero adjustments shown in Fig. 11·30. The differential action affords a linearity commensurate with that of most portable meters. For a grid input voltage of 1 volt at full scale (as shown), very little resistance is needed between the cathodes; if the minimum full-scale input voltage can be 2 or 5 volts, the higher resistance

allowed between the cathodes will appreciably improve the linearity. Saturation limits the overload on the meter to about 100 per cent.

A more precise voltmeter, suitable for actuating a recording meter, is given in Fig. 11·63.   Current feedback is employed, the current in
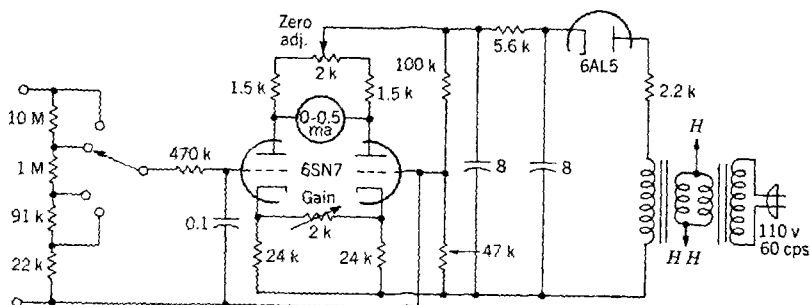

Fig. 11·62.—Simple vacuum-tube voltmeter.

the meter being measured by a stable resistance and a voltage proportional to it fed back to the auxiliary grid of the input differential amplifier.

The second stage is a pentode amplifier with its cathode returned to a cathode follower.   This arrangement permits differential input so that the full gain of the first stage is realized.   It also allows input voltages
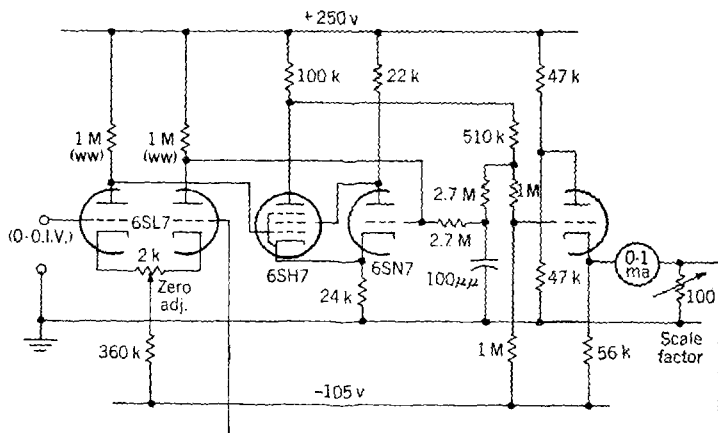

Fig. 11·63.—Voltmeter using high-gain amplifier.

at the plate level of the first stage; therefore, no dropping divider is needed at this point where the drift of such a divider would be felt much more than after the second stage.   Another feature permitted by the cathode follower in the second stage is the local positive feedback (by means of the 5.4-megohm resistor), which increases the linearity with

over-all negative feedback (Sec. 11·13). Connecting the pentode screen and triode plate together prevents the common dropping resistor from causing a reduction in gain; in addition, the reduction in screen voltage produces higher gain in the pentode. The output cathode follower is designed to limit the current in the 1-ma meter at about 2 ma in either direction; the two limits result from plate-current cutoff and from grid current in the input resistance.

The circuit may be used for full-scale input voltage as high as 10 volts, the limit depending on the value of the feedback resistor. The lower limit of full-scale input voltage is determined by the drift of the input tube. This limit is about 100 mv, and at this scale factor the drift may be noticeable over a period of hours. In order to minimize this drift,
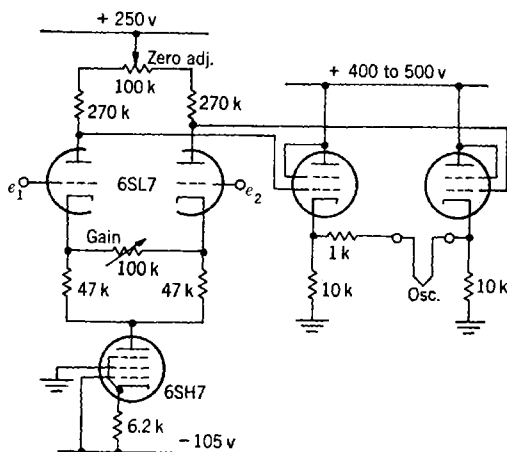

FIG. 11·64.—Two-stage oscillograph driver.

the input tube should be aged several hundred hours with the heater on. Selection of this tube is also recommended, as some tubes are subject to rapid erratic changes equivalent to input "noise" of 2 or 3 mv or more. This effect may be observed on the meter if the current-feedback resistor is reduced to 5 or 10 ohms.

Magnetic Oscillograph Driver.—The first stage of the driving circuit of Fig. 11·64 illustrates the use of a constant-current circuit in a differential amplifier to eliminate the input common-mode variation from the output voltage. The input voltage may be either differential or single-ended (with the unused input terminal fixed at the appropriate level) and may be at any level from −50 to +100 volts. Voltages beyond these limits may be reduced by the use of dividers. Thus, several driving circuits for measuring voltages at various levels may use the same power supply.

The gain may be varied from 6 to 25 ma per volt; large input voltages may be reduced by an input attenuator, but the amplifier gain should be set at minimum in such cases. The maximum output current is about ± 18 ma; it can be increased by the use of smaller cathode-follower cathode-return resistances and larger output tubes. The negative feedback is applied individually in this amplifier, because over-all feedback is impractical where the input level is so variable. Linearity is satisfactory for the purpose.

Servoamplifiers.—The servoamplifier in Fig. 11·65 is designed to drive the differently wound fields of a small instrument servo motor.
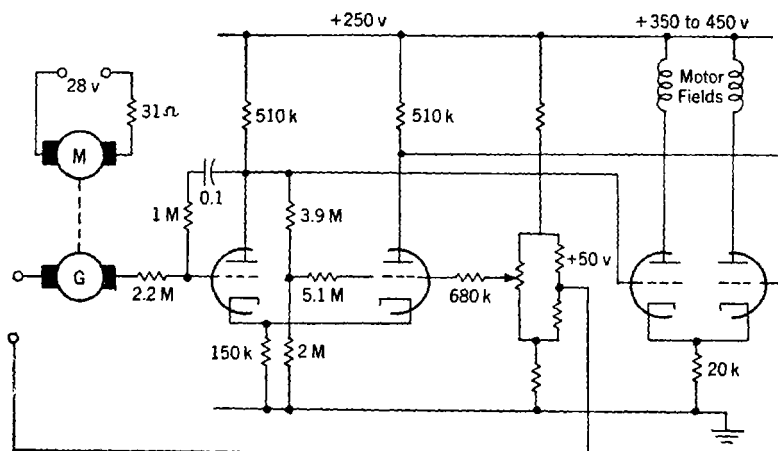

FIG. 11·65.—Velocity servo amplifier driving differentially wound motor fields.

The arrangement shown comprises a velocity servomechanism, wherein the motor drives a tachometer generator whose output voltage is subtracted from the input speed-control voltage. The difference voltage is the "error signal" to be amplified. No negative feedback is used in the amplifier itself, as this would detract from the linearizing effect of the over-all negative feedback, which includes the tachometer. On the other hand, local positive feedback is beneficial (Sec. 11·13), and this is provided in the first stage by the resistive coupling from the first plate to the second grid. This is designed for approximately infinite gain. The capacitive negative feedback, in combination with the resistances in series with the condenser, provides damping of oscillations of the servomechanism.

The input stage of the servoamplifier, which was shown in Fig. 11·42, employs a diode to balance the effect of heater-voltage variation on the cathode of the pentode. The zero-adjustment resistance and screen-grid

voltage divider do not impair this function, as the currents in the various parts of the divider are substantially constant. The diode current is two or three times the pentode current. A slight loss of gain results from the zero adjustment and diode variational resistance, but these are very small compared with the plate-load resistance. The small auxiliary resistors in series with grid and plate are designed to dampen possible high-frequency oscillation. The operation of the output stage has been explained in Sec. 11·9.

Current Regulator.—Where large currents are to be held constant, the constant-current circuits of Sec. 11·8 are inadequate. The power control and the voltage amplification must be done separately, (and the current must be fed back and compared with a standard). The
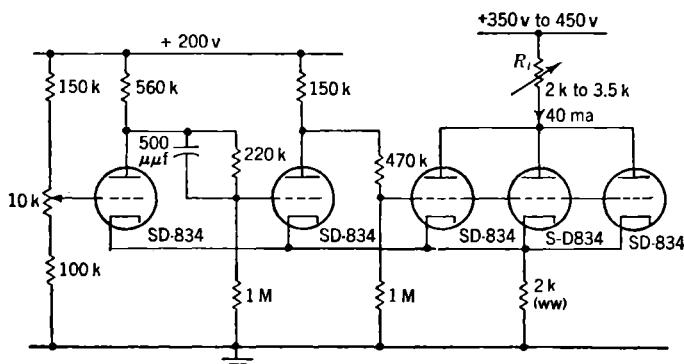

FIG. 11·66.—Current-regulating d-c amplifier.

function of the circuit of Fig. 11·66 employing an unregulated voltage source is to maintain a constant current in a load resistor whose resistance is subject to large variation. The load current flows in a constant resistance in the cathode circuit of the power stage, and the resulting voltage is fed back to the cathode of the input stage, whose grid receives the standard voltage. The cathode of the second voltage stage is connected to the same point, partly for convenience and partly because the slight cathode coupling between the first and second stages comprises positive feedback, increasing the gain. The condenser inhibits oscillation by decreasing phase retardation at high frequencies. If the 200-volt source changes, the output current changes in the same proportion. No compensation is provided for heater-voltage variation; a 10 per cent change amounts to an input drift of about 100 mv, which causes an output variation of 0.13 per cent.

11·15. Voltage-output Amplifiers. *Electrostatic Oscilloscope Driver.*— An electrostatic oscilloscope requires a large deflecting voltage, applied in a push-pull manner. The two-stage driver in Fig. 11·67 employs a

pair of pentodes to get a large voltage swing without using an excessive supply voltage. The pentodes are arranged in a differential amplifier with single-ended input and with considerable linearizing degeneration from resistance between the cathodes. The input stage illustrates the method of cancellation, with a series triode, of drift due to heater-voltage variation, as explained in Sec. 11·2. The cathode resistor may be adjusted to give the best cancellation (after the tube has been aged for several hours) by observing the output change upon variation of the heater voltage. Thereafter, the zero adjustment is made at the plate of the upper triode. The input grid resistor is designed to protect the grid from excessive current due to high input voltage.

Fig. 11·67.—Two-stage voltage amplifier for oscilloscope deflection.

Comparison Amplifier.—In d-c circuits and especially in computing devices, the need sometimes arises for a high-gain amplifier with two input terminals and a single output terminal, in which the output voltage depends only on the difference and not on the mean level of the input voltages. The general application is in the electronic solving of an algebraic equation. The two input voltages are the two equal parts of the equation, and the output voltage is the "unknown," which actuates other circuit elements involved in the equation in such a way as to force the two input voltages into equality. Thus there is an over-all negative feedback in the application of the amplifier, and linearization by means of negative feedback within the amplifier is superfluous and detrimental. On the other hand, local positive feedback, if designed toward obtaining infinite gain, can improve the performance.

In the amplifier of Fig. 11·68 freedom from the effect of the common-mode variation is obtained by the use of a differential input stage with a constant-current triode as the common-cathode return.   As the input voltage level is changed from 0 to +80 volts, the input voltage difference needs to be changed by less than 50 mv, for the majority of 6SU7 tubes, in order to hold the output voltage constant.

In addition to permitting the use of both output terminals of the first stage, the differential arrangement of the second stage provides a linearizing effect and allows the maximum output voltage range, as in Fig. 11·33c.   It also provides a simple means for local positive feedback,
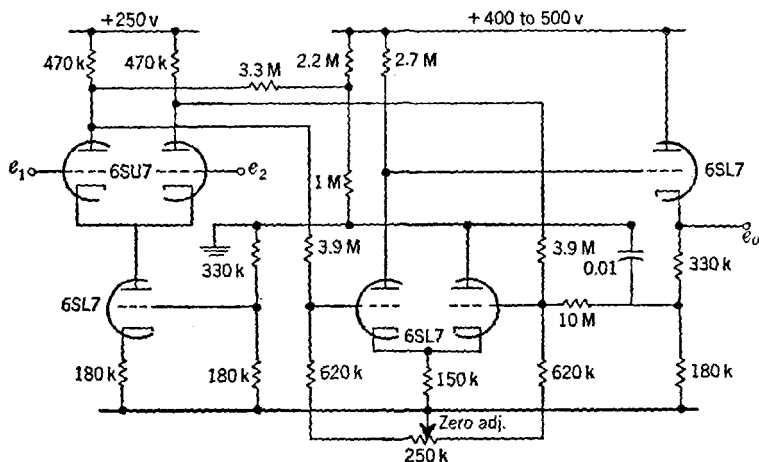
Fig. 11·68.—Two-stage amplifier with differential input and single-terminal cathode-follower output.

which is applied by resistive coupling from the cathode-follower output terminal to the cathode-follower grid in the second stage.   The condenser in this network prevents oscillation by decreasing the positive feedback at high frequencies.

The output voltage range is from somewhat below zero to considerably above 250 volts.   Without the positive feedback the gain of the amplifier is about 500, with very good linearity.   With the positive feedback operative the output voltages traverse at least 250 volts while the input difference changes not more than 50 mv.

The resistance coupling between the high-voltage supply and the plate of the first tube is used to cancel the effect on the output of variations in this voltage.   The +250 volts and −150-volt supplies are regulated.   Heater-voltage variations are, of course, balanced by the differential arrangement.

The comparison amplifier may be used as a driver, similar to a cathode follower but having much greater precision, by connecting the output terminal to one of the input terminals (e_2 in Fig. 11·68). It may also be used as a linear, adjustable-gain amplifier, by feeding back any desired fraction of the output voltage through a resistance attenuator.

Inverting Amplifier.—For inverting a signal voltage (i.e., obtaining the negative of it, not its reciprocal), the simple degenerative triode circuit of Fig. 11·8 may be employed with plate and cathode resistors nearly equal. Besides having only medium precision, such an arrangement is limited in application because the output voltage contains a large

Fig. 11·69.—Voltage inverter employing a two-stage amplifier with cathode-follower output.

additive constant. This constant voltage is directly affected by any supply-voltage variation, and both it and the gain are affected by output load.

More precise voltage inversion, without these disadvantages, is accomplished by a high-gain amplifier whose input voltage is the average of the voltage to be inverted and the amplifier output voltage. The amplifier is sensed to invert, so the averaging process is done by negative feedback. If the gain is high enough so that the amplifier input voltage is negligible, and if the averaging resistors are equal, the output voltage is the negative of the input voltage.

In Fig. 11·69 a two-stage amplifier is made to invert by the use of cathode-follower input to the second stage. Positive feedback is applied in the second stage by means of resistive coupling between the first plate and the second grid. The potential is dropped to the grid of the output

cathode follower, without attenuation, by the use of a triode constant-current circuit. This not only prevents attenuation but also permits a greater range of output voltages than would be possible with an ordinary divider. The input stage illustrates the use of a first grid as a diode plate to eliminate drift due to cathode change. For voltage inversion without change in scale factor R_1 and R_2 are equal. Amplification or attenuation may be obtained by other ratios between these resistances. The condenser-resistor feedback at the input stage prevents oscillation.

Voltage Level Changer.—Figure 11·70 illustrates a use of the push-pull quality of the output of a differential amplifier with a single input terminal. It is a simple device for providing a voltage that remains at a fixed potential difference above a variable input voltage. A possible
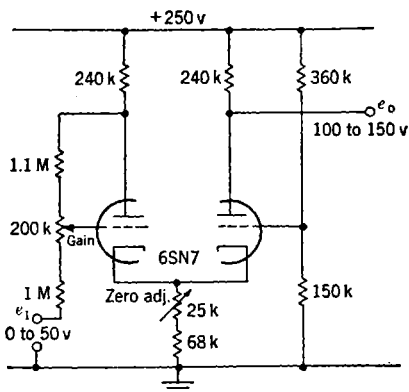
Fig. 11·70.—Differential amplifier arranged as a level shifter.

application might be to hold the screen of a pentode at a certain voltage with respect to the cathode regardless of movement of the latter. The resistance negative feedback from the first plate to the first grid is adjusted so that the gain from the input terminal to this plate is -1. Since the output is push-pull, the gain to the other plate is $+1$, but with a shift in voltage level.

11·16. A Galvanometer-photoelectric Tube Feedback Amplifier.— The minimum input scale factor of direct-coupled amplifiers is limited by drift of the characteristics of the input tube. The minimum significant input-voltage variation, with ordinary vacuum tubes, is seldom much less than 1 mv, even under short-time laboratory conditions.

The usual method of amplifying or measuring d-c voltages, too small for a direct-coupled amplifier, is to convert the direct current to alternating current by a modulator that is not subject to drift and to amplify the resultant voltage with an a-c amplifier. Depending on the output

requirements, the amplifier output voltage may or may not be demodulated. Such modulation systems are outside the scope of this chapter.

(a)

(b)

(c)

(d)

FIG. 11·71.—Feedback arrangements with galvanometer phototube input. (a) Voltage input, voltage output; (b) voltage input, current output; (c) current input, current output; (d) current input, voltage output.
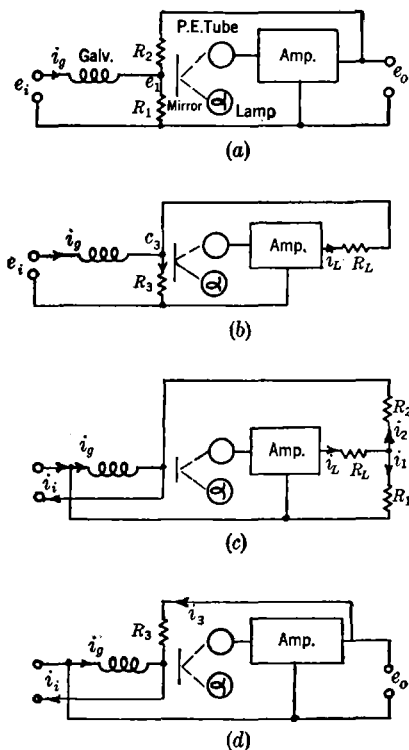
An alternative to the a-c conversion method employs a device that is inserted ahead of a direct-coupled amplifier and is capable of increasing the amplitude of the input signal without drift. Linearity in the device itself is not needed if it is possible to arrange over-all negative feedback.

Such a scheme is afforded by the combination of a mirror galvanometer with a photoelectric tube. The general arrangement is shown in Fig. 11·71. A current in the galvanometer in one direction turns the mirror and alters the light received by the phototube in such a way as to raise the amplifier output voltage, and a current in the galvanometer in the other direction lowers the amplifier output voltage in a similar manner. A very small current is required to turn the mirror far enough to obtain full amplifier output voltage in either sense.

The negative feedback through a potential divider, illustrated in Fig. 11·71a, acts to reduce the galvanometer current by making e_1 almost equal to e_i. If equality is attained and i_g is zero, the output voltage becomes

$$e_0 = \frac{R_1 + R_2}{R_1} e_i. \tag{119}$$

If i_g is not negligible, the output voltage is

$$e_0 = \frac{R_1 + R_2}{R_1} (e_i - Ri_g), \tag{120}$$

where R is the parallel combination of R_1 and R_2, plus the galvanometer resistance, plus any series input resistance between e_i and the gal-

vanometer.   For good linearity Ri_g should be small in comparison with e_i. Ideally, therefore, the galvanometer should have no mechanical restoring torque and no friction.

The series feedback of Fig. 11·71b, where the output current is measured by being passed through a fixed resistance R_3, gives an output current of

$$i_L = \frac{e_i}{R_3} \tag{121}$$

if i_g is negligible.

Figure 11·71c is an arrangement for current amplification.   If i_g is negligible, so that $i_i = i_2$, the output current is

$$i_L = \frac{R_1 + R_2}{R_1} i_i. \tag{122}$$

Also, the input voltage is zero, which is a requirement of an ideal current-measuring device.

The output voltage of the arrangement of Fig. 11·71d is

$$e_0 = R_3 i_i \tag{123}$$

if the galvanometer current is assumed to be zero.   If the requirements are such that this resistance is too large for practical application, e_0 may be attenuated before being fed back as a current through R_3.

In the circuits of Figs. 11·71b and 11·71c, which have current feedback, the output current is not affected by the load resistance, the output voltage adjusting itself to compensate for any variations thereof.

By the use of a suspension galvanometer with a minimum restoring torque, the galvanometer current required for full output may be made a very small fraction of a microampere.   In Figs. 11·71a and 11·71b the galvanometer and feedback resistances can be low [e.g., approximately 1000 ohms for R in Eq. (120)]; therefore if the input circuit resistance is also low, voltages well below 1 mv may be amplified with good linearity and without interference from drift.   Drift of the amplifier itself is of negligible importance because it is offset by a very small shift of the mirror position.

The device actually comprises a servomechanism and is therefore subject to mechanical oscillation under some conditions.   This oscillation may be obviated by negative feedback through a condenser, in addition to the resistive negative feedback and at a somewhat higher amplification.

The rapidity of response does not depend on mechanical restoring torque.   The latter can (and should) be zero; the negative feedback will always act to hold the galvanometer rigidly in its proper position. Any change of input voltage, which results in a galvanometer current and a turn of the mirror, is immediately counteracted by feedback of the

proper amount to cause the mirror to assume its appropriate new position. Any small error in this position results in a large restorative galvanometer current.   Since the total movement is small and the electrical restoring torque is large, the rate of response of the device is much higher than that of a galvanometer as such.   The frequency response is limited by the inertia, the movement required, and the condenser feedback needed to prevent oscillation.

The circuit of Fig. 11·72 has been used successfully as a micro-ammeter.[1]   The feedback is as illustrated in Fig. 11·71c, where R_1 and R_2

FIG. 11·72.—Microammeter based on Fig. 11·71c.

are in the form of an ayrton shunt to give several scale factors.   The condenser feedback to prevent hunting is derived at a point of higher gain than the resistive feedback, by employing a tap on the plate-load resistor for the latter.   No heater-voltage compensation is needed, as this effect is not noticeable; most of the gain of the system is ahead of the amplifier tube.

A differential phototube is employed, in a manner that affords very high output voltage to the amplifier grid for a small shift of the light beam.   At the low voltages employed, the rate of change of anode-to-cathode voltage with respect to illumination, for a given current in each part of the phototube, is high.   With the two parts connected in series as shown, the two currents are equal (if output current to the

[1] Shepard Roberts, personal communication.

amplifier grid is neglected), with the result that an increase of light to one cathode and a decrease of light to the other gives the same output voltage that would be obtained if an infinite load resistor were used for each section. A change of total light does not change the voltage level as it would if load resistors were used.

The electrical isolation between the amplifier input terminals and the galvanometer permits the former to be at a lower potential than the latter, so that the output signal need not have its potential lowered for the feedback.

The effective input resistance is much lower than that of the galvanometer alone, because the feedback tends to keep the voltage across the galvanometer at zero regardless of input current.

11·17. D-c Amplifier Analysis.—This section deals with the analysis of the performance of a d-c amplifier from inspection of its circuit dia-
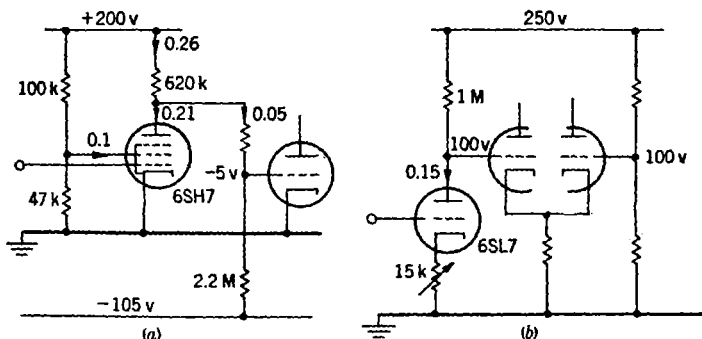
Fig. 11·73.—Computation of tube-operating conditions, based on an estimated grid potential for the following tube.

gram. The primary object is to determine the approximate voltages and currents at every part of the circuit in order to ascertain if each tube is being used in a satisfactory manner. Other objects of analysis may be the determination of linearity, effect of variation of supply and filament voltages, effects of the incidence of extremes of tube and resistor tolerances, etc.

Because of the rather large possible discrepancies between nominal and actual resistance values, only a fairly rough computation is justified in most instances. Simple application of Ohm's law, Thévenin's theorem, and the principles of the foregoing sections as applied to individual stages are sufficient to describe adequately the d-c aspects of any multistage amplifier.

An estimate of the grid-cathode potential in a tube is a good starting point for determining conditions in the preceding stage. From a knowledge of the tube type and the plate load the grid bias can be estimated

within 2 or 3 volts. (For a low-μ power tube a consultation of its characteristic curves may be advisable.) The computation proceeds backward in the circuit from this point. Some specific examples will illustrate the method; no attempt will be made to formulate a general set of rules.

In Fig. 11·73a the bias for the following tube has been estimated as −5 volts. The potential across the lower part of the divider is thus 100 volts; therefore (if there is no grid current) the pentode plate potential must be 1/2.2 of 100 above −5, or +40 volts, and the divider current is about 0.05 ma. The current in the plate resistor is $\frac{160}{620} = 0.26$ ma, and the plate current is $0.26 − 0.05 = 0.21$ ma. Screen current will be about one-fourth to one-half of this, for example, 0.1 ma. The resistance of the screen bleeder as viewed from the screen is 32,000 ohms, so its regulation due to screen current is only 3 volts. Screen potential is about 60 volts. The grid potential of the pentode is found, from characteristic curves, to be about −2.0 volts (subject to some variation from tube to tube). This bias is ample to ensure against positive grid current.

From the characteristic curves, the transconductance, at $i_p = 0.21$ ma and $e_s = 60$ volts, is found to be about 0.8 ma/volt. The value of μ_s, or $−\partial e_s/\partial e_g$, is 35. The effective load resistance is the parallel resistance of 620,000 ohms and 3.2 megohms, or 520,000 ohms. The gain, from Eq. (27), is therefore (assuming $n = 0.5$)

$$\mathcal{G} = \frac{-(520)(0.8)}{1 + \dfrac{(0.5)(32)(0.8)}{35}} = \frac{-420}{1 + 0.37} = -300.$$

(The second term in the denominator is due to screen-bleeder resistance; if this were reduced by a factor of 4 the gain would be 380, an increase that might be worth the extra current drain.) The output divider to the following grid attenuates the gain to about 200.

In Fig. 11·73b the output to the differential amplifier will evidently be in the neighborhood of +100 volts. Since there is no divider, the plate current is 150 volts/1000k = 0.15 ma. From Fig. 11·73a, the grid bias appears to be about −1.6 volts. If the cathode resistor is for the purpose of zeroing the input potential for an output of 100 volts, its value will be 1.6/0.15 = 10.7 kilohms. Plate resistance and amplification factor at $i_p = 0.15$ ma, $e_p = 100$ volts, are 130,000 ohms and 65 (approximately); therefore the gain is [Eq. (9)]

$$\mathcal{G} = -\frac{65}{1 + \dfrac{130 + (66)(10.7)}{1000}} = -\frac{65}{1 + 0.83} = -35.$$

In both (a) and (b) of Fig. 11·73, an increase of 10 per cent in heater voltage will require a lowering of input potential of about 0.1 volt if the

output is to be held constant; otherwise the output will change by this amount times the gain.

Since the output variation required for full-range operation of the following tube is only a few volts at most, the preceding computations should suffice. If the required output range of the tube in question is great, the computation must be repeated for each limit. This usually is necessary only at the output stage.

The output of the amplifier of Fig. 11·69 is supposed to range from −50 to +50 volts with respect to ground. At the lower limit, the cathode follower has a plate-cathode voltage of 300, a plate current of 1 ma, and a grid bias, therefore, of −4 volts; its grid potential will then be −54 volts. The current in the constant-current circuit comprising the lower part of the divider feeding the cathode follower is about 75 volts/1 megohm = 0.075 ma; the amplifier plate therefore is at a level 150 volts above the cathode-follower grid, or +96 volts. Current in the plate resistor is $(250 − 96)/560 = 0.275$ ma, so that the plate current is this quantity minus 0.075 or 0.2 ma. The grid of this triode is in the neighborhood of +10 volts, and the cathode will be slightly above the grid; therefore the current in the common cathode resistor is 0.32 ma, and 0.12 ma is left in the first triode of the differential amplifier. At the upper output limit, the amplifier plate is at +199 volts, plate current is 0.09 ma, and plate current in the other triode is 0.32 minus this, or 0.23 ma. In this condition the latter triode still has ample plate-cathode voltage (about 90 volts) to ensure that the grid current is negligible in view of the input resistance.

The average plate resistance of each of the differential amplifier triodes is about 120,000 ohms. (One increases and the other decreases as the output range is traversed.) The first triode is a kind of cathode follower driving the cathode of the second. The plate-supply divider adds 110,000 ohms to the r_p of this cathode follower with a μ of 65, and Eq. (40) gives a gain from the first grid to the second plate of 40, without the positive feedback. Thus, for 1 volt of input, the second plate moves 40 volts. The first plate moves $\frac{110}{560}$ of this, or 8 volts in the opposite direction, since a certain change of current in the 560,000-ohm plate resistor implies an equal and opposite change in the other. This 8 volts of plate movement causes a 0.8-volt displacement at the second grid, so that for the assumed 1 volt of total differential input only 0.2 volt is needed at the first grid; the positive feedback increases the gain by a factor of 5.

In determining the d-c levels in a multistage amplifier the logical direction for the computation to proceed is generally from the output toward the input. An important exception is in the case of differential amplifiers, where the double-ended signal is transmitted as such from stage to stage.

For example, in the servoamplifier of Fig. 11·65, the level of the grids of the output stage cannot be estimated as in the foregoing examples; this level is determined by conditions in the first stage. The input grids are at about $+50$ volts, and the cathodes ride a volt or two about this, so the sum of the plate currents is 0.35 ma. At balance, then, the plates will be at $250 - (510)(0.17) = 160$ volts, approximately. About 8 or 9 ma will flow in the output-tube cathode resistor, and this is divided between the two motor fields according to the amount of unbalance.

In Fig. 11·64 the first step is to compute the current in the constant-current pentode. If the i_p and i_s vs. e_g curves of the pentode are available, a simple method is to draw a line from the origin with a slope of $-1/R_k$, as shown in Fig. 11·74. The point at which this line intersects the curve of total cathode current, for the screen-cathode potential in
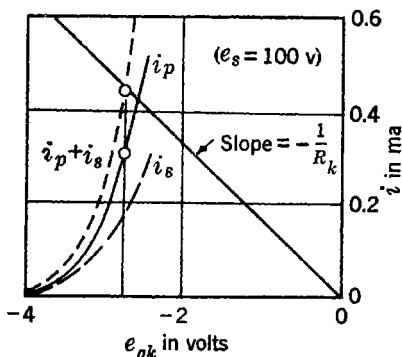
Fig. 11·74.—Determination of constant-current pentode current.

question (about 100 volts in this instance), will indicate the bias. In the example this is -2.8 volts, and the plate current is therefore 0.3 ma. This current flows in the two plates of the differential amplifier, regardless of the input level, so that the average of the two plate potentials is $250 - (320)(0.15) = 200$. The cathodes of the power tubes will ride at about 20 volts above their grids; therefore, at balance, the current in each of these tubes will be $\frac{220}{10} = 22$ ma. This is also the approximate current available for the load at saturation, i.e., when either tube is cut off.

When part of an amplifier is in differential form but the output is not, the computation may have to converge from both ends. The circuits of Figs. 11·63 and 11·68 are examples of this situation. In the latter, the grid of the output cathode follower is required to travel from about -8 volts to nearly $+250$ volts. This range gives the limits of plate potential (from ground) and current in the preceding triode; but to find its cathode level and the limits of current in the other triode of

this differential pair, a fresh start must be made at the input stage. Here, the constant-current circuit holds the total cathode current at 0.3 ma. At balance, each plate current will be 0.15 ma. With no plate current each plate would be at about 210 volts because of the dividers to the following stage input. The effective plate-load resistance is 420,000 ohms, including the dividers, so the plates will be at

$$210 - (420)(0.15) = 145 \text{ volts.}$$

Thus, the grids of the second stage are at -103 volts, and its total plate current is about 0.33 ma.

Computation of gain may be made either by application of the gain formulas (with care to evaluate the variational tube parameters in the vicinity of the existing conditions) or by recomputation of the input for two assumed values of output. The same alternatives exist in the determination of the effects of positive and negative feedback. Linearity may be estimated by means of a gain calculation at each extreme of output and application of Eq. (102) or an appropriate variation thereof.

For determination of the effect of possible extremes of resistor values and vacuum-tube characteristics, at least two computations of the conditions must be made, after a qualitative inspection to decide which extreme of each component will result in a displacement of a given sense.

In ascertaining the shift resulting from changes of d-c supply potentials, the appropriate derivatives of the amplifier formulas [e.g., Eqs. (3), (7), (15), and (26)] may be applied. Great care must be used to consider the effect on every electrode of each tube, e.g., the screen grids of pentodes. Perhaps a surer method is a recalculation of the amplifier input for a certain output, at each of two extremes of the supply voltage; the output shift would be this value times the gain.

CHAPTER 12

AMPLIFIER SENSITIVITY

By E. J. Schremp

12·1. Introduction.—There is always a limit to the number of vacuum-tube stages that may be used profitably in a vacuum-tube amplifier. Various factors enter in determining what this limit of useful amplification will be. The one that is treated in this chapter and is perhaps most common is that arising from the existence of electrical noise or spontaneous current and voltage fluctuations in the early stages of an amplifier and in the signal source itself. These small fluctuations, generally indiscernible under other circumstances, can always be detected in a circuit that is attached to or forms a part of the first few stages of a very high gain amplifier. Therefore, in order to detect signals entering the input of such an amplifier, input signals larger than a certain minimum size are required; input signals of smaller size could not be recognized even after amplification, because they would be masked by the amplified noise which is already observed. Because further amplification would increase by the same factor the noise and signal amplitudes observed at the output, such further amplification is useless in detecting input signals smaller than the minimum size. This notion of a "minimum detectable signal" may be taken, for the present, as a qualitative definition of the term "amplifier sensitivity." Later on in this chapter, this term is dealt with quantitatively, and the factors upon which it depends and the ways in which it may be improved are illustrated.

The existence, in all kinds of circuit elements, of residual or spontaneous current and voltage fluctuations, although not a macroscopic phenomenon and therefore not a matter of familiar experience, nevertheless represents a microscopic condition that prevails in all material bodies and in all forms of energy. That is to say, electrical noise is but one manifestation of the statistical fluctuations that occur on a microscopic scale in all forms of matter and energy in which it is found that the constituent molecules, ions, electrons, and even photons are in a perpetual state of more or less random motion. The simplest and earliest-known example of this state of chaotic motion underlying all matter may be observed, with the aid of a microscope, in a clear liquid containing a suspension of very fine particles. These particles will be observed to dance around in the liquid, in a permanent state of random zigzag motion.

The smaller the particles are the more violent their motion becomes and the more accurately they simulate the motions of the liquid molecules themselves. By this simple experiment in "Brownian movement" (so-called after the English botanist Brown who first observed this phenomenon in 1827), the existence and movement of molecules are proved directly. Similar movements of electrons and ions in electrically conducting materials are therefore to be expected; and in consequence of their movements, spontaneous current and voltage fluctuations must occur in such materials.

In general, any current or voltage fluctuation having the characteristics just described represents a certain variety of electrical noise. The distinguishing properties of such fluctuations are (1) their microscopic character, (2) their permanence, and (3) their intrinsic randomness (which may be of different types or degrees). Accordingly, the physical origin of such fluctuations must always involve large aggregations of small charged particles, moving under the influence of their own mutual forces or of other forces of comparable complexity or randomness. As long as the charged particles concerned are about the size of electrons or ions, their motions under such forces will be sufficiently random to give rise to current and voltage fluctuations of the type designated here as electrical noise.

12·2. Thermal Noise. *Definition.*—Of the various kinds of electrical noise, the simplest and perhaps also the most important is that called "thermal noise." Before defining it, however, it is well to consider first a little more deeply the meaning of the general term "noise" as it applies to electrical networks. In many applications of network theory it is sufficient to regard electricity as the flow of a continuous charged fluid. However, as already indicated by the "atomistic" line of thought in the preceding paragraphs, electric charge is known always to be localized in discrete charged particles, which are either electrons or ions. Every electric current, then, is the result of the motions of such electrons or ions. Now the motion of any one charged particle sets up a convection and displacement current, each of which is of the nature of a current pulse. Every electric current, therefore, is in actuality a sequence of such current pulses. Although in many cases, because of the immense number of charged particles in motion, the electric current is apparently continuous, it is nevertheless subject to fluctuations, for it must always remain a sequence of current pulses, perhaps very closely spaced, but never spaced closely enough so that the successive pulses join together continuously. The resulting small fluctuations in current are called "noise currents."

The nature of these noise currents is governed by the nature of the motions of the electrons or ions responsible for the observed current.

But the mechanical motions of these electrons or ions are coupled with other forms of energy, such as the mechanical energy of molecular heat motion, the mechanical energy of a dynamo, the chemical energy of an electrolytic cell, or the radiant energy incident on an antenna or photocell. In order to describe the character of the electronic or ionic motions and thereby to describe the character of their resulting noise currents, it is necessary, then, to specify something of the nature and conditions of the other forms of energy that interact with these electrons or ions.

If, in particular, the electrons or ions are in thermal equilibrium with all the other forms of energy with which they might be coupled, then the noise due to them is called "thermal noise." Thus, in a circuit that is in thermal equilibrium, while the average net current is zero at every point, there are electronic or ionic motions in every direction. These motions are coupled with the thermal agitation of the molecules of the circuit, and hence there will be current fluctuations, which under this special circumstance are called "thermal-noise currents."

Observation.—One way of observing thermal noise is with the aid of a very sensitive galvanometer. With such an instrument it is observed that the galvanometer mirror suspension never comes precisely to rest, no matter how carefully all external sources of vibration are eliminated. One cause of this residual random motion of the galvanometer suspension is the bombardment of the mirror by air molecules, a process completely analogous to the phenomenon of Brownian movement mentioned above. But even if the suspension were placed in an evacuated chamber, some residual random motion of the mirror would still remain. This would be due to the flow of thermal-noise currents in the moving coil. Although these currents are extremely small, they are sufficient to cause observable rotations of the moving coil, provided that the coil suspension is of very small moment of inertia.

Thermal noise can also be observed with the aid of a high-gain radio receiver. When this method is used, an audible noise, due to small current fluctuations in the first few stages of the receiver, can be heard in the loud-speaker. For this reason, such current fluctuations were originally called "noise currents." These current fluctuations are, in part, thermal-noise currents in the input circuit of the first stage. This may be verified by observing the reduction in audible noise that occurs when the receiver input is short-circuited.

Measurement of Open-circuit Thermal-noise Voltage.—Some of the principal features of thermal noise will now be investigated quantitatively and from an experimental approach. A parallel resonant circuit will be selected as the object upon which to make measurements; and these measurements will be made with an r-f amplifier with a thermocouple meter capable of reading the mean-squared output voltage of the

amplifier at its output. The experimental arrangement is shown in Fig. 12·1. The parallel resonant circuit is composed of the ohmic resistance R, the essentially loss-free coil L, and the essentially loss-free capacitance C (inclusive of the amplifier input capacity); M is the thermocouple output meter, and the entire parallel resonant input circuit is assumed to be in thermal equilibrium at an absolute temperature T.

There are certain reasons why a parallel resonant circuit is preferable to any other circuit as the object for measurement: (1) The r-f amplifier has an unavoidable input capacity which would shunt the circuit to be

FIG. 12·1.—RLC-input circuit, r-f amplifier, and output meter M.

measured, thereby ruling out the possibility of studying any other equally simple circuit such as a series resonant circuit, and (2) the study of a parallel-resonant circuit leads to an understanding of the separate thermal-noise behavior of a resistance R, an inductance L, and a capacitance C.

Using an r-f amplifier as the measuring instrument is, in effect, the same as attaching across the parallel-resonant circuit a sensitive voltmeter of infinite internal impedance (if as previously remarked, its input capacity is thought of as belonging to the parallel resonant circuit). Consequently, an r-f amplifier is especially well suited to the measurement of the *open-circuit thermal-noise voltage* across the parallel resonant circuit, which shall hereafter be designated as a certain unknown function of time $V(t)$.

There are several possible methods of indicating the output response of the r-f amplifier to this thermal-noise voltage $V(t)$. For example, the r-f amplifier might be followed by a frequency converter and an audio amplifier and loud-speaker, in which case the thermal-noise voltage $V(t)$ would be indicated by an audible noise. Or the r-f amplifier might be followed by a frequency converter or detector and a video amplifier and oscillograph. In this case, the thermal-noise voltage $V(t)$ would be indicated by a visible trace on the oscillograph. Or again, the r-f amplifier might be followed by an averaging device, such as a thermocouple, in which case the thermal-noise voltage $V(t)$ would be indicated by a d-c meter. Since the last type of indication is quantitative, whereas the other types are generally only qualitative, the thermocouple meter M has been used here as the output indicating device.

For the experimental arrangement of Fig. 12·1 to yield a correct quantitative measure of the thermal-noise voltage $V(t)$, three experimental conditions must be fulfilled:

1. The input capacitance of the r-f amplifier (which composes a part
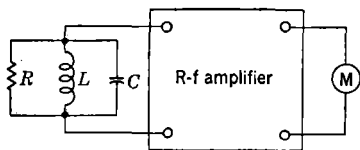
of the total capacitance C of the parallel-resonant input circuit) must be frequency-independent and must be the only contribution of the r-f amplifier to the input admittance.

2. The r-f amplifier must be linear.
3. The extraneous noise output from other sources of noise within the amplifier must be independent of the input termination of the r-f amplifier.

If the open-circuit voltage $V(t)$ is now regarded as existing only within a unit interval of time and as vanishing outside this time interval, then it may be assumed that it is quadratically integrable or, in other words, that

$$\int_{-\infty}^{+\infty} V^2(t) \, dt = \int_{-\frac{1}{2}}^{+\frac{1}{2}} V^2(t) \, dt = \overline{V^2} \tag{1}$$

is finite and equal to the mean-square value of $V(t)$. Accordingly, $V(t)$ possesses a Fourier integral representation

$$V(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} v(j\omega) e^{j\omega t} \, d\omega \tag{2}$$

with a Fourier transform

$$v(j\omega) = \int_{-\infty}^{+\infty} V(t) e^{-j\omega t} \, dt \tag{3}$$

which is a function of the angular frequency

$$\omega = 2\pi f \tag{4}$$

and with a *spectrum*

$$\begin{aligned} S(\omega) &= v^2(j\omega) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} V(t) V(t') e^{j\omega(t'-t)} \, dt \, dt' \\ &= \int_{-\infty}^{+\infty} R(\tau) \cos \omega\tau \, d\tau, \end{aligned} \tag{5}$$

where $R(\tau)$ is the *correlation function*

$$R(\tau) = \int_{-\infty}^{+\infty} V(t) V(t + \tau) \, dt. \tag{6}$$

Moreover, we can write

$$\overline{V^2} = \int_{-\infty}^{+\infty} V^2(t) \, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(\omega) \, d\omega = R(0) \tag{7}$$

as a consequence of Eqs. (1), (5), and (6). As the unit of time is indefinitely increased [i.e., as more and more of the time extension of $V(t)$ is compressed into the unit time interval], both $S(\omega)$ and $R(\tau)$ approach

a limit, and these limit functions $S(\omega)$ and $R(\tau)$ are regarded as the spectrum and correlation function of $V(t)$, respectively.

If the r-f amplifier is tuned to a band-center frequency ω_0 and possesses, from input to output, a complex voltage gain $G(j\omega,\omega_0)$, then the output voltage, $V_0(t,\omega_0)$, due to thermal noise in the input circuit, will possess the Fourier integral representation

$$V_0(t,\omega_0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} G(j\omega,\omega_0)v(j\omega)e^{j\omega t}\, d\omega \tag{8}$$

with a Fourier transform equal to

$$G(j\omega,\omega_0)v(j\omega) = \int_{-\infty}^{+\infty} V_o(t,\omega_0)e^{j\omega t}\, dt. \tag{9}$$

$V_o(t,\omega_o)$ will not vanish outside exactly the same time interval as does $V(t)$, except in the limit as the unit of time is indefinitely increased (in the sense of the preceding paragraph). But in this limiting case, $\overline{V_o}(t,\omega_0)$ will exist only within the same unit interval of time in which $V(t)$ exists and will vanish outside this time interval, so that it, too, may be assumed to be quadratically integrable or, in other words, that

$$\int_{-\infty}^{+\infty} V_0^2(t,\omega_0)\, dt = \int_{-\frac{1}{2}}^{+\frac{1}{2}} V_0^2(t,\omega_0)\, dt = V_0^2(\omega_o) \tag{10}$$

is finite and equal to the mean squared value of $V_o(t,\omega_o)$. This result is concordant with the assumption that the Fourier integral representation (8) and the Fourier transform (9) of $V_o(t,\omega_o)$ both exist.

Now the spectrum of $V_o(t,\omega_o)$ will be

$$S_o(\omega,\omega_o) = K(\omega,\omega_o)S(\omega), \tag{11}$$

where

$$K(\omega,\omega_o) = |G(k\omega,\omega_o)|^2 \tag{12}$$

so that, in analogy with (7),

$$\overline{V_0^2(\omega_o)} = \int_{-\infty}^{+\infty} V_0^2(t,\omega_o)\, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_o(\omega,\omega_o)\, d\omega \tag{13}$$

or

$$\overline{V_0^2(\omega_o)} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} K(\omega,\omega_o)S(\omega)\, d\omega. \tag{14}$$

This last equation may be regarded as an integral equation to be solved for the unknown spectrum function $S(\omega)$, in which the kernel $K(\omega,\omega_o)$ is the squared magnitude of the amplifier voltage gain function and the function $\overline{V_0^2(\omega_o)}$ is the mean squared output voltage due to thermal noise in the input circuit—both of the latter functions being experimentally measurable, as is shown below.

The measurement of the kernel $K(\omega,\omega_o)$ may be carried out by apply-

ing across the input terminals of the r-f amplifier a known voltage of frequency ω from a standard signal generator. Care must be taken that the output impedance of the signal generator is so compensated that the net impedance terminating the amplifier input terminals is identical with that of the original parallel resonant input circuit of Fig. 12·1. Then, with this known input voltage adjusted to a magnitude well above the noise level, the mean squared output voltage response to this signal may be directly measured by the thermocouple meter M (with the possible insertion of a frequency-independent attenuator of known attenuation). Thus the ratio of this mean squared output voltage to the mean squared input signal voltage provides the value of $K(\omega,\omega_0)$ for the signal frequency ω and the amplifier tuned frequency ω_0, and $K(\omega,\omega_o)$ can be measured as a function of both of its arguments ω and ω_o, which can be varied at will.

The measurement of $\overline{V_o^2(\omega_o)}$ may also be carried out in the following way. With the original arrangement of Fig. 12·1, let $M_o(\omega_o)$ be the reading of the thermocouple meter M when the parallel resonant input circuit is open-circuited, that is, when it is in the condition shown in Fig. 12·1. Then $M_o(\omega_o)$ will be equal to the mean squared output voltage due to all sources of noise within the amplifier. On the other hand, let $M_s(\omega_o)$ be the reading of the thermocouple meter M when the parallel resonant input circuit is short-circuited. Then $M_s(\omega_o)$ will be equal to the mean squared output voltage due to all sources of noise within the amplifier except that due to thermal noise in the input circuit. Since the thermal noise in the input circuit is uncorrelated with any other source of noise, it will add to the output with any other source of noise in the mean square; and since it has been assumed that the output response to all other sources of noise is independent of the input terminating impedance of the amplifier,

$$\overline{V_o^2(\omega_o)} = M_o(\omega_o) - M_s(\omega_o). \tag{15}$$

That is to say, $\overline{V_o^2(\omega_o)}$ is directly measurable as the difference in readings of the thermocouple output meter M, without and with the r-f input terminals short-circuited, respectively. Thus $\overline{V_o^2(\omega_o)}$ may be measured as a function of its argument ω_o, which can be varied at will.

Consider now what happens to the integral Eq. (14) as the bandwidth of the r-f amplifier is narrowed indefinitely. The spectrum $S(\omega)$ is unaffected in this process, but the kernel $K(\omega,\omega_o)$ becomes indefinitely narrower, becoming in the limit proportional to a delta function. More precisely, in the limit,

$$\frac{K(\omega,\omega_o)}{\int_{-\infty}^{+\infty} K(x,\omega_o)\,dx} = \delta(\omega - \omega_0), \tag{16}$$

where

$$\delta(\omega - \omega_o) = 0 \qquad \text{if } \omega \neq \omega_o \qquad (17a)$$
$$= \infty \qquad \text{if } \omega = \omega_o, \qquad (17b)$$

$$\int_{-\infty}^{+\infty} \delta(\omega - \omega_o) \, d\omega = 1, \qquad (17c)$$

$$\int_{-\infty}^{+\infty} \delta(\omega - \omega_o) f(\omega) \, d\omega = f(\omega_o) \qquad (17d)$$

if $f(\omega)$ is any *continuous* function of ω.   At the same time, the observed mean squared thermal-noise output voltage $\overline{V_o^2(\omega_o)}$ undergoes a change, approaching in the limit

$$\overline{V_o^2(\omega_o)} = \left(\frac{1}{2\pi}\right) \int_{-\infty}^{+\infty} K(x,\omega_o) \, dx \int_{-\infty}^{+\infty} \delta(\omega - \omega_o) S(\omega) \, d\omega. \qquad (18)$$

If the unknown spectrum $S(\omega)$ is assumed to be *continuous*, then from Eqs. (17d) and (18) it must be given by

$$\bar{S}(\omega_o) = \frac{2\pi \overline{V_o^2(\omega_o)}}{\displaystyle\int_{-\infty}^{+\infty} K(\omega,\omega_o) \, d\omega}, \qquad (19)$$

where the foregoing limiting condition of an infinitely narrow bandwidth is imposed.   Thus we have found *one* solution of the original integral Eq. (14), which is *continuous*, and experimentally determinable, according to Eq. (19), as the ratio of the measured mean squared thermal-noise output voltage $\overline{V_o^2(\omega_o)}$ to the measured gain-bandwidth integral $(1/2\pi) \int_{-\infty}^{+\infty} K(\omega,\omega_o) \, d\omega$, in the limit when the amplifier bandwidth is infinitely narrow.   In practice, this continuous spectrum $\bar{S}(\omega_o)$ is well approximated even when the experimentally observed bandwidth, though relatively very narrow, is still finite.   Thus, it is possible to measure the continuous spectrum $\bar{S}(\omega_o)$ as a function of its argument ω_o, which may be varied at will.

When so measured, for different values of R, L, C in Fig. 12·1 and for different absolute equilibrium temperatures T, the continuous spectrum $\bar{S}(\omega)$ will be found equal to

$$S(\omega) = \frac{2kTR}{1 + R^2 \left(\omega C - \dfrac{1}{\omega L}\right)^2} \qquad (20)$$

where k is Boltzmann's constant, 1.38×10^{-23} joules/°K.   Now, by Eqs. (2) and (5), such a spectrum corresponds to a single voltage pulse $U(t)$, initiated at the time $t = 0$, and behaving thereafter in accordance with the relation

$$U(t) = (2kTR)^{1/2}(2\pi)^{-1} \int_{-\infty}^{+\infty} \frac{e^{j\omega t}\, d\omega}{1 + R\left(j\omega C + \dfrac{1}{j\omega L}\right)}$$

$$= (2kTR)^{1/2}(RC)^{-1}e^{-t/2RC}[\cos \omega_c t - (2RC\omega_c)^{-1} \sin \omega_c t], \quad (21)$$

where

$$\omega_c^2 = \frac{1}{LC} - \frac{1}{(2RC)^2}. \quad (22)$$

But $U(t)$ is evidently not the actual thermal-noise voltage $V(t)$, because $V(t)$, if observed, would appear more like a sequence of such pulses $U(t)$ distributed at random in time and in amplitude.

Indeed, the spectrum of $V(t)$ is quite another solution of the integral Eq. (14). It is a function $S(\omega) \neq \bar{S}(\omega)$, *everywhere discontinuous*, for which we can say only that

$$\int_{-\infty}^{+\infty} \delta(\omega - \omega_o)S(\omega)\, d\pi = \bar{S}(\omega_o). \quad (23)$$

Because the number of such solutions $S(\omega)$ is infinite, neither the spectrum $S(\omega)$ nor $V(t)$ itself can be determined solely from the integral equation (14) alone.

Apart from its relation to the elementary pulse $U(t)$, however, the measured spectrum $\bar{S}(\omega)$ may be interpreted as the statistical spectrum of $V(t)$. Experimentally it is the frequency average

$$\bar{S}(\omega_0) = \int_{-\infty}^{+\infty} \delta(\omega - \omega_o)S(\omega)\, d\omega, \quad (24)$$

of the spectrum $S(\omega)$ of $V(t)$, taken over an infinitely narrow frequency band. From another viewpoint, it is effectively an ensemble average,[1] at one frequency ω_0, of a great number of spectra $S(\omega)$, corresponding to a great number of samples of $V(t)$. Finally, according to the equation [see Eq. (7)]

$$\overline{V^2} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(\omega)\, d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(\omega)\, d\omega, \quad (25)$$

it can be spoken of as the mean square of the thermal-noise voltage $V(t)$ per unit-frequency interval and can be written

$$\bar{S}(\omega) = \frac{d\overline{V^2}}{df} \qquad \left(df = \frac{d\omega}{2\pi}\right). \quad (26)$$

The result of the foregoing measurements, then, is that the thermal-

[1] An ensemble average of a set of functions $f_n(x)$, (where $n = 1, 2, \cdots, N$) taken at any point x_0 is defined by the expression $\overline{f(x_0)} = (1/N) \displaystyle\sum_n f_n(x_0)$.

noise voltage $V(t)$, appearing across the open-circuited parallel resonant circuit of Fig. 13·1, has a statistical spectrum

$$\frac{d\overline{V^2}}{df} = \bar{S}(\omega) = \overline{|v(j\omega)|^2} = \frac{2kTR}{\left[1 + R^2\left(\omega C - \frac{1}{\omega L}\right)^2\right]} \tag{27}$$

and a total mean square

$$\overline{V^2} = 2kTR \int_{-\infty}^{+\infty} \frac{df}{1 + R^2\left(\omega C - \frac{1}{\omega L}\right)^2} \tag{28}$$

$$= \frac{kT}{C}$$

Localization of Thermal-noise Emf.—In the parallel resonant circuit of Fig. 12·1, the source of thermal noise, or the thermal-noise emf, may be localized by extrapolating the measured statistical spectrum, Eq. (27), and mean square, Eq. (28), of $V(t)$ to two limiting cases: (1) an isolated resistance R (assuming L infinite and C zero) and (2) an isolated LC-circuit (assuming R infinite).

In the first case, $V(t)$ becomes $V_R(t)$, the open-circuit thermal-noise voltage of R, with a statistical spectrum [see Eq. (27)].

$$\frac{d\overline{V_R^2}}{df} = \overline{|v_R(j\omega)|^2} = 2kTR. \tag{29}$$

But then, according to Thévenin's theorem,[1] R must contain an internal series thermal-noise emf

$$E_R(t) = V_R(t), \tag{30}$$

independent of the termination of R, with a statistical spectrum

$$\frac{d\overline{E_R^2}}{df} = \overline{|e_R(j\omega)|^2} = 2kTR \tag{31}$$

and a total mean square $\overline{E_R^2}$ which is infinite. This result, Eq. (31), is evidently applicable to any resistance R occurring in any physical circuit configuration as long as R itself is in thermal equilibrium. The formula, Eq. (31), discovered by Johnson[2] in 1928, is known as Johnson's "thermal-noise formula."

In the second case, $V(t)$ becomes $V_{LC}(t)$, the open-circuit thermal-noise voltage of the LC-circuit, with a total mean square [*cf.* Eq. (28)]

$$\overline{V_{LC}^2} = \frac{kT}{C} \tag{32}$$

[1] For a statement of Thévenin's theorem, see, for example, Everitt, *Communication Engineering*, McGraw-Hill, New York, 1937, pp. 47–49.

[2] J. B. Johnson, *Phys. Rev.*, **32**, 97 (1928).

and a statistical spectrum [*cf.* Eqs. (27) and (28)]

$$\frac{d\overline{V_{Lc}^2}}{df} = \overline{|v_{Lc}(j\omega)|^2} = \frac{kT}{2C}\,\delta(f + f_c) \qquad \left(f_c = \frac{\omega_c}{2\pi}\right) \tag{33}$$

which vanishes everywhere but at the resonant frequency

$$\omega_c = (LC)^{-\frac{1}{2}}. \tag{34}$$

Application of Thévenin's theorem to this case produces indeterminate results because an *a priori* choice exists between a series emf on the one hand and a circulating current in the LC-circuit on the other, or certain combinations of the two. Fundamentally, however, there can be no emf in the LC-circuit, because such a circuit constitutes a conservative system, and the observed open-circuit voltage must be construed as a true potential difference arising exclusively from the flow of a circulating current. Hence, $V_{Lc}(t)$ must be regarded as a sinusoidal potential difference

$$V_{Lc}(t) = \frac{LdI_L(t)}{dt} = \frac{Q_c(t)}{C} = V_0 \cos \omega_c(t - t_0), \tag{35}$$

developed by a vestigial thermal-noise current

$$I_L(t) = -\frac{dQ_c(t)}{dt} = I_0 \sin \omega(t - t_0) \tag{36}$$

of arbitrary amplitude

$$I_0 = \frac{V_0}{\omega_c L} \tag{37}$$

permanently circulating in the LC-circuit, as shown in Fig. 12·2.

It is meaningless, of course, to speak of the temperature of this circuit or of the statistical spectrum of its open-circuit voltage $V_{Lc}(t)$, because $V_{Lc}(t)$ has a mean square
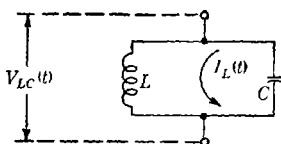
Fig. 12·2.—Isolated LC circuit.

$$\overline{V_{Lc}^2} = \frac{V_0^2}{2} = \frac{(\omega_c L I_0)^2}{2} \tag{38}$$

which is quite arbitrary. But in the sense of an ensemble of such circuits, successively disconnected from a given resistance R in thermal equilibrium, the quantity $V_0^2/2$ has an ensemble average that is

$$\frac{\overline{V_0^2}}{2} = \frac{kT}{C}, \tag{39}$$

where T is the temperature of R. Only in this sense does a definite meaning attach to the statement that $V_{Lc}(t)$ has a statistical spectrum which vanishes everywhere but at the resonant frequency ω_c and yet

has a mean square

$$\overline{V_{Lc}^2} = \frac{kT}{C}. \tag{40}$$

It may be concluded, therefore, with respect to the general parallel resonant circuit of Fig. 12·1, that a thermal-noise emf resides in R, but not in L or C. More generally, with respect to any network that is in thermal equilibrium and contains only R, L, and C elements, it may be concluded that all of the thermal-noise emf's reside in the R elements alone and are given by Eq. (31). It may be concluded further that no such emf resides in any L or C element of such a network.

Still more generally, if elements of mutual inductance M are included in any such network, it is still possible to conclude that all of the thermal-noise emf's reside in the R elements alone. The reason for this is that if these dissipative R elements are removed, the resulting network again constitutes a conservative system in which the observed thermal-noise voltages are true potential differences arising exclusively from the flow of circulating currents. These vestigial thermal-noise currents circulate permanently in the network, appearing as combinations of sinusoids of arbitrary amplitude, with certain discrete frequencies determined by the L, C, and M parameters of the network.

To summarize, then, in any network composed exclusively of R, L, C, and M elements, which is in thermal equilibrium, the resistances R and not the reactive elements L, C, and M are the seats of thermal-noise emf. Each thermal-noise emf is determinable, from the value of its associated resistance R, in terms of Eq. (31).

Harmonic Analysis of Thermal Noise.—In any network composed exclusively of R, L, C, and M elements, whose structure in terms of these elements is completely known and which is in thermal equilibrium, it is possible to write Kirchhoff's equations once the emf's of thermal noise have been localized and evaluated. In the following discussion, the

Fig. 12·3.—Thermal noise variables in *RLC* circuit.

parallel resonant circuit is taken as a representative example. The various thermal-noise variables are defined by Fig. 12·3.

If Fourier integral representations are assumed for all of the thermal noise variables, thus

$$X(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} x(j\omega)e^{-i\omega t}\, d\omega \tag{41a}$$

$$x(j\omega) = \int_{-\infty}^{+\infty} X(t)e^{-i\omega t}\, dt, \tag{41b}$$

where

$$X(t) = E_R(t),\ V(t),\ I_R(t),\ I_L(t),\ I_C(t),\ Q_C(t) \qquad (42a)$$

$$x(j\omega) = e_R(j\omega),\ v(j\omega),\ i_R(j\omega),\ i_L(j\omega),\ i_C(j\omega),\ q_C(j\omega) \qquad (42b)$$

then Kirchhoff's circuit equations can be written in either of two forms. The first form is as differential equations in the time domain, thus

$$V(t) = E_R(t) - R I_R(t) = \frac{L\,dI_L(t)}{dt} = \frac{Q_C(t)}{C}, \qquad (43a)$$

$$I_R(t) = I_L(t) + I_C(t), \qquad (43b)$$

$$I_C(t) = \frac{dQ_C(t)}{dt}. \qquad (43c)$$

Alternatively, they may be written as algebraic equations in the frequency domain, thus

$$v(j\omega) = e_R(j\omega) - R i_R(j\omega) = j\omega L i_L(j\omega) = \frac{q_C(j\omega)}{C}, \qquad (44a)$$

$$i_R(j\omega) = i_L(j\omega) + i_C(j\omega), \qquad (44b)$$

$$i_C(j\omega) = j\omega q_C(j\omega). \qquad (44c)$$

Now, Eqs. (44) are susceptible of analysis by the usual methods of a-c network theory. Thus, by manipulation of Eqs. (44), it is found that

$$\frac{e_R(j\omega)}{1 + R\left(j\omega C + \dfrac{1}{j\omega L}\right)} = v(j\omega) = \frac{i_R(j\omega)}{\left(j\omega C + \dfrac{1}{j\omega L}\right)} = j\omega L i_L(j\omega)$$
$$= \frac{i_C(j\omega)}{j\omega C} = \frac{q_C(j\omega)}{C}. \qquad (45)$$

These equations suffice to determine the statistical spectra of all the thermal-noise variables. For if, in accordance with Johnson's formula, Eq. (31), we write

$$\frac{\overline{dE_R^2}}{df} = \overline{|e_R(j\omega)|^2} = 2kTR \qquad (46)$$

then, from Eq. (2·45), it follows that

$$\frac{\overline{dV^2}}{df} = \overline{|v(j\omega)|^2} = \frac{2kTR}{1 + R^2\left(\omega C - \dfrac{1}{\omega L}\right)^2} \qquad (47a)$$

$$\frac{\overline{dI_R^2}}{df} = \overline{|i_R(j\omega)|^2} = \frac{2kTR\left(\omega C - \dfrac{1}{\omega L}\right)^2}{1 + R^2\left(\omega C - \dfrac{1}{\omega L}\right)^2} \qquad (47b)$$

$$\frac{\overline{dI_L^2}}{df} = \overline{|i_L(j\omega)|^2} = \frac{\dfrac{2kTR}{\omega^2 L^2}}{1 + R^2 \left(\omega C - \dfrac{1}{\omega L}\right)^2} \tag{47c}$$

$$\frac{\overline{dI_C^2}}{df} = \overline{|i_C(j\omega)|^2} = \frac{2kTR\omega^2 C^2}{1 + R^2 \left(\omega C - \dfrac{1}{\omega L}\right)^2} \tag{47d}$$

$$\frac{\overline{dQ_C^2}}{df} = \overline{|q_C(j\omega)|^2} = \frac{2kTRC^2}{1 + R^2 \left(\omega C - \dfrac{1}{\omega L}\right)^2}. \tag{47e}$$

It will be observed that Eq. (47a), the statistical spectrum of $V(t)$, agrees with the measured spectrum, Eq. (27), as it should.

Further, from Eqs. (46) and (47), we are led to the following mean square values of the thermal-noise variables:

$$\overline{E_R^2} = \int_{-\infty}^{+\infty} \overline{|e_R(j\omega)|^2} \, df = \infty, \tag{48}$$

and

$$\overline{V^2} = \int_{-\infty}^{+\infty} \overline{|v(j\omega)|^2} \, df = \frac{kT}{C}, \tag{49a}$$

$$\overline{I_R^2} = \int_{-\infty}^{+\infty} \overline{|i_R(j\omega)|^2} \, df = \infty, \tag{49b}$$

$$\overline{I_L^2} = \int_{-\infty}^{+\infty} \overline{|i_L(j\omega)|^2} \, df = \frac{kT}{L}, \tag{49c}$$

$$\overline{I_C^2} = \int_{-\infty}^{+\infty} \overline{|i_C(j\omega)|^2} \, df = \infty, \tag{49d}$$

$$\overline{Q_C^2} = \int_{-\infty}^{+\infty} \overline{|q_C(j\omega)|^2} \, df = kTC. \tag{49e}$$

In the present example, as seen from Eqs. (48), (49b), and (49d), there are three thermal-noise variables whose mean square values are infinite. They are $E_R(t)$, $I_R(t)$, and $I_C(t)$. Because these three variables are not quadratically integrable, they do not conform strictly to the conditions required to possess Fourier integral representations. This difficulty is due to a slightly incorrect extrapolation of the experimental formula (27) toward infinite frequencies. It will be shown below that the quantum theory of thermal noise requires that the spectrum of any thermal-noise variable shall fall off at very high frequencies and shall vanish at infinite frequency. The quantum theory, therefore, requires that every thermal-noise variable shall have a finite mean square value and hence shall possess a properly behaving Fourier integral representa-

tion. For practical purposes, however, the statistical spectra and the mean squares given by the present methods (whenever the latter are finite) are accurate enough.

The discussion of the statistical spectra and time averages of the thermal-noise variables in Fig. 12·3 has not been as rigorous as it should be. A more careful consideration of them follows here. From the differential equations (43), it can be concluded that all six thermal-noise variables $X(t)$ [cf. Eq. (42a)] are causally related in some way, as yet not adequately described. On the assumption that these thermal-noise variables all possess Fourier integral representations, their causal connections have been expressed by Eqs. (45). These equations show that the Fourier transforms $x(j\omega)$ [cf. Eq. (42b)] of all the thermal-noise variables are expressible as functions of any one Fourier transform, say $e_R(j\omega)$. Thus, the general expression

$$x(j\omega) = \phi_x(j\omega)e_R(j\omega) \tag{50}$$

can be written, where

$$\phi_x(j\omega) = \frac{\partial x(j\omega)}{\partial e_R(j\omega)} \tag{51}$$

is independent of $e_R(j\omega)$. Hence, from Eq. (41a),

$$X(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\partial x(j\omega)}{\partial e_R(j\omega)} e_R(j\omega)e^{j\omega t}\, d\omega \tag{52}$$

can be written. But, from Eq. (41b), the relation

$$e_R(j\omega) = \int_{-\infty}^{+\infty} E_R(\tau)e^{-j\omega\tau}\, d\tau \tag{53}$$

is clear, so that

$$X(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} E_R(\tau)\, d\tau \int_{-\infty}^{+\infty} \frac{\partial x(j\omega)}{\partial e_R(j\omega)} e^{j\omega(t-\tau)}\, d\omega, \tag{54}$$

or

$$X(t) = \int_{-\infty}^{+\infty} X'[E_R(t),\, (t-\tau)]\, E_R(\tau)\, d\tau, \tag{55}$$

where

$$X'[E_R(t),\, (t-\tau)] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\partial x(j\omega)}{\partial e_R(j\omega)} e^{j\omega(t-\tau)}\, d\omega \tag{56}$$

is the "functional derivative" of $X(t)$ with respect to $E_R(t)$. That is to say, all the thermal-noise variables are expressible as "functionals" of any one thermal-noise variable, say $E_R(t)$, in accordance with Eq. (55). The functional derivative $X'[E_R(t),\, (t-\tau)]$, when it refers to a current variable, is related to what is frequently called an "indicial admittance."

The physical meaning of the functional derivative, Eq. (56), is readily

seen from Eq. (55), for if

$$E_R(t) = \delta(t - t_0),\tag{57}$$

then from Eq. (55) it is found that

$$X(t) = X'[E_R(t), (t - t_0)].\tag{58}$$

In other words, the functional derivative of $X(t)$ with respect to $E_R(t)$ is the response $X(t)$ to a unit impulse function $E_R(t)$. In general, it is a pulse initiated at the time t_0 [or the time τ in Eq. (55)], which damps out at a certain rate in the future time $t > t_0$. For example, it is found that for $t \geqq \tau$

$$v'[E_R(t), (t - \tau)] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{j\omega(t-\tau)}\, d\omega}{1 + R\left(j\omega C + \dfrac{1}{j\omega L}\right)}$$

$$= (RC)^{-1} e^{-\frac{(t-\tau)}{2RC}} [\cos \omega_c(t - \tau) + (2RC\omega_c)^{-1} \sin \omega_c(t - \tau),\tag{59}$$

where

$$\omega_c^2 = \frac{1}{LC - 1/(2RC)^2}\tag{60}$$

and similar transients can be found for the functional derivatives of the other variables $X(t)$, except that, in the cases of $E_R(t)$, $I_R(t)$, and $I_C(t)$, the functional derivatives contain an additional part that is proportional to a delta function $\delta(t - \tau)$. But, for the rest, all the functional derivatives behave as Eq. (59), with a damping constant equal to $2RC$.

Assuming that $E_R(t)$, the actual thermal-noise emf, exists only within the unit-time interval $(-\frac{1}{2} \leqq t \leqq +\frac{1}{2})$ and vanishes outside it, it is possible to write

$$X(t) = \int_{-\frac{1}{2}}^{+\frac{1}{2}} X'[E_R(t), (t - \tau)]E_R(\tau)\, d\tau\tag{61}$$

and conclude that $x(t)$ vanishes for $t < -\frac{1}{2}$ and damps out with a time constant $2RC$ for $t > +\frac{1}{2}$. But now, as the unit time interval is made indefinitely large compared with $2RC$, the tails of the pulses of the type described by Eq. (59) and accordingly the tail of $X(t)$ itself for $t > +\frac{1}{2}$ become infinitely short in time duration. In the limit, this leads to the result that all of the thermal-noise variables $X(t)$ vanish outside the unit time interval $(-\frac{1}{2} \leqq t \leqq +\frac{1}{2})$. Accordingly, the mean square value of any variable $X(t)$ may be reckoned with respect to the foregoing unit time interval and may be written in the following alternative forms:

$$\overline{X^2} = \int_{-\frac{1}{2}}^{+\frac{1}{2}} X^2(t)\, dt = \int_{-\infty}^{+\infty} X^2(t)\, dt.\tag{62}$$

But since, in general,

$$\int_{-\infty}^{+\infty} X^2(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |x(j\omega)|^2\, d\omega,\tag{63}$$

it is also possible to write

$$\overline{X^2} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |x(j\omega)|^2 \, d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \overline{|x(j\omega)|^2} \, d\omega, \tag{64}$$

a general result that expresses the mean square of a variable $X(t)$ as an integral taken over its true spectrum $|x(j\omega)|^2$ or, equally well, over its statistical spectrum $\overline{|x(j\omega)|^2}$. Similarly, it is possible to speak of the statistical spectrum as the mean square value of the variable per unit-frequency interval, in accordance with the relation

$$\frac{d\overline{X^2}}{df} = \overline{|x(j\omega)|^2}. \tag{65}$$

Now that the results of Eqs. (64) and (65) have been established on a somewhat more rigorous basis, Eqs. (46) to (49), which were insufficiently explained above, can be better justified and also better understood.

Another interesting result of the analysis just dealt with above is that the statistical spectrum of any thermal-noise variable $X(t)$, which is a functional of the thermal-noise emf $E_R(t)$, is proportional to the true spectrum of the functional derivative $X'[E_R(t), (t - \tau)]$; provided that the statistical spectrum of $E_R(t)$ is a constant (as it is in the present case, according to Johnson's formula). This may be readily confirmed by comparing Eqs. (56) and (47). If the meaning of the functional derivative is kept in mind, it is possible to recognize in this the important result that the statistical spectrum of any thermal-noise variable $X(t)$, which is a functional of $E_R(t)$, is identical, to within a constant, with the true spectrum of the response function $X(t)$ to a unit impulse function $E_R(t)$.

Still another interesting result of the foregoing analysis is that if

$$E_R(t) = \sum_{-\infty}^{+\infty} A_n \delta(t - t_n), \tag{66}$$

then, by Eq. (55),

$$X(t) = \sum_{-\infty}^{+\infty} A_n X'[E_R(t), (t - t_n)], \tag{67}$$

or, in other words, a transient pulse expansion for any thermal noise variable $X(t)$ is found.

Electromagnetic Energy of Thermal Noise.—If the differential equations (43a) are multiplied by $I_R(t)$, the result is

$$V(t)I_R(t) = E_R(t)I_R(t) - RI_R^2(t) = I_R(t) \, L \frac{d}{dt} I_L(t) = \frac{I_R(t)Q_C(t)}{C}. \tag{68}$$

With the aid of Eq. (43b), $I_R(t)$ can be eliminated from the last two expres-

sions above, with the result that

$$V(t)I_R(t) = \frac{d}{dt}\frac{LI_L^2(t)}{2} + LI_c(t)\frac{d}{dt}I_L(t) \tag{69a}$$

$$= \frac{d}{dt}\frac{Q_c^2(t)}{2C} + \frac{I_L(t)Q_c(t)}{C}. \tag{69b}$$

Also,

$$\begin{aligned}
V(t)I_R(t) &= V(t)[I_L(t) + I_c(t)] \\
&= I_L(t)L\frac{d}{dt}I_L(t) + \frac{I_c(t)Q_c(t)}{C} \\
&= \frac{d}{dt}\left[\frac{LI_L^2(t)}{2} + \frac{Q_c^2(t)}{2C}\right].
\end{aligned} \tag{70}$$

The results of Eqs. (69) and (70) are equivalent, for it will be observed that since by Eq. (43a)

$$\frac{Q_c(t)}{C} = L\frac{d}{dt}I_L(t), \tag{71}$$

it follows that

$$\frac{d}{dt}\frac{LI_L^2(t)}{2} = \frac{I_L(t)Q_c(t)}{C} \tag{72a}$$

$$\frac{d}{dt}\frac{Q_c^2(t)}{2C} = LI_c(t)\frac{d}{dt}I_L(t). \tag{72b}$$

Now the result to be emphasized here is that it is possible to write

$$V(t)I_R(t) = E_R(t)I_R(t) - RI_R^2(t) = \frac{d}{dt}\left[\frac{LI_L^2(t)}{2} + \frac{Q_c^2(t)}{2C}\right]. \tag{73}$$

This result is an expression of Poynting's theorem, as it applies to the electromagnetic energy of thermal noise in the parallel resonant circuit of Fig. 13·3. Here it is seen that the circuit elements L and C are the *seat* of thermal-noise energy, where it is stored in the two forms of electromagnetic free energy

$$F_L(t) = \frac{LI_L^2(t)}{2}, \tag{74a}$$

$$F_c(t) = \frac{Q_c^2(t)}{2C}. \tag{74b}$$

It is seen, further, that the emf $E_R(t)$ is the source of this energy, producing it at a rate equal to $E_R(t)I_R(t)$, and that the circuit element R is the sink of this energy, dissipating it at a rate of $RI_R^2(t)$.

Similarly, in any circuit or portion of a circuit in thermal equilibrium, the *seat* of thermal-noise energy in electromagnetic form is the electromagnetic field associated with its positive, frequency-independent elements of inductance and capacitance; the *sinks* of this energy are its

positive, frequency-independent elements of resistance; and the *sources* of this energy are the thermal-noise emf's internal to and in series with these elements of resistance.

The time average of $V(t)I_R(t)$ is considered below:

$$\overline{V(t)I_R(t)} = \int_{-\frac{1}{2}}^{+\frac{1}{2}} V(t)I_R(t)\, dt = \int_{-\infty}^{+\infty} V(t)I_R(t)\, dt. \qquad (75)$$

By Parseval's theorem, this may be written

$$\overline{V(t)I_R(t)} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} v(j\omega)i_R(-j\omega)\, d\omega. \qquad (76)$$

But by Eq. (45) it may also be written

$$\overline{V(t)I_R(t)} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |e_R(j\omega)|^2 \frac{\left(j\omega C + \dfrac{1}{j\omega L}\right) d\omega}{1 + R^2\left(\omega C - \dfrac{1}{\omega L}\right)^2}, \qquad (77)$$

or again, since $|e_R(j\omega)|^2$ may be replaced in the integrand by $\overline{|e_R(j\omega)|^2}$,

$$\overline{V(t)I_R(t)} = \frac{kTR}{\pi} \int_{-\infty}^{+\infty} \frac{\left(j\omega C + \dfrac{1}{j\omega L}\right) d\omega}{1 + R^2\left(\omega C - \dfrac{1}{\omega L}\right)^2}. \qquad (78)$$

Now this integral must vanish, inasmuch as its integrand is an odd function of ω. Therefore, we have the result that

$$\overline{V(t)I_R(t)} = 0. \qquad (79)$$

According to Eq. (73), this means that, on the average, the rates of production and dissipation of electromagnetic energy of thermal noise are equal, that is,

$$\overline{E_R(t)I_R(t)} = \overline{RI_R^2(t)}, \qquad (80)$$

or in other words, that the average net rate of change of electromagnetic energy of thermal noise is zero.

$$\overline{\frac{d}{dt}[F_L(t) + F_C(t)]} = 0. \qquad (81)$$

It will be observed that these results conform with the condition that there be thermal equilibrium.

Next it is important to find the average rate of production (and the equal rate of dissipation) of the electromagnetic energy of thermal noise. From Eqs. (80) and (49b), we see that this average rate is infinite, that is,

$$\overline{RI^2(t)} = \infty. \qquad (82)$$

On the other hand, from Eq. (47b), it is clear that the average rate per unit-frequency interval is finite and equal to

$$\frac{d}{df}[\overline{RI_R^2(t)}] = \frac{2kTR^2\left(\omega C - \frac{1}{\omega L}\right)^2}{1 + R^2\left(\omega C - \frac{1}{\omega L}\right)^2} = 2kT\left[1 - \frac{1}{1 + R^2\left(\omega C - \frac{1}{\omega L}\right)^2}\right].$$

(83)

But, as was said above, the quantum theory of thermal noise requires that the average rate of production [Eq. (82)] be finite and that the average rate per unit-frequency interval [Eq. (83)] fall off to zero at infinite frequency. Equation (82), therefore, cannot be taken as correct, and Eq. (83) can be used only at frequencies that are not extremely high.

It was seen in Eq. (81) that the average rate of change of electromagnetic energy of thermal noise is zero. The average value of this electromagnetic energy is given by Eq. (49), from which it is found that

$$\overline{F_L(t)} = \frac{\overline{LI_L^2(t)}}{2} = \frac{kT}{2},$$

(84a)

$$\overline{F_C(t)} = \frac{\overline{Q_C^2(t)}}{2C} = \frac{kT}{2}.$$

(84b)

These results, Eqs. (84), are of general validity, in the sense that they are not restricted to the parallel resonant circuit of Fig. 12·3. They imply that in any inductance L or capacitance C, imbedded in a network in thermal equilibrium but not mutually coupled through its electromagnetic field with other elements, there resides an average amount of electromagnetic energy of thermal noise equal to $kT/2$. This general result was first deduced, theoretically, by DeHaas-Lorentz in 1913.

Classical Statistical Mechanics of Thermal Noise.—A discussion of the parallel resonant circuit of Fig. 12·3 from a thermodynamic viewpoint follows. The material components of this system may be regarded as an essentially loss-free metallic coil of inductance L, an essentially loss-free metallic condenser of capacitance C, and a dissipative element of resistance R, all of which are positive and linear, at least for small signals, and of negligible frequency variation over a reasonably large range of frequencies. The dissipative element R may be a metallic or electrolytic conductor, a dielectric or hysteresis loss, or a radiation resistance, and so on, provided it behaves in accordance with the conditions prescribed above.

It is assumed that the circuit is in thermal equilibrium and isolated. Alternatively, it might be assumed to be in thermal equilibrium with its surroundings. The former condition was chosen, so that the total energy U of the system would be finite and constant.

From a thermodynamic viewpoint, the total quantity of mechanical heat energy Q that resides in the molecular motions of the material of the whole circuit must be considered as a part of the total energy U of the system. The total free energy F residing in the system must also be considered as another part of the total energy U. This free energy usually includes the electromagnetic energy of thermal noise just discussed, as well as whichever of the following energy forms may be present: the mechanical energy of moving parts of the circuit, the electrochemical energy of an electrolyte, the electromagnetic energy of a radiation field, and similar energies. Accordingly, for the complete thermodynamic system, we may write

$$U = Q + F = \text{const.,} \tag{85a}$$

$$T = \frac{Q}{S} = \text{const.,} \tag{85b}$$

where S is the total *entropy* of the system.

The total free energy F of the system being considered can now be written

$$F = F(I_L, Q_c, x, y, z, \cdots, w) \tag{86a}$$

$$= F_L(I_L) + F_c(Q_c) + f(x, y, z, \cdots, w), \tag{86b}$$

where

$$F_L(I_L) = \frac{L I_L^2}{2}, \tag{87a}$$

$$F_c(Q_c) = \frac{Q_c^2}{2C}, \tag{87b}$$

are the contributions due to electromagnetic energy of thermal noise, which are continuous functions of the respective parameters I_L, Q_c, and where $f(x, y, z, \ldots, w)$ represents any other forms of free energy that may reside within the dissipative element R, which are continuous functions of certain other unspecified parameters $x, y, z, \ldots, w$.

A necessary and sufficient condition for *thermodynamic equilibrium*, with U and T constant, is that the entropy S shall be a maximum or, as seen from Eq. (85), that the total free energy F shall be a minimum. But in order that F may assume its minimum value F_o, it is necessary first, as seen from Eqs. (86) and (87), that $I_L = Q_c = 0$ and, second, again as seen from Eqs. (86) and (87), that $x, y, z, \ldots, w$ shall assume critical values $x_o, y_o, z_o, \ldots, w_o$ which make $f(x, y, z, \ldots, w)$ a minimum. Then, in the neighborhood of thermodynamic equilibrium, the deviation of the total free energy F may be written, approximately,

$$F - F_0 = \frac{L I_L^2}{2} + \frac{Q_c^2}{2C} + a(x - x_0)^2 + b(y - y_0)^2 + c(z - z_0)^2$$
$$+ \cdots + p(w - w_0)^2 \tag{88}$$

where a, b, c, ... , p are positive constants and where no cross-product terms exist, because it is assumed that the parameters x, y, z, . . . , w, like I_L and Q_c, are independent.

In actuality, true thermodynamic equilibrium cannot be maintained, because every individual interaction process between forms of heat energy Q and free energy F causes a finite change in the distribution of the total energy U in these two energy forms.   Thermodynamic equilibrium can be maintained on the average, however, and this state is called "statistical equilibrium."   But statistical equilibrium is attended by fluctuations in the heat energy Q and compensatory fluctuations in the free energy F, in view of Eq. (85a).

The fluctuations of the total free energy F, away from its minimum value F_0, are governed by the thermodynamic probability W, which is related to the entropy S of the system by the equation

$$W = W_0 e^{(S-S_0)/k}, \tag{89}$$

where k is Boltzmann's constant.   Although W is fundamentally a function of the entropy S alone, it may be specified completely in terms of the mechanical parameters of the heat energy Q, for by Eqs. (85b) and (89) it can be written

$$W = W_0 e^{(Q-Q_0)/kT}. \tag{90}$$

In this form it expresses the probability for any possible distribution of the heat energy Q.   But, according to Eq. (85a), we can write

$$U = Q + F = Q_0 + F_0 = \text{const.}, \tag{91}$$

and therefore we can rewrite Eq. (90) in the form

$$W = W_0 e^{(F_0-F)/kT}. \tag{92}$$

When written in this form, W expresses the probability for any possible distribution of the total free energy F.

Thus, with the aid of the thermodynamic probability W, the mean deviation $\overline{F - F_o}$ of the total free energy F away from its equilibrium value F_o can be evaluated.   In particular, it is possible to evaluate the mean deviation

$$\overline{F - F_0}^x = \overline{a(x - x_0)^2} \tag{93}$$

of the free energy associated with the fluctuations in any one parameter x on which the total free energy depends.   Thus, the probability that I_L, Q_c, x, y, z, . . . , w lie in the range (I_L, $I_L + dI_L$; $Q_c + dQ_c$; x, $x + dx$; y, $y + dy$; z, $z + dz$; · · · ; w, $w + dw$) is given by

$$W \, dI_L \, dQ_c \, dx \, dy \, dz \, \cdots \, dw, \tag{94}$$

and hence, when integrated over the entire domain of these variables,

$$\int \int \int \int \int \cdots \int W \, dI_L \, dQ_c \, dx \, dy \, dz \cdots dw = 1 \quad (95a)$$

$$\int \int \int \int \int \cdots \int a(x - x_0)^2 W \, dI_L \, dQ_c \, dx \, dy \, dz \cdots dw$$
$$= \overline{a(x - x_0)^2}. \quad (95b)$$

But, from Eqs. (88) and (92),

$$W = W_0 \exp - \left[\frac{L I_L^2}{2} + \frac{Q_c^2}{2C} + a(x - \dot{x}_0)^2 + b(y - y_0)^2 + c(z - z_0)^2 \right.$$
$$\left. + \cdots + \frac{p(w - w_0)^2}{kT} \right]. \quad (96)$$

Hence, dividing Eq. (95b) by Eq. (95a), it is found that

$$\overline{a(x - x_0)^2} = \frac{\int_{-\infty}^{+\infty} a(x - x_0)^2 \, e^{-a(x-x_0)^2/kT} \, dx}{\int_{-\infty}^{+\infty} e^{-a(x-x_0)^2/kT} \, dx} \quad (97)$$

Now, if

$$u = \left(\frac{a}{kT} \right)^{\frac{1}{2}} (x - x_0), \quad (98a)$$

$$du = \left(\frac{a}{kT} \right)^{\frac{1}{2}} dx, \quad (98b)$$

then Eq. (97) becomes

$$\overline{a(x - x_0)^2} = kT \frac{\int_{-\infty}^{+\infty} u^2 e^{-u^2} \, du}{\int_{-\infty}^{+\infty} e^{-u^2} \, du} \quad (99)$$

or

$$\overline{a(x - x_0)^2} = \frac{kT}{2}. \quad (100)$$

Similarly,

$$\frac{\overline{L I_L^2}}{2} = \frac{\overline{Q_c^2}}{2C} = \overline{a(x - x_0)^2} = \overline{b(y - y_0)^2} = \overline{c(z - z_0)^2} = \cdots$$
$$= \overline{p(w - w_0)^2} = \frac{kT}{2}, \quad (101)$$

or, in other words, the mean fluctuation in free energy associated with each independent parameter is equal to $kT/2$. This result is an extension of the theorem of equipartition of energy, which applies to molecular heat motions and states that the mean fluctuation in the kinetic energy of molecular heat motion is $kT/2$ for each distinct degree of freedom.

Comparison of the first two equations (101) with Eqs. (84) shows complete agreement between the former, which are theoretically deduced, and the latter, which are experimentally observed. It must be remem-

bered that Expressions (84) are *time averages*, however, whereas Expressions (101) are *ensemble averages*. But in a physical system of the sort dealt with here, the molecular motions as well as the time variations of the parameters of free energy are so complicated as to yield essentially an *ergodic* system. For such a system time averages and ensemble averages are the same

It is important to observe, also, that the first two equations (101) give the mean fluctuation in electromagnetic free energy of thermal noise and that Eq. (84) gives the mean value of this energy. Both of these terms have the same meaning in the present case, however, because the equilibrium value of the electromagnetic free energy of thermal noise is zero.

Integral Equations for the Spectrum of the Thermal-noise emf.—In the discussion leading to Eqs. (84), the experimental conclusion was reached that in any inductance L or capacitance C in thermal equilibrium, not mutually coupled through its electromagnetic field with other circuit elements, there resides an average amount of electromagnetic energy of thermal noise equal to $kT/2$. In other words, it was experimentally shown that

$$\frac{\overline{LI_L^2(t)}}{2} = \frac{\overline{Q_C^2(t)}}{2C} = \frac{kT}{2}. \tag{102}$$

The same conclusion was reached theoretically in the discussion leading to Eqs. (101). Now, taking this result as a point of departure, let us inquire what follows concerning the spectrum of the thermal-noise emf $E_R(t)$ in a parallel resonant circuit.

First, by virtue of Eqs. (62) and (63), Eq. (102) may be rewritten thus:

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} |i_L(j\omega)|^2 \, d\omega = \frac{kT}{L} \tag{103a}$$

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} |q_C(j\omega)|^2 \, d\omega = kTC. \tag{103b}$$

Further, utilizing the relations of Eq. (45) for a parallel resonant circuit, for such a circuit this becomes

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} |e_R(j\omega)|^2 \frac{d\omega}{\omega^2} \bigg/ \left[1 + R^2 \left(\omega C - \frac{1}{\omega L} \right)^2 \right] = kTL \tag{104a}$$

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} |e_R(j\omega)|^2 \frac{d\omega}{1 + R^2 \left(\omega C - \frac{1}{\omega L} \right)^2} = \frac{kT}{C}, \tag{104b}$$

which constitutes a pair of simultaneous integral equations that are satisfied by the unknown spectrum of the thermal-noise emf $E_R(t)$.

Now if

$$\omega_0^2 LC = 1, \tag{105a}$$

$$\omega_0 RC = Q, \tag{105b}$$

$$\frac{\omega}{\omega_0} = \eta, \tag{105c}$$

then Eqs. (104) simplify to

$$\frac{kTR}{Q} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |e_R(j\omega_0\eta)|^2 \frac{d\eta}{\eta^2} \Big/ \left[1 + Q^2 \left(\eta - \frac{1}{\eta} \right)^2 \right] \tag{106a}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |e_R(j\omega_0\eta)|^2 \frac{d\eta}{1 + Q^2 \left(\eta - \frac{1}{\eta} \right)^2}, \tag{106b}$$

or to

$$\frac{kTR}{Q} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |e_R(j\omega_0\eta)|^2 \frac{d\eta}{1 + Q^2 \left(\eta - \frac{1}{\eta} \right)^2} \tag{107a}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left| e_R \left(\frac{j\omega_0}{\eta} \right) \right|^2 \frac{d\eta}{1 + Q^2 \left(\eta - \frac{1}{\eta} \right)^2}. \tag{107b}$$

In other words, for a parallel resonant circuit, the thermal-noise emf $E_R(t)$ must have a spectrum $|e_R(j\omega)|^2$ that, for $\omega = \omega_0\eta$ *and for* $\omega = \omega_0/\eta$, satisfies the integral equation

$$\frac{kTR}{Q} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |e_R(j\omega)|^2 \frac{d\eta}{1 + Q^2 \left(\eta - \frac{1}{\eta} \right)^2}. \tag{108}$$

Now consider η as the imaginary part of the complex variable

$$\zeta = \xi + j\eta, \tag{109}$$

and reexpress Eq. (107) as contour integrals, taken counterclockwise around the right half-plane of ζ, thus:

$$\frac{kTR}{Q} = \frac{1}{2\pi j} \oint |e_R(\omega_0\zeta)|^2 \frac{d\zeta}{1 - Q^2 \left(\zeta + \frac{1}{\zeta} \right)^2} \tag{110a}$$

$$= \frac{1}{2\pi j} \oint \left| e_R \left(\frac{\omega_0}{\zeta} \right) \right|^2 \frac{d\zeta}{1 - Q^2 \left(\zeta + \frac{1}{\zeta} \right)^2}. \tag{110b}$$

One particular solution $e_R(\omega_0\zeta)$ may be assumed which possesses no poles in the right half-plane of ζ and for which $e_R(\omega_0/\zeta)$ therefore also

possesses no poles in the right half-plane of ζ, since the substitution of $1/\zeta$ for ζ maps the right half-plane of ζ into itself.   In this case, the poles of both integrands of Eqs. (110) in the right half-plane of ζ are the same and occur when

$$Q\left(\zeta + \frac{1}{\zeta}\right) = 1 \tag{111}$$

or when

$$2Q\zeta = 1 + \sqrt{1 - (2Q)^2}, \tag{112}$$

where the pole corresponding to the positive sign in Eq. (112) may be designated by ζ_1 and the pole corresponding to the negative sign by ζ_2. Then

$$\frac{-kTR}{Q} = Res\left[\frac{|e_R(\omega_0\zeta)|^2}{1 - Q^2\left(\zeta + \frac{1}{\zeta}\right)^2}\right]_{\zeta=\zeta_1} + Res\left[\frac{|e_R(\omega_0\zeta)|^2}{1 - Q^2\left(\zeta + \frac{1}{\zeta}\right)^2}\right]_{\zeta=\zeta_2}$$

$$= \frac{\frac{|e_R(\omega_0\zeta_1)|^2\zeta_1^2}{(1 - \zeta_1^2)} + \frac{|e_R(\omega_0\zeta_2)|^2\zeta_2^2}{(1 - \zeta_2^2)}}{2Q}, \tag{113a}$$

and similarly

$$\frac{-kTR}{Q} = Res\left[\frac{\left|e_R\left(\dfrac{\omega_0}{\zeta}\right)\right|^2}{1 - Q^2\left(\zeta + \dfrac{1}{\zeta}\right)^2}\right]_{\zeta=\zeta_1} + Res\left[\frac{\left|e_R\left(\dfrac{\omega_0}{\zeta}\right)\right|^2}{1 - Q^2\left(\zeta + \dfrac{1}{\zeta}\right)^2}\right]_{\zeta=\zeta_2}$$

$$= \frac{\left|e_R\left(\dfrac{\omega_0}{\zeta_1}\right)\right|^2\dfrac{\zeta_1^2}{(1 - \zeta_1^2)} + \left|e_R\left(\dfrac{\omega_0}{\zeta_2}\right)\right|^2\dfrac{\zeta_2^2}{1 - \zeta_2^2}}{2Q}. \tag{113b}$$

Now it will be observed, from Eq. (112), that

$$\zeta_1\zeta_2 = 1. \tag{114}$$

Hence Eq. (113) may also be written thus:

$$-2kTR = \frac{|e_R(\omega_0\zeta_1)|^2\zeta_1^2}{1 - \zeta_1^2} + \frac{\left|e_R\left(\dfrac{\omega_0}{\zeta_1}\right)\right|^2}{\zeta_1^2 - 1} \tag{115a}$$

$$= \frac{\left|e_R\left(\dfrac{\omega_0}{\zeta_1}\right)\right|^2\zeta_1^2}{1 - \zeta_1^2} + \frac{|e_R(\omega_0\zeta_1)|^2}{\zeta_1^2 - 1}, \tag{115b}$$

or

$$2kTR(1 - \zeta_1^2) = \left|e_R\left(\frac{\omega_0}{\zeta_1}\right)\right|^2 - \zeta_1^2|e_R(\omega_0\zeta_1)|^2 \tag{116a}$$

$$= |e_R(\omega_0\zeta_1)|^2 - \zeta_1^2\left|e_R\left(\frac{\omega_0}{\zeta_1}\right)\right|^2, \tag{116b}$$

or

$$\left[\left|e_R\left(\frac{\omega_0}{\zeta_1}\right)\right|^2 - 2kTR\right] - \zeta_1\left[|e_R(\omega_0\zeta_1)|^2 - 2kTR\right] = 0, \qquad (117a)$$

$$[|e_R(\omega_0\zeta_1)|^2 - 2kTR] - \zeta_1^2\left[\left|e_R\left(\frac{\omega_0}{\zeta_1}\right)\right|^2 - 2kTR\right] = 0, \qquad (117b)$$

or, eliminating $|e_R(\omega_0/\zeta_1)|^2$,

$$[|e_R(\omega_0\zeta_1)|^2 - 2kTR] (1 - \zeta_1^4) = 0. \qquad (118)$$

Therefore, regardless of the value of ζ_1 (and including the exceptional values that are roots of $\zeta_1^4 = 1$), it follows that throughout the right half-plane,

$$|e_R(\omega_0\zeta_1)|^2 = \left|e_R\left(\frac{\omega_0}{\zeta_1}\right)\right|^2 = 2kTR. \qquad (119)$$

This leads to the result that if the spectrum $|e_R(\omega_0\zeta)|^2$ of the thermal-noise emf $E_R(t)$ in a parallel resonant circuit is assumed to possess no poles in the right half-plane of ζ, then it must be a constant throughout the right half-plane of ζ and equal to

$$|e_R(\omega_0\zeta)|^2 = 2kTR, \qquad (120)$$

and, in particular, considered as a limit, it must be a constant everywhere along the imaginary axis of the ζ-plane and equal to

$$|e_R(j\omega)|^2 = 2kTR. \qquad (121)$$

This result does not imply, however, that $|e_R(\omega_0\zeta)|^2$ is a constant throughout the *left* half-plane of ζ unless it is further assumed that $e_R(\omega_0\zeta)$ is an analytic function throughout the ζ-plane.

It can be said, therefore, that the only function which satisfies the integral Eqs. (104) and at the same time possesses no simple poles in the right half-plane of ζ is a constant, given by Eq. (120). This solution corresponds to the measured statistical spectrum (31); but at the same time it corresponds to the true spectrum of a single impulsive emf

$$E_R(t) = E_0\delta(t - t_0) \qquad (122)$$

and not to the true spectrum of the thermal-noise emf $E_R(t)$.

There is another way of writing Eqs. (107) to apply to the complex ζ-plane, which is to be preferred to the formulation (110). Equations (107) can be written

$$\frac{kTR}{Q} = \frac{1}{2\pi j} \oint \frac{e_R(\omega_0\zeta)e_R(-\omega_0\zeta)\,d\zeta}{1 - Q^2\left(\zeta + \dfrac{1}{\zeta}\right)^2} \qquad (123a)$$

$$= \frac{1}{2\pi j} \oint \frac{e_R\left(\dfrac{\omega_0}{\zeta}\right) e_R\left(\dfrac{-\omega_0}{\zeta}\right) d\zeta}{1 - Q^2\left(\zeta + \dfrac{1}{\zeta}\right)^2} \qquad (123b)$$

where the integrals are taken around the same contour as before. If, now, a solution $e_R(\omega_0\zeta)$ that possesses no simple poles in the right half-plane of ζ is assumed, it does not follow that $e_R(\omega_0\zeta)e_R(-\omega_0\zeta)$ will also be free of poles in the right half-plane of ζ. For this to be so, it would have to be assumed that $e_R(\omega_0\zeta)$ possesses no simple poles *either* in the right *or* left half-plane of ζ. However, it is still admissible for $e_R(\omega_0\zeta)$ to possess poles along the axis of imaginaries, provided the contour integrals of Eqs. (123) do not enclose any of these poles. In this case, Eqs. (123) lead, in complete analogy with the preceding analysis, to the result that

$$e_R(\omega_0\zeta)e_R(-\omega_0\zeta) = 2kTR, \qquad (124)$$

or, in other words, that

$$e_R(\omega_0\zeta) = \pm(2kTR)^{1/2}, \qquad (125)$$

everywhere in the complex plane of ζ, except possibly along the imaginary axis. But if there are any deviations from Eq. (125) along the imaginary axis, these must be of a nonanalytic character, consisting of a distribution of singularities which is everywhere dense.

A more satisfactory picture of the solution $e_R(\omega_0\zeta)$, therefore, is that it is an analytic function everywhere off the imaginary axis of ζ and is a constant given by Eq. (125); but on the imaginary axis, $e_R(\omega_0\zeta)$ may or may not be analytic. If it is analytic, then it is given by Eq. (125) along the imaginary axis, and it represents the statistical spectrum of $E_R(t)$ or alternatively the true spectrum of the impulsive emf [Eq. (122)]. If it is not analytic, however, then its dense distribution of singularities along the imaginary axis will be quite arbitrary, and it will represent the true spectrum of $E_R(t)$, but not uniquely.

Now, in addition to the analytic solution [Eq. (125)], which is entirely free from singularities, there also exist other analytic solutions $e_R(\omega_0\zeta)$ possessing simple poles. Thus, if Eq. (123) is written in the three forms

$$\frac{kTR}{Q} = \frac{1}{2\pi j} \oint \frac{e_R(\omega_0\zeta)e_R(-\omega_0\zeta)\, d\zeta}{1 - Q^2\left(\zeta + \dfrac{1}{\zeta}\right)^2} \qquad (126a)$$

$$= \frac{1}{2\pi j} \oint \frac{e_R\left(\dfrac{\omega_0}{\zeta}\right) e_R\left(\dfrac{-\omega_0}{\zeta}\right) d\zeta}{1 - Q^2\left(\zeta + \dfrac{1}{\zeta}\right)^2} \qquad (126b)$$

$$= -\frac{1}{2\pi j} \oint \frac{e_R(\omega_0\zeta)e_R(-\omega_0\zeta)\,d\zeta}{\zeta^2 1 - Q^2\left(\zeta + \dfrac{1}{\zeta}\right)^2}, \tag{126c}$$

where the third of these equations is obtained from the second by replacing ζ by $1/\zeta$, and if $e_R(\omega_0\zeta)$ is assumed to be such that

$$e_R(\omega_0\zeta)e_R(-\omega_0\zeta) = a_0 + \sum_n a_n/(\zeta^2 - \zeta_n^2) \tag{127}$$

then from Eqs. (126a) and (126c), we can write

$$0 = \frac{1}{2\pi j} \oint \frac{e_R(\omega_0\zeta)e_R(-\omega_0\zeta)\left(1 + \dfrac{1}{\zeta^2}\right)d\zeta}{1 - Q^2\left(\zeta + \dfrac{1}{\zeta}\right)^2} \tag{128a}$$

$$= a_0\phi_0 + \sum_n a_n\phi_n, \tag{128b}$$

where

$$\phi_0 = -\left(\frac{1}{2Q}\right)\left(\frac{\zeta_1^2 + 1}{\zeta_1^2 - 1} + \frac{\zeta_2^2 + 1}{\zeta_2^2 - 1}\right) \tag{129a}$$

$$\phi_n = \phi_n' + \phi_n'' \tag{129b}$$

$$\phi_n' = -\frac{1}{2Q}\left[\frac{\zeta_1^2 + 1}{(\zeta_1^2 - 1)(\zeta_1^2 - \zeta_n^2)} + \frac{\zeta_2 + 1}{(\zeta_2^2 - 1)(\zeta_2^2 - \zeta_n^2)}\right] \tag{129c}$$

$$\phi_n'' = \frac{\left(\dfrac{1}{2\zeta_n^2}\right)\left(\zeta_n + \dfrac{1}{\zeta_n}\right)}{1 - Q^2\left(\zeta_n + \dfrac{1}{\zeta_n}\right)^2}, \tag{129d}$$

and where it will be noted, from Eq. (112), that

$$\zeta_1\zeta_2 = 1, \tag{130a}$$

$$\zeta_1^2 + 1 = \frac{\zeta_1}{Q} = \left(\frac{1}{2Q^2}\right)(1 + r) \qquad r = \sqrt{1 - (2Q)^2}, \tag{130b}$$

$$\zeta_2^2 + 1 = \frac{\zeta_2}{Q} = \left(\frac{1}{2Q^2}\right)(1 - r), \tag{130c}$$

$$\zeta_1^2 - 1 = \left(\frac{1}{2Q^2}\right)r(r + 1), \tag{130d}$$

$$\zeta_2^2 - 1 = \left(\frac{1}{2Q^2}\right)r(r - 1), \tag{130e}$$

$$\zeta_1^2 + \zeta_2^2 = \frac{1}{Q^2} - 2, \tag{130f}$$

$$\zeta_1^2 - \zeta_2^2 = \frac{r}{Q^2}, \tag{130g}$$

$$(\zeta_1^2 - \zeta_n^2)(\zeta_2^2 - \zeta_n^2) = 1 - \zeta_n^2(\zeta_1^2 + \zeta_2^2) + \zeta_n^4$$
$$= \left(\frac{\zeta_n}{Q}\right)^2 \left[1 - Q^2\left(\zeta_n + \frac{1}{\zeta_n}\right)^2\right]. \quad (130h)$$

Hence,

$$\phi_0 = -\left(\frac{1}{2Q}\right)\left(\frac{1}{r} - \frac{1}{r}\right) = 0, \quad (131a)$$

$$\phi_n' = -\left(\frac{1}{2Qr}\right)\left(\frac{1}{\zeta_1^2 - \zeta_n^2} - \frac{1}{\zeta_2^2 - \zeta_n^2}\right)$$

$$= \frac{\left(\frac{1}{2Q\zeta_n^2}\right)}{1 - Q^2\left(\zeta_n + \frac{1}{\zeta_n}\right)^2}, \quad (131b)$$

and therefore

$$\phi_0 = 0, \quad (132a)$$

$$\phi_n = \frac{\left(\frac{1}{2Q\zeta_n^2}\right)}{1 - Q\left(\zeta_n + \frac{1}{\zeta_n}\right)}. \quad (132b)$$

Consequently, Eq. (128) becomes

$$0 = \sum_n \frac{a_n}{\zeta_n^2}\left[1 - Q\left(\zeta_n + \frac{1}{\zeta_n}\right)\right]^{-1}. \quad (133)$$

It is clear that $e_R(\omega_0\zeta)$ may have any number of simple poles except one and that the positions of these poles may be anywhere in the ζ-plane, provided the coefficients ζ_n in Eq. (127) satisfy the single constraint, Eq. (133). With any given positional distribution of poles, however, and with all coefficients a_n fixed but one, the remaining coefficient is a function of Q, according to Eq. (133). Therefore, such solutions $e_R(\omega_0\zeta)$ as are here described are not independent of the Q of the parallel resonant circuit and hence cannot represent the true emf $E_R(t)$.

Quantum Statistics of Thermal Noise.—The theoretical result that the equation

$$\overline{|e_R(j\omega)|^2} = 2kTR \quad (134)$$

holds for the statistical spectrum of the thermal-noise emf $E_R(t)$ in series with a resistance R at an equilibrium temperature T needs to be reexamined if it is thought to apply to extremely high frequencies.

In the first place, the picture of a parallel resonant *RLC*-circuit is, strictly speaking, appropriate only to a range of frequencies below which radiation will occur. Obviously, it is physically impossible to construct a parallel resonant circuit in such a way as to confine the electromagnetic

fields, at extremely high frequencies, to the interior of the inductance L and capacitance C. In discussing extremely high frequencies, therefore, the parallel resonant circuit can no longer serve as a model; another circuit model, of a simple type that admits of the existence of propagated electromagnetic waves, must be found.

Perhaps the simplest such model[1] is the circuit shown in Fig. 12·4. Let the concentric line there shown be terminated at each end by a resistance R. If each of these resistances has the physical form of a solid
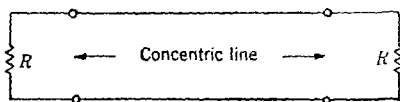
FIG. 12·4.—Concentric line terminated at both ends with a resistance R.

disk of material of appropriate conductivity, so that it closes completely its end of the concentric line, all electromagnetic radiation approaching either end of the line is either absorbed or reflected, provided the disks are thick enough or of enough resistivity to attenuate completely the outwardly propagated waves. This ensures that whatever electromagnetic fields exist in the interior of the line do not leak outside the line.

Such a circuit constitutes a physical system that can be regarded as completely isolated and can be studied under conditions of thermal equilibrium, just as the parallel resonant circuit was studied above as an isolated physical system in thermal equilibrium. As before, the resistances R are seats of molecular heat motion, each having a total heat energy Q. Other forms of free energy might reside in these resistances R, as assumed in the previous case, but that assumption need not again be made here. Just as the existence of electromagnetic free energy in the fields of the inductance L and the capacitance C was assumed for the parallel resonant circuit, the existence of electromagnetic energy in the form of electromagnetic waves propagated back and forth parallel to the axis of the concentric line is assumed here.

It is necessary, then, to find a set of independent parameters on which this electromagnetic energy depends, analogous to the current I_L in the inductance L and the charge Q_C of the capacitance C in the case previously treated. Although there were but two degrees of freedom of free energy (I_L, Q_C) in the former case, there are an infinite number in that now under discussion. These correspond to the infinite number of standing waves that could be set up along the axis of the concentric line, were it short-circuited and—a tacit assumption for this discussion—nondissipative.

Thus, if l is the length of the line and v the velocity of propagation of the electromagnetic waves along its length, then the longest possible wavelength of the standing waves in the short-circuited case would be

$$\lambda_1 = 2l, \tag{135}$$

[1] *Cf.* also H. Nyquist, *Phys. Rev.*, **32**, 110 (1928).

and the lowest frequency would be

$$\nu_1 = \frac{v}{\lambda_1} = \frac{v}{2l}. \tag{136}$$

But in addition to this frequency ν_1, all integral multiples of ν_1 are usually present. Hence two degrees of freedom (electric and magnetic) may be associated with each frequency.

$$\nu_n = n\nu_1. \tag{137}$$

Now if the fundamental frequency ν_1 is made very small by making the length l of the line very large, the number of harmonic frequencies contained in any arbitrarily small frequency interval $(\nu,\ \nu + d\nu)$ can be maintained very large. Assuming that this is the case, we may speak henceforth of the electromagnetic energy dU_ν associated with all degrees of freedom in the frequency interval $(\nu,\ \nu + d\nu)$ and similarly of the power dP_ν delivered across any cross section of the line in either direction by the corresponding traveling waves.

The number of degrees of freedom of electromagnetic energy in the interval of frequency $(\nu,\ \nu + d\nu)$ is

$$dN_\nu = \frac{2d\nu}{\nu_1} = \left(\frac{4l}{v}\right)d\nu, \tag{138}$$

and so, on the basis of classical statistical mechanics, which assigns to each degree of freedom an average energy equal to $kT/2$, the corresponding average electromagnetic energy would be

$$dU_\nu = \left(\frac{kT}{2}\right)dN_\nu = \left(\frac{2l}{v}\right)kT\,d\nu. \tag{139}$$

If this energy, residing in standing waves in the case of a short-circuited line, can be thought of as shared by equal and opposite traveling waves, then the average power passing any cross section of the line in either direction is half of the fraction of dU_ν that is contained in v units of length of the line. In other words, the average power in either direction equals

$$dP_\nu = kTd\nu. \tag{140}$$

The result, Eq. (140), is applicable, according to the above derivation, only to the short-circuited line, on the assumption that the standing electromagnetic waves existing therein are in thermal equilibrium at a temperature T. In particular, however, it is applicable at any cross section of a line that extends indefinitely in both directions. But in this latter case the impedance of the line, looking in either direction, is a pure resistance equal to $\sqrt{L/C}$, if L and C are the inductance and capacitance

per unit length of the line, respectively. Therefore, if in Fig. 12·4 both resistances R are equal to

$$R = \sqrt{\frac{L}{C}} \tag{141}$$

then at any cross section of the line in Fig. 12·4 the conditions will be indistinguishable from those which would exist in the case of a short-circuited line extending infinitely far in both directions, provided the

Fig. 12·5.—Equivalent representation of resistance R as a source of thermal noise, with any load impedance Z_L.

physical system of Fig. 12·4 is in thermal equilibrium. Hence, it can be concluded that each resistance R in Fig. 12·4 emits and absorbs an amount of power of average value

$$dP_\nu = kTd\nu \tag{142}$$

in every frequency interval $(\nu, \nu + d\nu)$ when matched into the line and when in thermal equilibrium with the line at a temperature T.

The power emitted by each resistance R should not be dependent on the nature of the load to which it is matched, provided the latter is in thermal equilibrium with it. Either resistance R, then, may be represented as a source of power, as shown in Fig. 12·5, with a load impedance

$$Z_L = R, \tag{143}$$

whose nature need not be specified except that it should be matched. In Fig. 12·5 the power emitted by R is also the power absorbed by Z_L. But if E_ν is the emf giving rise to this emitted power in the frequency interval $(\nu, \nu + d\nu)$, then the current flowing in Z_L is

$$I_\nu = \frac{E_\nu}{R + Z_L} = \frac{E_\nu}{2R}, \tag{144}$$

and the power absorbed by Z_L is

$$dP_\nu = I_\nu^2 R = \frac{E_\nu^2}{4R}. \tag{145}$$

Therefore, by Eq. (142),

$$E_\nu^2 = 4kTRd\nu. \tag{146}$$

This, then, is the average value of the square of the thermal-noise emf residing in R for the frequency interval $(\nu, \nu + d\nu)$. But the ratio $E_\nu^2/d\nu$ is just twice the statistical spectrum $\overline{|e_R(j\omega)|^2}$ the value of which was found above by other methods. (Negative as well as positive frequencies are employed conventionally for the latter and only positive frequencies for the former.) The result

$$\overline{|e_R(j\omega)|^2} = 2kTR \tag{147}$$

is therefore found again.

Actually, this result is not strictly correct, as might already be inferred from the nature of Eq. (139) for the total electromagnetic energy in the frequency interval $(\nu,\ \nu + d\nu)$. This is because the integral of this expression diverges when taken over all frequencies, indicating that the total electromagnetic energy for all frequencies, or for all degrees of freedom, is infinite. In fact, however, the total energy in the present isolated physical system is finite. That the heat energy is finite is assured by the fact that the number of degrees of freedom of the molecular motions, though large, is yet finite. But the electromagnetic energy has an infinite number of degrees of freedom, and the result given above, to the effect that each degree of freedom (electric or magnetic) of electromagnetic energy possesses an average energy equal to $kT/2$, must therefore be reexamined.

If the short-circuited line is reconsidered from the point of view of quantum theory, associating *quanta* of radiant energy $h\nu_s$ (where h is Planck's constant) with the standing electromagnetic waves of frequency ν_s, then a statistical consideration of these quanta, involving the following quantities, must be made.

$$u_s = h\nu_s = \text{energy of each quantum of frequency } \nu_s.$$

Z_{ns} = number of modes of oscillation of frequency ν_s containing n quanta.

$Z_s = \sum_n Z_{ns}$ = total number of modes of oscillation of frequency ν_s.

$N_s = \sum_n n Z_{ns}$ = total number of quanta of frequency ν_s.

$N = \sum_s N_s$ = total number of quanta of all frequencies.

$U_s = \sum_n n u_s Z_{ns}$ = total energy of quanta of frequency ν_s.

$U = \sum_s U_s$ = total energy of quanta of all frequencies.

The probability of a state in which, at the frequency ν_s, Z_{0s} modes of oscillation contain no quanta, Z_{1s} modes contain one quantum, Z_{2s} modes contain two quanta, and in general Z_{ns} modes contain n quanta, is

$$W_s = \frac{\prod\limits_s Z_s!}{\prod\limits_n Z_{ns}!} \tag{148}$$

Therefore the complete probability, or thermodynamic probability, of any distribution of quanta at all frequencies is

$$W = \prod_s W_s = \frac{\prod_s Z_s!}{\prod_n \prod_s Z_{ns}!} \tag{149}$$

For the short-circuited line, thermodynamic equilibrium therefore corresponds to the maximum value of Eq. (149) or to the maximum value of the corresponding entropy expression

$$S = k \ln W \tag{150a}$$

$$= k \sum_s \ln Z_s! - k \sum_n \sum_s \ln Z_{ns}! \tag{150b}$$

Now, by Stirling's approximation,

$$\ln x! = x \ln x - x \qquad (x \gg 1); \tag{151}$$

we can therefore write

$$S = k \sum_s Z_s \ln Z_s - k \sum_n \sum_s Z_{ns} \ln Z_{ns}. \tag{152}$$

When S and therefore also W are a maximum with respect to variations in Z_{ns}, we have

$$0 = \delta S = k \sum_s \delta Z_s \ln Z_s - k \sum_n \sum_s \delta Z_{ns} \ln Z_{ns}. \tag{153}$$

But because Z_s is a constant fixed by the geometry of the concentric line, we also have

$$0 = \delta Z_s = \sum_n \delta Z_{ns}, \tag{154}$$

and therefore Eqs. (153) and (154) become

$$0 = \sum_n \sum_s \delta Z_{ns} \ln Z_{ns}, \tag{155a}$$

$$0 = \sum_n \delta Z_{ns}. \tag{155b}$$

In addition, the total number N of quanta and the total energy U of the quanta are constants. Therefore

$$0 = \delta N = \sum_n \sum_s n \delta Z_{ns}, \tag{156a}$$

$$0 = \delta U = \sum_n \sum_s n u_s \delta Z_{ns}. \tag{156b}$$

If now we multiply Eq. (155a) by unity, Eq. (155b) by $-\ln \alpha_s$, Eq. (156a) by β, and Eq. (156b) by ν, where α_s, β_s, ν are undetermined constants, and if we add the four resulting expressions, the coefficient of δZ_{ns} in the resulting sum must vanish; that is,

$$0 = \ln Z_{ns} - \ln \alpha_s + n(\beta + \nu u_s). \tag{157}$$

Therefore we may write

$$\frac{Z_{ns}}{\alpha_s} = e^{-n(\beta + \nu u_s)}, \tag{158}$$

and hence

$$\frac{Z_s}{\alpha_s} = \sum_n e^{-n(\beta + \nu u_s)}$$
$$= [1 - e^{-(\beta + \nu u_s)}]^{-1} - 1. \tag{159a}$$

$$\frac{N_s}{\alpha_s} = \sum_n n e^{-n(\beta + \nu u_s)}$$

$$= - \frac{\partial}{\partial(\beta + \nu u_s)} \sum_n e^{-n(\beta + \nu u_s)}$$

$$= - \frac{\partial}{\partial(\beta + \nu u_s)} [1 - e^{-(\beta + \nu u)_s}]^{-1}$$
$$= e^{-(\beta + \nu u_s)}[1 - e^{-(\beta + \nu u_s)}]^{-2}. \tag{159b}$$

Consequently,

$$\frac{N_s}{Z_s} = [e^{(\beta + \nu u_s)} - 1]^{-1}. \tag{160}$$

If we call $e^\beta = B$ and write $u_s = h\nu_s$, this becomes

$$\frac{N_s}{Z_s} = (Be^{\nu h\nu_s} - 1)^{-1}. \tag{161}$$

When this result is compared with Expression (138), it is found that for the frequency interval $(\nu_s, \nu_s + d\nu_s)$

$$Z_s = \frac{2l}{v} d\nu_s. \tag{162}$$

Furthermore, the total average energy of the quanta of frequency ν_s can be written

$$U_s = N_s h\nu_s, \tag{163}$$

or, by Eqs. (161) and (162),

$$U_s = \frac{2l}{v} \frac{h\nu_s \, d\nu_s}{Be^{\nu h\nu_s} - 1}. \tag{164}$$

But this result, for small values of the frequency ν_s, should agree with Expression (139). Therefore, Eq. (164), in the notation of Eq. (139), can be written thus:

$$
\begin{aligned}
dU_\nu &= \frac{2l}{v} \frac{h\nu \, d\nu}{Be^{\nu h\nu} - 1} \\
&= \frac{2l}{v} \frac{h\nu \, d\nu}{(B - 1) + B\nu h\nu}.
\end{aligned}
\tag{165}
$$

Equating Eqs. (139) and (165), we find that

$$\frac{h\nu}{kT} = (B - 1) + B\nu h\nu. \tag{166}$$

Since equality (166) must hold for all low frequencies, it must follow that

$$B = 1, \tag{167a}$$

$$\nu = \frac{1}{kT}. \tag{167b}$$

Therefore, the correct expression (164) for the total average energy of the quanta in the frequency interval $(\nu, \ \nu + d\nu)$ becomes

$$dU_\nu = \frac{2l}{v} \frac{h\nu \, d\nu}{e^{h\nu/kT} - 1}, \tag{168}$$

and the corresponding expression for the average power passing any cross section of the line in either direction becomes

$$dP_\nu = \frac{h\nu \, d\nu}{e^{h\nu/kT} - 1}. \tag{169}$$

Finally, the corresponding statistical spectrum $\overline{|e_R(j\omega)|^2}$ of the thermal-noise voltage $E_R(t)$ in either resistance R in Fig. 12·4 or 12·5 becomes

$$\overline{|e_R(j\omega)|^2} = \frac{2Rh\nu}{e^{h\nu/kT} - 1}. \tag{170}$$

Actually, for room temperature and for frequencies up to and well beyond the visible region, Expression (170) is indistinguishable from the usual low-frequency expression (47).

Nyquist's Thermal-noise Theorem.—Expression (147) or its more exact quantum theoretical extension (170) gives the statistical spectrum of the thermal-noise emf $E_R(t)$ in series with an ohmic, frequency-independent resistance R. This result may be extended to apply to any passive two-terminal network that is in thermal equilibrium, as was first

shown by Nyquist.[1]   To show this, let Z_L in Fig. 12·5 be an arbitrary
passive two-terminal network, whose impedance Z_L may be expressed as
a complex function of frequency

$$Z_L = Z(j\omega) = R(j\omega) + jX(j\omega). \tag{171}$$

If Z_L is in thermal equilibrium with R in Fig. 12·5, then the average
thermal-noise power absorbed by Z_L and arising from R in any frequency
interval $(\nu, \nu + d\nu)$ must equal the average thermal-noise power absorbed
by R and arising from Z_L in that frequency interval.   Hence, if

$$\overline{|e_R(j\omega)|^2}\, d\nu = 2kTR\, d\nu \tag{172}$$

is the mean squared thermal-noise emf for this frequency interval which
is in series with the resistance R, and if

$$\overline{|e_Z(j\omega)|^2}\, d\nu = 2kTR'\, d\nu \tag{173}$$

is the corresponding mean squared thermal-noise emf supposed to be in
series with Z_L, then the power absorbed by Z_L and arising from R is

$$\begin{aligned}
dP_Z &= \frac{\overline{|e_R(j\omega)|^2}\, d\nu R(j\omega)}{[R + R(j\omega)]^2 + X^2(j\omega)} \\
&= \frac{2kT\, d\nu R R(j\omega)}{[R + R(j\omega)]^2 + X^2(j\omega)},
\end{aligned} \tag{174}$$

while the power absorbed by R and arising from Z_L is

$$\begin{aligned}
dP_R &= \frac{\overline{|e_Z(j\omega)|^2}\, d\nu R}{[R + R(j\omega)]^2 + X^2(j\omega)} \\
&= \frac{2kT\, d\nu R' R}{[R + R(j\omega)]^2 + X^2(j\omega)}.
\end{aligned} \tag{175}$$

But since $dP_Z = dP_R$, we find that $R' = R(j\omega)$ and hence that

$$\overline{|e_Z(j\omega)|^2} = 2kTR(j\omega). \tag{176}$$

This result, Eq. (176), is the analytical statement of Nyquist's theorem,
in the usual form corresponding to
the low-frequency approximation
to Eq. (170).   According to this
theorem, any passive two-termi-
nal network in thermal equilib-
rium at the temperature T may
be represented in the sense of
Thévenin's theorem as an active
generator of thermal noise, as

FIG. 12·6.—Two equivalent representa-
tions of Nyquist's noise theorem: (a) in terms
of impedance; (b) in terms of admittance.

shown in (a) of Fig. 12·6, where its impedance $Z(j\omega)$ is given by Eq. (171)
and the Fourier transform $e_Z(j\omega)$ of its thermal-noise emf $E_Z(t)$ is given
in the mean square by Eq. (176).

[1] H. Nyquist, *Phys. Rev.*, **32**, 110 (1928).

An alternative expression of Nyquist's theorem is possible on an admittance basis.   Let

$$Y(j\omega) = \frac{1}{Z(j\omega)} = G(j\omega) + jB(j\omega) \tag{177}$$

be the corresponding admittance of Z_L in Fig. 12·5, expressed as a complex function of frequency.   Then if

$$i_r(j\omega) = \frac{e_z(j\omega)}{Z(j\omega)} = e_z(j\omega)Y(j\omega), \tag{178}$$

the quantity $i_r(j\omega)$ represents the Fourier transform of the constant-current generator $I_r(t)$ which, placed with $Y(j\omega)$ as shown in (b) of Fig. 12·6, acts as an equivalent source of thermal-noise power in the sense of Norton's theorem (the admittance analogue of Thévenin's theorem). But, according to Eq. (178),

$$\overline{|i_r(j\omega)|^2} = \overline{|e_z(j\omega)|^2}[G^2(j\omega) + B^2(j\omega)]$$
$$= 2kTR(j\omega)[G^2(j\omega) + B^2(j\omega)]; \tag{179}$$

and therefore, since Eq. (177) gives the result that

$$R(j\omega) = \frac{G(j\omega)}{G^2(j\omega) + B^2(j\omega)}, \tag{180}$$

it follows that

$$\overline{|i_r(j\omega)|^2} = 2kTG(j\omega). \tag{181}$$

This result, Eq. (181), is another analytical statement of Nyquist's theorem, in terms of the admittance representation of (b) of Fig. 12·6 and in the form corresponding to the low-frequency approximation to Eq. (170).

   A Generalization of Nyquist's Thermal-noise Theorem.[1]—For passive networks in thermal equilibrium that have three or more terminals, it would appear possible, in theory, to represent the complete thermal-noise behavior by applying Nyquist's theorem independently to each component element of the network.   In practice, however, networks whose actual component elements are unknown are often found.   In such cases a network must be represented in terms of apparent component elements whose unreality is commonly manifested by their failure to be passive. Consequently, for the general M-terminal passive network in thermal equilibrium, a complete thermal-noise representation lies beyond the province of Nyquist's original theorem.   In the following discussion, an analogous but more general theorem which represents completely the thermal-noise behavior of any passive network in thermal equilibrium is stated.   The proof of this theorem is given elsewhere.

   [1] *Cf.* also E. J. Schremp, to be published.

As a preliminary to the statement of this theorem, certain restrictions that are conventionally imposed must be removed from the formulation (181) of Nyquist's theorem.

Actually, the current $i_Y(j\omega)$ that occurs in (*b*) of Fig. 12·6 and in Eq. (181) is the complex Fourier transform in the Fourier integral resolution

$$I_Y(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} i_Y(j\omega) e^{j\omega t}\, d\omega \tag{182}$$

of the time-varying thermal-noise current impressed across the passive admittance $Y(j\omega)$. This time-varying current is experimentally observable and is, therefore, quite naturally taken to be a *real* function of time, so that between the Fourier transforms at frequencies ω and $-\omega$ the relation

$$i_Y(-j\omega) = i_Y^*(j\omega) \tag{183}$$

is assumed to hold. Consequently, Nyquist's theorem (181) may also be written in either of the two forms

$$\overline{i_Y(j\omega)i_Y^*(j\omega)} = 2kTG(j\omega), \tag{184a}$$

$$\overline{i_Y(j\omega)i_Y(-j\omega)} = 2kTG(j\omega). \tag{184b}$$

The first form is always equivalent to the form (181), but the second form is equivalent only if the relation (183) holds. As will presently be seen, the conventional formulation of Nyquist's theorem as exemplified by the relations (181) and (184a) must be abandoned in what follows, in favor of the formulation (184b); and the conventional relation (183) must be abandoned in favor of a more general relation. The same kind of changes must be made in Eq. (176).

It is necessary in what follows to bear in mind certain properties of the admittance $Y(j\omega)$ whose real part $G(j\omega)$ occurs in Eq. (181). This admittance is always a function of the argument $j\omega$ and never of any other argument containing j or ω. Consequently it always follows that

$$Y(-j\omega) = Y^*(j\omega). \tag{185}$$

Hence, since

$$G(j\omega) = \tfrac{1}{2}[Y(j\omega) + Y^*(j\omega)], \tag{186}$$

Eqs. (184b) can be rewritten in the form

$$\overline{i_Y(j\omega)i_Y(-j\omega)} = kT[Y(j\omega) + Y(-j\omega)]. \tag{187}$$

By similar reasoning, Eq. (176) can be replaced by the relation

$$\overline{e_Z(j\omega)e_Z(-j\omega)} = kT[Z(j\omega) + Z(-j\omega)]. \tag{188}$$

In what follows, Eqs. (187) and (188) are taken as the fundamental formulations of Nyquist's theorem. Or, using the complex frequency,

$$\lambda = \sigma + j\omega, \tag{189}$$

these two equations can be expressed to apply to the whole λ-plane, thus:

$$\overline{i_Y(\lambda)i_Y(-\lambda)} = kT[Y(\lambda) + Y(-\lambda)] \qquad (190a)$$

$$\overline{e_Z(\lambda)e_Z(-\lambda)} = kT[Z(\lambda) + Z(-\lambda)]. \qquad (190b)$$

The complete thermal-noise behavior of a given passive network in thermal equilibrium can be described, when its complete physical structure is known, by applying Nyquist's theorem in the form (190a),

$$\overline{i_{mn}(\lambda)i_{mn}(-\lambda)} = -kT[y_{mn}(\lambda) + y_{mn}(-\lambda)] \qquad (191)$$

independently to each passive branch element $-y_{mn}(\lambda)$. This is the internodal formulation of Nyquist's theorem, because it is expressed in terms of the internodal currents impressed between different nodes m and n. These internodal currents have the following noteworthy properties. First, because $i_{mn}(\lambda)$ is, by definition, the current flowing from node m to node n, we can write

$$i_{mn}(\lambda) = -i_{nm}(\lambda). \qquad (192)$$

Second, because of the incoherence of the thermal agitations of any two distinct passive admittances, we can write

$$\overline{i_{mn}(\lambda)i_{rs}(-\lambda)} = 0 \qquad (193)$$

for any two distinct nodal pairs (m,n) and (r,s).

On the other hand, it is equally possible to express Nyquist's theorem in terms of the nodal thermal-noise currents

$$i_m(\lambda) = \sum_n i_{nm}(\lambda) \qquad (m \neq n), \qquad (194)$$

each of which represents the net thermal-noise current flowing into a given node m. Consider the average product

$$\overline{i_m(\lambda)i_r(-\lambda)} = \sum_n \sum_s \overline{i_{nm}(\lambda)i_{sr}(-\lambda)} \qquad (m \neq n, \ r \neq s). \qquad (195)$$

As a consequence of the incoherency properties (193), this average product reduces to either of the expressions

$$\overline{i_m(\lambda)i_r(-\lambda)} = -\overline{i_{mr}(\lambda)i_{mr}(-\lambda)} \qquad (m \neq r) \qquad (196a)$$

$$= \sum_n \overline{i_{mn}(\lambda)i_{mn}(-\lambda)} \qquad (m = r \neq n), \qquad (196b)$$

depending on whether the nodes m and r are different or not. Therefore, by using Eqs. (191), we can write

$$\overline{i_m(\lambda)i_r(-\lambda)} = kT[y_{mr}(\lambda) + y_{mr}(-\lambda)] \qquad (m \neq r) \tag{197a}$$

$$= -kT \sum_n [y_{mn}(\lambda) + y_{mn}(-\lambda)] \qquad (m = r \neq n). \tag{197b}$$

But if

$$y_{mm}(\lambda) = -\sum_n y_{mn}(\lambda) \qquad (m \neq n) \tag{198}$$

is called the total admittance from node m to all other nodes, then Eq. (197) can be written in the single form

$$\overline{i_m(\lambda)i_r(-\lambda)} = kT[y_{mr}(\lambda) + y_{mr}(-\lambda)], \tag{199}$$

which is valid both when the nodes m and r are distinct and when they are the same.   Equation (199), then, is the nodal formulation of Nyquist's theorem.

In considering the problem of representing the actual thermal-noise behavior in a passive M-terminal network the physical structure of which is unknown, the existence of this actual physical structure may certainly be assumed.   It may be supposed to be represented by a network containing $N \geqq M$ nodes and having nodal thermal-noise currents $i_m(\lambda)$ ($m = 1, 2, \cdots, N$) that obey Eqs. (199).   If $N = M$, the problem would not exist, because the actual physical structure of the M-terminal network would then be known.   On the other hand, if $N > M$, and if node N is internal to the network, then the current $i_N(\lambda)$ and the branch admittances $-y_{mN}(\lambda)$ are not usually susceptible to measurement.

Because this is so, the current $i_N(\lambda)$ and the branch admittances $-y_{mN}(\lambda)$ may be thought of as quantities that may be varied in certain ways, as long as these variations do not conflict with the observable properties of the thermal noise in the network.   Specifically, $i_N(\lambda)$ can be changed to

$$i_N(\lambda,\alpha) = [1 - \alpha(\lambda)]i_N(\lambda), \tag{200}$$

where $\alpha(\lambda)$ is any complex number if $i_m(\lambda)$ and $y_{mn}(\lambda)$ are also changed as follows:

$$i_m(\lambda,\alpha) = i_m(\lambda) - \alpha(\lambda) \frac{y_{mN}(\lambda)}{y_{NN}(\lambda)} i_N(\lambda) \tag{201a}$$

$$y_{mn}(\lambda,\alpha) = y_{mn}(\lambda) - \alpha(\lambda) \frac{y_{mN}(\lambda)}{y_{NN}(\lambda)} y_{Nn}(\lambda), \tag{201b}$$

as may be seen in detail by referring to the proof of this result, published elsewhere.[1]

Equations (201) form a continuous group of transformations, with the complex parameter α, that transform the actual physical network into an

[1] E. J. Schremp, to be published.

infinite variety of equivalent networks. Let us then consider the expression $\overline{i_m(\lambda,\alpha)i_r(-\lambda,\alpha)}$ which, for $\alpha = 0$, is given by Nyquist's theorem as expressed by Eqs. (199). By Eq. (201a),

$$
\begin{aligned}
i_m(\lambda,\alpha)i_r(-\lambda,\alpha) &= i_m(\lambda)i_r(-\lambda) \\
&- \alpha(\lambda)\,\frac{y_{mN}(\lambda)}{y_{NN}(\lambda)}\,i_N(\lambda)i_r(-\lambda) \\
&- \alpha(-\lambda)\,\frac{y_{rN}(-\lambda)}{y_{NN}(-\lambda)}\,i_N(-\lambda)i_m(\lambda) \\
&+ \alpha(\lambda)\alpha(-\lambda)\,\frac{y_{mN}(\lambda)y_{rN}(-\lambda)}{y_{NN}(\lambda)y_{NN}(-\lambda)}\,i_N(\lambda)i_N(-\lambda).
\end{aligned}
\tag{202}
$$

Therefore, by Eq. (199), this gives

$$
\begin{aligned}
\frac{\overline{i_m(\lambda,\alpha)i_r(-\lambda,\alpha)}}{kT} &= [y_{mr}(\lambda) + y_{mr}(-\lambda)] \\
&- \alpha(\lambda)\,\frac{y_{mN}(\lambda)}{y_{NN}(\lambda)}\,[y_{Nr}(\lambda) + y_{Nr}(-\lambda)] \\
&- \alpha(-\lambda)\,\frac{y_{rN}(-\lambda)}{y_{NN}(-\lambda)}\,[y_{mN}(\lambda) + y_{mN}(-\lambda)] \\
&+ \alpha(\lambda)\alpha(-\lambda)\,\frac{y_{mN}(\lambda)y_{rN}(-\lambda)}{y_{NN}(\lambda)y_{NN}(-\lambda)}\,[y_{NN}(\lambda) + y_{NN}(-\lambda)].
\end{aligned}
\tag{203}
$$

But this can be written simply

$$
\overline{i_m(\lambda,\alpha)i_r(-\lambda,\alpha)} = kT[y_{mr}(\lambda,\alpha) + y_{mr}(-\lambda,\alpha)],
\tag{204}
$$

because, equating the right-hand sides of Eqs. (203) and (204) and using Eq. (201b), we get

$$
\begin{aligned}
&y_{mr}(\lambda) - \alpha(\lambda)\,\frac{y_{mN}(\lambda)}{y_{NN}(\lambda)}\,y_{Nr}(\lambda) \\
&+ y_{mr}(-\lambda) - \alpha(-\lambda)\,\frac{y_{mN}(-\lambda)}{y_{NN}(-\lambda)}\,y_{Nr}(-\lambda) \\
&= y_{mr}(\lambda) + y_{mr}(-\lambda) \\
&- \alpha(\lambda)\,\frac{y_{mN}(\lambda)}{y_{NN}(\lambda)}\,[y_{Nr}(\lambda) + y_{Nr}(-\lambda)] \\
&- \alpha(-\lambda)\,\frac{y_{rN}(-\lambda)}{y_{NN}(-\lambda)}\,[y_{mN}(\lambda) + {}_{mN}(-\lambda)] \\
&+ \alpha(\lambda)\alpha(-\lambda)[y_{mN}(\lambda)y_{rN}(-\lambda)]\left[\frac{1}{y_{NN}(\lambda)} + \frac{1}{y_{NN}(-\lambda)}\right].
\end{aligned}
\tag{205}
$$

Or, noting that the y's are bilateral,

$$
0 = y_{mN}(\lambda)y_{rN}(\lambda)\left\{\frac{\alpha(\lambda)}{y_{NN}(\lambda)} + \frac{\alpha(-\lambda)}{y_{NN}(-\lambda)} - \alpha(\lambda)\alpha(-\lambda)\left[\frac{1}{y_{NN}(\lambda)} + \frac{1}{y_{NN}(-\lambda)}\right]\right\}.
\tag{206}
$$

Therefore, the result in Eq. (204) holds as long as $\alpha(\lambda)$ obeys the constraint

$$0 = \alpha(\lambda)[1 - \alpha(-\lambda)]/y_{NN}(-\lambda) + \alpha(-\lambda)[1 - \alpha(\lambda)]/y_{NN}(\lambda). \quad (207)$$

This result is of considerable significance; for it shows that Eqs. (204) constitute for the representation of thermal noise an invariant expression that holds for any value of the continuous parameter $\alpha(\lambda)$ that obeys the constraint in Eq. (207). Moreover, an analogous result holds with respect to any other internal node of the original physical network or of any transformed variety of this network corresponding to a certain value of $\alpha(\lambda)$. Expression (204) is therefore an explicit method for finding all possible equivalent representations of the given network for which, in terms of the appropriate thermal-noise nodal currents $i_m(\lambda; \alpha_1, \alpha_2, \ldots)$ and branch admittances $-y_{mr}(\lambda; \alpha_1, \alpha_2, \ldots)$, it is possible to express the complete thermal-noise behavior in the invariant form

$$\overline{i_m(\lambda; \alpha_1, \alpha_2, \cdots)i_r(-\lambda; \alpha_1, \alpha_2, \cdots)} = kT[y_{mr}(\lambda; \alpha_1, \alpha_2, \cdots) + y_{mr}(-\lambda; \alpha_1, \alpha_2, \cdots)]. \quad (208)$$

In particular, if $\alpha(\lambda) = 1$ in Eqs. (201), then it is seen that

$$i_N(\lambda,\alpha) = 0, \quad (209a)$$
$$y_{mN}(\lambda,\alpha) = 0, \quad (209b)$$

or, in other words, that the node N is completely eliminated. But by iterating the transformation in Eqs. (201) with respect to all internal nodes in succession and setting

$$\alpha_1(\lambda) = \alpha_2(\lambda) = \cdots = \alpha_{N-M}(\lambda) = 1 \quad (210)$$

one arrives at the M-node or external representation of the given M-terminal network itself. For this representation Eq. (208) assumes the specialized form

$$\overline{i_m(\lambda; 1, 1, \cdots)i_r(-\lambda; 1, 1, \cdots)} = kT[y_{mr}(\lambda; 1, 1, \cdots) + y_{mr}(-\lambda; 1, 1, \cdots)]. \quad (211)$$

The content of this generalization of Nyquist's theorem is thus expressed, in the general case, by Eq. (208). But in addition to this nodal formulation of the result, it is possible also to give an internodal formulation. The existence of mutually incoherent internodal thermal-noise currents that, for any representation of the given network, satisfy relations analogous to Eqs. (192) to (196) can again be assumed. Therefore, the result in Eq. (208) can also be expressed in the internodal form

$$\overline{i_{mr}(\lambda; \alpha_1, \alpha_2, \cdots)i_{mr}(-\lambda; \alpha_1, \alpha_2, \cdots)} = -kT[y_{mr}(\lambda; \alpha_1\alpha_2, \cdots) + y_{mr}(-\lambda; \alpha_1, \alpha_2, \cdots)], \quad (212)$$

or, what is the same thing, in the form

$$\overline{i_{mr}(\lambda;\,\alpha_1,\,\alpha_2,\,\cdots\,)i_{mr}(-\lambda;\,\alpha_1,\,\alpha_2,\,\cdots\,)}$$
$$= -2kT g_{mr}(\lambda;\,\alpha_1,\,\alpha_2,\,\cdots\,),\quad(213)$$

where $-g_{mr}(\lambda;\,\alpha_1,\,\alpha_2,\,\cdots\,)$ is the effective branch conductance linking nodes m and r.

For

$$\alpha_1(\lambda) = \alpha_2(\lambda) = \cdots = 0,\qquad(214)$$

the result in Eq. (213) is identical with Nyquist's original theorem, as it applies to the actual physical structure of the given network; whereas, for general values of $\alpha_1(\lambda),\,\alpha_2(\lambda),\,\ldots$, consistent with the constraints of the form in Eq. (207), the result in Eq. (212) is a formal extension of Nyquist's theorem to any equivalent representation of the given network wherein the effective branch conductances $-g_{mr}(\lambda;\,\alpha_1,\,\alpha_2,\,\cdots\,)$ may or may not be passive elements. Because, in general, $-g_{mr}(j\omega;\,\alpha_1,\,\alpha_2\,\cdots\,)$ is not necessarily positive for all real frequencies ω, it follows that we cannot, in general, write

$$i_{mr}(-j\omega;\,\alpha_1,\,\alpha_2,\,\cdots\,) = i_{mr}(j\omega;\,\alpha_1,\,\alpha_2,\,\cdots\,).\qquad(215)$$

However, from Eqs. (213) it can be seen that it is sufficient, instead, to write

$$i_{mr}(-j\omega;\,\alpha_1,\,\alpha_2,\,\cdots\,) = \pm i_{mr}(j\omega;\,\alpha_1,\,\alpha_2,\,\cdots\,),\qquad(216)$$

where the positive sign is taken if the conductance $-g_{mr}(j\omega;\,\alpha_1,\,\alpha_2,\,\cdots\,)$ is positive and the negative sign if this conductance is negative.

The significance of this change from the conventional restriction in Eq. (215) to the weaker restriction in Eq. (216) is apparent from the Fourier integral representation

$$I_{mr}(t;\,\alpha_1,\,\alpha_2,\,\cdots\,) = \frac{1}{2\pi}\int i_{mr}(j\omega;\,\alpha_1,\,\alpha_2,\,\cdots\,)e^{j\omega t}\,d\omega\qquad(217)$$

of the time-varying thermal-noise current impressed across the apparent branch admittance $-y_{mr}(j\omega;\,\alpha_1,\,\alpha_2,\,\cdots\,)$. This change means that, although the time-varying thermal-noise currents $I_{mr}(t;\,0,\,0,\,\ldots\,)$ envisaged by Nyquist's theorem are always real functions of time, the corresponding currents $I_{mr}(t;\,\alpha_1,\,\alpha_2,\,\ldots\,)$ envisaged by the generalized theorem stated, are here, in general, complex functions of time. This result is not surprising; just as the real character of these currents in the conventional sense is required by the fact that they are experimentally observable (or at least are assumed to have a real existence), so their complex character in the generalized sense just described reflects the fact that these currents need not be experimentally observable (and indeed need not have any real existence).

On the other hand, the nodal thermal-noise currents

$$I_m(t; \alpha_1, \alpha_2, \cdots) = \sum_n I_{nm}(t; \alpha_1, \alpha_2, \cdots) \qquad (m \neq n) \quad (218)$$

are always real, although they are observable only if they refer to nodes that are also terminals of the given network. To show that this is so, consider Eqs. (208) for $m = r$:

$$\overline{i_m(j\omega; \alpha_1, \alpha_2, \cdots)i_m(-j\omega; \alpha_1, \alpha_2, \cdots)} = 2kTg_{mm}(j\omega; \alpha_1, \alpha_2, \cdots). \quad (219)$$

By definition [see Eq. (198)], $g_{mm}(j\omega; \alpha_1, \alpha_2, \ldots)$ is precisely that conductance between the node m and all other nodes short-circuited together. This conductance is always positive for a passive network. It follows, therefore, that

$$i_m(-j\omega; \alpha_1, \alpha_2, \cdots) = i_m(j\omega; \alpha_1, \alpha_2, \cdots) \quad (220)$$

and that $I_m(t; \alpha_1, \alpha_2, \ldots)$ is invariably a real function of time. This real character of the nodal currents $I_m(t; \alpha_1, \alpha_2, \ldots)$ is not incompatible with either the sometimes complex character of the internodal currents $I_{mr}(t; \alpha_1, \alpha_2, \ldots)$ or the sometimes nonpassive character of the apparent branch admittances $-y_{mr}(j\omega; \alpha_1, \alpha_2, \ldots)$, because no inconsistency arises in either Eqs. (218) or (198) under these circumstances.

The Quasi-thermal Noise in Passive Networks That Are Not in Thermal Equilibrium.—Under certain circumstances the theory of thermal noise can be applied to situations in which true thermal equilibrium does not hold. For example, a passive circuit in which the different resistive elements are at different temperatures is not, strictly speaking, in true thermal equilibrium even if the different resistive elements are maintained at their respective temperatures by some artificial means. But if they are so maintained, the noise power that each individual resistance emits may be taken equal to that which would occur in the case of equilibrium at its particular temperature, although the noise power absorbed by the same resistance will not be equal to that emitted, as in true thermal equilibrium.

Thus, if a linear passive network is considered to have a nonuniform temperature distribution of such a character that each of its resistance elements R_n ($n = 1, 2, \ldots, N$) is maintained at a constant absolute temperature T_n, then Nyquist's theorem can be applied to each resistance R_n at its appropriate temperature T_n. In other words, a series thermal-noise emf with a mean square per unit-frequency interval equal to

$$\overline{|e_n(j\omega)|^2} = 2kT_nR_n \quad (221)$$

may be associated with each such resistance.

On this premise, it is then possible, as was first shown by Williams,[1] to find an analytical expression for the mean square value of the open-circuit "thermal" noise voltage appearing across any two points of the given network. To find this expression, let the given network be represented as shown in Fig. 12·7, where we consider two terminal pairs (m,m), (n,n). Let the terminal pair (m,m) represent any two points

FIG. 12·7.—Arbitrary linear passive network showing one component resistance R_n at temperature T_n, regarded as input with terminals (n,n); and any two points of the network (m,m) regarded as output terminals. S is a short-circuiting switch.

of the network across which it is desired to evaluate the open-circuit thermal-noise voltage, and let the terminal pair (n,n) represent the terminals of any one component resistance R_n. Now let $Y_{nm}(j\omega)$ be the transfer admittance between the generator $e_n(j\omega)$ and the output terminals (m,m), short-circuited by the switch S. The mean square value per unit-frequency interval of the short-circuit thermal-noise current $i_n(j\omega)$ which will flow through the switch S, because of the generator $e_n(j\omega)$, can be expressed in terms of this transfer admittance. The result, from Eq. (221), is

$$\overline{|i_n(j\omega)|^2} = 2kT_nR_n|Y_{nm}(j\omega)|^2. \tag{222}$$

Since a similar result holds for each resistance R_n, the net short-circuit thermal-noise current

$$i(j\omega) = \sum_n i_n(j\omega) \tag{223}$$

has a mean square per unit-frequency interval equal to

$$\overline{|i(j\omega)|^2} = \sum_n \sum_k \overline{i_n(j\omega)i_k(-j\omega)} \tag{224a}$$

$$= \sum_n \sum_k Y_{nm}(j\omega)Y_{km}(-j\omega)\overline{e_n(j\omega)e_k(-j\omega)}. \tag{224b}$$

But since, for different resistances R_n and R_k, the thermal-noise voltages $e_n(j\omega)$ and $e_k(j\omega)$ are uncorrelated, i.e.,

$$\overline{e_n(j\omega)e_k(-j\omega)} = 0 \qquad (n \neq k), \tag{225}$$

Eq. (224b) reduces to

$$\overline{|i(j\omega)|^2} = \sum_n |Y_{nm}(j\omega)|^2\overline{|e_n(j\omega)|^2} \tag{226a}$$

[1] F. C. Williams, *Jour. Inst. Elec. Eng.*, **81**, 751 (1937).

$$= \sum_n 2kT_nR_n|Y_{nm}(j\omega)|^2. \tag{226b}$$

Knowing now the short-circuit current $i(j\omega)$, the mean square thermal-noise voltage appearing across the open-circuited terminal pair (m,m) can be evaluated by an application of Thevenin's theorem. According to this theorem, this open-circuit voltage $e(j\omega)$ is related to the short-circuit current $i(j\omega)$ by the equation

$$e(j\omega) = Z(j\omega)i(j\omega), \tag{227}$$

where $Z(j\omega)$ is the driving-point impedance at the terminal pair (m,m). Thus, by Eqs. (226b) and (227), it is found that

$$\overline{|e(j\omega)|^2} = \sum_n 2kT_nR_n|Y_{nm}(j\omega)|^2|Z(j\omega)|^2, \tag{228}$$

which is the desired result.

In the special case of true thermal equilibrium, when all the resistances R_n are at a common temperature T, Eq. (228) reduces to

$$\overline{|e(j\omega)|^2} = 2kT \sum_n R_n|Y_{nm}(j\omega)|^2|Z(j\omega)|^2. \tag{229}$$

But by Nyquist's theorem, it should be possible to write this

$$\overline{|e(j\omega)|^2} = 2kTR(j\omega), \tag{230}$$

where $R(j\omega)$ is the real part of $Z(j\omega)$. This equivalence will hold if

$$R(j\omega) = \sum_n R_n|Y_{nm}(j\omega)|^2|Z(j\omega)|^2. \tag{231}$$

It is shown below that Eq. (231) is identically true, following a proof given by Williams.

Referring again to the network of Fig. 12·7, let $Y_{mn}(j\omega)$ be the transfer admittance from the terminals (m,m) to the terminals of the generator $e_n(j\omega)$. Now, since the given network is composed entirely of bilateral elements,

$$Y_{mn}(j\omega) = Y_{nm}(j\omega). \tag{232}$$

Therefore, if a voltage

$$e = E \sin \omega t \tag{233}$$

is impressed across the terminals (m,m), the power dissipated in the resistance R_n is, on the average,

$$P_n = \frac{R_n E^2 |Y_{mn}(j\omega)|^2}{2} \tag{234a}$$

$$= \frac{R_n E^2 |Y_{nm}(j\omega)|^2}{2}, \tag{234b}$$

and the total power dissipated in the network will be

$$P = \sum_n P_n = \frac{E^2}{2} \sum_n R_n |Y_{nm}(j\omega)|^2. \tag{235}$$

But if $Z(j\omega)$ is written

$$Z(j\omega) = R(j\omega) + jX(j\omega), \tag{236}$$

the same total power can be written also

$$P = \frac{\frac{E^2}{2} R(j\omega)}{|Z(j\omega)|^2}. \tag{237}$$

Therefore, if expressions in Eqs. (235) and (237) are equated, it is found that

$$R(j\omega) = \sum_n R_n |Y_{nm}(j\omega)|^2 |Z(j\omega)|^2, \tag{238}$$

which is the same result as Eq. (231).

This result, in combination with the result in Eq. (229), provides, in effect, an independent proof of Nyquist's theorem.

12·3. Shot Noise. *Definition.*—In the general definition of noise in Sec. 12·1 it was stated that the essential characteristic of any source of noise is that it involves large aggregations of small charged particles moving under the influence of their own mutual forces or of other forces of comparable complexity or randomness. Rather general equilibrium conditions can be used to evaluate the special case of thermal noise just treated, without considering in detail the behavior of the individual charged particles. Another fundamental type of noise, which, in contrast to thermal noise, emphasizes much more than individual behavior of each discrete charged particle, is considered here. This type of noise, called "shot noise," is of prime importance not only because of its frequent occurrence in practice but also because the study of it opens up an important body of theory, complementary to the theory of thermal noise.

The simplest example of shot noise, which has been called "pure shot noise," is found in a temperature-limited diode, in which electrons are emitted from a hot cathode and collected at an anode in a completely unilateral flow. The flow may be one in which the electron trajectories are parallel, as in a planar diode, or it may have various other geometries.

such as that in which the trajectories are radial, as in a concentric-cylinder diode. In any case, the essential feature is that the flow of electrons is unilateral, starting at one surface and ending at another. The result of requiring that the diode be temperature-limited is, in the first place, that every electron emitted by the cathode goes to the anode. Secondly, each electron moves at so great a velocity v that for any finite current density,

$$j = \rho v, \tag{239}$$

the associated charge density ρ will be negligibly small. As a result, the potential distribution V between the two electrodes, determined by Poisson's equation

$$\nabla^2 V = -4\pi\rho, \tag{240}$$

will be a solution of Laplace's equation

$$\nabla^2 V = 0 \tag{241}$$

depending only on the diode geometry and the potential difference and not upon the electrons themselves.

Actually, the situation may not be so simple because, even if the charge density ρ is negligible, the time variation of the potential V may not be negligible. It might, therefore, be better to write instead of Eq. (241) the wave equation

$$\nabla^2 V - \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2} = 0· \tag{242}$$

More than this, the vector potential A, which may also act upon the electrons and obeys the D'Alembert equation

$$\nabla^2 A - \frac{1}{c^2}\frac{\partial^2 A}{\partial t^2} = -\frac{4\pi j}{c}, \tag{243}$$

must be considered.

These more exact considerations may be neglected, at least except at very high frequencies, however, because in the Lorentz equation for the force F acting on each electron, i.e.,

$$F = e\left(E + \frac{v \times H}{c}\right), \tag{244}$$

the electric field

$$E = -\nabla V - \frac{1}{c}\frac{\partial A}{\partial t} \tag{245}$$

is to a good approximation given by

$$E = -\nabla V, \tag{246}$$

where V is the static potential distribution given by Eq. (241), and the

magnetic field

$$H = \nabla \times A \qquad (247)$$

is so small in comparison to E that it may be neglected in Eq. (244).

That is to say, then, that as far as their low-frequency components are concerned, the electron trajectories in a temperature-limited diode are determined solely by the static scalar potential distribution V and not by the electrons themselves. The flight of each electron from cathode to anode, therefore, may be regarded as an independent event. More than this, it is usually permissible to assume that the different electrons exhibit, in their flight, no difference in behavior or in their effect upon the circuit.

Hence, it becomes possible to analyze the shot noise in the temperature-limited diode by considering a single typical electron k, emitted at the cathode at a time t_k, pursuing a typical trajectory and giving rise to a typical response in the diode circuit. The extension of the analysis to all emitted electrons is then a matter of summing all these similar responses, which differ only in respect to the emission time t_k. Since these emission times t_k have a Poisson probability distribution, a statistical treatment of them is capable of yielding average values of the resulting shot-noise effects.

From this simple example of shot noise, it can be seen that the typical characteristic of shot-noise phenomena in general is one that emphasizes the electrodynamic properties of the individual electrons (or, in some cases, other charged particles) and that the typical methods of treatment of shot noise should commence, whenever possible, with the selection of groups of similar, independent electrons.

Observation of Shot Noise in Diodes.—This form of noise, like thermal noise, is readily observed with the aid of a high-gain radio receiver. If a temperature-limited diode is placed across the input terminals of the receiver, the intensity of the audible noise in the loud-speaker will increase as the direct current in the diode is increased, being a minimum when no diode current flows at all. The increase in audible noise with increasing diode current is thus an indication of the shot noise contributed separately by the diode.

Shot Noise in a Parallel Resonant Circuit. Measurement of Open-circuit Shot-noise Voltage.—A parallel resonant circuit, with which a temperature-limited diode is in parallel, can be used to measure open-circuit shot-noise voltage. Such a diode contains an inherent capacitance, which can be included as part of the total capacitance C of the parallel resonant circuit. Under temperature-limited conditions, the plate resistance of the diode will be so high that it may be neglected. Therefore the total shunt resistance R of the parallel resonant circuit may be regarded as containing no contribution from the diode, except possibly

due to small ohmic losses, in which case all of the resistance is represent-
able as a single shunt resistance R, passive and frequency-independent
and in thermal equilibrium at a temperature T.

Following further the procedure of measurement used earlier in
conjunction with thermal noise, let the above parallel resonant circuit
containing the diode be attached to the input terminals of a high-gain r-f
amplifier. As before, let the output indicating device be a thermo-
couple meter M, calibrated to read the mean square output voltage of
the amplifier. The experimental arrangement is then as shown in Fig.
12·8, where R is the net parallel resistance of the parallel resonant input
circuit, L is an essentially loss-free coil, and C is the net parallel capaci-
tance of the input circuit (inclusive of the diode capacitance and the
input capacitance of the first amplifier tube) and where the whole input
circuit (excluding the diode and the input amplifier tube) is in thermal
equilibrium at an absolute temperature T.

Under these conditions the open-circuit noise voltage will be a function
of time $V(t)$, which contains two uncorrelated parts, $V_1(t)$ and $V_2(t)$.

$$V(t) = V_1(t) + V_2(t), \tag{248}$$

where $V_1(t)$ is the open-circuit thermal-noise voltage, which was pre-
viously analyzed in detail, and where $V_2(t)$ is the open-circuit shot-noise
voltage, which is examined below.

If now, pursuing the same method of measurement employed above in
connection with thermal noise, the statistical spectrum $\bar{S}(\omega)$ of the open-
circuit noise voltage $V(t)$ is measured as a function of the parameters R,
L, and C of the resonant circuit and as a function of the temperature T
and the diode direct current I, it is found that

$$\bar{S}(\omega) = \frac{2kTR + eIR^2}{1 + R^2\left(\omega C - \dfrac{1}{\omega L}\right)^2} \tag{249}$$

where k is Boltzmann's constant and e is the electronic charge.

Because the noise voltages $V_1(t)$ and $V_2(t)$ are statistically inde-
pendent, it follows that their statistical spectra $\bar{S}_1(\omega)$ and $\bar{S}_2(\omega)$ add
linearly, according to the relation

$$\bar{S}(\omega) = \bar{S}_1(\omega) + \bar{S}_2(\omega). \tag{250}$$

When Eqs. (249) and (250) are compared, and it is remembered that

$$\bar{S}_1(\omega) = \frac{2kTR}{1 + R^2\left(\omega C - \dfrac{1}{\omega L}\right)^2}, \tag{251}$$

then

$$\tilde{S}_2(\omega) = \frac{eIR^2}{1 + R^2\left(\omega C - \dfrac{1}{\omega L}\right)^2} \tag{252}$$

is found as the experimental result for the statistical spectrum of the *open-circuit shot-noise voltage* in Fig. 12·8.

The open-circuit shot-noise voltage $V_2(t)$, then, possesses a statistical spectrum equal to

$$\frac{d\overline{V_2^2}}{df} = \frac{eIR^2}{1 + R^2\left(\omega C - \dfrac{1}{\omega L}\right)^2}, \tag{253}$$

and a total mean square equal to

$$\overline{V_2^2} = \int_{-\infty}^{+\infty} \frac{eIR^2 \displaystyle\int_{-\infty}^{+\infty} df}{1 + R^2\left(\omega C - \dfrac{1}{\omega L}\right)^2}$$

$$= \frac{eIR}{2C}. \tag{254}$$

Localization of the Shot-noise Constant-current Generator.—If, as in the analysis of thermal noise, the respective limiting cases of Eq. (253) for an isolated resistance R and an isolated LC-circuit are considered here, it is found that in the first case ($L = \infty$, $C = 0$)

$$\frac{d\overline{V_2^2}}{df} = eIR^2 \tag{255}$$

and in the second case ($R = \infty$)

$$\frac{d\overline{V_2^2}}{df} = \frac{eI}{\left(\omega C - \dfrac{1}{\omega L}\right)^2}. \tag{256}$$

In neither of these limiting cases is there any evidence of a source of noise intrinsic to the elements involved (R in the first case or L and C in the second) as was found in the case of thermal noise. Rather, there is evidence of an external source of noise, depending on the diode current I. The appropriate representation of this source, as seen from either the
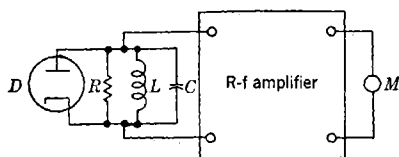
Fig. 12·8.—Temperature-limited shot noise in a parallel resonant circuit as measured with an r-f amplifier.

limiting Eqs. (255) and (256) or the general Eq. (253), is in the form of a

shunt constant-current generator $I(t)$ with a mean square per unit-frequency interval equal to

$$\frac{\overline{dI^2}}{df} = eI. \tag{257}$$

Such a current generator, impressed across the RLC-circuit, does, in fact, lead to the result of Eq. (253). Similarly, if it is impressed across the isolated resistance R or the isolated LC-circuit, it leads to the result of Eq. (255) or (256), respectively.

Hence Eq. (257) gives the experimental determination of the source of shot noise in a temperature-limited diode. This result agrees with that which was first deduced theoretically by Schottky[1] in 1918.

Theory of Temperature-limited Shot Noise.—A theoretical proof of the result Eq. (257) is given here. To do this, the individual electrons emitted by the cathode and the effect of the passage of these electrons upon the diode circuit will be examined. The effect of the passage of any one electron k upon the diode circuit may be expressed in terms of the total current $I_k(t)$ which flows through a short-circuiting external path between anode and cathode as a result of this passage. From the Maxwell equation

$$\nabla \times H = \frac{4\pi j + \dfrac{\partial E}{\partial t}}{c}, \tag{258}$$

where H is magnetic-field strength, j is current density, and E is electric-field strength, it follows that

$$\nabla \cdot \left(j + \frac{1}{4\pi} \frac{\partial E}{\partial t} \right) = 0. \tag{259}$$

If Eq. (259) is integrated across the area A normal to the lines of flow of the total current density, $j + (1/4\pi)\partial E/\partial t$, it is found that the integrated total current

$$I_k(t) = \int j \cdot dA + \frac{1}{4\pi} \int \frac{\partial E}{\partial t} \cdot dA \tag{260}$$

is the same at all cross sections of the current flow at the same instant t. That is, $I_k(t)$ is divergenceless, which means that the passage of the electron k sets up a circulatory current, flowing from anode to cathode within the diode and returning from cathode to anode through the external short-circuiting path. Moreover, this current $I_k(t)$ is the same at every point of this circulatory path at any instant t. The part of the path within the diode corresponds to the constant-current generator of Eq. (257) as long as the potential difference induced between the elec-

[1] W. Schottky, *Ann. phys.* **57**, 541 (1918); **68**, 157 (1922).

trodes does not react on the electron in its flight.   It is justifiable to
assume this condition, at least at low frequencies.

If Eq. (259) is integrated with respect to the time t, an expression
having the form of a charge results:

$$Q_k(t) = \int I_k(t)\, dt = \int \int j \cdot dA\, dt + \frac{1}{4\pi} \int E \cdot dA. \qquad (261)$$

Here the first term refers to the fraction of the electrons's own charge $-e$
which is deposited on the anode at any time t.   This fraction, of course, is
either zero or one, according to whether the electron has reached the
anode or not.   If the first term is called $Q'_k(t)$, then $Q'_k(t)$ is a function of
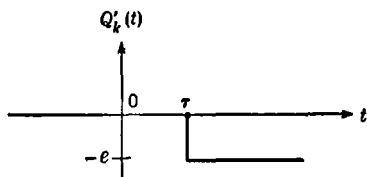
FIG. 12·9.—Convection charge due to
the passage of single electron k from
cathode to anode.
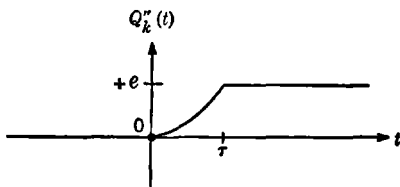
FIG. 12·10.—Induced charge due to the
passage of a single electron k from cathode
to anode.

time having the form shown in Fig. 12·9 in which. for convenience,
the emission time t_k has been set equal to zero.

The second term in Eq. (261) can be written

$$Q''_k(t) = \frac{1}{4\pi} \int E \cdot dA. \qquad (262)$$

It is $1/4\pi$ times the total electric flux incident on the anode at the time t,
and hence it is the net charge induced on the anode by the electron.   A
qualitative picture of what this induced charge is at any time t, corre-
sponding to the various successive positions of the electron in the cathode-
anode space, can be obtained.   If, as assumed above, the electron
departs from the cathode at the time $t = t_k = 0$, then at that time and
at all previous times the charge $Q''_k(t)$ induced on the anode is zero, as
indicated in Fig. 12·10.   As t increases beyond $t = 0$, however, an
increasing fraction of the opposite charge $+e$ will be induced on the
anode.   This fraction will increase monotonically until, at the transit
time τ, the whole charge $+e$ is induced.   But precisely at the time τ and
for all times thereafter, the electron comes to rest in front of and infinitely
close to the anode.   Accordingly, for $t \geqq \tau$, the induced charge remains
equal to $+e$, as shown in Fig. 12·10.

The charge $Q'_k(t)$, as represented in Fig. 12·9, applies strictly to a
plane in front of and infinitely close to the anode but not to the anode

itself, whereas the charge $Q_k''(t)$, as represented in Fig. 12·10, applies to the anode itself. As far as the anode itself is concerned, the charge $Q_k'(t)$ should be represented as always zero, because for the purpose of establishing $Q_k''(t)$, it has been assumed that the electron stops just before reaching the anode. Accordingly, the total charge $Q_k(t)$ is equal to $Q_k''(t)$ when computed at the anode. This would *not* be the case, however, were the total charge $Q_k(t)$ to be computed at some other plane within the cathode-anode space.

There has been much confusion as to the interpretation of the charge $Q_k(t)$ resulting from the passage of a single electron k. It has been interpreted by some writers as just the induced charge $Q_k''(t)$, corresponding to the displacement current, and by others (particularly in the low-frequency limiting case) as just the charge $Q_k'(t)$, corresponding to the convection current. But it was shown above that it is actually the total charge $Q_k(t) = Q_k'(t) + Q_k''(t)$, corresponding to the total current $I_k(t)$ given by Eq. (260). In the present case, $Q_k'(t) = 0$ and $Q_k(t) = Q_k''(t)$, but this is merely an accidental result of the particular choice of plane for the evaluation of the total charge $Q_k(t)$.

The confusion occurring here disappears as soon as the exact nature of the electron's trajectory is recognized. This trajectory, in the present case, is specified by saying that for $t \leqq 0$ the electron is at rest in front of and infinitely close to the cathode, for $0 \leqq t \leqq \tau$ the electron is in motion from cathode to anode, and for $t \geqq \tau$ the electron is at rest in front of and infinitely close to the anode. The electron trajectory, described as a function $x_k(t)$, is therefore broken in character at $t = 0$ and $t = \tau$, as is shown qualitatively in Fig. 12·10, and this broken character is the point that resolves the confusion.

The analytic expression for $Q_k(t)$ in the interval $0 < t < \tau$ is ascertained next. At this time, as at all other times t, $Q_k(t) = Q_k''(t)$, the induced charge on the anode. This induced charge is easily calculated when the electron is permanently at rest at a point x somewhere between the cathode and anode. Actually, however, the electron, when at this point x, is not at rest. It is necessary, then, to ask what effect, if any, results from its motion. From the viewpoint of special relativity theory, the electron's charge $-e$ is invariant under a Lorentz transformation, and so also is the net induced charge $+e$ on the anode and cathode. But what can be said about that part of the induced charge which is on the anode alone? The answer can be found from Expression (262). For not only is the integral (262) an invariant when taken over both anode and cathode, but its integrand itself is invariant. This integrand can be written

$$dQ_k''(t) = \frac{1}{4\pi} E_x \, dy \, dz. \tag{263}$$

But in this expression E_x, dy, and dz are all unaffected by a Lorentz transformation in the x-direction.   Therefore, $dQ_k''(t)$ is Lorentz-invariant, as is its integral $Q_k''(t)$ also, when taken over the anode alone.   This result means that $Q_k''(t)$ may be computed from the point of view that the electron is at rest at a point x and hence may be computed as a function

Fig. 12·11.—Potential distribution in a plane parallel diode under space charge conditions.

$$Q_k''(t) = f(x) \qquad (264)$$

and then, using the actual dependence of x on t, written as a function of t.

Now from electrostatics it is known that

$$f(x) = e\frac{x}{d}, \qquad (265)$$

where d is the cathode-anode spacing (see Fig. 12·11).   On the other hand, x varies with time in a way approximately prescribed by the equation of motion

$$\frac{md^2x}{dt^2} = -eE, \qquad (266)$$

where m is the electronic mass and $E = const.$ is the electrostatic field in the diode space, due to the fixed potential difference $V = Ed$.   Therefore the electron, starting at rest, has a trajectory given by

$$x = \frac{-eE}{m}\frac{t^2}{2} = Kt^2, \qquad (267)$$

where K is a positive constant (since E is intrinsically negative).   Accordingly, from Eqs. (264), (265), and (267),

$$Q_k''(t) = \frac{eK}{d}t^2 \qquad (0 < t < \tau) \qquad (268)$$

From Eq. (267),

$$d = K\tau^2, \qquad (269)$$

so Eq. (268) may be written as

$$Q_k''(t) = e\left(\frac{t}{\tau}\right)^2, \qquad (270)$$

which gives $Q_k'' = e$ when $t = \tau$, as it should.

Thus, by differentiating Eq. (270) with respect to t, it is found that

$$I_k''(t) = 2\frac{e}{\tau}\frac{t}{\tau} \qquad (0 < t < \tau) \qquad (271)$$

is the displacement current flowing into the anode.   For $t < 0$ and for $t > \tau$, the displacement current $I_k''(t)$ vanishes (see Fig. 12·10).   This is also the total current $I_k(t)$ that would be detected in a short-circuiting ammeter placed between anode and cathode, and this current, therefore, is to be identified as the contribution to Eq. (257), due to a single electron.

The behavior of the Fourier transform of $I_k(t)$ is examined next. This is

$$i_k(j\omega) = \int_0^\tau I_k(t)e^{-i\omega t}\,dt, \tag{272}$$

or, from Eq. (271),

$$i_k(j\omega) = \frac{2e}{\theta^2} \int_0^\theta ue^{-iu}\,du$$
$$= \frac{2e}{\theta^2}\left[(\cos\theta - 1 + \theta\sin\theta) - j(\sin\theta - \theta\cos\theta)\right], \tag{273}$$

where $\theta = \omega\tau$ is the transit angle.   As $\theta \to 0$, this tends to

$$i_k(j\omega) = e. \tag{274}$$

Hence, for infinitely small transit time τ, $I_k(t)$ can be written

$$I_k(t) = \frac{e}{2\pi}\int_{-\infty}^{+\infty} e^{i\omega t}\,d\omega, \tag{275}$$

and this is e times a delta-function of time, zero everywhere except at the emission time $t = 0$, but with a time integral equal to e.

The spectrum of the current pulse $I_k(t)$ is given, in general, by

$$|i_k(j\omega)|^2 = \left(\frac{2e}{\theta^2}\right)^2\left[(\cos\theta - 1 + \theta\sin\theta)^2 + (\sin\theta - \theta\cos\theta)^2\right]$$
$$= \left(\frac{2e}{\theta^2}\right)^2\left[\theta^2 + 2(1 - \cos\theta - \theta\sin\theta)\right]. \tag{276}$$

As $\theta \to 0$, this tends to

$$|i_k(j\omega)|^2 = e^2, \tag{277}$$

a result in agreement with Eq. (274).   Hence, for infinitely small transit time τ, $I_k(t)$ has a constant spectrum equal to e^2.

Hitherto the time $t = 0$ has been taken as the time of emission of a single electron from the cathode.   The case in which electrons are emitted at random times t_k is now considered.   Then the total current $I_k(t)$ due to the kth electron, emitted at time t_k, has to be written

$$I_k(t) = 2\frac{e}{\tau}\frac{t - t_k}{\tau} \qquad (t_k < t < t_k + \tau). \tag{278}$$

The Fourier transform $i_k(j\omega)$ is now

$$i_k(j\omega) = \int_{t_k}^{t_k+\tau} I_k(t)e^{-j\omega t}\, dt \tag{279a}$$

$$= \frac{2e}{\theta^2} \int_{u_k}^{u_k+\theta} (u - u_k)e^{-ju}\, du \tag{279b}$$

where $\theta = \omega\tau$, $u = \omega t$, $u_k = \omega t_k$. But if $v = u - u_k$, Eq. (279b) becomes

$$i_k(j\omega) = \frac{2e}{\theta^2}\, e^{-j\omega t_k} \int_0^\theta ve^{-jv}\, dv. \tag{280}$$

That is, it is possible to write

$$i_k(j\omega) = i(j\omega)e^{-j\omega t_k}, \tag{281}$$

where $i(j\omega)$ is given by Eqs. (271) to (273).

Hence for the total current due to all electrons,

$$I(t) = \sum_k I_k(t) \tag{282a}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} i(j\omega)\, d\omega\, e^{j\omega t} \sum_k e^{-j\omega t_k} \tag{282b}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} i(j\omega)e^{j\omega t}\, d\omega S(\omega), \tag{282c}$$

where $S(\omega)$ is the random function

$$S(\omega) = \sum_k e^{-j\omega t_k}. \tag{283}$$

To find average values, either time averages of $I(t)$ or ensemble averages of $S(\omega)$ may be taken, since these two kinds of average are here equivalent. Thus, for the time average of $I(t)$ it is possible to write

$$\overline{I(t)} = I = \frac{1}{2\pi} \int_{-\infty}^{+\infty} i(j\omega)e^{j\omega t}\, d\omega \overline{S(\omega)}, \tag{284}$$

where $\overline{S(\omega)}$ is the ensemble average of $S(\omega)$. Now when $\omega \neq 0$, $\overline{S(\omega)} = 0$, but the integral of $\overline{S(\omega)}$ is finite, like that of a delta-function. Hence, taking the value of the integrand of Eq. (284) for $\omega = 0$,

$$\frac{1}{2\pi} \int \overline{S(\omega)}\, d\omega = \frac{I}{e}, \tag{285}$$

where I is the average or direct current flowing in the diode.

Now consider the expression for the square of the instantaneous total

current.   From Eq. (282c) this will be of the form

$$I^2(t) = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} i(j\omega)i(j\omega')e^{j(\omega+\omega')t} \, d\omega \, d\omega' S(\omega)S(\omega'). \quad (286)$$

Thus, for the time average of $I^2(t)$, it is possible to write

$$\overline{I^2(t)} = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} i(j\omega)i(j\omega')e^{j(\omega+\omega')t} \, d\omega \, d\omega' \overline{S(\omega)S(\omega')} \quad (287)$$

where $\overline{S(\omega)S(\omega')}$ is the ensemble average of $S(\omega)S(\omega')$ for any values of ω and ω'.   In general, for independently selected values of ω and ω',

$$\overline{S(\omega)S(\omega')} = \overline{S(\omega)} \ \overline{S(\omega')} = 0 \quad (288)$$

as long as neither ω nor ω' vanishes and also provided $(\omega + \omega')$ does not vanish.   This means that the double integral of Eq. (287) contributes nothing except along the three loci

$$\omega = 0, \quad (289a)$$
$$\omega' = 0, \quad (289b)$$
$$\omega + \omega' = 0. \quad (289c)$$

Taking the first locus, Eq. (289a), and using the result, Eq. (285),

$$\overline{I^2(t)}_{\omega=0} = \frac{I}{2\pi} \int_{-\infty}^{+\infty} i(j\omega')e^{j\omega't} \, d\omega' \overline{S(\omega')} \quad (290)$$

is found for the contribution along this locus.   But the integrand of this result vanishes unless $\omega' = 0$, so that the two loci, Eqs. (289a) and (289b), likewise contribute nothing except at their common point of intersection, $\omega = \omega' = 0$.   At this common point of intersection, Eq. (290) becomes

$$\overline{I^2(t)}_{\omega=\omega'=0} = I^2. \quad (291)$$

That is to say, the contribution to Eq. (287) at the origin $\omega = \omega' = 0$ is just the square of the average or direct current.

The only remaining contribution is along the locus of Eq. (289c), now excluding the origin.   But along this locus, $\omega' = -\omega$, $S(\omega)S(\omega')$ becomes, according to Eq. (283),

$$S(\omega)S(\omega') = \sum_m \sum_n e^{j\omega(t_m-t_n)}. \quad (292)$$

The ensemble average of this may be readily shown to be identical with $\overline{S(0)}$, which behaves like the improper part of the delta-function.   Hence the contribution to Eq. (287) along the locus Eq. (289c) can be written

$$\overline{I^2(t)}_{\omega+\omega'=0} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} i(j\omega)i(-j\omega) \, d\omega \, \frac{1}{2\pi} \int \overline{S(0')} \, d\omega', \quad (293)$$

where the second integral, evaluated at the point $\omega' = -\omega$, gives the result

$$\frac{1}{2\pi} \int S(0)\, d\omega' = \frac{I}{e}. \tag{294}$$

Hence we have, for the a-c fluctuations in the diode current, a mean square given by (dropping the subscript $\omega + \omega' = 0$)

$$\overline{I^2(t)} = \frac{1}{2\pi} \frac{I}{e} \int_{-\infty}^{+\infty} |i(j\omega)|^2\, d\omega, \tag{295}$$

or, for the mean square fluctuation per unit-frequency interval,

$$\frac{d\overline{I^2}}{df} = \frac{I}{e} |i(j\omega)|^2. \tag{296}$$

But from Eq. (276), this becomes

$$\frac{d\overline{I^2}}{df} = eI \frac{2}{\theta^2} [\theta^2 + 2(1 - \cos\theta - \theta\sin\theta)]. \tag{297}$$

Expression (297) is the desired result. It gives the statistical spectrum of temperature-limited shot noise, or so-called "pure shot noise," for any value of the transit angle θ. At low frequencies, when θ is negligible, the result (297) reduces to

$$\frac{d\overline{I^2}}{df} = eI, \tag{298}$$

which agrees with the experimental result (257) for the low-frequency case.

Space-charge Reduction of Shot Noise.[1]—In the case of the temperature-limited diode just treated, the conditions were of the greatest possible simplicity: (1) Space charge was completely negligible so that there was no interaction among the electrons. (2) All electrons had essentially the same trajectory, ending at the anode. (3) The effects of different electrons were additive after the fashion of mutually independent random events.

The question of what happens at low frequencies when the diode potential difference is not sufficiently great to eliminate space charge is considered next. In this case, each electron moves in an electrostatic field that is due in part to the diode potential difference and also in part to other emitted electrons. The potential distribution between cathode and anode is now established from Poisson's equation

$$\nabla^2 V = -4\pi\rho, \tag{299}$$

D. O. North, *RCA Rev.*, **4**, 4, April 1940; **5**, 1, July 1940; also W. Schottky, *Wiss. Veröff. Siemens-Werke*, **16**(2), 1 (1937); F. Spenke, *Wiss. Veröff. Siemens-Werke*, **16**(2), 19 (1937).

where ρ is the electronic space-charge density, a function of position in the interelectrode space.   Moreover, in contrast to the temperature-limited case, the spread in emission velocities of the different electrons must also be taken into account.   In the temperature-limited case, the potential difference was so great that this velocity spread was of negligibly small importance.

In a parallel plane diode, the potential distribution, that is, the solution V of Poisson's equation (299), is in general a function of the distance x between cathode and anode, of the form shown in Fig. 12·11.   There is usually a potential minimum, at a distance x_m from the cathode (d is the anode-cathode distance) and with a negative value V_m with respect to the cathode (V_a is the anode-cathode potential difference).

Unless an emitted electron has an initial kinetic energy in excess of $-eV_m$, it will not go to the anode but will return to the cathode.   Hence, in general, there are two groups of emitted electrons: (1) those of lower energy than $-eV_m$ which return to the cathode (α-electrons) and (2) those of higher energy than $-eV_m$ which are collected at the anode (β-electrons).

The distribution in kinetic energy of the emitted electrons is given by the Maxwell-Boltzmann law, which, in this case, can be written

$$\frac{dI}{dE} = I_s e^{-E}, \tag{300}$$

where

$$E = \frac{eV}{kT}, \tag{301a}$$

$$dE = \frac{e}{kT}\, dV, \tag{301b}$$

and where eV is the initial kinetic energy of the emitted electrons, k is Boltzmann's constant, T is the absolute cathode temperature, and dI is the emission current corresponding to the range of initial kinetic energies edV.   The meaning of I_s is found by integrating Eq. (300) over all emission energies from zero to infinity; it is simply the total emission current.

According to Fig. 12·11, the current collected at the anode is just

$$I = \int_{E_m}^{\infty} I_s e^{-E} dE = I_s e^{-E_m}, \tag{302}$$

where

$$E_m = -\frac{eV_m}{kT} \tag{303}$$

is the lower integration limit for the β-electrons.

It might be supposed that the effect of the space charge and in par-

ticular of the potential minimum is to separate the electrons into α- and β-groups in a fixed way and that the problem of ascertaining the resulting noise could be handled by treating the α- and β-groups of electrons as independent sources of noise, more or less along the lines employed above for the temperature-limited case.   But this would be to overlook the fact that *each individual electron alters momentarily the potential minimum* V_m *by its entry into the space-charge region.*   This fact is, at bottom, the sole cause of the space-charge reduction in the diode shot noise.

For suppose that a β-electron of (normalized) kinetic energy

$$ E = \frac{eV}{kT} $$

is allowed to traverse the potential minimum.   When it does so, it alters momentarily the space-charge density ρ and therefore also the potential distribution $V(x)$ *everywhere* in the diode space, according to Poisson's equation (299).   Along with this change, there is of course a change in the potential minimum V_m.   Hence, by Eq. (302) there will be a departure from the equilibrium current I collected at the anode.   This effect may be expressed in the following way.

If there were no such effect on the potential minimum, the line of thought of the preceding paragraph could be followed.   Then it could be said that the effect of the given β-electron would be to set up in the external anode-cathode short-circuit path a total current pulse $I_\beta^0(t)$ with a Fourier transform which, at low frequencies and disregarding its time phase, can be written

$$ i_\beta^0(j\omega) = e. \tag{304} $$

Now, however, a correlated correction term must be included with $i_\beta^0(j\omega)$ due to the departure from the equilibrium value of the direct diode current I, caused by $i_\beta^0(j\omega)$.   This departure from equilibrium is proportional to $i_\beta^0(j\omega)$, with a proportionality factor that can be written $-\delta_\beta$, signifying by the subscript β that it refers to a β-electron (of a given energy) and signifying by the minus sign that the correction term is usually of a compensatory nature.   Therefore the net current developed in the external short-circuiting path by the passage of the given β-electron must be assigned a (low-frequency) Fourier transform equal to

$$ i_\beta(j\omega) = i_\beta^0(j\omega)(1 - \delta_\beta) = e(1 - \delta_\beta). \tag{305} $$

Similarly, if an α-electron of (normalized) kinetic energy $E = eV/kT$ is emitted into the space-charge region, it will also cause a departure from the equilibrium current I to the anode, even though it does not traverse the potential minimum.   If there were no such effect on the equilibrium current I, the line of thought of the preceding two paragraphs

could again be followed. The effect of the given α-electron could be said to induce in the anode a displacement current $I_\alpha^0(t)$ with a Fourier transform which, at low frequencies and disregarding its time phase, can be written

$$i_\alpha^0(j\omega) = ef(j\omega), \tag{306}$$

where $f(j\omega)$ expresses the actual frequency dependence of the resulting displacement current, which must tend to zero with the frequency. Here again, however, a correlated correction term must be included with $i_\alpha^0(j\omega)$ due to the departure from the equilibrium value of the direct diode current I, caused by $i_\alpha^0(j\omega)$. This departure from equilibrium leads, in analogy with Eq. (305), to a net current developed in the external short-circuiting path by the passage of the given α-electron, whose (low-frequency) Fourier transform can be written

$$i_\alpha(j\omega) = e[f(j\omega) - \delta_\alpha]. \tag{307}$$

But since the low-frequency approximation for $f(j\omega)$ is zero, this becomes

$$i_\alpha(j\omega) = -e\delta_\alpha. \tag{308}$$

Now Equations (305) and (308) refer to a single electron, of given initial energy and of the α- or β-type, respectively. Actually, within any velocity class, even though considered infinitesimal, there will be a large number of emitted electrons. These may be treated in the same manner as the temperature-limited case. Following that treatment for the β-electrons of energy E, the corresponding mean square current fluctuation per unit-frequency interval is, in analogy with Eq. (296),

$$\frac{d\overline{I_\beta^2(E)}}{df} = \frac{dI(E)}{e} |i_\beta(j\omega)|^2 = edI(E)(1 - \delta_\beta)^2. \tag{309}$$

If

$$\gamma_\beta = 1 - \delta_\beta, \tag{310}$$

and if Eq. (309) is integrated over all β-electron emission energies, the total mean squared current fluctuation per unit frequency interval due to all β-electrons is found to be

$$\frac{d\overline{I_\beta^2}}{df} = eI\overline{\gamma_\beta^2}, \tag{311}$$

where

$$\overline{\gamma_\beta^2} = \frac{1}{I} \int_{E_m}^\infty \gamma_\beta^2 dI(E) = \frac{\displaystyle\int_{E_m}^\infty \gamma_\beta^2 e^{-E}\, dE}{\displaystyle\int_{E_m}^\infty e^{-E}\, dE} \tag{312}$$

is the average value of γ_β^2 over all β-electron energies.

Similarly, for the α-electrons of energy E, the mean squared current fluctuation per unit-frequency interval is

$$\frac{d\overline{I_\alpha^2(E)}}{df} = \frac{dI(E)}{e} |i_\alpha(j\omega)|^2 = edI(E)\delta_\alpha^2 \tag{313}$$

Letting

$$\gamma_\alpha = -\delta_\alpha \tag{314}$$

and integrating Eq. (313) over all α-electron emission energies, the total mean squared current fluctuation per unit-frequency interval due to all α-electrons is the expression

$$\frac{d\overline{I_\alpha^2}}{df} = eI\overline{\gamma_\alpha^2}, \tag{315}$$

where

$$\overline{\gamma_\alpha^2} = \frac{1}{I} \int_0^{E_m} \gamma_\alpha^2 \, dI(E) = \left(\frac{I_s}{I} - 1\right) \frac{\int_0^{E_m} \gamma_\alpha^2 e^{-E} \, dE}{\int_0^{E_m} e^{-E} \, dE}, \tag{316}$$

is $(I_s/I - 1)$ times the average value of γ_α^2 over all α-electron energies.

Finally, adding the mutually independent results (311) and (315), the total mean squared anode current fluctuations per unit-frequency interval, due to both α- and β-electrons, is the expression

$$\frac{d\overline{I^2}}{df} = eI\Gamma^2, \tag{317}$$

where

$$\Gamma^2 = \overline{\gamma_\alpha^2} + \overline{\gamma_\beta} \tag{318}$$

is the so-called "space-charge reduction factor."

It was intended here only to indicate the logical processes leading up to the result (317) and to give a qualitative description of the way in which the space-charge reduction factor Γ^2 should be evaluated.   For the actual evaluation of Γ^2 the reader is referred to other sources.

Shot Noise in Negative-grid Triodes.—Qualitatively, a negative-grid triode resembles closely the diode with space charge in so far as the low-frequency analysis of its shot noise is concerned.   The space-charge reduction factor Γ^2 is, in principle, arrived at in exactly the same manner as has just been indicated for the diode.   The exact calculation of Γ^2 has been carried out by North and others for normal triode operation, for tubes with a fairly high amplification factor μ.   An approximate expression for Γ^2 thus obtained is

$$\Gamma^2 = \frac{\theta}{\sigma} \frac{2kTg_m}{eI_p}, \tag{319}$$

where T is the cathode absolute temperature, g_m is the mutual conductance, I_p is the direct current to the plate, σ is a geometrical factor usually ranging from $\frac{1}{2}$ to 1, and θ is a space-charge factor for which a good approximation is given by the formula

$$\theta = 3\left(1 - \frac{\pi}{4}\right), \tag{320}$$

Accordingly, from Eqs. (317) and (319), the statistical spectrum of the plate shot-noise current is given by

$$\frac{d\overline{I_p^2}}{df} = \overline{|i_p(j\omega)|^2} = \left(\frac{\theta}{\sigma}\right)2kTg_m. \tag{321}$$

This shot-noise current, with a Fourier transform $i_p(j\omega)$, is to be regarded as a nodal current, flowing out of the plate under the condition that all three nodes (cathode c, grid g, plate p) are mutually short-circuited. If $i_c(j\omega)$, $i_g(j\omega)$, and $i_p(j\omega)$ are written for the nodal shot-noise currents flowing out of the cathode, grid, and plate, respectively, then in general

$$i_c(j\omega) + i_g(j\omega) + i_p(j\omega) = 0. \tag{322}$$

At low frequencies the statistical spectrum of $i_p(j\omega)$ is given by Eq. (321). Furthermore, at low frequencies,

$$i_g(j\omega) = 0, \tag{323}$$

since the grid, being negative, collects no electrons and therefore receives only an induced current that has no low-frequency components. Consequently, Eq. (322) becomes at low frequencies

$$i_c(j\omega) + i_p(j\omega) = 0, \tag{324}$$

whence it is seen that the cathode nodal shot-noise current has the same statistical spectrum as the plate nodal shot-noise current, namely,

$$\frac{d\overline{I_c^2}}{df} = \overline{|i_c(j\omega)|^2} = \left(\frac{\theta}{\sigma}\right)2kTg_m \tag{325}$$

Instead of dealing with nodal shot-noise currents, which are in general mutually correlated, it is preferable to find an equivalent set of internodal shot-noise currents, so chosen as to be statistically independent. The procedure for doing this has already been described in the section on the generalization of Nyquist's thermal-noise theorem. For there it was shown that if $i_m(j\omega)$ $(m = 1, 2, \cdots, N)$ are the nodal noise currents in a given network, then an equivalent set of statistically independent inter-

nodal noise currents may be set up through the defining relations

$$\overline{|i_{mn}(j\omega)|^2} = -\overline{i_m(j\omega)i_n(-j\omega)} \qquad (m \neq n) \tag{326}$$

provided the quantities on the right-hand side of Eq. (326) are positive and real.

Consequently, applying this procedure to the three nodal shot-noise currents $i_c(j\omega)$, $i_g(j\omega)$, and $i_p(j\omega)$ for the negative-grid triode at low frequencies and remarking that $i_g(j\omega)$ here vanishes, it is found that

$$\overline{|i_{cg}(j\omega)|^2} = -\overline{i_c(j\omega)i_g(-j\omega)} = 0 \tag{327a}$$

$$\overline{|i_{gp}(j\omega)|^2} = -\overline{i_g(j\omega)i_p(-j\omega)} = 0 \tag{327b}$$

$$\overline{|i_{pc}(j\omega)|^2} = -\overline{i_p(j\omega)i_c(-j\omega)} = \overline{|i_p(j\omega)|^2} = \left(\frac{\theta}{\sigma}\right) 2kTg_m. \tag{327c}$$

Hence, for a negative-grid triode at low frequencies, there is only one statistically independent internodal shot-noise generator, given by Eq. (327c).

Thus, regardless of the nature of the circuit in which the negative-grid triode is placed, that is, regardless of the impedances placed in the cathode, grid, and plate leads, the response to the shot noise generated by the triode may be found by regarding the source of shot noise as an internodal constant-current generator placed between cathode and plate and having the statistical spectrum given by Eq. (327c).

It has been a common practice to replace the true internodal shot-noise current source Eq. (327c) by an equivalent shot-noise emf in the grid-cathode input circuit. The advantages of this usage are dubious and are apt to lead to a confusion of the true state of affairs, and hence this practice is less to be recommended than the practice of using the representation (327c). However, since this device is so commonly applied, it will be briefly described here.

The current $i_{pc}(j\omega)$ in Eq. (327c) may be related to a fictitious cathode-grid voltage $e_{cg}(j\omega)$ by the equation

$$i_{pc}(j\omega) = g_m e_{cg}(j\omega). \tag{328}$$

Hence the shot noise arising in the triode, actually as the current source $i_{pc}(j\omega)$, may, for certain purposes only, be regarded as arising as an emf $e_{cg}(j\omega)$ in series with whatever impedance $z_{cg}(j\omega)$ is located in the grid-cathode branch. This emf $e_{cg}(j\omega)$ is then said to have a statistical spectrum which equals, according to Eqs. (327c) and (328),

$$\overline{|e_{cg}(j\omega)|^2} = \left(\frac{\theta}{\sigma}\right) \frac{2kT}{g_m}. \tag{329}$$

It is further found convenient by many writers, though again not subscribed to here, to regard the emf $e_{cg}(j\omega)$ as a thermal-noise emf, due to an effective equivalent resistance R_{eff}. If T_0 is room temperature (it will be recalled that T is here cathode temperature), then one might accordingly write

$$\overline{|e_{cg}(j\omega)|^2} = 2kT_0R_{\text{eff}}. \tag{330}$$

Hence, identifying Eq. (329) with Eq. (328), we find that

$$R_{\text{eff}} = \frac{\theta}{\sigma}\frac{T}{T_0}\frac{1}{g_m}. \tag{331}$$

Thus, the true source of shot noise in the triode is here replaced by a fictitious one, namely, a resistance R_{eff} at room temperature T_0, placed in series with the branch impedance $z_{cg}(j\omega)$, which is regarded as emitting thermal noise but for all other purposes is regarded as absent from the circuit.

When the shot noise generated at very high frequencies in a negative-grid triode is considered, the detailed analysis becomes much more difficult. Actually, no really satisfactory solution of the high-frequency case has yet been obtained. However, it is possible to carry out a very approximate calculation of the first-order effects of the electron transit time, assuming that the transit angle is small. The basis for this approximate analysis rests upon the assumption that the effect of the α-electrons is negligible and that the effect of the β-electrons is representable in terms of an equivalent group of β-electrons, all having a common velocity equal to the average velocity $\bar{v}$ of the actual Maxwell-Boltzmann distribution of β-electrons.

As an indication of the approximate validity of this assumption, an independent derivation[1] of the low-frequency formula (321) is carried out by means of this simplified model of a triode tube.

First, the average velocity $\bar{v}$ of the β-electrons is defined by the formula

$$\bar{v} = \frac{\displaystyle\int_{E_m}^{\infty} v_m \, dI(E)}{\displaystyle\int_{E_m}^{\infty} dI(E)}, \tag{332}$$

where v_m is the velocity of the β-electrons at the potential minimum when their initial velocity is $v(E)$. That is,

$$\tfrac{1}{2}mv_m^2 = (E - E_m)kT. \tag{333}$$

Hence,

$$v_m = \left[(E - E_m)\left(\frac{2kT}{m}\right)\right]^{1/2} \tag{334}$$

[1] *Cf.* also C. J. Bakker, *Physica*, **8**, 23 (1941).

Now, from Eq. (300),

$$dI(E) = I_s e^{-E} dE, \tag{335}$$

where I_s is the total cathode emission current. Hence,

$$\bar{v} = \left(\frac{2kT}{m}\right)^{1/2} \frac{\int_{E_m}^{\infty} (E - E_m)^{1/2} e^{-E} \, dE}{\int_{E_m}^{\infty} e^{-E} \, dE} \tag{336}$$

Multiplying numerator and denominator by e^{E_m},

$$\tilde{v} = \left(\frac{2kT}{m}\right)^{1/2} \frac{\int_0^{\infty} u^{1/2} e^{-u} \, du}{\int_0^{\infty} e^{-u} \, du} \tag{337}$$

is found, where

$$u = E - E_m. \tag{338}$$

But

$$\int_0^{\infty} u^{1/2} e^{-u} \, du = \Gamma\left(\frac{3}{2}\right) = \frac{\pi^{1/2}}{2}, \tag{339a}$$

$$\int_0^{\infty} e^{-u} \, du = \Gamma(1) = 1. \tag{339b}$$

Therefore

$$\tilde{v} = \left(\frac{\pi kT}{2m}\right)^{1/2}. \tag{340}$$

This, then, is the single equivalent initial velocity which will be assumed for all the β-electrons in a triode. By neglecting the α-electrons and assuming for the β-electrons this common "emission" velocity $\bar{v}$ from a fictitious "cathode" located at the potential minimum, we arrive at a model in which the electron trajectories are all the same, all electrons "emitted" being collected at the anode, and in which the field and therefore the electron acceleration is zero at the "cathode."

Now any individual electron on this model will move in accordance with the equation of motion

$$ma = -eE, \tag{341}$$

where a is its acceleration at any point x between cathode and anode and where E is the electric field that it experiences at the point x. Now the time rate of change of E, taken along the trajectory of the moving electron, is

$$\frac{dE}{dt} = \left(\frac{\partial E}{\partial x}\right)v + \frac{\partial E}{\partial t}, \tag{342}$$

where v is the electron velocity at the point x, in the x-direction. But if the field E is a function of x and not of y or z, as it will be in a planar diode

or similarly constructed triode (except very near the grid wires), it is possible to write

$$\frac{\partial E}{\partial x} = \nabla \cdot E = 4\pi\rho, \tag{343}$$

where ρ is the space-charge density.   Therefore

$$\frac{dE}{dt} = 4\pi\rho v + \frac{\partial E}{\partial t}. \tag{344}$$

But if I is the total current density at any distance x from the cathode, then it is possible to write

$$I = \rho v + \left(\frac{1}{4\pi}\right)\frac{\partial E}{\partial t}. \tag{345}$$

Hence

$$\frac{dE}{dt} = 4\pi I, \tag{346}$$

and consequently, from Eq. (341),

$$\frac{da}{dt} = -4\pi\left(\frac{e}{m}\right)I. \tag{347}$$

Now, according to Eq. (346),

$$\frac{\partial I}{\partial x} = \left(\frac{1}{4\pi}\right)\left(\frac{d}{dt}\right)\left(\frac{\partial E}{\partial x}\right), \tag{348}$$

or, according to Eq. (343)

$$\frac{\partial I}{\partial x} = \frac{d\rho}{dt}. \tag{349}$$

But since the electronic charges are conserved along their trajectories, $d\rho/dt$ must vanish, and therefore

$$\frac{\partial I}{\partial x} = 0. \tag{350}$$

Hence, I is the same at all points x, at a given time t, or, in other words, a function of the time t alone.   If we assume a steady-state current in the tube, then we may assume that I is constant in time.   This result therefore also holds in Eq. (347), where I is to be evaluated along the trajectory of the moving electron.   Therefore, da/dt is a constant, and

$$a(t) = -4\pi\left(\frac{e}{m}\right)It, \tag{351}$$

the initial acceleration being zero.   Integrating again,

$$v(t) = -2\pi\left(\frac{e}{m}\right)It^2 + \bar{v}, \tag{352}$$

where, according to assumption $\bar{v}$ is the electron's initial velocity. Integrating once again,

$$x(t) = -\left(\frac{2\pi}{3}\right)\left(\frac{e}{m}\right)It^3 + \bar{v}t, \tag{353}$$

if $x(0) = 0$ is the "cathode" position. If d is the diode spacing (or the equivalent diode or cathode-grid spacing in the triode), and if τ is the corresponding electron transit time, then

$$d = -\left(\frac{2\pi}{3}\right)\left(\frac{e}{m}\right)I\tau^3 + \bar{v}. \tag{354}$$

This result gives the total current I in terms of the spacing d, the initial electron velocity $\bar{v}$, and the transit time τ.

This result corresponds to the equilibrium total current in the equivalent diode but says nothing about the current fluctuations that we seek to evaluate. But this formula should indicate that if there are such current fluctuations δI, these fluctuations should be subject to Eq. (354) as a constraint. The source of such fluctuations δI must be sought as residing in a fluctuation $\delta\bar{v}$ in the emission velocity $\bar{v}$ of the electrons. According to Eq. (354), since d and τ are fixed, δI and $\delta\bar{v}$ must be related thus:

$$0 = -\left(\frac{2\pi}{3}\right)\left(\frac{e}{m}\right)\tau^3\,\delta I + \tau\,\delta\bar{v}, \tag{355}$$

or

$$\frac{\delta\bar{v}}{\delta I} = \left(\frac{2\pi}{3}\right)\left(\frac{e}{m}\right)\tau^2. \tag{356}$$

Referring to Eq. (332), it becomes possible to assign an origin to the fluctuations $\delta\bar{v}$. For the result

$$dI(E) = I_s e^{-E}\,dE \tag{357}$$

is true only on the average, and in every range of initial energy dE there actually are fluctuations $\delta I(E)$ in the emission current. These fluctuations will cause fluctuations $\delta\bar{v}$, according to Eq. (332), not only because of the changes in the integrands of Eq. (332) but also because of associated changes in the potential minimum and therefore in the lower integration limit E_m, this change being a small quantity δE_m. Hence, from Eq. (332),

$$\bar{v} + \delta\bar{v} = \frac{\displaystyle\int_{E_m}^{\infty} v_m\,dI(E) + v_m\,\delta I(E) - \bar{v}\left(\dfrac{dI}{dE}\right)E_m\,\delta E_m}{\displaystyle\int_{E_m}^{\infty} dI(E) + \delta I(E) - \left(\dfrac{dI}{dE}\right)E_m\,\delta E_m} \tag{358}$$

is the augmented initial velocity due ultimately to a fluctuation $\delta I(E)$

in the emission of β-electrons of energy in the range dE at E, in the *actual* Maxwell-Boltzmann distribution. Accordingly,

$$\delta \bar{v} = \frac{\displaystyle\int_{E_m}^{\infty} v_m \, dI(E) - \bar{v} \int_{E_m}^{\infty} dI(E) + (v_m - \bar{v}) \, \delta I(E)}{\displaystyle\int_{E_m}^{\infty} dI(E) + \delta I(E) - \left(\frac{dI}{dE}\right) E_m \, \delta E_m}. \tag{359}$$

Now, using Eq. (332) and neglecting the variations $\delta I(E)$ and δE_m in the denominator of Eq. (359), in the first approximation, it is found that

$$\begin{aligned}\delta \bar{v} &= \frac{(v_m - \bar{v}) \, \delta I(E)}{\displaystyle\int_{E_m}^{\infty} dI(E)} \\ &= \frac{(v_m - \bar{v}) \, \delta I(E)}{I}.\end{aligned} \tag{360}$$

Hence, from Eq. (356),

$$\frac{\delta I}{\delta I(E)} = \left(\frac{3}{2\pi}\right)\left(\frac{m}{e}\right)\frac{v_m - \bar{v}}{I\tau^2} \tag{361}$$

gives the fluctuation in I due to the fluctuation $\delta I(E)$ in the emission of β-electrons of energy in the range dE at E.

Now $\delta I(E)$ is a true shot current fluctuation, with a statistical spectrum

$$\left(\frac{d}{df}\right)|\overline{\delta I(E)|^2} = e \, dI(E). \tag{362}$$

Therefore the statistical spectrum of δI, due to the β-electrons of energy in the range dE at E, is, from Eqs. (361) and (362),

$$\left(\frac{d}{df}\right)\overline{|\delta I|^2} = \left(\frac{3}{2\pi}\right)^2 \left(\frac{m}{e}\right)^2 \left(\frac{(v_m - \bar{v})^2 e \, dI(E)}{I^2\tau^4}\right), \tag{363}$$

and the statistical spectrum of the total fluctuation in the equivalent diode current, due to β-electrons of all energies, is

$$\frac{d\overline{I^2}}{df} = \left(\frac{3}{2\pi}\right)^2 \left(\frac{m}{e}\right)^2 \left(\frac{e}{I^2\tau^4}\right) \int_{E_m}^{\infty} (v_m - \bar{v})^2 \, dI(E). \tag{364}$$

The integral in Eq. (364) is

$$\begin{aligned}\int_{E_m}^{\infty} (v_m - \bar{v})^2 \, dI(E) &= \int_{E_m}^{\infty} v_m^2 \, dI(E) - 2\bar{v} \int_{E_m}^{\infty} v_m \, dI(E) + \bar{v}^2 \int_{E_m}^{\infty} dI(E) \\ &= \int_{E_m}^{\infty} v_m^2 \, dI(E) - \bar{v}^2 I\end{aligned} \tag{365}$$

or, from Eqs. (334) and (340),

$$
\begin{aligned}
\int_{E_m}^{\infty} (v_m - \bar{v})^2 \, dI(E) &= \left(\frac{2kT}{m}\right) I_s \int_{E_m}^{\infty} (E - E_m)e^{-B} \, dE - \left(\frac{\pi kT}{2m}\right) I \\
&= \left(\frac{2kTI}{m}\right)\left(\int_0^{\infty} ue^{-u} \, du - \frac{1}{4}\pi\right) \\
&= \left(\frac{2kTI}{m}\right)\left(1 - \frac{1}{4}\pi\right).
\end{aligned}
\tag{366}
$$

Hence

$$
\frac{dI^2}{df} = \left(\frac{3}{2\pi}\right)^2 \left(\frac{m}{e}\right)\left(\frac{2kT}{I\tau^4}\right)\left(1 - \frac{1}{4}\pi\right).
\tag{367}
$$

Expression (367) may be simplified by making use of the value of the a-c conductance g of the equivalent diode. The derivation of the expression for g proceeds from Eq. (346), where I now includes a small a-c component. Because of the length of this derivation only the result is given here:

$$
\frac{1}{g} = \left(\frac{4\pi^2}{3}\right)\left(\frac{e}{m}\right) I\tau^4,
\tag{368}
$$

where I is the equilibrium direct current as previously defined. Inserting this expression in Eq. (367), it is found that

$$
\frac{d\overline{I^2}}{df} = 3\left(1 - \frac{1}{4}\pi\right)(2kTg).
\tag{369}
$$

Here the space-charge factor in Eq. (320) first introduced in conjunction with Formula (319) may be identified. Futhermore if the relation

$$
g_m = \sigma g
\tag{370}
$$

between the triode mutual conductance g_m and the conductance g of the equivalent diode is used, then Eq. (370) becomes

$$
\frac{d\overline{I^2}}{df} = \left(\frac{\theta}{\sigma}\right) 2kTg_m,
\tag{371}
$$

which is identical with the result of Eq. (321). Hence the present model of a triode tube seems justified, at least under operating conditions where θ is given by Eq. (320), as usually occurs under normal triode operation.

This model is now used to evaluate in the first approximation, the shot-noise current fluctuations that are induced in the grid of a triode tube. To do this, recall that the fluctuation current δI given by Eq. (361) is the *total* current fluctuation incident upon the grid. Of this total current, one part (the convection current) flows to the plate, and another part (the displacement current) flows out the grid lead. The latter part is

given by

$$\delta I_g = \left(\frac{1}{4\pi}\right)\left(\frac{\partial E}{\partial t}\right), \tag{372}$$

where E is evaluated at the grid, or in other words at the "anode" of the equivalent diode. E is given by Eq. (341) in terms of the electron acceleration a at the grid, thus:

$$E = -\left(\frac{m}{e}\right)a, \tag{373}$$

while da/dt is given by Eq. (347):

$$\frac{da}{dt} = -4\pi\left(\frac{e}{m}\right)I. \tag{374}$$

But now, in contrast to the previous treatment of this equation, it must be assumed that I is no longer constant in time, but is of the form

$$I = I_0 + \delta I, \tag{375}$$

where I_0 is the constant equilibrium value of I and δI is the time-varying fluctuation given by Eq. (361), which, for any particular angular frequency ω, may be assumed to be a sinusoidal function of the time t. This leads to a change in the initial conditions of Eq. (374), so that upon integration of this equation one gets, instead of Eq. (351), the result (good only to the first order)

$$a = -4\pi\left(\frac{e}{m}\right)\tau\left(\frac{I_0 + 2\delta I}{3}\right) + \frac{2\delta\bar{v}}{\tau}, \tag{376}$$

where δI and $\delta\bar{v}$ are both sinusoidal functions of time (for any single noise frequency) that are correlated with each other according to Eq. (356), viz.,

$$\frac{\delta\bar{v}}{\delta I} = \left(\frac{2\pi}{3}\right)\left(\frac{e}{m}\right)\tau^2. \tag{377}$$

If we eliminate $\delta\bar{v}$ from Eq. (376), it is found that

$$a = -4\pi\left(\frac{e}{m}\right)\tau\left(\frac{I_0 + \delta I}{3}\right). \tag{378}$$

Now, since δI is of the form $e^{j\omega t}$,

$$\frac{-\partial a}{\partial t} = j\omega 4\pi\left(\frac{e}{m}\right)\left(\frac{\tau\,\delta I}{3}\right). \tag{379}$$

Hence, from Eq. (373),

$$\frac{\partial E}{\partial t} = 4\pi\left(\frac{j\omega\tau}{3}\right)\delta I, \tag{380}$$

and from Eq. (372),

$$\delta I_g = \left(\frac{j\omega\tau}{3}\right)\delta I. \tag{381}$$

This result is true for the noise components at the frequency ω and shows that the induced grid-noise components at any frequency ω are (within the limits of validity of this proof) in quadrature with the total noise current fluctuation δI. It is further to be noted that this result refers not only to a single frequency ω but also to a single emission energy E of the β-electrons in the actual Maxwell-Boltzmann distribution, in accordance with Eq. (361). But the integration over this distribution of emission energies E is to be carried out just as before, so that for the statistical spectrum of the induced grid shot-noise current due to all β-electrons the result

$$\frac{d\overline{I_g^2}}{df} = \left(\frac{\omega\tau}{3}\right)^2\frac{d\overline{I^2}}{df} \tag{382}$$

is found. Or, using Eq. (371),

$$\frac{d\overline{I_g^2}}{df} = \left(\frac{\omega\tau}{3}\right)^2\left(\frac{\theta}{\sigma}\right)2kTg_m. \tag{383}$$

This result may be further simplified by introducing the expression for the a-c electronic loading of the equivalent diode. This loading $1/R_e$ may also be evaluated by integrating Eq. (374) under first-order a-c initial conditions. Because of the length of the derivation, only the result

$$\frac{1}{R_e} = \left(\frac{g}{20}\right)(\omega\tau)^2 \tag{384}$$

is given here. Introducing this result into Eq. (383), it is found that

$$\frac{d\overline{I_g^2}}{df} = \left(\frac{20\theta}{9}\right)\left(\frac{2kT}{R_e}\right) \tag{385}$$

From Eq. (383) it is seen that the induced grid noise, in the first approximation, increases proportionally to the square of the frequency, so that at low frequencies it is negligible, and Eqs. (327) give an adequate description of the source of shot noise in a negative-grid triode. But at higher frequencies, calling

$$-I_c(t) = I_p(t) + I_g(t) \tag{386}$$

the total shot-noise current incident upon the grid and

$$-i_c(j\omega) = i_p(j\omega) + i_g(j\omega) \tag{387}$$

its Fourier transform,

$$\frac{d\overline{I_c^2}}{df} = \overline{|i_c(j\omega)|^2} = \left(\frac{\theta}{\sigma}\right) 2kTg_m \tag{388a}$$

$$\frac{d\overline{I_g^2}}{df} = \overline{|i_g(j\omega)|^2} = \left(\frac{\omega\tau}{3}\right)^2 \left(\frac{\theta}{\sigma}\right) 2kTg_m \tag{388b}$$

$$\frac{d\overline{I_p^2}}{df} = \overline{|i_p(j\omega)|^2} = \left[1 + \left(\frac{\omega\tau}{3}\right)^2\right]\left(\frac{\theta}{\sigma}\right) 2kTg_m \tag{388c}$$

must be written for the statistical spectra of the three nodal shot-noise currents $I_c(t)$, $I_p(t)$, $I_p(t)$.

Although Eqs. (388) indicate the presence of three nodal current generators $i_c(j\omega)$, $i_g(j\omega)$, $i_p(j\omega)$, these three currents are correlated with one another. Accordingly, there should exist another equivalent representation of a negative-grid triode in which, even at moderately high frequencies, there is just one statistically independent source of shot noise. Such a representation will now be found.[1]

Llewellyn has shown[2] that a negative-grid triode may be represented at high frequencies in the manner shown by Fig. 12·12, where C_{cg}, C_{gp}, C_{pc} are the low-frequency interelectrode capacitances, r_p is the low-frequency plate resistance, and

FIG. 12·12.—Representation of a negative-grid triode at moderately high frequencies.

$$i_{pc} = g_m(e_g - e_c), \tag{389}$$

where g_m is the low-frequency transconductance. Here there enter in addition two small series resistances r_{cg} and r_{gp}, which are negligible at low frequencies but not at high, and are responsible for the electronic loading of the triode. For the case where the grid-plate transit time is negligible, their values are

$$r_{cg} = \left(\frac{9}{80g_m}\right)\left(\frac{1 + \mu}{1 + \mu + \dfrac{4y}{3}}\right), \tag{390a}$$

$$r_{gp} = -\left(\frac{1}{5g_m}\right)\frac{y^2}{1 + \mu + \dfrac{4y}{3}}, \tag{390b}$$

where μ is the low-frequency amplification factor and y is the ratio of grid-plate to cathode-grid spacing. It is seen that r_{cg} is positive, r_{gp} is negative, and both are frequency-independent.

---

[1] *Cf.* also E. J. Schremp, to be published.

[2] F. B. Llewellyn, *Bell System Tech. Jour.*, **15**, 575 (1936).

If the nodal shot-noise currents $i_c(j\omega)$, $i_g(j\omega)$, $i_p(j\omega)$ are incorporated into Fig. 12·12, then these will all three be subject to the relations [see Eqs. (381) and (386)]

$$i_c(j\omega) = i, \tag{391a}$$

$$i_g(j\omega) = \left(\frac{j\omega\tau}{3}\right) i, \tag{391b}$$

$$i_p(j\omega) = -\left(\frac{1 + j\omega\tau}{3}\right) i, \tag{391c}$$

where i is the single statistically independent shot-noise current whose proper location is to be found.

If a fourth node n, artificially introduced into the network of Fig. 12·12, is used, then it may be assumed that i is an internodal current located between the nodes c and n. Furthermore, if $-y_{cn}$, $-y_{gn}$, $-y_{pn}$ are arbitrary bilateral branch admittances linking cathode, grid, and plate with the fictitious node n, then, by Eqs. (201) (in which $\alpha = 1$), we can write

$$i_c = i_c^0 - \left(\frac{y_{cn}}{y_{nn}}\right) i_n^0, \tag{392a}$$

$$i_g = i_g^0 - \left(\frac{y_{gn}}{y_{nn}}\right) i_n^0, \tag{392b}$$

$$i_p = i_p^0 - \left(\frac{y_{pn}}{y_{nn}}\right) i_n^0, \tag{392c}$$

where, to conform with the desired shot-noise representation, one may set

$$i_c^0 = -i_n^0 = i, \tag{393a}$$

$$i_g^0 = i_p^0 = 0. \tag{393b}$$

Thus it is found that

$$i_c = \left(\frac{1 + y_{cn}}{y_{nn}}\right) i, \tag{394a}$$

$$i_g = \left(\frac{y_{gn}}{y_{nn}}\right) i, \tag{394b}$$

$$i_p = \left(\frac{y_{pn}}{y_{nn}}\right) i, \tag{394c}$$

and hence, from Eqs. (391), that

$$\frac{y_{cn}}{y_{nn}} = 0, \tag{395a}$$

$$\frac{y_{gn}}{y_{nn}} = \frac{j\omega\tau}{3} \tag{395b}$$

$$\frac{y_{pn}}{y_{nn}} = -\left(\frac{1 + j\omega\tau}{3}\right). \tag{395c}$$

This result means that the node n is effectively a point "tapped in" on the grid-plate impedance

$$\frac{-1}{y_{gp}} = r_{gp} + \frac{1}{j\omega C_{gp}}. \tag{396}$$

The result then should be

$$\frac{1}{y_{gp}} = \frac{1}{y_{gn}} + \frac{1}{y_{np}}, \tag{397}$$

or, using Eqs. (395),

$$\frac{y_{nn}}{y_{gp}} = \frac{\dfrac{3}{j\omega\tau} - 1}{1 + \dfrac{j\omega\tau}{3}}$$

$$= \frac{3}{j\omega\tau\left(1 + \dfrac{j\omega\tau}{3}\right)}. \tag{398}$$

But from Eqs. (396),

$$-y_{gp} = \frac{j\omega C_{gp}}{1 + j\omega r_{gp}C_{gp}}, \tag{399}$$

so

$$-y_{nn} = \frac{\left(\dfrac{3C_{gp}}{\tau}\right)}{(1 + j\omega r_{gp}C_{gp})\left(\dfrac{1 + j\omega\tau}{3}\right)}. \tag{400}$$

Hence, from Eq. (395b),

$$\begin{aligned}
\frac{-1}{y_{gn}} &= (1 + j\omega r_{gp}C_{gp})\left(\frac{1 + j\omega\tau}{3}\right)\left(\frac{3}{j\omega\tau}\right)\left(\frac{\tau}{3C_{gp}}\right) \\
&= \left(r_{gp} + \frac{1}{j\omega C_{gp}}\right)\left(\frac{1 + j\omega\tau}{3}\right) \\
&= \frac{\left(r_{gp} + \dfrac{\tau}{j\omega C_{gp}}\right) + j\omega\left(\dfrac{r_{gp}\tau}{3}\right) + 1}{j\omega C_{gp}},
\end{aligned} \tag{401}$$

while, from Eqs. (396), (397), and (401),

$$+\frac{1}{y_{pn}} = \frac{\tau}{3C_{gp}} + j\omega\left(\frac{r_{gp}\tau}{3}\right). \tag{402}$$

Thus, the branch impedance $-1/y_{gn}$ may be written

$$-\frac{1}{y_{gn}} = r_{gn} + j\omega L_{gn} + \frac{1}{j\omega C_{gp}}, \tag{403}$$

where

$$r_{gn} = r_{gp} + \frac{\tau}{3C_{gp}} \tag{404}$$

is a frequency-independent resistance that is negative or positive according as $\tau < |3r_{gp}C_{gp}|$ or $\tau > |3r_{gp}C_{gp}|$ and where

$$L_{gn} = \frac{r_{gp}\tau}{3} \tag{405}$$

is a negative, frequency-independent inductance. Similarly, the branch

Fig. 12·13.—Noise representation of a negative-grid triode at moderately high frequencies.

impedance $-1/y_{pn}$ can be written

$$\frac{-1}{y_{pn}} = r_{pn} + j\omega L_{pn}, \tag{406}$$

where

$$r_{pn} = \frac{-\tau}{3C_{gp}} \tag{407}$$

is a negative, frequency-independent resistance and where

$$L_{pn} = \frac{-r_{gp}\tau}{3} \tag{408}$$

is a positive, frequency-independent inductance.

Therefore, one complete equivalent network representation of a negative grid triode at moderately high frequencies, in which there is but one statistically independent shot-noise current generator, is as shown in Fig. 12·13. Here the shot-noise constant-current generator i has a statistical spectrum that, in the first approximation, is a constant equal to the low-frequency spectrum:

$$|i|^2 = \frac{\theta}{\sigma} 2kTg_m. \tag{409}$$

The current generator i_{pc} is, as before, given by the relation

$$i_{pc} = g_m(e_g - e_c). \tag{410}$$

The representation of Fig. 12·13 must not be expected to hold when the cathode-grid transit angle becomes large or when the grid-plate transit angle is too large to neglect completely. Theoretically, it would be possible to modify the present representation so as to render these limitations unnecessary; but it would be of little value to do this, because when these limitations are removed, the basic assumption of the present
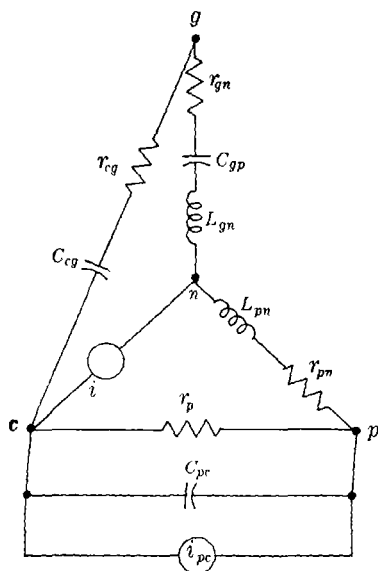
analysis becomes questionable, namely, that only the β-electrons need be considered, and all have a common initial velocity.

Once it becomes necessary to account for the Maxwell-Boltzmann distribution of electron emission velocities, the possibility no longer exists of representing the shot noise at very high frequencies in terms of a single statistically independent current generator, in the manner of Fig. 12·13. Rather, at least two independent generators then appear to be necessary.

Shot Noise in Pentodes.[1,2]—A suppressor-type pentode operated with its control-grid negative is effectively a tetrode, since the suppressor grid is ordinarily at the cathode d-c potential and collects no electrons. Its usual function is merely to suppress the flow of secondary electrons from the plate to the screen grid.

In a suppressor-type pentode the total cathode direct current I_c divides between the screen grid and the plate. If I_s is the direct current flowing to the screen grid and I_p the direct current flowing to the plate, then the direct current relation

$$I_c = I_s + I_p \tag{411}$$

exists.

Since the theory of shot noise in a pentode at low frequencies can be easily extended to the case of a multicollector amplifier tube with N collectors, the general case in which there are N collector electrodes is dealt with here. In this case, Eq. (411) becomes

$$I_c = \sum_1^N I_n. \tag{412}$$

In order to examine the detailed behavior of the shot-noise fluctuations in the case of an N-collector tube, it is necessary to return to the shot-noise analysis in a diode with space charge, modifying it in such a way as to account for the new configuration of collector electrodes.

In the absence of any compensating action at the virtual cathode (potential minimum), a pure shot current fluctuation i_n^0 having a statistical spectrum equal to

$$|i_n^0|^2 = eI_n \tag{413}$$

will flow into each collector n. These fluctuation currents will be statistically independent, that is,

$$\overline{i_m^0 i_n^0} = 0 \qquad (m \neq n). \tag{414}$$

Correspondingly, there will flow out of the cathode a pure shot current

---

[1] R. Q. Twiss and E. J. Schremp, Abstract, Spring 1946 meeting, American Physical Society, Cambridge.

[2] D. O. North, *RCA Rev.*, **5**, 2, October 1940.

fluctuation i_c^0, given by

$$i_c^0 = \sum_1^N i_n^0. \tag{415}$$

Therefore, from Eqs. (414) and (415),

$$\overline{i_c^0 i_n^0} = \overline{|i_n^0|^2} = eI_n \tag{416a}$$

$$\overline{|i_c^0|^2} = \sum_1^N \overline{|i_n^0|^2} = e \sum_1^N I_n = eI_c. \tag{416b}$$

In the presence of a compensating action at the virtual cathode, the shot current fluctuation i_n that now flows into the nth collector will consist of three parts:

$$i_n = i_n' + i_n'' + i_n'''. \tag{417}$$

The first part i_n' can be written

$$i_n' = \left[1 - \frac{I_n}{I_c}(1 - \gamma_\beta) \right] i_n^0. \tag{418}$$

This part is due to the β-electrons (strictly, it refers only to a single velocity group) flowing to the nth collector. The second term in Eq. (418) indicates a reduction due to the compensating action at the virtual cathode. This reduction term is similar to the compensating term δ_β in Eqs. (305) and (310) for the diode case but differs from it by the factor I_n/I_c. This factor I_n/I_c is required in the present case, since the compensating current at the nth collector is, on the average, proportional to the direct current I_n to the nth collector.

The second part i_n'' may be written

$$i_n'' = -\frac{I_n}{I_c}(1 - \gamma_\beta)(i_c^0 - i_n^0). \tag{419}$$

This part is due to the β-electrons flowing to all other collectors but the nth. It is a purely compensatory current, completely analogous to the second term of Eq. (418).

The third part i_n''' can be written

$$i_n''' = +\left(\frac{I_n}{I_c}\right)\gamma_\alpha i_\alpha^0. \tag{420}$$

This part is due to the α-electrons (again of a single velocity group) and arises from their effect upon the virtual cathode. Here again i_α^0 may be regarded as a pure shot-noise fluctuation current that would exist if there were no virtual cathode and would have a statistical spectrum (for all

α-velocity classes) equal to

$$\overline{|i_\alpha^0|^2} = e(I_s - I_c), \tag{421}$$

where I_s is the total cathode emission current.   If the quantity

$$i_c''' = \sum_{1}^{N} i_n''' \tag{422}$$

is defined, then, from Eq. (420),

$$i_c''' = \gamma_\alpha i_\alpha^0, \tag{423}$$

and hence it follows that

$$\overline{|i_c'''|^2} = \overline{\gamma_\alpha^2}\,\overline{|i_\alpha^0|^2}, \tag{424}$$

where $\overline{\gamma_\alpha^2}$ is the average value of γ_α^2 over all α-electron energies.   But according to Eq. (316), if $\overline{\gamma_\alpha^2}$ is multiplied by the quantity $(I_s/I_c - 1)$, the average value of γ_α^2 over all β-electron energies is obtained.   In the diode case, this was called $\overline{\gamma_\alpha^2}$.   Thus Eq. (424) becomes

$$\overline{|i_c'''|^2} = \frac{\overline{\gamma_\alpha^2}\,\overline{|i_\alpha^0|^2}}{\left(\dfrac{I_s}{I_c} - 1\right)} \tag{425}$$

or, using Eq. (421)

$$\overline{|i_{c_i}'''|^2} = \overline{\gamma_\alpha^2} e I_c. \tag{426}$$

But this expression is identical with the contribution of the α-electrons to the statistical spectrum of the shot-noise current in a diode, as given by Eq. (315).   Hence the result

$$\overline{\gamma_\alpha^2}\,\overline{|i_\alpha^0|^2} = \overline{\gamma_\alpha^2} e I_c \tag{427}$$

is obtained.   Furthermore, i_α^0 is statistically independent of the fluctuation currents i_n^0 $(n = 1, 2, \cdots, N)$ and hence, by Eq. (415), also of i_c^0.

Again, in the presence of a compensating action at the virtual cathode, the shot-current fluctuation i_c that flows out of the cathode consists of two parts:

$$i_c = i_c' + i_c'''. \tag{428}$$

The first part i_c' can be written

$$i_c' = \gamma_\beta i_c^0. \tag{429}$$

It is due to the β-electrons (again of a single velocity group) leaving the cathode and may be identified, from Eqs. (418) and (419), with the expression

$$i_c' = \sum_{1}^{N} (i_n' + i_n''). \tag{430}$$

The second part i_c''' is given by Eq. (423) and is due to the α-electrons (of a single velocity group) leaving the cathode. Finally, from Eqs. (422) and (430), we can write

$$i_c' + i_c''' = \sum_1^N (i_n' + i_n'' + i_n'''), \tag{431}$$

or, from Eqs. (417) and (428),

$$i_c = \sum_1^N i_n. \tag{432}$$

This last result was, of course, to be expected. It expresses the fact that the sum of all the shot-noise nodal currents flowing out of the tube must vanish.

Thus, from Eqs. (417) to (420), one can write

$$i_n = i_n^0 - \left(\frac{I_n}{I_c}\right)(1 - \gamma_\beta)i_c^0 + \left(\frac{I_n}{I_c}\right)\gamma_\alpha i_\alpha^0, \tag{433}$$

while, from Eqs. (423), (428), and (429),

$$i_c = \gamma_\beta i_c^0 + \gamma_\alpha i_\alpha^0. \tag{434}$$

Now consider the expression $\overline{i_m i_n}$ for $m \neq n$. From Eq. (433),

$$\overline{i_m i_n} = -\left(\frac{I_n}{I_c}\right)(1 - \overline{\gamma_\beta})\overline{i_m^0 i_c^0} - \left(\frac{i_m}{I_c}\right)(1 - \overline{\gamma_\beta})\overline{i_n^0 i_c^0}$$
$$+ \frac{I_m I_n}{I_c^2}\overline{(1 - \gamma_\beta)^2}\,\overline{|i_c^0|^2} + \left(\frac{I_m I_n}{I_c^2}\right)\overline{\gamma_\alpha^2}\,\overline{|i_\alpha^0|^2}, \tag{435}$$

or, from Eqs. (416) and (427),

$$\frac{\overline{i_m i_n}}{e} = -2\left(\frac{I_m I_n}{I_c}\right)(1 - \overline{\gamma_\beta}) + \left(\frac{I_m I_n}{I_c}\right)\overline{(1 - \gamma_\beta)^2} + \left(\frac{I_m I_n}{I_c}\right)\overline{\gamma_\alpha^2}$$
$$= \left(\frac{I_m I_n}{I_c}\right)(\overline{\gamma_\alpha^2} + \overline{\gamma_\beta^2} - 1), \tag{436}$$

or, remembering that

$$\Gamma^2 = \overline{\gamma_\alpha^2} + \overline{\gamma_\beta^2}, \tag{437}$$

one finds

$$\overline{i_m i_n} = -e\left(\frac{I_m I_n}{I_c}\right)(1 - \Gamma^2) \qquad (m \neq n). \tag{438}$$

Again, consider the expression $\overline{i_n i_c}$. From Eqs. (433) and (434),

$$\overline{i_n i_c} = \overline{\gamma_\beta}\,\overline{i_n^0 i_c^0} - \left(\frac{I_n}{I_c}\right)\overline{(1 - \gamma_\beta)\gamma_\beta}\,\overline{|i_c^0|^2} + \left(\frac{I_n}{I_c}\right)\overline{\gamma_\alpha^2}\,\overline{|i_\alpha^0|^2} \tag{439}$$

or, from Eqs. (416) and (427),

$$\frac{\overline{i_n i_c}}{e} = \overline{\gamma_\beta} I_n - \overline{(1 - \gamma_\beta)\gamma_\beta} I_n + \overline{\gamma_\alpha^2} I_n$$

$$= (\overline{\gamma_\alpha^2} + \overline{\gamma_\beta^2}) I_n \tag{440}$$

or

$$\overline{i_n i_c} = e I_n \Gamma^2. \tag{441}$$

The results (438) and (441) indicate that all of the nodal shot-noise currents are mutually correlated. But if instead of employing nodal currents, an equivalent set of internodal shot-noise currents is introduced, it turns out as in the case of the generalization of Nyquist's thermal noise theorem that this latter set of internodal shot-noise currents may be so chosen as to be mutually uncorrelated. For let

$$i_c = \sum_n i_{cn}, \qquad (n = 1, 2, \cdots, N) \tag{442a}$$

and

$$i_n = i_{cn} + \sum_r i_{rn} \qquad (r \neq n). \tag{442b}$$

Now if it is assumed that all of these internodal currents i_{mn} are mutually uncorrelated, that is, that

$$\overline{i_{mn} i_{rs}} = 0, \qquad (mn) \neq (rs), \tag{443}$$

then, from Eqs. (442),

$$\overline{i_m i_n} = \overline{i_{nm} i_{mn}} = - \overline{|i_{mn}|^2}, \tag{444a}$$

$$\overline{i_n i_c} = \overline{i_{cn} i_{cn}} = + \overline{|i_{cn}|^2}, \tag{444b}$$

whence, from Eqs. (438) and (441),

$$\overline{|i_{cn}|^2} = e I_n \Gamma^2, \tag{445a}$$

$$\overline{|i_{mn}|^2} = e \left(\frac{I_m I_n}{I_c} \right) (1 - \Gamma^2) \qquad (m \neq n). \tag{445b}$$

Thus between the cathode and any collector n there is a shot-noise internodal current generator i_{cn} whose statistical spectrum is that of a diode with a direct current I_n, and between any two collectors m and n there is a *partition noise* internodal current generator i_{mn} whose statistical spectrum is always real and positive and vanishes when either collector draws no direct current or when $\Gamma^2 = 1$. Moreover, the shot-noise generators i_{cn} and the partition noise generators i_{mn} are mutually uncorrelated, either among themselves or with each other.

The question may be raised as to the validity of this apparently arbitrary choice of internodal current generators, subject to the inco-

herency conditions of Eq. (443).   For if one considers the expression for the potential difference e_{mn} between any two nodes m and n, this will be of the form

$$e_{mn} = \tfrac{1}{2} \sum_r \sum_s Z_{rs}^{mn} i_{rs}, \tag{446}$$

where Z_{rs}^{mn} is the transfer impedance between the input nodal pair (rs) and the output nodal pair (mn); and the question may be raised as to whether or not the value of e_{mn} is independent of the choice of internodal currents i_{rs} when this choice is subject to Eq. (442).

The question may be settled in the affirmative in the following way.[1] Since

$$Z_{rs}^{mn} = -Z_{sr}^{mn}, \tag{447}$$

one can write

$$Z_{rs}^{mn} = Z_r^{mn} - Z_s^{mn}, \tag{448}$$

and therefore Eq. (446) becomes

$$e_{mn} = \tfrac{1}{2} \sum_r \sum_s (Z_r^{mn} i_{rs} - Z_s^{mn} i_{rs}). \tag{449}$$

But if

$$i_r = \sum_s i_{rs} \tag{450}$$

is written for the nodal current at node r, then Eq. (449) becomes

$$\begin{aligned} e_{mn} &= \frac{1}{2}\left(\sum_r Z_r^{mn} i_r + \sum_s Z_s^{mn} i_s \right) \\ &= \sum_r Z_r^{mn} i_r. \end{aligned} \tag{451}$$

Furthermore, if use is made of the relation

$$\sum_r i_r = 0 \tag{452}$$

to eliminate one particular node, k say, then Eq. (451) can be written

$$\begin{aligned} e_{mn} &= \sum_r (Z_r^{mn} - Z_k^{mn}) i_r \\ &= \sum_r Z_{rk}^{mn} i_r. \end{aligned} \tag{453}$$

[1] E. J. Schremp, to be published.

This last result is expressed in terms of the original nodal currents and hence must be the correct expression for e_{mn}. But it has just been seen that Eq. (453) is completely equivalent to Eq. (446) as long as Eqs. (450)

hold true.   Therefore Eq. (446) must be valid, regardless of the choice of internodal currents, as long as the choice is subject to Eq. (450).

To begin with, a representation of a pentode for high frequencies, corresponding to Llewellyn's representation (Fig. 12·12) of a triode, can be found.   This pentode representation is shown in Fig. 12·14, where C_{cg}, C_{gp}, C_{pc}, C_{sc}, C_{sg}, C_{sp} are the low-

FIG. 12·14.—Representation of a pentode at moderately high frequencies.

frequency interelectrode capacitances, r_p is the low-frequency plate resistance, r_s is the analogous low-frequency screen resistance, and

$$i_{pc} = g_{pg}(e_g - e_c) + g_{ps}(e_s - e_c), \tag{454a}$$
$$i_{sc} = g_{sg}(e_g - e_c), \tag{454b}$$

where g_{pg}, g_{ps}, g_{sg} are the low-frequency transconductances between plate and grid, plate and screen, and screen and grid, respectively.   Here there occur, in series with the grid capacitances C_{gc}, C_{gp}, C_{gs}, three small resistances r_{gc}, r_{gp}, r_{gs}, which are negligible at low frequencies but not at high and are responsible for the electronic loading of the pentode.

It is here assumed that the grid-screen and grid-plate branch impedances are of the form

$$\frac{-1}{y_{gs}} = r_{gs} + \frac{1}{j\omega C_{gs}} \tag{455a}$$

$$\frac{-1}{y_{gp}} = r_{gp} + \frac{1}{j\omega C_{gp}}. \tag{455b}$$

To find the values of r_{gs} and r_{gp}, which correspond to the assumed forms, Eq. (455), for the branch admittances $-y_{gs}$ and $-y_{gp}$, the case where the pentode is connected as a triode can be considered, i.e., where screen and plate are short-circuited together.   Then there will be a new grid-plate branch admittance

$$-y_{gp}^0 = -(y_{gs} + y_{gp}), \tag{456}$$

and this will be of the form

$$\frac{-1}{y_{gp}^0} = r_{gp}^0 + \frac{1}{j\omega(C_{gs} + C_{gp})}, \tag{457}$$

where r_{gp}^0 is given by Llewellyn's formula Eq. (391b). Accordingly, one must set

$$\frac{j\omega(C_{gs} + C_{gp})}{1 + j\omega r_{gp}^0(C_{gs} + C_{gp})} = \frac{j\omega C_{gs}}{1 + j\omega r_{gs}C_{gs}} + \frac{j\omega C_{gp}}{1 + j\omega r_{gp}C_{gp}} \tag{458}$$

or

$$j\omega C_{gs}\left[\frac{1}{1 + j\omega r_{gp}^0(C_{gs} + C_{gp})} - \frac{1}{1 + j\omega r_{gs}C_{gs}}\right]$$
$$+ j\omega C_{gp}\left[\frac{1}{1 + j\omega r_{gp}^0(C_{gs} + C_{gp})} - \frac{1}{1 + j\omega r_{gp}C_{gp}}\right] = 0. \tag{459}$$

The only way of satisfying this condition at all frequencies is to set

$$1 + j\omega r_{gp}^0(C_{gs} + C_{gp}) = 1 + j\omega r_{gs}C_{gs} = 1 + j\omega r_{gp}C_{gp} \tag{460}$$

or in other words to let

$$r_{gp}^0(C_{gs} + C_{gp}) = r_{gs}C_{gs} = r_{gp}C_{gp} \tag{461}$$

Then,

$$r_{gs} = r_{gp}^0\left(1 + \frac{C_{gp}}{C_{gs}}\right) \tag{462a}$$

$$r_{gp} = r_{gp}^0\left(1 + \frac{C_{gs}}{C_{gp}}\right) \tag{462b}$$

Thus the following may be taken as the expressions for r_{gc}, r_{gs}, r_{gp}:

$$r_{gc} = \left(\frac{9}{80g_m}\right)\left(\frac{1 + \mu}{1 + \mu + \frac{4y}{3}}\right), \tag{463a}$$

$$r_{gs} = -\left(1 + \frac{C_{gp}}{C_{gs}}\right)\left(\frac{1}{5g_m}\right)\left(\frac{y^2}{1 + \mu + \frac{4y}{3}}\right), \tag{463b}$$

$$r_{gp} = -\left(1 + \frac{C_{gs}}{C_{gp}}\right)\left(\frac{1}{5g_m}\right)\left(\frac{y^2}{1 + \mu + \frac{4y}{3}}\right), \tag{463c}$$

where μ and g_m are the low-frequency amplification factor and transconductance of the pentode when connected as a triode and where y is the ratio of the grid-screen to cathode-grid spacing. These results give a positive grid-cathode resistance r_{gc}, given by the triode formula, and a negative grid-screen resistance r_{gs} and grid-plate resistance r_{gp}, given by slight modifications of the triode formula.

In order to incorporate in this representation the sources of shot and partition noise, notice the fact that according to Formulas (445), a pentode at low frequencies is analogous to two triodes, with statistically

independent shot-noise current generators i_{cs} and i_{cp}, together with a statistically independent partition noise current generator i_{sp}.

Therefore at relatively high frequencies the noise representation of a pentode will be a simple extension of the triode representation shown in

FIG. 12·15.—Noise representation of a pentode at moderately high frequencies.

Fig. 12·13. This extension of that representation is shown in Fig. 12·15, where

$$r_{gm} = r_{gs} + \frac{\tau}{3C_{gs}}, \qquad (464a)$$

$$r_{gn} = r_{gp} + \frac{\tau}{3C_{gp}}, \qquad (464b)$$

$$L_{gm} = \frac{r_{gs}\tau}{3}, \qquad (464c)$$

$$L_{gn} = \frac{r_{gp}\tau}{3}, \qquad (464d)$$

$$r_{sm} = \frac{-\tau}{3C_{gs}}, \qquad (464e)$$

$$r_{pn} = \frac{-\tau}{3C_{gp}}, \qquad (464f)$$

$$L_{sm} = \frac{-r_{gs}\tau}{3}, \qquad (464g)$$

$$L_{pn} = \frac{-r_{gp}\tau}{3}. \qquad (464h)$$

Here the three tube noise current generators have statistical spectra given by

$$\overline{|i_{cm}|^2} = eI_s\Gamma^2 \tag{465a}$$

$$\overline{|i_{cn}|^2} = eI_p\Gamma^2 \tag{465b}$$

$$\overline{|i_{sp}|^2} = e\frac{I_sI_p}{I_c}(1 - \Gamma^2), \tag{465c}$$

whereas the two current generators i_{sc} and i_{pc} are again given by

$$i_{sc} = g_{sg}(e_g - e_c), \tag{466a}$$

$$i_{pc} = g_{pg}(e_g - e_c) + g_{ps}(e_s - e_c). \tag{466b}$$

In the triode case, Γ^2 is given by

$$\Gamma^2 = \frac{\theta}{\sigma}\frac{2kTg_m}{eI_c}. \tag{467}$$

But it is approximately true that

$$\frac{g_m}{I_c} = \frac{g_{sg}}{I_s} = \frac{g_{pg}}{I_p}, \tag{468}$$

so that Eqs. (465) can also be written in the form

$$\overline{|i_{cm}|^2} = \frac{\theta}{\sigma}2kTg_{sg}, \tag{469a}$$

$$\overline{|i_{cn}|^2} = \frac{\theta}{\sigma}2kTg_{pg}, \tag{469b}$$

$$\overline{|i_{sp}|^2} = \frac{g_{sg}}{g_m}\frac{g_{pg}}{g_m}\left(eI_c - \frac{\theta}{\sigma}2kTg_m\right). \tag{469c}$$

12·4. The Logical Distinction between Thermal Noise and Shot Noise.—Now that the two principal kinds of noise—thermal noise and shot noise—have been discussed in some detail, it will be profitable to summarize the distinctive features of each kind of noise.

The essential characteristic of thermal noise is that it exists only in a physical system that is in strict temperature equilibrium. Thermal noise is therefore usually an idealization, but there are, in practice, many situations where temperature equilibrium may be assumed to hold approximately, and in such situations it is convenient and useful to describe the noise as thermal.

On the other hand, shot noise is characterized as random electrical fluctuations due to the unilateral flow of electrons or other charged particles through some part of a circuit. Of course, practically every kind of noise, including thermal noise, may be regarded as a form of shot noise, provided allowance is made for the various possible trajectories

of the charged particles and each group of particles having the same trajectory (but different time phases) is treated as a separate unilateral flow. But it is usually more convenient to treat thermal noise as a phenomenon subject to analysis by statistical mechanics rather than to treat it as a special form of shot noise—which, nevertheless, it is.

Therefore, whenever any form of noise is encountered that does not have the characteristics, at least approximately, of thermal noise, it is probable that the most convenient approach to its analysis will be to regard that form of noise as a superposition of shot-noise fluctuations, taking into account the unilateral flow of each group of charged particles having a common type of trajectory.

In the last section not only the so-called "pure shot noise" in a temperature-limited diode but also every other form of noise associated with thermionically emitted electrons has been treated as a form of shot noise. In so doing, the meaning of the term "shot noise" may have been broadened beyond what it was originally meant to convey. But this extension of meaning and its further extension to include thermal noise and, indeed, almost all forms of noise seem justified by the fact that as long as a certain degree of randomness (which, by definition, is a characteristic of all forms of noise) exists in the motions of the charged particles responsible for the noise, of whatever kind observed, then the methods of analysis of the "pure shot effect" are applicable thereto. These methods may, however, as in the case of space charge reduced noise and partition noise, have to include a treatment of the mutual correlations of different charged particles or, as in the case of induced grid noise, have to include a treatment of the finite inertia of the charged particles.

There have been published several treatments of thermal noise as a form of shot noise, based upon considerations of the trajectories of electrons in a metallic conductor. Since none of these treatments has the generality of the treatment by statistical mechanics, they are not included here. But one particular example of the equivalence of the two points of view will be given here, whose merit is that it clearly exhibits the necessary conditions for the noise to be thermal and yet represents a very simple kind of shot noise.

This example is that of a bilateral diode.[1] Let both the cathode and plate of a diode be thermionic emitters at the same temperature T. The cathode and plate need not be of the same material, nor need they have the same total emission current. However, the fact that both cathode and plate are at the same temperature, which is constant in time, indicates that the interior of the diode is a system in thermal equilibrium. For even if the surroundings are at a lower temperature, for example, the temperature T of the electrodes may be maintained by d-c heating of the

[1] *Cf.* also D. O. North, *RCA Rev.,* **4**, 4, April 1940.

electrodes, and this heater power completely dissipated by external radiation to the surroundings. Thus, even though the diode may not be in thermal equilibrium with its surroundings but rather in a state of quasi-equilibrium characterized by a flow of power from the heater batteries to the surroundings, via radiation, nevertheless the partial system consisting of the interior of the diode will be in thermal equilibrium at the electrode temperature T.

Now if the diode electrodes are externally short-circuited, the cathode will have a total emission current I_c^0, of which only a part, I_c, will be collected by the plate, on the average, and the plate will have a total emission current I_p^0, of which only a part, I_p, will be collected by the cathode, on the average. But since there is thermal equilibrium, there can be no net average flow of current from cathode to plate or vice versa; for the existence of any net average flow of current in either direction would heat up the collector electrode and cool off the emitter electrode.

Hence, on the average,

$$I_c = I_p = I, \tag{470}$$

but also fluctuation currents are superposed upon both I_c and I_p, with average values that are zero. It is these fluctuation currents which are to be identified with the thermal-noise current fluctuations in the diode, whose average spectrum, according to statistical mechanics, is

$$\overline{|i|^2} = 2kTg, \tag{471}$$

where g is the diode conductance.

If these current fluctuations are regarded as shot current fluctuations associated with the cathode and anode emission currents I_c^0 and I_p^0, the average spectra of these two shot current fluctuations are, respectively,

$$\overline{|i_c|^2} = eI_c\Gamma_c^2 = eI\Gamma_c^2, \tag{472a}$$
$$\overline{|i_p|^2} = eI_p\Gamma_p^2 = eI\Gamma_p^2, \tag{472b}$$

and the average spectrum of the net shot current fluctuations is

$$\begin{aligned}\overline{|i|^2} &= \overline{|i_c|^2} + \overline{|i_p|^2} \\ &= eI(\Gamma_c^2 + \Gamma_p^2). \end{aligned} \tag{473}$$

Therefore the values of the average equilibrium current I from either electrode to the other and of the space-charge reduction factors Γ_c^2 and Γ_p^2 for the two emission currents must be determined.

First Γ_c^2 and Γ_p^2 must be evaluated. To do this, consider what the compensatory effect will be upon the pure shot current fluctuation produced by the emission of a single electron, of any energy, from the cathode. This electron will cause a change in the equilibrium potential distribution between the cathode and anode and a consequent change

in the average equilibrium currents I_c and I_p. But since both I_c and I_p are governed by the same Maxwell-Boltzmann distribution of emission velocities, and since the new equilibrium potential distribution therefore reduces both I_c and I_p by the same amount, the new equilibrium values of I_c and I_p must again be equal. Therefore the emission of one electron from the cathode does not cause any compensatory current fluctuation in either direction but causes only a pure shot current fluctuation—and this only if it succeeds in reaching the plate. Therefore $\Gamma_c^2 = 1$. By the same reasoning, $\Gamma_p^2 = 1$. Hence Eq. (473) becomes

$$\overline{|i|^2} = 2eI. \tag{474}$$

Returning to Eq. (471), the diode conductance g can be evaluated. To do this, let the external short circuit between cathode and plate be replaced by a voltage E, such that the plate, say, is positive with respect to the cathode. This voltage difference will cause I_c to be greater than I_p, by an amount given, according to the Maxwell-Boltzmann distribution, by the ratio

$$\frac{I_c}{I_p} = e^{eE/kT} \tag{475}$$

According to this expression, any variation in E, with the temperature T held fixed, must lead to variations in I_c and I_p given by

$$\frac{\delta I_c}{I_c} - \frac{\delta I_p}{I_p} = \frac{e\,\delta E}{kT}, \tag{476}$$

from which it follows that

$$\frac{1}{I_c}\frac{\partial I_c}{\partial E} - \frac{1}{I_p}\frac{\partial I_p}{\partial E} = \frac{e}{kT}. \tag{477}$$

In particular, when $E = 0$ this becomes

$$\frac{\partial(I_c - I_p)}{\partial E} = \frac{eI}{kT} \qquad (E = 0), \tag{478}$$

for then Eq. (470) holds. But by definition, the diode conductance g is

$$g = \frac{\partial(I_c - I_p)}{\partial E} \qquad (E = 0), \tag{479}$$

so that the result is that

$$g = \frac{eI}{kT}. \tag{480}$$

But according to this result Eqs. (471) and (474) are completely equivalent. Thus the complete equivalence of the thermal-noise and shot-noise points of view for the case of a diode in thermal equilibrium has been shown.

12·5. Other Types of Tube Noise.[1] *Cathode Flicker Effect.*—At audio frequencies the noise generated by vacuum tubes with oxide-coated cathodes exceeds greatly the magnitude predicted by the ordinary shot effect previously discussed. This departure is observed both under temperature-limited and space-charge operating conditions. It is also observed, at still lower frequencies, when the cathode is a tungsten filament. This effect was first pointed out by Johnson[2] and subsequently named "flicker effect" by Schottky,[3] who at the same time presented a theoretical analysis of the effect.

Experimentally, what is observed is an excess production of noise, superposed upon the usual shot effect, increasing in most cases with the square of the emission current (or temperature-limited plate current) but decreasing with the inverse square of the frequency. Because of its frequency dependence, the flicker effect becomes negligible at frequencies exceeding a few kilocycles per second.

The explanation of this phenomenon offered by Johnson and Schottky is that at the surface of the cathode, the emissivity fluctuates in time as a result of local changes in the surface condition of the cathode. These fluctuations in emissivity are thought to arise from the appearance or departure, at the cathode surface, of a single foreign atom or a small group of such atoms in the molecular condition.

The average time duration τ of one of these foreign atoms at the surface is the quantity assumed to be responsible for the falling off of this effect with increasing frequency. The total cathode emission current may therefore be thought of as itself fluctuating with an average period equal to τ. Hence, for frequencies $\omega \gg 1/\tau$, the components of these fluctuations in emissivity may be expected to fall of rapidly, and the exact analysis shows that they fall off as $1/\omega^2$.

The fact that the flicker effect increases with the square of the emission current is explained by the property that during its period of existence at the cathode surface, any foreign atom will cause a steady change in current that is proportional to the total current. Hence it follows that the mean square current fluctuations are proportional to the square of the total current.

Under space-charge conditions flicker effect is thought to be reduced even more than the true shot effect. The reduction, for flicker effect, is probably of the order of magnitude that follows from assuming, as Llewellyn first assumed[4] erroneously for shot effect, that the plate current

[1] For a more extensive treatment, see E. B. Moullin, *Spontaneous Fluctuations of Voltage*, Oxford University Press.

[2] J. B. Johnson, *Phys. Rev.*, **26**, 71 (1925).

[3] Moullin, *loc. cit.*

[4] Moullin, *loc. cit*

fluctuations are linearly reduced below the emission current fluctuations by the ratio $\partial I / \partial J$, where I is the plate current and J the emission current.

Positive-ion Noise.—Positive ions may exist in a vacuum tube either as a result of emission from the cathode or as a result of collision ionization of residual gas in the tube. Because the net positive-ion current is exceedingly small, its noise contribution is negligible in most modern vacuum tubes, even under space-charge conditions where the action of a single positive ion on the virtual cathode may multiply its current pulse many fold. A simple test of the possible existence of positive-ion noise is to measure the grid gas current. If this does not exceed a few hundredths of a microampere, positive-ion noise should be negligible.

Ballantine[1] was the first to analyze the effects of collision ionization on the noise in tubes. His result is that the mean square noise voltage due to this cause is proportional to the five-thirds power of the plate current.

Thompson and North,[2] in a subsequent investigation of this effect, showed that the fluctuations due to collision ionization are proportional to the grid gas current, giving the complete dependence on tube geometry and electrode potentials.

Positive-ion noise is more variable with frequency than the ordinary shot effect, for the transit times of positive ions are hundreds of times larger than those of electrons. Thus, at frequencies above 10^5 cycles transit time effects for positive ions already have to be considered.

Secondary-emission Noise.—When an electron emitted by the cathode of a multicollector tube reaches any particular collector electrode, the possibility exists that one or more secondary electrons may be emitted by that collector. If the potential distribution in the tube is not intentionally adjusted to retard this secondary emission, there will generally be a flow of secondary electrons between the various electrodes. The secondary electrons will be correlated in time with the primaries and will generally lead to departures in the noise behavior of the tube from that which has been heretofore described.

Suppose, for example, that in a screen-grid tube secondary electrons are produced at the screen, which flow to the plate. The most fundamental way of describing this situation is the following. As in the discussion of temperature-limited shot noise, the total cathode emission current can be written as

$$I_0(t) = e \int_{-\infty}^{+\infty} e^{i\omega t} \, df \sum_n e^{-i\omega t_n} \tag{481}$$

where t_n is the emission time of the nth emitted electron. Similarly,

[1] Moullin, *loc. cit.*

[2] B. J. Thompson and D. O. North, *RCA Rev.*, **5**, 3, January 1941.

the total current flowing to the screen can be written

$$I_s(t) = \int_{-\infty}^{+\infty} e^{j\omega t} \, df \sum_n q_{sn} e^{-j\omega t_n}, \qquad (482)$$

where q_{sn} is the charge delivered to the screen as a result of the emission of the nth electron from the cathode. In the same way again, the total current flowing to the plate can be written

$$I_p(t) = \int_{-\infty}^{+\infty} e^{j\omega t} \, df \sum_n q_{pn} e^{-j\omega t_n}, \qquad (483)$$

where q_{pn} is the charge delivered to the plate as a result of the emission of the nth electron from the cathode.

Consider now the average value of the product $I_s(t)I_p(t)$. This will be

$$\overline{I_s(t)I_p(t)} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{j(\omega+\omega')t} \, df \, df' \, \overline{\sum_{mn} q_{sm} q_{pn} e^{-j(\omega t_m + \omega' t_n)}}, \quad (484)$$

where the average of the double sum in the integrands of Eq. (484) is an ensemble average. As has previously been seen in the case of the temperature-limited diode, there is a d-c contribution to such an expression, evaluated by setting $\omega = \omega' = 0$. In this case the average of the double sum becomes

$$\sum_{mn} \overline{q_{sm} q_{pn}} = \sum_m \overline{q_{sm}} \sum_n \overline{q_{pn}}. \qquad (485)$$

But this contribution, as seen from Eqs. (482) and (483), leads simply to the product of the average values of $I_s(t)$ and $I_p(t)$; so that

$$\begin{aligned} \overline{I_s(t)I_p(t)} = \ & \overline{I_s(t)} \; \overline{I_p(t)} \\ & + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{j(\omega+\omega')t} \, df \, df' \, \overline{\sum_{mn} q_{sm} q_{pn} e^{-j(\omega t_m + \omega' \Sigma t_n)}}, \end{aligned} \quad (486)$$

where it is to be understood that the double integration excludes the contribution for $\omega = \omega' = 0$.

Again as previously seen, for arbitrary values of ω and ω' the ensemble average vanishes for the double sum in the integrand of Eq. (486). However, if $\omega + \omega' = 0$, this is no longer the case. For then Eq. (486) becomes

$$\overline{I_s(t)I_p(t)} - \overline{I_s(t)I_p(t)} = df' \int_{-\infty}^{+\infty} df \, \overline{\sum_{mn} q_{sm} q_{pn} e^{-j\omega(t_m - t_n)}} \quad (487)$$

Unless $t_m = t_n$, the ensemble average of the double sum in Eq. (487) will again vanish. Hence Eq. (487) can be written as

$$\overline{I_s(t)I_p(t)} - \overline{I_s(t)}\,\overline{I_p(t)} = df' \int_{-\infty}^{+\infty} df \sum_m \overline{q_{sm}q_{pm}}, \tag{488}$$

or, what is the same thing,

$$\frac{d}{df}\,\overline{I_s(t)I_p(t)} = df' \sum_m \overline{q_{sm}q_{pm}}, \tag{489}$$

where the average product $\overline{q_{sm}q_{pm}}$ is to be evaluated in the sense that q_{sm} and q_{pm} are the charges delivered to the screen and plate, respectively, as the result of the emission of the mth electron from the cathode.

Suppose now that the result of Eq. (489) is applied to the special case where the screen and plate are tied together (and hence may be regarded as identical) and where temperature-limited conditions obtain. This procedure will, of course, exclude momentarily any secondary emission effects.   As a result, however,

$$I_s(t) = I_p(t) = I_0(t), \tag{490}$$

and

$$q_{sm} = q_{pm} = e_m = e. \tag{491}$$

Correspondingly, Eq. (489) becomes

$$\frac{d}{df}\,I_0^2(t) = df' \sum_m e_m^2. \tag{492}$$

But for this case

$$\frac{d}{df}\,\overline{I_0^2(t)} = eI_0, \tag{493}$$

where I_0 is the average value of $I_0(t)$.   Hence, from Eq. (492),

$$df' \sum_m e_m^2 = eI_0. \tag{494}$$

But Eq. (489) can be written as

$$\left(\frac{d}{df}\right)\overline{I_s(t)I_p(t)} = df'\left(\overline{\frac{q_s q_p}{e^2}}\right)\sum_m e_m^2, \tag{495}$$

and hence, from Eq. (494),

$$\left(\frac{d}{df}\right)\overline{I_s(t)I_p(t)} = eI_0\left(\overline{\frac{q_s q_p}{e^2}}\right), \tag{496}$$

where $\overline{q_s q_p}$ is the average product of the charges delivered to the screen

and plate, respectively, as the result of the emission of *all* electrons from the cathode.

Equation (496) is thus a general formula, applicable to all cases, whether involving secondary emission or space-charge reduction or partition noise effects. In order to apply it, one must only be sure to interpret correctly the meaning of the average product $\overline{q_s q_p}$.

Thus, suppose again that the tube is in a temperature-limited condition but that the screen and plate are regarded as now distinct electrodes. In this case, if there is no secondary emission from the screen, one may assign q_s/e the two values 0,1 and q_p/e the two corresponding values 1,0. But the probability that $q_s/e = 0$ is I_p/I_0, and the probability that $q_s/e = 1$ is I_s/I_0. Hence

$$\left(\overline{\frac{q_s q_p}{e^2}}\right) = \left(\frac{I_p}{I_0}\right)(0)_s(1)_p + \left(\frac{I_s}{I_0}\right)(1)_s(0)_p = 0, \tag{497}$$

and therefore, in this case,

$$\left(\frac{d}{df}\right)\overline{I_s(t)I_p(t)} = 0. \tag{498}$$

That is to say, for the temperature-limited case, in the absence of secondary emission, there is no noise current generator between screen and plate, and hence no partition noise.

But if there were secondary emission from the screen, then in the temperature-limited case one would have to assign q_s/e the two values 0,1 and q_p/e the two corresponding values 1, δ, where δ is the *average* number of secondaries produced per primary. Then, instead of Eq. (497),

$$\left(\overline{\frac{q_s q_p}{e^2}}\right) = \left(\frac{I_p}{I_0}\right)(0)_s(1)_p + \left(\frac{I_s}{I_0}\right)(1)_s(\delta)_p = \left(\frac{I_s}{I_0}\right)\delta. \tag{499}$$

Then, placing this result in Eq. (496),

$$\left(\frac{d}{df}\right)\overline{I_s(t)I_p(t)} = eI_s\delta, \tag{500}$$

which means that there would be a noise current generator between screen and plate, acting like a pure shot current generator in which the emission of charges of magnitude $e\delta$, associated with the screen direct current I_s, takes place from screen to plate.

Again, consider the tube under space-charge conditions, with screen and plate tied together and therefore no secondary emission. In this case q_s/e and q_p/e both take on the value γ_α for the α-electrons and the value γ_β for the β-electrons. Therefore

$$\overline{\frac{q_s q_p}{e^2}} = \overline{\gamma_\alpha^2} + \overline{\gamma_\beta^2}, \tag{501}$$

where the averages on the right-hand side are taken over all emission energies.   That is, using the notation of the preceding section,

$$\overline{\overline{\gamma_\alpha^2}} = \left(\frac{1}{I_0}\right) \int_0^\infty \gamma_\alpha^2 \, dI_\alpha(E) = \left(\frac{1}{I_0}\right) \int_0^{E_m} \gamma_\alpha^2 \, dI(E) \qquad (502a)$$

$$\overline{\overline{\gamma_\beta^2}} = \left(\frac{1}{I_0}\right) \int_0^\infty \gamma_\beta^2 \, dI_\beta(E) = \left(\frac{1}{I_0}\right) \int_{E_m}^\infty \gamma_\beta^2 \, dI(E) \qquad (502b)$$

But the averages of γ_α^2 and γ_β^2 have been written previously with respect to the β-electron energies only, thus:

$$\overline{\overline{\gamma_\alpha^2}} = \left(\frac{1}{I_c}\right) \int_0^{E_m} \gamma_\alpha^2 \, dI(E) \qquad (503a)$$

$$\overline{\overline{\gamma_\beta^2}} = \left(\frac{1}{I_c}\right) \int_{E_m}^\infty \gamma_\beta^2 \, dI(E) \qquad (503b)$$

Since, however, the integrals in Eqs. (502) and (503) are respectively equal, it follows that

$$\overline{\overline{\gamma_\alpha^2}} I_0 = \overline{\gamma_\alpha^2} I_c \qquad (504a)$$

$$\overline{\overline{\gamma_\beta^2}} I_0 = \overline{\gamma_\beta^2} I_c \qquad (504b)$$

so that Eq. (501) becomes

$$\overline{\frac{q_s q_p}{e^2}} = (\overline{\gamma_\alpha^2} + \overline{\gamma_\beta^2}) \left(\frac{I_c}{I_0}\right) \qquad (505)$$

and Eq. (496) becomes [noting that $I_s(t) = I_p(t) = I_c(t)$]

$$\left(\frac{d}{df}\right) \overline{I_c^2} = e I_c \Gamma^2, \qquad (506)$$

which is the same as the result earlier established for this case.

Finally, consider the case where the screen and plate are distinct but without any secondary emission taking place.   In this case q_s/e takes on the value $[1 - (I_s/I_c)(1 - \gamma_\beta)]$ and q_p/e takes on the value

$$-\left(\frac{I_p}{I_c}\right)(1 - \gamma_\beta)$$

if an emitted electron goes to the screen; q_s/e takes on the value

$$-\left(\frac{I_s}{I_c}\right)(1 - \gamma_\beta)$$

and q_p/e takes on the value $[1 - (I_p/I_c)(1 - \gamma_\beta)]$ if an emitted electron goes to the plate; and q_s/e takes on the value $(I_s/I_c)\gamma_\alpha$ and q_p/e takes on the value $(I_p/I_c)\gamma_\alpha$ if an emitted electron goes back to the cathode.

Therefore

$$\frac{\overline{q_s q_p}}{e^2} = -\left(\frac{1}{I_0}\right) \int_0^\infty \left\{\left[1 - \left(\frac{I_s}{I_c}\right)(1 - \gamma_\beta)\right]\left(\frac{I_p}{I_c}\right)(1 - \gamma_\beta)\, dI_s(E)\right.$$
$$+ \left(\frac{I_s}{I_c}\right)(1 - \gamma_\beta)\left[1 - \left(\frac{I_p}{I_c}\right)(1 - \gamma_\beta)\right] dI_p(E)$$
$$\left. - \left(\frac{I_s I_p}{I_c^2}\right)\gamma_\alpha^2\, dI_\alpha(E)\right\} \quad (507)$$

$$= -\left(\frac{1}{I_0}\right)\left\{\left[\left(\frac{I_p}{I_c}\right)(1 - \overline{\gamma_\beta}) - \left(\frac{I_s I_p}{I_c^2}\right)\overline{(1 - \gamma_\beta)^2}\right] I_s\right.$$
$$+ \left[\left(\frac{I_s}{I_c}\right)(1 + \overline{\gamma_\beta}) - \left(\frac{I_s I_p}{I_c^2}\right)\overline{(1 - \gamma_\beta)^2}\right] I_p$$
$$\left. - \left(\frac{I_s I_p}{I_c^2}\right)\overline{\gamma_\alpha^2}(I_0 - I_c)\right\} \quad (508)$$

$$= -\left(\frac{1}{I_0}\right)\left(\frac{I_s I_p}{I_c}\right)[2(1 - \overline{\gamma_\beta}) - \overline{(1 - \gamma_\beta)^2} - \overline{\gamma_\alpha^2}] \quad (509)$$

$$= -\left(\frac{1}{I_0}\right)\left(\frac{I_s I_p}{I_c}\right)(1 - \overline{\gamma_\beta^2} - \overline{\gamma_\alpha^2}), \quad (510)$$

so that Eq. (496) becomes

$$\left(\frac{d}{df}\right)\overline{I_s(t)I_p(t)} = -e\left(\frac{I_s I_p}{I_c}\right)(1 - \overline{\gamma_\alpha^2} - \overline{\gamma_\beta^2})$$
$$= -e\left(\frac{I_s I_p}{I_c}\right)(1 - \Gamma^2). \quad (511)$$

This result agrees with the result formerly obtained for the partition noise current generator between screen and plate.

These few examples suffice to show how the general formula Eq. (496) may be applied to any particular case. The case where both partition noise and secondary emission noise exist together in a pentode with space charge is left untreated. To deal with this case requires merely a minor change in the values just previously assumed for q_s/e and q_p/e, involving the quantity δ which was introduced in the temperature-limited case. The cases of amplifier tubes having special secondary electron emitting electrodes and of electron multiplier tubes are also not treated, although such cases also may be treated with the aid of equations analogous to Eq. (496).

From the discussion above, it is seen that the processes of secondary emission and of space-charge reduction are in a way just different manifestations of one fundamental process, for which Eq. (496) provides the basic formal treatment.

Microphonics.—A still further form of noise occurring in amplifier tubes, but one that strictly speaking might better be described otherwise than as noise, is so-called microphonics. This microphonic "noise"

arises from mechanical vibrations of the tube electrodes. The effect is found to arise mostly at audio frequencies where, due to the vicinity of loud-speakers or other sources of vibration, the tube structure frequently commences to vibrate at certain resonant frequencies. Although the mechanical vibrations are confined to low frequencies, they may cause disturbances at radio frequencies through modulation effects. Microphonics can be reduced by the use of shock-absorbing tube mounts and by shielding the tubes from sound waves. They also depend to a great extent upon the tubes themselves, for out of a set of tubes of the same type it is often possible to select certain ones in which microphonic action is much less pronounced. In severe cases it is necessary to build special nonmicrophonic tubes.

12·6. Other Types of Input-circuit Noise.—Under ideal conditions, input-circuit noise is pure thermal noise, at an absolute temperature T that may usually be taken to be room temperature or about 63°F or 290°K.

However, departures from the ideal conditions are frequently encountered. Such departures always are due to the failure of the input circuit to be in true thermal equilibrium. For example, an input circuit that is completely passive will still not be in thermal equilibrium if various parts of that circuit are, for any reason, subjected to temperature gradients.

Again, the output of a preceding stage or component of a receiver is often regarded as the input circuit of a given following stage or component, and under such conditions one must inquire into the character of the preceding stage or component. For example, the preceding component which acts as input circuit may include a crystal or tube detector or frequency converter. Such an element is in general nonpassive and hence is incapable of exhibiting true thermal equilibrium. The same holds true if the preceding component is another amplifier stage or if it includes any form of loading due to a tube or other nonpassive element.

Still again, when the input circuit includes an antenna, the radiation resistance of the antenna is very seldom in a condition of strict temperature equilibrium. This is because the radiation field seen by the antenna couples the latter with sources of radiation of all descriptions, such as the terrain, the atmosphere, and even the sun and other astronomical bodies, and such sources of radiation are far from being in thermal equilibrium. However, it is frequently useful and a sufficiently good approximation to regard the antenna to be in thermal equilibrium with the outside space. Whether or not this is so, it is common practice to specify the over-all sensitivity of a receiver in terms of the thermal noise that would be produced by the antenna's radiation resistance if the latter were in true thermal equilibrium at room temperature.

Similarly, the sensitivity of any amplifier stage or component is, for

the sake of standardization, defined in terms of the thermal noise that its
effective signal source would produce if the latter were in thermal equi-
librium at room temperature.

**12·7. Amplifier Sensitivity: Definition and Theoretical Discussion of
Noise Figure, Available Power Gain, and Noise Temperature.**—At the
beginning of this chapter the manner in which the noise in an amplifier
sets a limit to the amount of useful amplification was discussed qualita-
tively. It was there indicated, again qualitatively, that as the level
of the noise is lowered, the amplifier is able to detect smaller and smaller
signals; in other words, the *sensitivity* of the amplifier is increased.

There is an absolute limit to the sensitivity that can be obtained with
an amplifier, which is set not by the amplifier itself but by the signal
source that feeds it. For the signal source of any amplifying system is
itself always a source of noise even if the remainder of the amplifier is not.
As has been seen in the preceding section, the input circuit of an amplifier
(which contains the signal source) is to be regarded, under ideal condi-
tions, as a source of thermal noise at room temperature. Of course, it is
theoretically possible to lower the temperature of a source of true thermal
noise indefinitely and thereby to eliminate completely its thermal noise.
But, practically speaking, the ultimate signal source of an amplifier,
whatever it may be, is usually beyond the control of the observer, as far
as the regulation of its "temperature" is concerned.

Thus, if the ultimate signal source is the radiation field seen by an
antenna, it is obviously impossible to suppress completely the sources of
noise in the surrounding terrain, atmosphere, and other more or less
remote seats of electromagnetic energy. Similarly, if the ultimate signal
source is a communication network, a microphone or television scanning
system, an ionization chamber, a photocell, or a counter, the ultimate
sources of noise residing in the signal source are partially or wholly
beyond the control of the observer.

Therefore, since in most cases the ultimate signal source approximates
in behavior a passive impedance in thermal equilibrium at room tem-
perature, it is reasonable to adopt the convention that the absolute limit
to an amplifier's over-all sensitivity is fixed by regarding its ultimate
signal source as a source of thermal noise at room temperature. If, in any
particular case, the ultimate signal source produces more or less noise
than this rule would assign, a corresponding correction can be made in
specifying the amplifier sensitivity for the particular case in question.

In this section, for the purpose of dealing quantitatively with the
problem of amplifier sensitivity, it will be convenient to introduce the con-
cepts of *noise figure, available power gain,* and *noise temperature.* These
quantities will be defined for an amplifier component network, and
methods will be presented for measuring them by means of electron tube

noise standards. For the sake of generality, the amplifier component network in question will be considered as the first of two networks in cascade, and the definitions will be phrased in terms of suitable output noise powers, registered in the output of the combined network.

Network Block Diagram.—In Fig. 12·16 the two component networks are shown in cascade, with their terminals (11), (22), and (33) located across the shunt output resistances R_1 of the signal source and R_2 of Network 1 and the shunt resistance R_3 of the power detector, respectively. The input impedances Z_1 of Network 1 and Z_2 of Network 2 include the reactances shunting their respective signal sources R_1 and R_2, and the
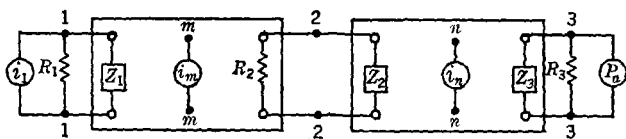
FIG. 12·16.—Two component networks in cascade.

output impedance Z_3 of Network 2 includes the shunt reactance of the power detector, which registers the power P_a absorbed in R_3.

With a thermal signal source of resistance R_1 at room temperature T, the sources of noise in the configuration of Fig. 12·16 are (1) a thermal-noise current i_1 whose statistical spectrum is

$$\overline{i_1^2} = \frac{2kT}{R_1} \tag{512}$$

and (2) an arbitrary number of *statistically independent* excess noise cur-rents i_m and i_n, impressed across in-ternal branches (mm) of Network 1 and (nn) of Network 2, whose sta-tistical spectra have arbitrary magni-tudes $\overline{i_m^2}$ and $\overline{i_n^2}$, respectively. Thus, if Network 1 is passive and dissipa-tionless, its shunt output resistance R_2 will be the reflection of R_1, and the

FIG. 12·17.—First network of Fig. 12·16 replaced by a thermal signal source of equal resistance R_2 at room temperature.

only noise current will be i_1. Otherwise the extra noise currents are all included in the i_m.

In Fig. 12·17 the first component network is replaced by a thermal signal source of equal resistance R_2 at room temperature T. In this configuration the terminal impedances remain unchanged, but the sources of noise are (1) a thermal-noise current i_2 whose statistical spectrum is

$$\overline{i_2^2} = \frac{2kT}{R_2} \tag{513}$$

and (2) the statistically independent excess noise currents i_n of Fig. 12·16. The output noise power in this case is P_b.

Output Noise Power.—In the two configurations of Figs. 12·16 and 12·17 the total output noise powers are, respectively,

$$P_a = P_1 + \sum_m P_m + \sum_n P_n, \tag{514a}$$

$$P_b = P_2 + \sum_n P_n, \tag{514b}$$

where

$$P_1 = \int_{-\infty}^{+\infty} P_1(\omega) \, df = \frac{\overline{i_1^2}}{R_3} \int_{-\infty}^{+\infty} |Z_{13}(j\omega)|^2 \, df, \tag{515a}$$

$$P_2 = \int_{-\infty}^{+\infty} P_2(\omega) \, df = \frac{\overline{i_2^2}}{R_3} \int_{-\infty}^{+\infty} |Z_{23}(j\omega)|^2 \, df, \tag{515b}$$

$$P_m = \int_{-\infty}^{+\infty} P_m(\omega) \, df = \frac{\overline{i_m^2}}{R_3} \int_{-\infty}^{+\infty} |Z_{m3}(j\omega)|^2 \, df, \tag{515c}$$

$$P_n = \int_{-\infty}^{+\infty} P_n(\omega) \, df = \frac{\overline{i_n^2}}{R_3} \int_{-\infty}^{+\infty} |Z_{n3}(j\omega)|^2 \, df, \tag{515d}$$

and where the transfer impedances Z_{13}, Z_{23}, Z_{m3}, Z_{n3} refer to the input terminals (11), (22), (mm), (nn), respectively, and to the output terminals (33).

Definition of Noise Figure, Available Power Gain, and Noise Temperature.—For an arbitrary network the noise figure F, available power gain $\mathcal{W}$, and noise temperature t are

F = total noise power output from network with a thermal signal source ÷ noise power output from thermal signal source,

$\mathcal{W}$ = noise power output from thermal signal source ÷ noise power output from thermal output impedance,

t = total noise power output from network with a thermal signal source ÷ noise power output from thermal output impedance.

In terms of the output noise power expressions of Eq. (515), these definitions lead to the following equations for the noise figure F_1, available power gain $\mathcal{W}_1$ and noise temperature t_1 of Network 1, and for the noise figures F_2 of Network 2 and F_{12} of the combined network:

$$F_1 = \frac{P_1 + \sum_m P_m}{P_1}, \tag{516a}$$

$$\mathcal{W}_1 = \frac{P_1}{P_2}, \tag{516b}$$

$$t_1 = \frac{P_1 + \sum_m P_m}{P_2} \tag{516c}$$

$$= F_1 \mathcal{W}_1, \tag{516d}$$

$$F_2 = \frac{P_1 + \sum_n P_n}{P_2}, \tag{516e}$$

$$F_{12} = \frac{P_1 + \left(\sum_m P_m + \sum_n P_n\right)}{P_1} \tag{516f}$$

$$= F_1 + \frac{F_2 - 1}{\mathcal{W}_1}, \tag{516g}$$

$$= \frac{t_1 + F_2 - 1}{\mathcal{W}_1}. \tag{516h}$$

Single-frequency Noise Figures, Available Power Gain, and Noise Temperature.—If, through suitable filtering, the output power detector in Figs. 12·16 and 12·17 is made to respond only to the noise powers $P(\omega) \Delta f$ within a narrow band Δf at the angular frequency $\omega = 2\pi f$, these single-frequency powers may be used in Eq. (516) to define and measure the single-frequency noise figure $F_1(\omega)$, available power gain $\mathcal{W}_1(\omega)$, and noise temperature $t_1(\omega)$ of Network 1, and the single-frequency noise figures $F_2(\omega)$ of Network 2 and $F_{12}(\omega)$ of the combined network. The explicit forms of these quantities are [*cf.* Eqs. (515) and (516)]

$$F_1(\omega) = 1 + \sum_m \left(\frac{\overline{i_m^2}}{\overline{i_1^2}}\right) \left|\frac{Z_{m3}(j\omega)}{Z_{13}(j\omega)}\right|^2, \tag{517a}$$

$$\mathcal{W}_1(\omega) = \left(\frac{\overline{i_1^2}}{\overline{i_2^2}}\right) |Z_{13}(j\omega)/Z_{23}(j\omega)|^2, \tag{517b}$$

$$t_1(\omega) = F_1(\omega)\mathcal{W}_1(\omega), \tag{517c}$$

$$F_2(\omega) = 1 + \sum_n \left(\frac{\overline{i_n^2}}{\overline{i_2^2}}\right) \left|\frac{Z_{n3}(j\omega)}{Z_{23}(j\omega)}\right|^2 \tag{517d}$$

$$F_{12}(\omega) = F_1(\omega) + \frac{F_2(\omega) - 1}{\mathcal{W}_1(\omega)} \tag{517e}$$

From these equations it can be shown that F_1, $\mathcal{W}_1$, t_1, F_2, and F_{12} are weighted means of their single-frequency analogues, with weight factors given by one or another of the normalized pass bands

$$B_1(\omega) = \frac{|Z_{13}(j\omega)|^2}{\int_{-\infty}^{+\infty} |Z_{13}(j\omega)|^2 \, df} \tag{518a}$$

$$B_2(\omega) = \frac{|Z_{23}(j\omega)|^2}{\int_{-\infty}^{+\infty} |Z_{23}(j\omega)|^2 \, df}, \tag{518b}$$

of the combined network or Network 2, respectively; thus

$$F_1 = \int_{-\infty}^{+\infty} F_1(\omega) B_1(\omega) \, df, \tag{519a}$$

$$\mathbb{W}_1 = \int_{-\infty}^{+\infty} \mathbb{W}_1(\omega) B_2(\omega) \, df, \tag{519b}$$

$$t_1 = \int_{-\infty}^{+\infty} t_1(\omega) B_2(\omega) \, df, \tag{519c}$$

$$F_2 = \int_{-\infty}^{+\infty} F_2(\omega) B_2(\omega) \, df, \tag{519d}$$

$$F_{12} = \int_{-\infty}^{+\infty} F_{12}(\omega) B_1(\omega) \, df. \tag{519e}$$

Noise Measurements with Electron Tube Noise Standards.—If, in the configuration of Fig. 12·16, electron tube noise standards are placed across the terminals (11) and (22), there will be impressed across these terminals pure shot-noise currents Δi_1 and Δi_2 whose statistical spectra are arbitrarily variable and known in absolute magnitude from the formulas

$$\overline{(\Delta i_1)^2} = eI_1, \tag{520a}$$

$$\overline{(\Delta i_2)^2} = eI_2, \tag{520b}$$

where I_1 and I_2 are the arbitrarily variable direct currents of the respective noise standards.

Furthermore, if the two noise standards are tuned to zero admittance for the frequencies passed by the combined network and Network 2, respectively, then the incremental output noise powers registered at the terminals (33), due to the respective shot-noise currents Δi_1 and Δi_2, will be

$$\Delta P_1 = \left(\frac{eI_1}{R_3}\right) \int_{-\infty}^{+\infty} |Z_{13}(j\omega)|^2 \, df \tag{521a}$$

$$\Delta P_2 = \left(\frac{eI_2}{R_3}\right) \int_{-\infty}^{+\infty} |Z_{23}(j\omega)|^2 \, df, \tag{521b}$$

where the transfer impedances $Z_{13}(j\omega)$ and $Z_{23}(j\omega)$ remain unmodified by the presence of the two electron tube noise standards; for if either noise standard has an admittance that is irremovable, it is generally possible to remove the equivalent amount of admittance from the networks themselves.

With two noise standards thus applied to the circuit configurations of Figs. 12·16 and 12·17, the quantities $\mathbb{W}_1$, t_1, F_2, and F_{12} can be directly

measured, and F_1 can be calculated from these quantities with the aid of Eq. (516d) or (516g).

Measurement of Available Power Gain $\mathfrak{W}_1$.—In the configuration of Fig. 12·16, bring the two noise standards across the terminals (11) and (22) alternately to arbitrary direct currents I_1 and I_2, and note the corresponding incremental output noise powers ΔP_1 and ΔP_2 [*cf.* Eqs. (521)]. Now from Eqs. (515) and (516)

$$\mathfrak{W}_1 = \left(\frac{\overline{i_1^2}}{\overline{i_2^2}}\right) \frac{\displaystyle\int_{-\infty}^{+\infty} |Z_{13}(j\omega)|^2 \, df}{\displaystyle\int_{-\infty}^{+\infty} |Z_{23}(j\omega)|^2 \, df}. \tag{522}$$

and hence, from Eqs. (512) and (513),

$$\mathfrak{W}_1 = \left(\frac{R_2}{R_1}\right) \frac{\displaystyle\int_{-\infty}^{+\infty} |Z_{13}(j\omega)|^2 \, df}{\displaystyle\int_{-\infty}^{-\infty} |Z_{23}(j\omega)|^2 \, df}. \tag{523}$$

Therefore, from Eqs. (521),

$$\mathfrak{W}_1 = \frac{I_2 R_2}{I_1 R_1} \frac{\Delta P_1}{\Delta P_2}. \tag{524}$$

Thus $\mathfrak{W}_1$ can be measured by means of Formula (524) by noting the ratio of the incremental output noise powers ΔP_1 and ΔP_2 corresponding to arbitrary direct currents I_1 and I_2 if the resistances R_1 and R_2 are known.

Measurement of Noise Temperature t_1.—In the configuration of Fig. 12·16, note the total output noise power

$$P_a = P_1 + \sum_m P_m + \sum_n P_n. \tag{525}$$

Then, changing to the configuration of Fig. 12·17, bring the noise standard across the terminals (22) to a direct current I_2 such that the total output noise power $P_b + \Delta P_2$ is the same as the total output noise power P_a in Fig. 12·16. Then, since

$$P_b = P_2 + \sum_n P_n, \tag{526}$$

it follows that

$$P_1 + \sum_m P_m = P_2 + \Delta P_2 \tag{527}$$

and

$$t_1 = \frac{P_1 + \displaystyle\sum_m P_m}{P_2} \tag{528a}$$

$$= 1 + \frac{\Delta P_2}{P_2}. \tag{528b}$$

Now, from Eqs. (513) and (515),

$$P_2 = \left(\frac{2kT}{R_2 R_3}\right) \int_{-\infty}^{+\infty} |Z_{23}(j\omega)|^2 \, df, \tag{529}$$

while, from Eqs. (521b)

$$\Delta P_2 = \left(\frac{e I_2}{R_3}\right) \int_{-\infty}^{+\infty} |Z_{23}(j\omega)|^2 \, df. \tag{530}$$

Hence

$$t_1 = 1 + \left(\frac{e}{2kT}\right) I_2 R_2. \tag{531}$$

For an assumed room temperature $T = 290°K$, $(e/2kT) = 20$ volts^{-1}; so that, for this assumed value of the room temperature,

$$t_1 = 1 + 20 I_2 R_2. \tag{532}$$

Thus, from the value of I_2 in amperes and that of R_2 in ohms, t_1 can be measured by means of Formula (532).

Measurement of Noise Figure F_2.—In the configuration of Fig. 12·17, bring the noise standard across the terminals (22) to a direct current I_2 such that the original output noise power

$$P_b = P_2 + \sum_n P_n \tag{533}$$

is just doubled.   Then

$$F_2 = 1 + \frac{\displaystyle\sum_n P_n}{P_2}, \tag{534a}$$

$$= \frac{P_b}{P_2} \tag{534b}$$

$$= \frac{\Delta P_2}{P_2}. \tag{534c}$$

Now

$$\Delta P_2 = \left(\frac{e I_2}{R_3}\right) \int_{-\infty}^{+\infty} |Z_{23}(j\omega)|^2 \, df, \tag{535a}$$

$$P_2 = \left(\frac{2kT}{R_2 R_3}\right) \int_{-\infty}^{+\infty} |Z_{23}(j\omega)|^2 \, df, \tag{535b}$$

so that

$$F_2 = \left(\frac{e}{2kT}\right) I_2 R_2. \tag{536}$$

Thus, at the room temperature assumed above, F_2 can be measured from the value of I_2 in amperes and that of R_2 in ohms by means of the formula

$$F_2 = 20I_2R_2. \tag{537}$$

Measurement of Noise Figure F_{12}.—In the configuration of Fig. 12·16, turn off the noise standard across the terminals (22) and bring the noise standard across the terminals (11) to a direct current I_1 such that the original output noise power

$$P_a = P_1 + \sum_m P_m + \sum_n P_n \tag{538}$$

is just doubled.   Then

$$F_{12} = 1 + \frac{\sum_m P_m + \sum_n P_n}{P_1}, \tag{539a}$$

$$= \frac{P_a}{P_1} \tag{539b}$$

$$= \frac{\Delta P_1}{P_1}. \tag{539c}$$

Now

$$\Delta P_1 = \left(\frac{eI_1}{R_3}\right) \int_{-\infty}^{+\infty} |Z_{13}(j\omega)|^2 \, df, \tag{540}$$

while, from Eqs. (512) and (515),

$$P_1 = \left(\frac{2kT}{R_1R_3}\right) \int_{-\infty}^{+\infty} |Z_{13}(j\omega)|^2 \, df. \tag{541}$$

Hence

$$F_{12} = \left(\frac{e}{2kT}\right) I_1R_1. \tag{542}$$

Thus, at the room temperature assumed above, F_{12} can be measured from the value of I_1 in amperes and that of R_1 in ohms by means of the formula

$$F_{12} = 20I_1R_1. \tag{543}$$

Limitations of the Foregoing Methods.—The usual definitions of mean noise figure, available power gain, and noise temperature break down whenever the signal source resistance R_1 varies with frequency through the pass band of the combined network or the output resistance R_2 of Network 1 varies with frequency through the pass band of Network 2. However, the definitions of single-frequency noise figure, available power gain, and noise temperature retain their meaning at any frequency ω at

which the frequency dependent signal source resistance $R_1(\omega)$ and output resistance $R_2(\omega)$ do not become negative, and these single-frequency quantities can therefore be measured by the foregoing methods at any such frequency, provided the ohmic resistances R_1 and R_2 employed in these measurements are made equal to $R_1(\omega)$ and $R_2(\omega)$, respectively. Furthermore, if $R_1(\omega)$ and $R_2(\omega)$ are positive at all frequencies passed by the networks, it is possible to generalize the definitions of mean noise figure, available power gain, and noise temperature by expressing these quantities as weighted means of their single-frequency analogues, in accordance with Eqs. (519).

If $R_1(\omega)$ becomes negative at any frequency ω, it becomes impossible to define $F_1(\omega)$, $\mathcal{W}_1(\omega)$, or $F_{12}(\omega)$ operationally in terms of physically realizable measurements. Similarly, if $R_2(\omega)$ becomes negative, it is impossible to define operationally $\mathcal{W}_1(\omega)$, $t_1(\omega)$, or $F_2(\omega)$. In the latter case $F_1(\omega)$ and $F_{12}(\omega)$ can be defined and $F_{12}(\omega)$ can be measured provided $R_1(\omega)$ is positive, but the dependence of $F_{12}(\omega)$ upon the excess noise of Network 2 can no longer be expressed in terms like those of Eqs. (516g) and (516h).

Strictly speaking, these limitations are not attributable to the particular methods of measurement but are inherent in any operational definition of noise figure, available power gain, and noise temperature. However, for theoretical purposes it is possible and also sometimes useful to generalize the definition of these quantities[1] in the same sense that attaches to our generalization of Nyquist's thermal-noise theorem.

Specification of Noise Figure, Available Power Gain, and Noise Temperature.—The single-frequency noise figure $F_1(\omega)$, available power gain $\mathcal{W}_1(\omega)$, and noise temperature $t_1(\omega)$ of a given network are not, in general, attributes of this network alone but are dependent upon the output impedance of the signal source and the input impedance of the following network. Therefore, along with measured values of these quantities, the corresponding values of these impedances should be specified.

Furthermore, the mean noise figure F_1, available power gain $\mathcal{W}_1$, and noise temperature t_1 of the given network are, in general, dependent upon the bandpass characteristic of the following network. Therefore, along with measured values of these quantities, the corresponding bandpass characteristic of the following network should be specified. In the special case when $F_1(\omega)$, $\mathcal{W}_1(\omega)$, or $t_1(\omega)$ is frequency-independent, the mean quantity is equal to the single-frequency quantity and is then independent of the bandpass characteristic of the following network.

12·8. Amplifier Sensitivity: Methods of Improvement by the Suppression of Tube Noise.[2]—In the design of an amplifier component net-

---

[1] E. J. Schremp, to be published.

[2] Cf. also E. J. Schremp, to be published.

work for operation at low signal levels, one objective is to reduce the noise figure to a minimum, subject to certain other practical requirements. From Eq. (517a) it is seen that to achieve this, it is necessary to minimize both the ratios $\overline{i_m^2}/\overline{i_1^2}$ and the ratios $|Z_{m3}(j\omega)/Z_{13}(j\omega)|^2$. The ratios $\overline{i_m^2}/\overline{i_1^2}$ depend upon the signal source resistance and upon such internal characteristics of the network as the resistance elements, tube types, and operating voltages used. The ratios $|Z_{m3}(j\omega)/Z_{13}(j\omega)|^2$ between the responses to excess noise and the response to signal depend exclusively upon the amplifying properties of the network. In order fully to minimize these ratios, it is not generally sufficient merely to maximize the signal transfer impedance $Z_{13}(j\omega)$. Rather, it is necessary also to manipulate (for example, by means of internal feedback) the excess noise transfer impedances $Z_{m3}(j\omega)$ so as to suppress wholly or partially the output excess noise, even at a possible sacrifice in the signal gain.

As each of the terms of Eq. (517a) of the form $(\overline{i_m^2}/\overline{i_1^2})|Z_{m3}(j\omega)/Z_{13}(j\omega)|^2$ tends to zero, the noise figure of the amplifier component network approaches unity. In this limiting case the network is said to be an ideal amplifier. Although any one term of this form may be made to vanish in four possible ways (i.e., by letting i_m or $Z_{m3}(j\omega)$ approach zero or by letting i_1 or $Z_{13}(j\omega)$ approach infinity, the two ways that correspond to the condition

$$\frac{Z_{m3}(j\omega)}{Z_{13}(j\omega)} = 0 \tag{544}$$

will be of most concern in what follows. When, for each independent excess noise source m, Eq. (544) is fulfilled, it is generally fulfilled at only one frequency, and therefore the amplifier network is ideal at only one frequency. It is, in general, possible to achieve the conditions of Eq. (544) for each noise source m in one or more ways by establishing constraints on the values of the branch admittances of the amplifier network, as implied by Eq. (544) and more explicitly indicated when one has at hand the explicit expressions for the transfer impedances $Z_{m3}(j\omega)$ and $Z_{13}(j\omega)$ for the particular amplifier network in question. These conditions for an ideal amplifier are therefore, in a sense, analogous to a set of bridge balance conditions.

Thus, in order to deal adequately with the problem of suppressing tube noise in amplifiers, it is advisable to begin with a discussion of the various possible circuit arrangements that can be built around a single tube (here, for simplicity, assumed to be a triode) in such wise as to yield *no* response in the output to the shot noise generated by that tube. Hence the single-triode case will be considered here in enough detail to afford a fundamental understanding of our method of approach. Although other, more complicated cases are not treated here, the methods in such cases are but simple extensions of those described here.

As has been seen, the source of shot noise in a triode tube, at least at relatively low frequencies where transit time is unimportant, may be regarded as a constant-current generator impressed between the cathode and plate. Hence, those circuit arrangements containing a single triode in which the *transfer impedance* between this current generator and the output terminals *vanishes* will be considered first.

Accordingly, it is necessary to begin with a brief discussion, first of the *branch admittances* encountered in a general triode circuit and second of the *transfer impedances* encountered therein.

Branch Admittances of an Isolated Triode.—An *isolated* vacuum-tube triode (i.e., one without external circuit elements attached) constitutes a three-node network, in which the three electrodes c, g, p (cathode, grid, plate) are the three nodes. The *nodal* network equations for it can be written

$$i_m = \sum_\mu Y_\mu^m e_\mu \qquad (m, \mu = c, g, p) \tag{545}$$

where i_m is the net alternating current entering the node m, e_m is the alternating voltage of the node m, and

$$Y_n^m = \frac{\partial i_m}{\partial e_n} \qquad (m, n = c, g, p) \tag{546}$$

is the negative of the *branch admittance* between nodes m and n (which has a definite sense, for it may not be bilateral). These branch admittances satisfy the relations

$$\sum_\mu Y_\mu^m = \sum_\mu Y_m^\mu = 0 \qquad (m, \mu = c, g, p.) \tag{547}$$

They form a three-by-three array

$$\begin{Bmatrix} Y_c^c & Y_c^g & Y_c^p \\ Y_g^c & Y_g^g & Y_g^p \\ Y_p^c & Y_p^g & Y_p^p \end{Bmatrix} \tag{548}$$

in which, according to Eq. (547), the sums of all rows and all columns are zero. Therefore, the diagonal admittances Y_c^c, Y_g^g, Y_p^p are derivable from the nondiagonal admittances (the actual branch admittances) through the relations [*cf.* Eq. (547)]

$$-Y_c^c = Y_c^g + Y_c^p = Y_g^c + Y_p^c, \tag{549a}$$
$$-Y_g^g = Y_g^p + Y_g^c = Y_p^g + Y_c^g, \tag{549b}$$
$$-Y_p^p = Y_p^c + Y_p^g = Y_c^p + Y_g^p. \tag{549c}$$

Now if

$$C_{cg} = C_{gc}, \tag{550a}$$

$$C_{gp} = C_{pg}, \tag{550b}$$

$$C_{pc} = C_{cp}, \tag{550c}$$

are the (bilateral) interelectrode capacitances of the triode tube and g_m and r_p are its mutual conductance and plate resistance, respectively, it may be readily deduced from the defining Eqs. (546) or from the low-frequency approximation to Fig. 12·12 that the branch admittances $-Y_n^m$ of an isolated triode possess the values set forth in the following array:

$$-Y_n^m =$$

$n \diagdown m$	c	g	p
c		$j\omega C_{gc}$	$g_m + \dfrac{1}{r_p} + j\omega C_{pc}$
g	$g_m + j\omega C_{gc}$		$-g_m + j\omega C_{pg}$
p	$\dfrac{1}{r_p} + j\omega C_{pc}$	$j\omega C_{gp}$	

$$\tag{551}$$

The diagonal elements Y_c^c, Y_g^g, Y_p^p are not exhibited in this array; but as previously mentioned, they are to be found with the aid of Eqs. (549). It will be noted that none of the nondiagonal admittances (the actual branch admittances) is bilateral; for instead of the relations

$$Y_g^c - Y_c^g = Y_p^g - Y_g^p = Y_c^p - Y_p^c = 0 \tag{552}$$

which characterize bilateral branch admittances, we find the relations

$$Y_g^c - Y_c^g = Y_p^g - Y_g^p = Y_c^p - Y_p^c = g_m. \tag{553}$$

Branch Admittances of a General Three-node Triode Network.—The branch admittances $-Y_n^m$ just discussed above are intrinsic to a triode. If across every nodal pair (m,n) an external admittance of arbitrary value $-\Delta Y_n^m$ is attached, then one has for branch admittances the negatives of the quantities

$$Y_n^{m\prime} - Y_n^m + \Delta Y_n^m \qquad (m, n = c, g, p), \tag{554}$$

and the resulting network will constitute what may be called a *general three-node triode network*.

If every external branch admittance $-\Delta Y_n^m$ is bilateral, i.e., if

$$\Delta Y_n^m - \Delta Y_m^n = 0 \qquad (m, n = c, g, p), \tag{555}$$

then the resulting network will be referred to as *bilaterally loaded*.

Transfer Impedances of an Isolated Triode: Definition.—The nodal network equations (545) may be rewritten in a way that expresses the nodal voltages e_m in terms of the nodal currents i_m, thus:

$$e_m = \sum_\mu Z_\mu^m i_\mu \qquad (m, \mu = c, g, p). \tag{556}$$

If, further, e_{mn} is defined as the internodal voltage difference

$$e_{mn} = e_m - e_n, \tag{557}$$

and if

$$Z_r^{mn} = Z_r^m - Z_r^n, \tag{558}$$

then, from Eq. (556),

$$e_{mn} = \sum_\mu Z_\mu^{mn} i_\mu. \tag{559}$$

Again, if it is assumed that the nodal currents i_m are expressible solely in terms of internodal constant-current generators i_{nm}, flowing from node n to node m, according to the relations

$$i_m = \sum_\mu i_{\mu m} \qquad (m, \mu = c, g, p), \tag{560}$$

then Eq. (559) can be rewritten thus:

$$e_{mn} = \sum_\mu \sum_\nu Z_{\mu\nu}^{mn} i_{\nu\mu}. \tag{561}$$

Finally, if

$$Z_{rs}^{mn} = Z_r^{mn} - Z_s^{mn}, \tag{562}$$

and if it is noted that the dummy subscripts μ, ν in Eqs. (561) are interchangeable and that

$$i_{mn} = -i_{nm}, \tag{563}$$

it follows that Eq. (561) can be written in the form

$$e_{mn} = -\tfrac{1}{2} \sum_\mu \sum_\nu Z_{\mu\nu}^{mn} i_{\mu\nu} \qquad (m, n, \mu, \nu = c, g, p). \tag{564}$$

There has thus been introduced a set of coefficients Z_{rs}^{mn}, each of which possesses a definite physical meaning; namely, if, between a certain nodal pair (r,s) a constant-current generator i_{rs} is impressed, then there will result across any other (or the same) nodal pair (m,n) a voltage difference

$$e_{mn} = -Z_{rs}^{mn} i_{rs}. \tag{565}$$

The coefficient Z_{rs}^{mn} is, therefore, the negative of the *transfer impedance* between the input nodal pair (r,s) and the output nodal pair (m,n).

If Eqs. (545) and (556) are considered together,

$$i_m = \sum_\mu Y_\mu^m e_\mu \tag{566a}$$

$$e_m = \sum_\mu Z_\mu^m i_\mu \tag{566b}$$

and are substituted one in the other, it is found that

$$i_m = \sum_\mu \sum_\nu Y_\mu^m Z_\nu^\mu i_\nu = \sum_\nu i_\nu \sum_\mu Y_\mu^m Z^\mu, \tag{567a}$$

$$e_m = \sum_\mu \sum_\nu Z_\mu^m Y_\nu^\mu e_\nu = \sum_\nu c_\nu \sum_\mu Z_\mu^m Y^\mu. \tag{567b}$$

Now let

$$\sum_\mu Y_\mu^m Z_\nu^\mu = \delta_{m\nu} + A_{m\nu}, \tag{568a}$$

$$\sum_\mu Z_\mu^m Y_\nu^\mu = \delta_{m\nu} + B_{m\nu}, \tag{568b}$$

where $\delta_{m\nu}$ is the Kronecker symbol, for which

$$\delta_{m\nu} = 0 \qquad (m \neq \nu) \tag{569a}$$
$$= 1 \qquad (m = \nu) \tag{569b}$$

and where $A_{m\nu}$, $B_{m\nu}$ are to be determined. On substituting Eq. (568) in Eq. (567), it is found that $A_{m\nu}$, $B_{m\nu}$ must be such that

$$\sum_\nu A_{m\nu} i_\nu = 0, \tag{570a}$$

$$\sum_\nu B_{m\nu} e_\nu = 0. \tag{570b}$$

The only way for $A_{m\nu}$ to be other than zero is for a linear dependency to exist among the nodal currents i_m and similarly for the $B_{m\nu}$ and the nodal voltages e_m. Such a linear relation does exist for the i_m and is of the form

$$\sum_\nu i_\nu = 0, \tag{571}$$

whence it can be concluded that $A_{m\nu}$ may be any constant A_m. But

ordinarily no other such linear relation exists among the i_m, and no such relation at all exists among the e_m. Hence $A_{m\nu} = A_m$, $B_{m\nu} = 0$, and therefore

$$\sum_\mu Y_\mu^m Z_\nu^\mu = \delta_{m\nu} + A_m, \tag{572a}$$

$$\sum_\mu Z_\mu^m Y_\nu^\mu = \delta_{m\nu}. \tag{572b}$$

Further, according to Eq. (558),

$$\sum_\mu Z_\mu^{mn} Y_r^\mu = \delta_{mr} - \delta_{nr}, \tag{573}$$

and, in view of Eq. (547),

$$\sum_\mu Y_\mu^m Z_r^{\mu n} = \delta_{mr} + A_m. \tag{574}$$

Hence, by Eqs. (562) and (547),

$$\sum_\mu Z_{\mu s}^{mn} Y_r^\mu = \delta_{mr} - \delta_{nr}, \tag{575a}$$

$$\sum_\mu Y_\mu^m Z_{rs}^{\mu n} = \delta_{mr} - \delta_{ms}, \tag{575b}$$

or, by a slight change of subscript notation,

$$\sum_\mu Y_\mu^m Z_{rs}^{\mu n} = \sum_\mu Z_{\mu n}^{rs} Y_m^\mu = \delta_{mr} - \delta_{ms} \ (m, n, r, s, \mu = c, g, p). \tag{576}$$

Either set of Eqs. (576) allows the transfer impedances $-Z_{rs}^{mn}$ to be evaluated in terms of the branch admittances $-Y_n^m$ for an isolated triode.

Transfer Impedances of an Isolated Triode: Evaluation.—If in Eq. (576) m takes on successively the values r, s, t, where $r, s, t = c, g, p$ in any *cyclic* order, then Eq. (576) may be written thus:

$$\sum_\mu Y_\mu^r Z_{rs}^{\mu n} = \sum_\mu Z_{\mu n}^{rs} Y_r^\mu = 1, \tag{577a}$$

$$\sum_\mu Y_\mu^s Z_{rs}^{\mu n} = \sum_\mu Z_{\mu n}^{rs} Y_s^\mu = -1, \tag{577b}$$

$$\sum_\mu Y_\mu^t Z_{rs}^{\mu n} = \sum_\mu Z_{\mu s}^{rn} Y_t^\mu = 0. \tag{577c}$$

Of these three equations, only two are independent, since [*cf.* Eq. (547)]

$$Y_\mu^r + Y_\mu^s + Y_\mu^t = Y_r^\mu + Y_s^\mu + Y_t^\mu = 0. \tag{578}$$

Setting $n = s$ in Eq. (577), the result, written out in full, becomes

$$Y_r^r Z_{rs}^{rs} + Y_t^r Z_{rs}^{ts} = Z_{rs}^{rr} Y_r^r + Z_{ts}^{rs} Y_r^t = 1, \tag{579a}$$
$$Y_r^s Z_{rs}^{rs} + Y_t^s Z_{rs}^{ts} = Z_{rs}^{rs} Y_s^r + Z_{ts}^{rs} Y_s^t = -1, \tag{579b}$$
$$Y_r^t Z_{rs}^{rs} + Y_t^t Z_{rs}^{ts} = Z_{rs}^{rs} Y_t^r + Z_{ts}^{rs} Y_t^t = 0, \tag{579c}$$

where Eq. (579c) is also seen to follow from Eqs. (579a) and (579b) by simple addition.

Now it will be observed from Eq. (579c) that

$$Y_t^t Z_{rs}^{rs} = Y_t^r Z_{rs}^{st}, \tag{580a}$$
$$Y_t^r Z_{rs}^{rs} = Y_t^t Z_{st}^{rs}, \tag{580b}$$

or, in other words, that

$$\frac{Z_{rs}^{rs}}{Y_t^t} = \frac{Z_{rs}^{st}}{Y_r^t} = \frac{Z_{st}^{rs}}{Y_t^t} = \frac{1}{\Delta} = \text{invariant.} \tag{581}$$

These invariant relations can be put in the more concise form

$$\frac{Z_{rs}^{ab}}{Y_c^t} = \frac{1}{\Delta} = \text{invariant,} \tag{582}$$

where $a, b, c = r, s, t$ in any cyclic order. Written out in full, Eq. (582) states that

$$\frac{Z_{cg}^{cg}}{Y_p^p} = \frac{Z_{cg}^{gp}}{Y_c^p} = \frac{Z_{cg}^{pc}}{Y_g^p} = \frac{1}{\Delta}, \tag{583a}$$
$$\frac{Z_{gp}^{cg}}{Y_p^c} = \frac{Z_{gp}^{gp}}{Y_c^c} = \frac{Z_{gp}^{pc}}{Y_g^c} = \frac{1}{\Delta}, \tag{583b}$$
$$\frac{Z_{pc}^{cg}}{Y_p^g} = \frac{Z_{pc}^{gp}}{Y_c^g} = \frac{Z_{pc}^{pc}}{Y_g^g} = \frac{1}{\Delta}, \tag{583c}$$

and hence it shows that every transfer impedance is proportional to a certain *corresponding* branch admittance, with a proportionality factor $1/\Delta$ that is invariant for all transfer impedances. The invariant quantity Δ may be found by substituting Eq. (582) in Eq. (579). Thus Eq. (579) becomes

$$Y_r^r Y_t^t - Y_t^r Y_r^t = Y_r^r Y_t^t - Y_t^r Y_r^t = \Delta, \tag{584a}$$
$$Y_r^s Y_t^t - Y_t^s Y_r^t = Y_r^s Y_t^t - Y_t^s Y_r^t = -\Delta, \tag{584b}$$
$$Y_r^t Y_t^t - Y_t^t Y_r^t = Y_r^t Y_t^t - Y_t^t Y_r^t = 0, \tag{584c}$$

or, in other words,

$$\Delta = \Delta_s^s = \Delta_s^r = \Delta_r^s = \Delta_c^c = \Delta_c^g = \Delta_c^p \tag{585a}$$
$$= \Delta_g^c = \Delta_g^g = \Delta_g^p \tag{585b}$$
$$= \Delta_p^c = \Delta_p^g = \Delta_p^p, \tag{585c}$$

where Δ_n^m is the *cofactor* of Y_n^m in the determinant $|Y_n^m|$ of the array (548).

From Eq. (585) and the branch admittance array (551), it is found that

$$\Delta = j\omega C_{gp}(g_m - j\omega C_{gp}) + j\omega(C_{cg} + C_{gp})\left[\frac{1}{r_p} + j\omega(C_{gp} + C_{pc})\right] \quad (586)$$

Evidently Δ will vanish only when the frequency ω is zero. Accordingly, every transfer impedance of an isolated triode will be finite at frequencies $\omega \neq 0$, becoming neither infinite nor zero, as may be seen from Eq. (586) and the array (551). At zero frequency the branch admittances $-Y_c^g$, $-Y_g^g$, $-Y_p^g$ vanish, and accordingly, from Eqs. (583) and (586), the *corresponding* transfer impedances $-Z_{pc}^{gp}$, $-Z_{pc}^{pc}$, $-Z_{pc}^{cg}$ remain finite; whereas, since all the remaining branch admittances remain finite, the remaining transfer impedances all become infinite.

Therefore, in an isolated triode, all of the responses to shot noise, indicated by Z_{pc}^{gp}, Z_{pc}^{pc}, Z_{pc}^{cg}, are finite at all frequencies; so that here the condition of zero shot-noise output is never attainable (except, of course, at infinite frequency). On the other hand, the response to any signal placed across any nodal pair other than the plate-cathode pair becomes infinite at zero frequency, and hence the ratio of shot noise to signal output, at any output nodal pair, becomes zero at zero frequency.

Transfer Impedances of a General Three-node Triode Network.—The transfer impedances $-Z_{rs}^{mn}$ discussed in the last several paragraphs relate to an *isolated* triode. However, all of the general relations there set forth, for both branch admittances and transfer impedances, apply equally well to a general three-node triode network provided the (negative) branch admittances Y_n^m of an isolated triode are replaced by the branch admittances (negative)

$$Y_n^{m'} = Y_n^m + \Delta Y_n^m \quad (587)$$

of a general three-node triode network and provided the (negative) transfer impedances Z_{rs}^{mn} of an isolated triode are replaced by the transfer impedances (negative)

$$Z_{rs}^{mn'} = Z_{rs}^{mn} + \Delta Z_{rs}^{mn} \quad (588)$$

of a general three-node triode network. In this way, the new transfer impedances $-Z_{rs}^{mn}$ are defined by the relationships of Eqs. (583) and (584), now written with the Y's and Z's primed rather than unprimed. Thus, in particular,

$$\frac{Z_{rs}^{ab'}}{Y_c^{t'}} = \frac{1}{\Delta'} \quad (589)$$

and

$$Y_a^{r'}Y_b^{s'} - Y_b^{r'}Y_a^{s'} = \Delta', \quad (590)$$

where a, b, $c = r$, s, t in any cyclic order, thereby *defining* the $Z_{rs}^{mn'}$ in terms of the $Y_n^{m'}$.

This procedure of defining the $Z_{rs}^{mn'}$ *in terms of* the $Y_n^{m'}$ suggests that the branch admittances $-Y_n^{m'}$ are capable of independent variation [subject to the constraints of Eqs. (578) and (553)] whereas the $Z_{rs}^{mn'}$ are not. This is indeed the case, as will be found demonstrated in detail elsewhere.[1]

Conditions for Complete Suppression of Tube Noise in a Three-node Triode Network.—It is always possible, at a single frequency ω_0, to reduce to zero the tube noise voltage developed between any pair of electrodes in a three-node triode network. If the two electrodes in question are designated by r, s (where r, s, $t = c$, g, p in any cyclic order), then this condition of zero tube noise output is expressible in the form

$$Z_{pc}^{rs'} = 0. \tag{591}$$

Now, according to Eq. (589),

$$Z_{pc}^{rs'} = \frac{Y_t^{g'}}{\Delta'}. \tag{592}$$

Hence, as long as Δ' is not zero, the condition of Eq. (591) is equivalent to the condition

$$Y_t^{g'} = 0. \tag{593}$$

That is to say, if the tube noise voltage between nodes r and s is to be made zero, then the branch admittance $Y_t^{g'}$ between the grid g and the remaining node t must be zero.

There are three possible choices of output terminals, namely, (c,g), (g,p), (p,c). For each of these terminal pairs, then, the corresponding condition for zero tube noise output is

$$Y_p^{g'} = 0 \qquad (c,g), \tag{594a}$$
$$Y_c^{g'} = 0 \qquad (g,p), \tag{594b}$$
$$Y_g^{g'} = 0 \qquad (p,c). \tag{594c}$$

According to Eqs. (587) and the admittance array (551), these three conditions can be written, at the frequency $\omega = \omega_0$,

$$-\Delta Y_p^g = Y_p^g = -j\omega_0 C_{gp} \qquad (c,g), \tag{595a}$$
$$-\Delta Y_c^g = Y_c^g = -j\omega_0 C_{gc} \qquad (g,p), \tag{595b}$$
$$-\Delta Y_g^g = Y_g^g = +j\omega_0(C_{gp} + C_{gc}) \qquad (p,c). \tag{595c}$$

From Eqs. (595), it is clearly necessary that for each of these three cases, the corresponding external branch admittance (namely, $-\Delta Y_p^g$, $-\Delta Y_c^g$, or $+\Delta Y_g^g$) shall be purely reactive at the frequency ω_0. If in any one of these cases this external branch admittance happens to be arbitrarily assignable regardless of the external circuit arrangement, then it may

[1] E. J. Schremp, to be published.

be taken most conveniently to be an inductance that resonates, at the frequency ω_0, the appropriate one of the three capacitances C_{gp}, C_{gc}, $(C_{vp} + C_{pc})$.

On the other hand, if this external branch admittance happens to include a part that is required by the external circuit arrangement, then it must also include the exact negative of that part. For example, if the part required by the external circuit arrangement is a positive resistance representing the signal source, then the total external branch admittance in question must contain a compensating part that, at the frequency ω_0, assumes the form of an equal negative resistance. Such a compensation can generally be achieved by the introduction of a fourth node, linked to the original three nodes (c,g,p) by means of suitable passive branch admittance elements. However, the use of this device to eliminate tube noise must, in general, introduce a slight amount of excess thermal noise.

In connection with the three conditions of Eqs. (595) it is important to notice that not more than one of these three conditions should be simultaneously satisfied at the frequency ω_0. For if any two of these conditions are simultaneously satisfied, then by Eq. (578) the third will also be satisfied; and therefore, since

$$\Delta' = Y_c^{c'}Y_g^{g'} - Y_g^{c'}Y_c^{g'}, \tag{596}$$

it will follow that $\Delta' = 0$ at $\omega = \omega_0$. In this case, the tube noise voltage output will not be suppressed at any of the three terminal pairs (c,g), (g,p), (p,c). However, this case is again a favorable one from the point of view of signal-to-noise ratio, for it again corresponds to an infinite ratio of signal to noise, inasmuch as the voltage responses to signals applied elsewhere than between plate and cathode are all infinite.

MINIMAL NOISE CIRCUITS

By Richard Q. Twiss and Yardley Beers

13·1. Introduction.—The purpose of this chapter is to describe practical methods by which amplifiers can be designed and adjusted to detect the weakest possible signals. The subject of the detectability of weak signals involves many aspects, including some factors external to the amplifier, such as the kind of signal, the manner in which the information is presented to the observer, the time-duration of observation, and subjective factors concerning the observer himself. These matters can only be mentioned in passing here; complete details can be found in another book of this series.[1]

In the amplifier itself, the principal factor is the relative amount of noise produced in the amplifier and in the signal source as compared with a perfect amplifier of the same pass band, which contains no noise except in the signal source. In other words, for a given bandwidth the principal factor is the noise figure. A less important factor is the shape of the bandpass curve.

In this chapter attention is focused primarily upon methods of getting good noise figures. A few qualitative remarks concerning the effect of the pass band upon the detectability of weak signals are in order, however. In general, the spectrum of the input signal is distributed over a limited band of frequencies, while sources of noise produce noise over very much wider bands of frequencies. If the pass band of the amplifier is narrow compared with the signal spectrum, the detectability of signals is impaired because the output signal reaches only a fraction of its peak value. As the pass band is widened, the signals are reproduced more faithfully, and the reproduced signal reaches its peak value. At the same time, if the gain is held constant, both the total signal and noise output powers increase. An optimum width is ultimately reached. For still wider pass bands the output noise power continues to increase steadily, but the output signal power increases hardly at all; hence the detectability of weak signals deteriorates. The bandwidth can vary considerably in either direction from the optimum value, however, before the detectability of weak signals is greatly impaired.[2] From the point of

[1] See *Threshold Signals*, Vol. 24, Radiation Laboratory Series.

[2] With rectangular pulses, the optimum 3-db bandwidth is 1.2 times the reciprocal

view of detectability of weak signals, therefore, selection of the optimum bandwidth is not critical and may be determined by other factors, such as frequency instability of the signal source.

The first part of this chapter is concerned with the basic properties of the tube itself, such as noise figure, available power gain, and stability. In Secs. 13·9 and 13·10 this information is used as the basis of a discussion of the relative advantages of various possible tube configurations. It is shown that to obtain the best possible noise figure it is desirable to use a double-triode input combination consisting of a grounded-cathode triode followed by a grounded-grid triode.

Not very much attention is paid to the design of interstage coupling networks, because when the double triode input configuration, mentioned above, is used, these circuits do not have an important effect upon the noise figure. The input network that couples the source to the first tube is, however, of primary importance, because its design and adjustment can make a considerable difference in the noise figure. For narrowband receivers a straightforward single-tuned circuit can be used as input network. To obtain good noise figures with wider bandwidths, however, double-tuned circuits or feedback loading must be used.

From the previous paragraphs, it can be inferred that a good noise figure often can be obtained only at the cost of more critical adjustment, extra components, and compromises on other characteristics of the amplifier. The improved performance is usually worth the difficulty encountered in obtaining the best possible noise figure, but under some circumstances it is not.

For example, suppose that the amplifier is a radio receiver working under conditions in which atmospheric noise is predominant. The sources of noise within the amplifier would have to be very strong to influence the total noise level appreciably, and there would be little advantage in having an exceedingly good noise figure. In more quantitative terms, if the antenna were replaced by a resistor of value equal to the radiation resistance of the antenna, all the sources of noise could be represented (according to the definition of noise figure in Sec. 12·7) by a generator having an available power of $FkTB$. Part of this, kTB, is due to the resistor, and the other part, $(F - 1)kTB$, is due to the amplifier. (The noise figure F as used here is a dimensionless ratio.) Because the antenna is in a strong electric field produced by atmospheric disturbances, its available noise power is larger by a factor t (called the equivalent antenna temperature) than that of the resistor at room temperature.

pulse length, and the detectability of weak signals does not deteriorate more than 1 db until the bandwidth is either approximately three times or one-third the optimum. Further details are contained in *Threshold Signals*, Vol. 24, Radiation Laboratory Series.

Consequently, when the antenna is connected, all of the sources of noise combined have an available power

$$P_{an} = (t + F - 1)kTB. \tag{1}$$

The minimum detectable signal is proportional to P_{an}.

The meaning of Eq. (1) can be clarified by considering a numerical example. What is the improvement in minimum detectable signal if F is improved from 2.00 (3 db) to 1.59 (2 db) given (1) $t = 1$ and (2) $t = 10$? In Case (1), the quantity in the parentheses in Eq. (1) is equal to 2.00 when $F = 2.00$ and 1.59 when $F = 1.59$. The ratio of the two values of P_{an} is 0.79, or 1 db, indicating that a 1-db improvement in noise figure causes a 1-db improvement in minimum detectable signal. In Case (2), the values of the quantity in parentheses are respectively 11.00 and 10.59, and the ratio of the values of P_{an} is 0.96, or 0.17 db. Hence, in this case, a 1-db improvement in noise figure causes only a 0.17-db improvement in minimum detectable signal. This improvement in noise figure would hardly be worth while if an extra tube would have to be added in the amplifier to attain it.

Another situation to be considered is one in which the amplifier derives its signal from another network (which might contain other amplifiers or frequency converters). The noise figure of the combination is given by Eq. (12·516g) which is repeated here for convenience:

$$F = F_1 + \frac{(F_2 - 1)}{\mathcal{W}_1} \tag{2}$$

where F = noise figure of the combination,

    F_1 = noise figure of first network,

    F_2 = noise figure of the second network (the amplifier under consideration),

    $\mathcal{W}_1$ = available power gain of the first network.

(All quantities are expressed as pure ratios.) The second term on the right, which expresses the contribution of the amplifier under consideration, contains a factor of $\mathcal{W}_1$ in the denominator. Hence if $\mathcal{W}_1$, the gain of the first network, is large, any sources of noise within the second amplifier have little effect on the noise figure of the combination. It is, therefore, useless to expend much effort in making the noise figure of the second amplifier very good. If the gain of the first network is small, however, the reverse is true. Some other consequences of Eq. (2) are discussed below.

If the first network consists of a crystal converter and the second is the i-f amplifier of a receiver, $\mathcal{W}_1$ is low, and the noise figure of the amplifier following it is moderately important. According to convention, when discussing crystal converters, the noise figure of the crystal is

written as the product of the loss ratio $1/\mathcal{W}$ of the crystal and another parameter called the equivalent crystal temperature t.   The symbol t is defined as the ratio of the available noise power at the i-f terminals of the crystal to that of an ohmic resistor at room temperature.   The expression for the over-all noise figure Eq. $(12 \cdot 516h)$ is repeated here.

$$F = \frac{1}{\mathcal{W}} (F_2 + t - 1), \tag{3}$$

wherein all quantities are expressed as pure ratios.

It is illuminating to consider a numerical example.   What is the effect on the over-all noise figure when F_2 is improved from 2.00 (3 db) to 1.59 (2 db) if the crystal temperature has a typical value of 2?   In the first case, the quantity in the parentheses in Eq. (3) is 3.00, and in the second case it has a value of 2.59.   The improvement in the over-all noise figure is the ratio $3/2.59 = 1.16 = 0.64$ db (the gain having canceled out).   This is comparable to the 1-db improvement in F_2.   In this example, it has been supposed that the local oscillator has contributed no noise.   This assumption is valid in many practical cases, especially those in which the intermediate frequency is large compared with the signal frequency or in which a balanced mixer is employed.

Now consider what would have happened to this result if local oscillator noise had been present.   The local oscillator generates noise that is concentrated mainly in a frequency band surrounding the local oscillator frequency.   If the intermediate frequency is a small fraction of the radio frequency, an appreciable amount of this noise is at the signal frequency and is converted to the intermediate frequency.   This i-f noise produces an increase in the equivalent crystal temperature. Suppose that the local oscillator noise has caused the equivalent crystal temperature to increase from 2 to 6.   Then as F_2 is improved 1 db from 2.00 to 1.59, the over-all noise figure is improved only by the ratio $7/6.59 = 1.06 = 0.26$ db.   This relatively small improvement would not be worth while if it involved the addition of many parts.

13·2. Basic Noise-figure Considerations.—A multistage amplifier can be represented as a chain of boxes connected between the signal source and the indicating device.   Each box contains an electrical network, a pair of input terminals, and a pair of output terminals.   Boxes of two general types occupy alternate positions in the chain.   One type contains vacuum tubes, sources of d-c power, and decoupling filters.   The amplifier designer can choose the tube-type, whether to use a pentode or a triode; and the tube configuration (grounded cathode, grounded grid, or grounded plate).

The second type of box contains coupling networks consisting of reactances and resistances. . The designer is free to choose the form of

these networks to give the optimum noise figure, subject to the condition that the over-all transfer characteristics of the amplifier meet the required bandwidth specifications. The most important of these networks from the point of view of noise is the one between the signal source and the first tube. Basically, the network is an impedance transforming device. It is connected between the signal source and the input terminals of the first box, which contains a vacuum tube serving as a load for the signal source. This first network usually contains only those resistive elements which are unavoidable in practical reactances, because, as will become evident later, the noise figure deteriorates when dissipative elements are added to the input circuit. In certain special cases, however, it may be necessary to use dissipative elements to broaden the bandwidth, even at the cost of some deterioration in noise figure.

The signal source is not only the origin of the signal but also the origin of a certain amount of noise, partly due to the thermal-agitation effect within its resistive elements. Its internal impedance is the same when it is acting as a source of signal as when it is acting as a source of noise. The ratio of signal power P_s to noise power P_n, which the signal source dissipates in the load, is therefore independent of the value of the load impedance or the adjustment of the coupling network. This is true if the effect of the small unavoidable resistive elements within the network is neglected and it is assumed that the pass band of the network remains wide compared with that of the circuits which follow the first stage. The maximum values of P_s and P_n both occur when the load impedance is matched to the signal source and are independent of the value of the load. Thus, the addition of the coupling network to the signal source does not alter the available signal power P_s or noise power P_n of the signal source but only the value of the impedance to which these powers are supplied. The load impedance, which is, of course, the input impedance of the first tube, contains sources of noise within itself. The amount of noise that the load dissipates within itself usually depends upon the impedance mismatch between it and the signal source. Thus the ratio of signal power to total noise power P_n' dissipated within the load[1] varies with the adjustment of the first network.

Usually the adjustment that causes this ratio to be a maximum is not the same adjustment that gives maximum gain. · The latter condition is obtained with the signal source impedance matched to the load. Signal-to-noise ratio can usually be improved by proper mismatch of the signal

[1] The ratio of the available signal power to the available noise power at the output of the amplifier is equal to P_s/P_n' if it is assumed that the noise generated by the load impedance represents all the tube noise. If the amplifier were ideal and had no sources of noise within it, this ratio would be equal to P_s/P_n. The latter ratio divided by the former gives the noise figure according to the primary definition in Sec. (12·7).

source to the input of the first tube, at the expense of gain which can be restored by later stages of the amplifier.

In the preceding discussion, it has been implied that all the sources of noise within the amplifier can be represented as a single source of noise located in its input impedance. It will be seen later that if so desired, this representation can be achieved by certain simple artifices. The qualitative conclusion drawn from the previous argument is then perfectly valid. The previous discussion has also been simplified by the assumption that the bandwidth of the coupling network is large compared with the bandwidth of the later circuits in the amplifier. This assumption implies that the over-all bandwidth, as well as the values of the impedances of the signal source and load averaged over the pass band, are independent of the adjustment of the input network. (If this assumption is not valid, these considerations are complicated in an additional way: The available power of the signal source and of all noise sources depends upon the over-all bandwidth of the circuit between these sources and the output indicating device.) Thus, the bandwidth associated with the sources of noise in the output of the first tube depends only upon the coupling networks following the first tube. The available power of this tube, therefore, is independent of the input network, whereas the available signal and noise powers of the signal source vary if the input network is made narrow.

In order to calculate the noise figure of an amplifier in which the bandwidth of the coupling network is not large compared with the bandwidth of the later networks, it can be assumed that the pass band is broken up into narrow elements. The calculation for each element can then be performed as described above, and the results averaged with respect to the gain of the amplifier. The previous discussion can apply as an approximation, therefore, even to cases in which the input network has a narrow bandwidth.

Noise Representation.—A source of noise can be represented in a variety of ways. It can be represented as a conventional electrical generator, which can then be described in terms of its internal impedance or admittance and one of the following: electromotive force, current strength, or available power. Various other representations arise from applications of the formulas for shot and thermal noise; and conversely, sources of shot and thermal noise can be expressed in terms of conventional generators. For example, the thermal noise in a resistor R can be expressed as a voltage E in series with a noiseless resistor R, where the rms value of E is given by

$$E^2 = 4kTRB, \tag{4}$$

where k = Boltzmann's constant = 1.380×10^{-23} joule per degree Kelvin,

B = bandwidth in cycles per second,

T = room temperature in degrees Kelvin.

The available power W, the maximum power that this noise source can supply to an external load, is obtained when the external load is matched to the source:

$$W = \frac{E^2}{4R} = kTB; \tag{5}$$

$W = 4.00 \times 10^{-15}$ watt, for $B = 1$ Mc/sec, and $T = 290°$K.

According to the usual methods of transformation, this can be considered with equal validity as a current generator I in parallel with a noiseless conductance G where $G = 1/R$ and $I = GE$. When these quantities are substituted into Eq. (4) one obtains

$$I^2 = 4kTGB. \tag{6}$$

Conversely, a noise current I in parallel with a noiseless conductance G can be thought of as being the thermal noise due to the conductance at a fictitious temperature T_1, where T_1 is defined by $I^2 = 4kT_1GB$.

It is sometimes convenient also to represent an rms noise voltage E in terms of a parameter R_{eq}, called the equivalent noise resistance, in such a way that

$$E^2 = 4kTR_{eq}B. \tag{7}$$

By comparison with Eq. (4), it is clear that R_{eq} can be considered as the value of resistance whose thermal-agitation voltage is E when the temperature is room temperature T. It should be remembered that although R_{eq} has the dimensions of a resistance, it is only a parameter that gives the magnitude of an emf and does not imply the dissipation of power.

13·3. The Determination of the Noise Figure, Power Gain, and Other Characteristics of the First Stage.—A schematic diagram of a typical amplifier chain is shown in Fig. 13·1. The first stage consists of the
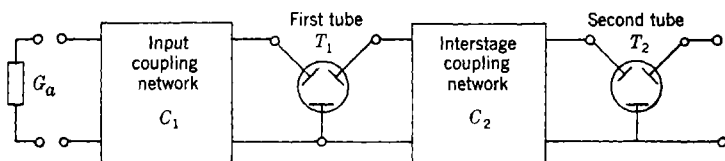
FIG. 13·1. —Conventional amplifier schematic diagram.

signal source, the first coupling network, and the first tube; the rth stage consists of the rth tube and the network coupling the rth with the $(r-1)$th tube.

If the effects of feedback from stage to stage are neglected, Eq. (2) can be expanded to express the over-all noise figure of the amplifier in the form

$$F = F_1 + \frac{F_2 - 1}{\mathfrak{W}_1} + \frac{F_3 - 1}{\mathfrak{W}_2} + \cdots + \frac{F_r - 1}{\mathfrak{W}_{r-1}} + \frac{F_R - 1}{\mathfrak{W}_r}, \quad (8)$$

where F_r = the noise figure of the rth stage when fed from a source
impedance equal to the output impedance of the $(r - 1)$th
stage,

$\mathfrak{W}_r$ = the available power gain of the first r stages,

F_R = the noise figure of the remainder of the amplifier succeeding
the rth stage.

In all practical cases, the series of Eq. (8) is rapidly convergent, so that,
at most, three and usually only two terms are significant.

It may be seen from Eq. (8) that if the available power gain and noise
figure of any individual stage is known, then the over-all noise figure of an
amplifier built of such stages can be found.

It is the purpose of this section to provide this information together
with other relevant data, such as input and output admittance, voltage
gain, and so on, that will be useful when discussing the problems of
amplifier design involved in getting an optimum noise figure.

The Input Coupling Network and Associated Noise Sources.—The first
stage is composed of two parts: the tube itself and the coupling network
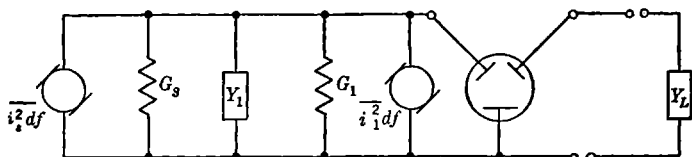
Fig. 13·2.—Equivalent circuit for input network.

together with the source. Two properties only of the input network are
of interest here: the power gain, and the admittance that the network
presents to the input terminals of the tube. It is shown in Sec. 13·11
that the circuit of Fig. 13·2 is an equivalent network, valid for any type
of passive four-terminal coupling as far as these two properties are
concerned. If

G_s = the transformed source conductance,

G_1 = the parallel combination of the network losses and the ohmic losses
across the input of the tube,

Y_1 = the susceptance presented by the network to the input terminals
of the tube,

the total input admittance Y_s presented to the tube is

$$Y_s = G_s + G_1 + jY_1. \quad (9)$$

In Fig. 13·2, Y_L is the load,[1] presumed noiseless, into which the tube is to
work.

[1] The noise sources associated with Y_L are looked upon as being part of the second
stage in the present division.

Associated with the input network are two statistically independent constant-current thermal-noise generators i_s and i_1, the mean square values of which for the frequency interval df are given by

$$\left.\begin{aligned}\overline{i_s^2}df &= 4kTG_s\,df, \\ \overline{i_1^2}df &= 4k\alpha TG_1\,df.\end{aligned}\right\} \tag{10}$$

The effective source temperature T is defined to be $290°\mathrm{K} = 62.6°\mathrm{F}$. In an ideal receiver the thermal-noise generator associated with the source conductance G_s would be the only noise generator in the amplifier.

This arbitrary choice of the magnitude of T is made here for reasons of numerical convenience, namely, that $e/2kT = 20$ for $T = 290°\mathrm{K}$, where e is the charge of an electron $= 1.60 \times 10^{-19}$ coulomb, and is not necessarily the actual temperature of the source. Thus, if the source is the radiation resistance of an antenna with effective temperature νT, ν may range from 0 to 100 or more. The over-all noise figure F of the system, as distinct from F_A, the noise figure of the amplifier defined as above, is given in this case by

$$F = \nu - 1 + F_A.$$

Although it might appear that difficulty would arise if the local temperature of the amplifier at the source were different from 290°, this difficulty is only apparent. The noise figure of the amplifier is *defined* to be the noise figure obtained when all the thermal-noise sources are at 290°K. Suitable corrections have to be made when it is required to find the noise output of the amplifier at other local temperatures.

In view of the above discussion it might be thought inevitable that α be taken equal to unity [see Eq. (10)]. In fact, this will usually be done in this chapter. Sometimes, however, it is convenient or even necessary to assume that α is not equal to unity. For example, some of the network loading may be produced by electronic means, so that i_1 is not a pure thermal-noise current. Alternatively, it might be desired to represent some or all of the other noise sources in the amplifier by a suitable modification of the effective temperature of G_1. For these reasons α will not be taken equal to unity at this stage.

Further discussion of the input coupling network and its associated noise sources is deferred until Sec. 13·14. The choice of the first tube and the location of the associated noise source will now be considered.

Equivalent Noise Representation of the First Tube.—In general, the main body of the amplifier consists of a sequence of pentodes, but some choice can often be exercised in deciding upon the circuit connections for the first or second tube, especially when sensitivity considerations are important. Thus, although it is quite usual to employ a grounded-cathode pentode as the first tube in an r-f or i-f amplifier, improved noise figures can often be obtained by the use of one or even two triode input

stages, the circuit connections of which can be chosen in a number of different ways. For example, grounded-cathode, grounded-grid, and grounded-plate triode input stages have

FIG. 13·3.—Equivalent noise representation for a pentode.

been tried at one time or another. Each of these choices possesses certain advantages.
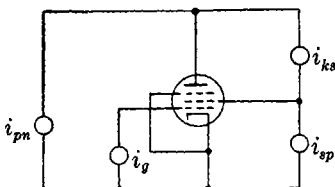
It was shown in Chap. 12 that the circuit of Fig. 13·3 provides an equivalent noise representation for a pentode, valid under the assumptions there stated, where the tube itself is now assumed noiseless and the sources of noise are four constant-current intranodal generators located as shown in the figure. The currents i_{pk}, i_{ks}, i_{sp} are statistically independent, i.e.,

$$\overline{i_{pk}i_{ks}} = \overline{i_{ks}i_{sp}} = \overline{i_{sp}i_{pk}} = 0, \tag{11}$$

but i_g is given by the equation

$$i_g = j\omega \frac{\tau}{3} (i_{pk} + i_{ks}), \tag{12}$$

where ω is the angular frequency and τ the grid-cathode transit time.

The cathode nodal noise generator i_k is defined by

$$i_k = i_{pk} + i_{ks}, \tag{13}$$

so that

$$i_g = j\omega \frac{\tau}{3} i_k$$

FIG. 13·4.—Equivalent noise representation for pentode with screen connected to cathode.

is 90° out of phase with the cathode nodal noise generator. The mean square values of i_{pk}, i_{ks}, i_{sp} in the frequency interval df are given by Eqs. (12·465).

The most important case is one in which the screen of the pentode is short-circuited to the cathode as far as all noise generators are concerned, so that the equivalent noise circuit assumes the simpler form of Fig. 13·4, where

$$i_p = i_{pk} + i_{sp}$$

and

$$i_g = j\omega \frac{\tau}{3} (i_{pk} + i_{ks}).$$

From Eqs. (12·465) and (12·382), it may be seen that

$$\left. \begin{aligned} \overline{i_p^2}\, df &= \left[2eI_p\Gamma^2 + \frac{2eI_pI_s}{I_k} (1 - \Gamma^2) \right] df, \\ \overline{i_g^2}\, df &= \frac{(\omega\tau)^2}{9} 2eI_k\Gamma^2\, df. \end{aligned} \right\} \tag{14}$$

More familiar expressions for $\overline{i_g^2}$ and $\overline{i_p^2}$ will now be obtained.

If G_r is the grid-cathode conductance due to transit-time loading, then $\overline{i_g^2}/G_r$ is independent of $\omega\tau$ and it is usual to regard i_g as being the thermal-noise current due to a conductance G_r at a temperature β times the room temperature, so that

$$\overline{i_g^2}\,df = 4k\beta TG_r\,df. \tag{15}$$

For tubes with oxide-coated cathodes and limiting space-charge reduction factors, β is approximately equal to 5.[1]

It is often convenient to regard i_p as the thermal-noise current due to a conductance $R_{eq}g_m^2$ at room temperature where g_m is the transconductance of the tube and R_{eq} the equivalent noise resistor, so that

$$\overline{i_p^2}\,df = 4kTR_{eq}g_m^2\,df, \tag{16}$$

where

$$R_{eq} = \dfrac{2eI_p\Gamma^2 + \dfrac{2eI_pI_s(1-\Gamma^2)}{I_k}}{4kTg_m^2}. \tag{17}$$

From some points of view it is desirable to preserve the distinction between shot and partition noise currents and write R_{eq} in the form

$$R_{eq} = R_{nc} + R_{ns}, \tag{17a}$$
$$R_{nc} = \frac{2eI_p\Gamma^2}{4kTg_m^2},$$

where

$$R_{ns} = \frac{2eI_pI_s(1-\Gamma^2)/I_p}{4kTg_m^2}. \tag{17b}$$

R_{nc} and R_{ns} are called the equivalent shot-noise resistor and the equivalent partition noise resistor respectively.

It has been shown[2] that for tubes with good space-charge reduction factor $\Gamma^2 \sim 0.05$ and a cathode temperature of $1000°C$,

$$R_{nc} \approx \frac{2.5}{g_m} \tag{17c}$$
$$R_{ns} \approx \frac{20I_s}{I_kg_m}.$$

In the above case, the equivalent noise circuits for the triode and the pentode are identical, the only difference being that R_{eq} is appreciably larger for the pentode. Accordingly, there is no need to treat the two cases separately.[3]

In the case of the usual low-pass amplifier, the induced grid noise

[1] C. J. Bakker, "Fluctuation and Electron Inertia," *Physica*, **8**, No. 1, January 1941; D. O. North and W. R. Ferris, "Fluctuations Induced in Vacuum Tube Grids at High Frequencies," *Proc. I.R.E.*, **29**, 49–50 (1942).

[2] W. A. Harris, *RCA Rev.*, **5**, 505–524, April 1941; **6**, 115–124 (July 1941).

[3] This statement will have to be modified when the effects at cathode-lead inductance on noise figure are considered in Sec. 13·12 and when the correlation of induced grid noise and shot noise is considered in Sec. 13·13.

generator i_g, of mean square magnitude proportional to tne square of the frequency, can generally be neglected. In the high-frequency bandpass case the mathematical analysis is complicated by the fact that this noise generator is correlated with the cathode shot-noise generator. When, however, the impedances presented to the electrodes of the tube are all purely resistive, this correlation may be ignored because the mean square effects produced by the generators at a subsequent stage in the amplifier add linearly.

If all the reactances associated with the first stage are midband resonant, then i_g and I_k are 90° out of phase at midband and add destructively at frequencies above resonance and constructively at frequencies below resonance. Hence, if the pass band of the input circuits is wide compared with the over-all pass band of the amplifier, the error in the noise figure will be small if it is assumed that i_g and i_k are uncorrelated over the *whole* amplifier pass band.[1] The general case will be taken up again in Sec. 13.13, and the possibility of utilizing the coherence between induced grid noise and cathode shot noise to obtain improved noise figures at very high frequencies will be considered.

In deriving the noise figure for the first stage, two additional simplifying assumptions will be made. These are that

1. The effects of lead inductance in the grounded electrode lead can be neglected.
2. Any reactance coupling the input terminals of the tube to the output terminals is neutralized, at least over the receiver pass band.

The validity and importance of these assumptions are discussed in Sec. 13.9.

Expressions for the noise figure, available power gain, and so on, of the first stage will be derived for three alternative connections: (1) grounded cathode, (2) grounded grid, and (3) grounded plate, as shown in Fig. 13.5, where the noise sources have been omitted in order not to complicate the diagrams excessively. The correct locations of the noise sources are given in Figs. 13.2 and 13.4, and their mean square magnitudes by Eqs. (9), (15), and (16).

It is necessary to sound a note of warning. Although the results in this section are based upon the equivalent circuit of Fig. 13.3, it must be emphasized that the theoretical and experimental basis for this circuit is not yet firmly established. At low frequencies, where the transit angle is so small that the contribution of induced grid noise is negligible, the equivalent circuit can be accepted fairly confidently, but it can be relied

[1] It is shown in Sec. 13.13 that if the variation of G_7 over the band is neglected and if the amplifier transfer characteristics are geometrically symmetrical about midband, then this error is zero.

upon far less at higher frequencies.   All that can fairly be claimed for the circuit of Fig. 13·3 is that it is the best available model.[1]   It gives results in good agreement with experiment for some tubes, such as the type 6AK5; and even if the values of the equivalent noise generators cannot
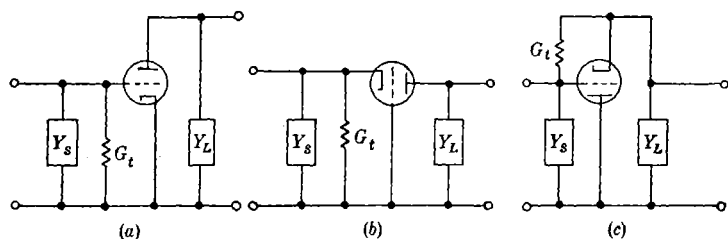
Fig. 13·5.—Alternative connections of first tube.

always be taken direct from the formulas, their general form and order of magnitude are correct.   Despite this, however, the divergence between theory and practice is considerable in the case of some tubes, like the type 6J4.   In such cases the equivalent noise generators must be found by experiment, as must the correlation between them, for the particular circuit arrangement under discussion.

General Relations for A Three-terminal Tube.—The machinery used to derive the required results will now be discussed before it is applied to the circuits of Fig. 13·5.

Expressions for the basic properties of these circuits will be found scattered throughout the technical literature and are usually proved by elementary methods.   Here they will be derived by a unified treatment based on the tube impedance theory developed in Chap. 12, the main results and the notation of which will, for convenience, be restated here.

Fig. 13·6.—The general externally loaded three-node network.

The basis of discussion is the general three-node network of Fig. 13·6, where the three electrodes k, g, and p (cathode, grid, plate) are the three nodes linked by three passive internodal admittances Y_{kg}, Y_{gp}, Y_{pk}.   Quantities Y_μ^m (m, $\mu = k$, g, p) called the branch admittances will be introduced, satisfying the relations

---

[1] The theoretical basis, which is due to C. J. Bakker, *op. cit.*, is essentially the same as that developed by F. B. Llewellyn and his associates at the Bell Telephone Laboratories.   The Llewellyn analysis is, however, more detailed and shows how the equivalent shot-noise resistance depends upon the frequency and the transit angles in the tube, a dependence that has been neglected in this chapter.

$$\sum_{\mu} Y_{\mu}^m = \sum_{m} Y_{\mu}^m = 0 \qquad (m, \mu = k, g, p). \tag{18}$$

They form a three-by-three array whose elements are shown in Table 13·1. The diagonal elements are not shown in this array but are obtained from Eq. (18). The Y_{μ}^m are not of direct interest but are defined because they are related to the transfer impedances $Z_{\mu\nu}^{\alpha\beta}$ by the simple relation

$$\frac{Z_{\beta\nu}^{\alpha\beta}}{Y_{\nu}^{\alpha}} = \frac{1}{\Delta} \qquad (\alpha, \beta, \nu = k, g, p), \tag{19}$$

where Δ is the cofactor of Y_m^m in the determinant $|Y_{\beta}^{\alpha}|$ of the array and where $Z_{\mu\nu}^{\alpha\beta}$ is the transfer impedance from terminals μ, ν to terminals α, β.

(To distinguish the various circuits of Fig. 13·5 one from another, the symbol for the grounded electrode is used throughout this section as a left-hand suffix wherever a danger of ambiguity arises. Thus $_gZ_{kg}^{gp}$ denotes the forward transfer impedance of the grounded-grid tube while $_kF$ denotes the noise figure of the grounded-cathode tube.)

In all three cases, expressions for Δ will be needed. These are given in Eqs. (20) and are calculated from Table 13·1 for the network of Fig. 13·5.

$$_k\Delta = \begin{vmatrix} _kY_g^g & _kY_g^p \\ _kY_p^g & _kY_p^p \end{vmatrix} = \begin{vmatrix} -(Y_s + G_\tau) & -g_m \\ 0 & -\left(Y_L + \dfrac{1}{r_p}\right) \end{vmatrix}$$
$$= (Y_s + G_\tau)\left(Y_L + \frac{1}{r_p}\right),$$

$$_g\Delta = \begin{vmatrix} _gY_g^g & _gY_g^p \\ _gY_p^g & _gY_p^p \end{vmatrix} = \begin{vmatrix} -(Y_s + G_\tau + Y_L) & (-g_m + Y_L) \\ Y_L & -\left(\dfrac{1}{r_p} + Y_L\right) \end{vmatrix}$$
$$= g_m'Y_L + (Y_s + G_\tau)\left(\frac{1}{r_p} + Y_L\right), \quad (20)$$

$$_p\Delta = \begin{vmatrix} _pY_g^g & _pY_g^p \\ _pY_p^g & _pY_p^p \end{vmatrix} = \begin{vmatrix} -(G_\tau + Y_s) & (-g_m + Y_s) \\ Y_s & -\left(Y_L + Y_s + \dfrac{1}{r_p}\right) \end{vmatrix}$$
$$= (g_m + G_\tau)Y_s + \left(Y_L + \frac{1}{r_p}\right)(Y_s + G_\tau),$$

where

$$g_m' = g_m + \frac{1}{r_p} = \frac{\mu + 1}{r_p} \tag{21}$$

and where μ is the amplification factor of the tube.   All the introductory material has now been assembled, and the principal properties of the three circuits can be written down.   The more familiar results are derived with a minimum of discussion, but the power gain and noise figure are treated in greater detail.

TABLE 13·1.—THE BRANCH ADMITTANCES Y_μ^m

m / μ	Cathode, k	Grid, g	Plate, p
Cathode, k	. . . .	Y_{kg}	$g_m + \dfrac{1}{r_p} + Y_{pk}$
Grid, g	$g_m + Y_{kg}$	. . . .	$-g_m + Y_{gp}$
Plate, p	$\dfrac{1}{r_p} + Y_{pk}$	Y_{gp}	. . . .

r_p is the plate resistance of the tube.

The Input Admittances.—The input admittances Y_{11} are simply the reciprocals of the input impedances $Z_{\mu\nu}^{\mu\nu}$, which can be calculated directly from Table 13·1 with the aid of Eqs. (20) and (21) to give

$$
\left.
\begin{aligned}
{}_kY_{11} &= \frac{1}{{}_kZ_{kg}^{kg}} = \frac{{}_k\Delta}{{}_kY_p^p} = \frac{\left(\dfrac{1}{r_p} + Y_L\right)(Y_s + G_r)}{\dfrac{1}{r_p} + Y_L} = Y_s + G_r. \\[2ex]
{}_gY_{11} &= \frac{1}{{}_gZ_{kg}^{kg}} = \frac{{}_g\Delta}{{}_gY_p^p} = \frac{g_m' Y_L + (Y_s + G_r)\left(\dfrac{1}{r_p} + Y_L\right)}{\dfrac{1}{r_p} + Y_L} = Y_s + G_r \\[2ex]
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \frac{g_m' Y_L}{Y_L + \dfrac{1}{r_p}} \\[2ex]
{}_pY_{11} &= \frac{1}{{}_pZ_{gp}^{gp}} = \frac{{}_p\Delta}{{}_pY_k^k} = \frac{(g_m + G_r)Y_s + \left(Y_L + \dfrac{1}{r_p}\right)(Y_s + G_r)}{g_m + \dfrac{1}{r_p} + G_r + Y_L} \\[2ex]
&\qquad\qquad\qquad\qquad = Y_s + \frac{G_r\left(Y_L + \dfrac{1}{r_p}\right)}{g_m' + G_r + Y_L}.
\end{aligned}
\right\}
\quad (22)
$$

The Output Admittances.—The output admittances Y_{22} are the reciprocals of the output impedances $Z_{\alpha\beta}^{\alpha\beta}$ and can be found in the same way as the input admittances.   They are given by

$$\left.\begin{aligned}
{}_kY_{gp}^{gp} = \frac{1}{{}_kZ_{gp}^{gp}} = \frac{{}_k\Delta}{{}_kY_k^k} &= \frac{\left(\dfrac{1}{r_p} + Y_L\right)(Y_s + G_r)}{Y_s + G_r} = \frac{1}{r_p} + Y_L, \\[2mm]
{}_oY_{gp}^{gp} = \frac{1}{{}_oZ_{gp}^{gp}} = \frac{{}_o\Delta}{{}_oY_k^k} &= \frac{G_m'Y_L + (Y_s + G_r)\left(\dfrac{1}{r_p} + Y_L\right)}{g_m' + Y_s + G_r} \\[2mm]
&= Y_L + \frac{\dfrac{1}{r_p}(Y_s + G_r)}{g_m' + Y_s + G_r}, \\[2mm]
{}_pY_{pk}^{ph} = \frac{1}{{}_pZ_{pk}^{ph}} = \frac{{}_p\Delta}{{}_pY_g^g} &= \frac{(g_m + G_r)Y_s + Y_L + \dfrac{1}{r_p}(Y_s + G_r)}{G_r + Y_s} \\[2mm]
&= Y_L + \frac{1}{r_p} + \frac{(g_m + G_r)Y_s}{Y_s + G_r}.
\end{aligned}\right\} \tag{23}$$

Voltage Gain.—The voltage gain $\mathcal{G}$ is equal to the ratio of the forward transfer impedance to the input impedance, so that $\mathcal{G}$ is given by

$$\mathcal{G} = \frac{Z_{\mu\nu}^{\alpha\beta}}{Z_{\mu\nu}^{\mu\nu}}. \tag{24}$$

The values of $\mathcal{G}$ derived from Eqs. (24) and (19), using Table 13·1, are given by

$$\left.\begin{aligned}
{}_k\mathcal{G} &= \frac{{}_kZ_{kg}^{pk}}{{}_kZ_{kg}^{kg}} = \frac{{}_kY_g^p}{{}_kY_p^p} = \frac{-g_m}{Y_L + \dfrac{1}{r_p}}, \\[2mm]
{}_o\mathcal{G} &= \frac{{}_oZ_{kg}^{gp}}{{}_oZ_{kg}^{kg}} = \frac{{}_oY_k^p}{{}_oY_k^p} = \frac{g_m'}{Y_L + \dfrac{1}{r_p}}, \\[2mm]
{}_p\mathcal{G} &= \frac{{}_pZ_{gp}^{pk}}{{}_pZ_{gp}^{gp}} = \frac{{}_pY_g^k}{{}_pY_k^k} = \frac{g_m + G_r}{g_m' + G_r + Y_L}.
\end{aligned}\right\} \tag{25}$$

Transfer Admittances.—The transfer admittances Y_{12} can be worked out directly from their definition as the reciprocal of $Z_{\mu\nu}^{\alpha\beta}$. They can also be derived, however, from Eq. (24), which may be put in the alternative form

$$Y_{12} = \frac{Y_{11}}{\mathcal{G}}. \tag{26}$$

Using Eqs. (22) and (20), we get

$$\left.\begin{aligned}
{}_kY_{12} &= \frac{{}_kY_{11}}{{}_k\mathcal{G}} = -\frac{(Y_s + G_r)\left(Y_L + \dfrac{1}{r_p}\right)}{g_m}, \\[2mm]
{}_oY_{12} &= \frac{{}_oY_{11}}{{}_o\mathcal{G}} = Y_L + \frac{(Y_s + G_r)\left(Y_L + \dfrac{1}{r_p}\right)}{g_m'}, \\[2mm]
{}_pY_{12} &= \frac{{}_pY_{11}}{{}_p\mathcal{G}} = \frac{Y_s(g_m' + G_r + Y_L)}{g_m + G_r} + \frac{G_r\left(Y_L + \dfrac{1}{r_p}\right)}{g_m + G_r}.
\end{aligned}\right\} \tag{27}$$

Available Power Gain.—By definition, the available power gain P_s from a current source I_s in parallel with a source admittance $G_s + jY$, where G_s and Y are real, is the maximum power that can be delivered to a load and is given by

$$P_s = \frac{I_s^2}{4G_s}. \tag{28}$$

Accordingly when the amplifier is driven by a current source I_s, Eq. (28) gives the available input power.

To calculate the available power gain of the first stage, it will first be necessary to find the power fed into the load Y_L when there is an input current source I_s and then to choose Y_L to make this power P_L a maximum. By definition the available power gain $\mathcal{W}$ is then given by

$$\mathcal{W} = \frac{P_{L,\text{max}}}{P_s}. \tag{29}$$

Now if $Z_{\mu\nu}^{\mu\nu}$ is the input impedance of the tube, the input voltage E_g is given by

$$E_g = I_s Z_{\mu\nu}^{\mu\nu}; \tag{30}$$

and if $\mathcal{G}$ is the voltage gain, the voltage developed across the load is

$$E_L = \mathcal{G} Z_{\mu\nu}^{\mu\nu} I_s = \frac{\mathcal{G} I_s}{Y_{11}}, \tag{31}$$

so that the power delivered to the load is

$$P_L = E_L^2 G_L = \frac{I_s^2 |\mathcal{G}|^2 G_L}{|Y_{11}^2|}, \tag{32}$$

where G_L is the conductive component of Y_L, which is to be chosen to make P_L a maximum.

From Eqs. (32), (25), and (22) it follows that

$$
\begin{aligned}
{}_kP_L &= G_L \left| \frac{g_m I_s}{\left(\dfrac{1}{r_p} + Y_L\right)(Y_s + G_r)} \right|^2, \\[2ex]
{}_gP_L &= G_L \left| \frac{g_m' I_s}{\left(\dfrac{1}{r_p} + Y_L\right)\left(Y_s + G_r + \dfrac{Y_L g_m'}{\dfrac{1}{r_p} + Y_L}\right)} \right|^2 \\[2ex]
&= G_L \left| \frac{g_m' I_s}{g_m' Y_L + \left(\dfrac{1}{r_p} + Y_L\right)(Y_s + G_r)} \right|^2, \\[2ex]
{}_pP_L &= G_L \left| \frac{(g_m + G_r) I_s}{(g_m' + G_r + Y_L)\left[Y_s + \dfrac{G_r\left(Y_L + \dfrac{1}{r_p}\right)}{g_m' + G_r + Y_L}\right]} \right|^2,
\end{aligned}
\tag{33}
$$

The power gain in the grounded-cathode case is a maximum when $Y_L = 1/r_p$, so that $_k\mathbb{W}$, the available power gain, is given by

$$_k\mathbb{W} = \frac{g_m^2 G_s r_p}{|Y_s + G_r|^2}. \tag{34a}$$

The expression for $_oP_{L,max}$ is quite complicated unless all the admittances involved are purely conductive. In this special case, the optimum value of Y_L is

$$Y_L = G_L = \frac{(G_s + G_1 + G_r)}{r_p(g_m' + G_s + G_1 + G_r)}, \tag{35}$$

and the available power gain $_o\mathbb{W}$ is given by

$$_o\mathbb{W} = \frac{g_m'^2 G_s r_p}{(g_m' + G_s + G_1 + G_r)(G_s + G_1 + G_r)}. \tag{34b}$$

In maximizing $_pP_L$ it is assumed, as is almost universally the case, that

$$\frac{G_r\left(Y_L + \dfrac{1}{r_p}\right)}{g_m' + G_r + Y_L} \ll Y_s, \tag{36}$$

so that the optimum value of Y_L is given simply by

$$Y_L = g_m' + G_r, \tag{37}$$

and the available power gain $_p\mathbb{W}$ by

$$_p\mathbb{W} = \frac{(g_m + G_r)^2 G_s}{(g_m' + G_r)|Y_s|^2}. \tag{34c}$$

In the special but important case in which Y_s is purely resistive and the two inequalities

$$G_r \ll G_s, \qquad G_s \ll g_m$$

are satisfied, the expressions for the available power gains assume the simple form

$$\left.\begin{array}{l} _k\mathbb{W} = \dfrac{\mu g_m}{G_s}, \\[2mm] _o\mathbb{W} = \mu + 1, \\[2mm] _p\mathbb{W} = \dfrac{\mu g_m}{(\mu + 1)G_s}, \end{array}\right\} \tag{38}$$

so that in this case

$$_k\mathbb{W} = _o\mathbb{W}\,_p\mathbb{W}. \tag{39}$$

The First Stage Noise Figure.—In Sec. 13·2 it was pointed out that all sources of noise within the amplifier can be represented by a single source of noise in the input impedance. This artifice will not be used, however,

in calculating the noise figure in this section. The total mean square noise voltage produced across the output terminals of the tube will be determined instead. The ratio of this quantity to the mean square noise voltage produced by the thermal-noise generator associated with the source conductance G_s is the noise figure of the first stage F_1.

In order to simplify the analysis, attention is confined to the so-called "single-frequency" noise figure, which is the ratio of the available noise powers, in the infinitesimal frequency interval df, delivered by the actual amplifier to those which might be delivered by an ideal amplifier. The over-all noise figure can then be found by the averaging process described in general terms in Sec. 14·2.[1]

It has been shown above that the noise sources of the first stage can be represented, under the simplifying assumptions stated, by four statistically independent constant-current noise generators: i_s and i_1 across the input terminals μ, ν; i_p across the plate-cathode terminals; i_g across the grid-cathode terminals.

Because these generators are statistically independent, the total mean square noise voltage $\overline{e_n^2}\, df$, in the frequency interval df, produced across the output terminals of the network is equal to the sum of the mean square noise voltages produced by these generators independently. Hence,

$$\overline{e_n^2}\, df = (\overline{i_s^2} + \overline{i_1^2})|Z_{\mu\nu}^{\alpha\beta}|^2\, df + \overline{i_g^2}|Z_{kg}^{\alpha\beta}|^2\, df + \overline{i_p^2}|Z_{pk}^{\alpha\beta}|^2\, df, \qquad (40)$$

while $\overline{e_s^2}\, df$, the mean square noise voltage produced across the output terminals of an ideal amplifier by the source noise generator, is given by

$$\overline{e_s^2}\, df = \overline{i_s^2}|Z_{\mu\nu}^{\alpha\beta}|^2\, df. \qquad (41)$$

Hence the single-frequency noise figure of the first stage is

$$F_1 = 1 + \frac{\overline{i_1^2}}{\overline{i_s^2}} + \frac{\overline{i_g^2}}{\overline{i_s^2}}\frac{|Z_{kg}^{\alpha\beta}|^2}{|Z_{\mu\nu}^{\alpha\beta}|^2} + \frac{\overline{i_p^2}}{\overline{i_s^2}}\frac{|Z_{pk}^{\alpha\beta}|^2}{|Z_{\mu\nu}^{\alpha\beta}|^2}. \qquad (42)$$

Utilizing Eqs. (10), (15) and (16) one gets

$$F_1 = 1 + \frac{\alpha G_1}{G_s} + \frac{\beta G_r}{G_s}|a^2| + \frac{g_m^2 R_{eq}}{G_s}|b^2|, \qquad (43)$$

where

$$a = \frac{Z_{kg}^{\alpha\beta}}{Z_{\mu\nu}^{\alpha\beta}} \qquad \text{and} \qquad b = \frac{Z_{pk}^{\alpha\beta}}{Z_{\mu\nu}^{\alpha\beta}}. \qquad (44)$$

[1] In quantitative terms, the over-all noise factor $\bar{F}$ is given by

$$\bar{F} = \frac{\displaystyle\int_0^\infty F A^2(f)\, df}{\displaystyle\int_0^\infty A^2(f)\, df}$$

where F is the single-frequency noise figure and $A(f)$ is the characteristic of voltage absolute value vs. frequency of the amplifier from source to output terminals.

TABLE 13·2.—THE PROPERTIES OF THE THREE TRIODE INPUT CIRCUITS OF FIG. 13.5

Property	Grounded cathode	Grounded grid	Grounded plate
Δ	$(Y_s + G_\tau)\left(Y_L + \dfrac{1}{r_p}\right)$	$g'_m Y_L + (Y_s + G_\tau)\left(Y_L + \dfrac{1}{r_p}\right)$	$(g_m + G_\tau)G_s + (Y_s + G_\tau)\left(Y_L + \dfrac{1}{r_p}\right)$
Input admittance	$Y_s + G_\tau$	$Y_s + G_\tau + \dfrac{g'_m Y_L}{Y_L + \dfrac{1}{r_p}}$	$Y_s + \dfrac{G_\tau\left(Y_L + \dfrac{1}{r_p}\right)}{g'_m + G_\tau + Y_L}$
Output admittance	$Y_L + \dfrac{1}{r_p}$	$Y_L + \dfrac{\dfrac{1}{r_p}(Y_s + G_\tau)}{g'_m + Y_s + G_\tau}$	$Y_L + \dfrac{1}{r_p} + \dfrac{(g_m + G_\tau)Y_s}{Y_s + G_\tau}$
Transfer admittance	$-\dfrac{(Y_s + G_\tau)\left(Y_L + \dfrac{1}{r_p}\right)}{g_m}$	$Y_L + \dfrac{(Y_s + G_\tau)\left(Y_L + \dfrac{1}{r_p}\right)}{g'_m}$	$Y_s\left(\dfrac{g'_m + G_\tau + Y_L}{g_m + G_\tau}\right) + \dfrac{G_\tau Y_L + \dfrac{1}{r_p}}{g_m + G_\tau}$
Voltage gain	$\dfrac{-g_m}{Y_L + \dfrac{1}{r_p}}$	$\dfrac{g'_m}{Y_L + \dfrac{1}{r_p}}$	$\dfrac{g_m + G_\tau}{g'_m + G_\tau + Y_L}$
Available power gain	$\dfrac{g_m^2 G_s r_p}{|Y_s + G_\tau|^2}$	$\dfrac{g'^2_m G_s r_p}{(g'_m + G_s + G_1 + G_\tau)(G_s + G_1 + G_\tau)}$ *	$\dfrac{(g_m + G_\tau)^2 G_s}{(g'_m + G_\tau)|Y_s|^2}$
Approximate available midband power gain	$\dfrac{\mu g_m}{G_s}$	$\mu + 1$	$\dfrac{\mu g_m}{(\mu + 1)G_s}$
Noise figure	$1 + \dfrac{\alpha G_1}{G_s} + \dfrac{\beta G_\tau}{G_s} + \dfrac{R_{eq}|Y_s + G_\tau|^2}{G_s}$	$1 + \dfrac{\alpha G_1}{G_s} + \dfrac{\beta G_\tau}{G_s} + \dfrac{R_{eq}\mu^2|Y_s + G_\tau|^2}{G_s(\mu + 1)^2}$	$1 + \dfrac{\alpha G_1}{G_s} + \dfrac{\beta G_\tau}{G_s}\left(\dfrac{\varsigma_m - G_s}{g_m + G_\tau}\right)^2 + \dfrac{R_{eq}g_m^2|Y_s + G_\tau|^2}{(g_m + G_\tau)^2}$

* When all reactances are tuned out.

Whenever numerical values are to be inserted into Eq. (43), α will be taken equal to 1 and β equal to 5.

It now remains to compute a and b for the three networks of Fig. 13·5, which is done with the aid of Eq. (19) and Table 13·1.

$$
\left.
\begin{aligned}
{}_k a &= \frac{Z_{kg}^{pk}}{Z_{kg}^{pk}} = 1, & {}_k b &= \frac{Z_{pk}^{pk}}{Z_{kg}^{pk}} = \frac{Y_g^g}{Y_g^p} = -\frac{Y_s + G_\tau}{g_m}, \\
{}_g a &= \frac{Z_{kg}^{gp}}{Z_{kg}^{gp}} = 1, & {}_g b &= \frac{Z_{pk}^{gp}}{Z_{kg}^{gk}} = \frac{Y_k^g}{Y_k^p} = \frac{Y_s + G_\tau}{g_m'}, \\
{}_p a &= \frac{Z_{kg}^{pk}}{Z_{gp}^{pk}} = \frac{Y_g^p}{Y_g^k} = -\frac{g_m - G_s}{g_m + G_\tau}, & {}_p b &= \frac{Z_{pk}^{pk}}{Z_{gk}^{pk}} = \frac{Y_g^g}{Y_g^k} = \frac{Y_s + G_\tau}{g_m + G_\tau},
\end{aligned}
\right\} \quad (45)
$$

so that the noise figures in the three cases are given by

$$
\left.
\begin{aligned}
{}_k F_1 &= 1 + \frac{\alpha G_1}{G_s} + \frac{\beta G_\tau}{G_s} + \frac{R_{eq}|Y_s + G_\tau|^2}{G_s}, \\
{}_g F_1 &= 1 + \frac{\alpha G_1}{G_s} + \frac{\beta G_\tau}{G_s} + \frac{R_{eq}\mu^2|Y_s + G_\tau|^2}{(\mu + 1)^2 G_s}, \\
{}_p F_1 &= 1 + \frac{\alpha G_1}{G_s} + \frac{\beta G_\tau |g_m - G_s|^2}{G_s(g_m + G_\tau)^2} + \frac{R_{eq}g_m^2|Y_s + G_\tau|^2}{|g_m + G_\tau|^2 G_s}.
\end{aligned}
\right\} \quad (46)
$$

All the results derived in this section are displayed in Table 13·2. It will be noticed [see Eq. (39)] that the available power gain of the grounded-cathode triode is approximately equal to the product of the power gains of the other two circuits. This is a most significant point, clearly displaying the advantage of the grounded-cathode configuration.

13·4. The Equivalent Noise Resistance of Practical Tubes.—In the previous section, expressions for the noise figure of the first stage were obtained in qualitative form. Before discussing these results with a view to optimum design of the input circuit and tube arrangement, it is desirable to provide a quantitative basis for the analysis by giving typical magnitudes for the various parameters appearing in Table 13·2.

The Equivalent Noise Resistor R_{eq}.—In Eq. (17) is given a theoretical expression for the equivalent noise resistance of a tube, which is applicable either to a triode or to a pentode the screen of which is connected to its cathode.

This equation has been utilized in setting up Table 13·3, in which the values of equivalent noise resistances and transconductances of common receiving tubes are given. These values are calculated from the nominal values of the transconductance and plate and screen currents given in the *RCA Tube Handbook*. Average values of input and output capacitances, which are valid for the case when the tube has its cathode grounded, are given also; these values do not include socket or wiring capacitances.

It must be emphasized once again that these theoretical values,

although valuable as guides to performance, should be supplemented by experimental results for any given tube type. Although agreement between theory and experiment is quite good for some tube types, such as the type 6AK5, it is much less satisfactory for others, such as the type 6J4. The reasons for this discrepancy are not yet understood. It has been vaguely conjectured that the gold-plated grid of the type 6AK5, with its low emission, may be responsible for its lower noise.

TABLE 13·3.—EQUIVALENT NOISE RESISTANCES OF RECEIVING TUBES

Tube	g_m, μmhos	R_{eq}, ohms	C_{in}, $\mu\mu$f	C_{out}, $\mu\mu$f
Triode amplifiers				
6AC7	11,250	220	11.0	4.0
6AK5	6,670	385	4.0	2.0
6C4	2,200	1,140	1.8	1.3
6F4	5,800	430	2.0	0.6
6J4	12,000	210	2.8	0.2
6J5*	2,600	960	3.4	3.6
6J6*	5,300	470	2.2	0.4
6SC7*	1,325	1,890	2.2	0.3
6SL7*	1,600	1,560	3.2	3.6
6SN7*	2,600	960	2.9	1.0
7F8*	5,650	440	2.8	1.4
9002	2,200	1,140	1.2	1.1
Sharp cutoff pentodes				
1L4	1,025	4,300	3.6	7.5
6AC7	9,000	720	11.0	5.0
6AG5	5,000	1,640	6.5	1.8
6AJ5	2,750	2,650	4.1	2.0
6AK5	5,000	1,880	4.0	2.4
6AS6	3,500	4,170	4.0	3.0
6SH7	4,900	2,850	8.5	7.0
6SJ7	1,650	5,840	6.0	7.0
9001	1,400	6,600	3.6	3.0
Remote cutoff pentodes				
1T4	750	20,000	3.5	7.3
6AB7	5,000	2,440	8.0	5.0
6SG7	4,700	4,000	8.5	7.0
6SK7	2,000	10,500	6.0	7.0
9003	1,800	13,000	3.4	3.0

* One unit of a dual triode tube.

Another point to bear in mind is that there is wide variation in the equivalent noise resistances of individual tubes of a given type. A plot of

cathode current vs. transconductance of a large number of tubes under fixed operating voltages shows that, in general, cathode current increases with transconductance. Both theoretically and experimentally the noise resistance is found to decrease as the transconductance and cathode current increase.

A few tubes of each type draw appreciably more current than typical tubes with the same transconductance; an even smaller number draw less current. The former are found to have particularly large noise resistances; in fact, a tube with medium transconductance and abnormally large cathode current generally leads to a worse noise figure than a tube that has minimum transconductance and normal current. Possible causes are the presence of gas, or defects in the control-grid winding such that it does not cover the entire useful length of the cathode structure. Tubes with abnormally low currents, on the other hand, usually have very low noise resistances. In the case of multigrid tubes a possible explanation is an accidental alignment of the wires of the various grids which reduces screen current without reducing transconductance.

Optimum operating voltages of the tube are found by experiment. In general, a tube should be operated to give as high a transconductance as possible; this means that it must have large plate and screen currents. Because the input stage is not required to handle very large signals, the first tube may be operated at moderately low plate and screen voltages, and a high transconductance may be obtained by the use of a low bias. If the bias is too small, however, grid current will flow and cause the noise figure to deteriorate seriously.

The Input Circuit Loss G_1.—In Sec. 13·3 the effective temperature of G_1 was taken to be αT in order to take into account the case in which some of the loading is electronic. In the present discussion, attention is confined to the simple case in which the damping arising from causes other than transit time is entirely due to pure resistive loss, so that α can be taken equal to unity.

In general, G_1 consists of two components: G_D, the loss of the input circuit, and G_o, the cold loss of the tube.

The cold loss G_o is due partly to leakage around the bulb of the tube and partly to losses in the tube socket. The magnitude of G_o varies with different tubes and different tube sockets, but by careful design it should be possible to keep it small in comparison with G_D.

If C_s is the total input capacity and f_0 the midband frequency, then

$$G_D = \frac{2\pi f_0 C_s}{Q}, \tag{47}$$

where the Q is that of the coil. The Q varies slowly with capacity and frequency and depends chiefly upon wire diameter and mechanical

construction.   In practical amplifiers, the Q's obtained range from 80 to 300.   In this chapter a Q of 150 is assumed whenever it is desired to insert numerical values.   Typical values of C_s range from 5 to 20 $\mu\mu f$ or even higher, so that at 60 Mc/sec G_D ranges between 10 and 50 μmhos for a Q of 150.

It can be observed from Eq. (47) that G_D varies directly with f_0 and C_s and inversely with Q.

The Transit-time Damping G_r.—Despite the great importance of induced grid noise very little reliable information is available as to the transit-time loading and equivalent noise temperature βT of the common i-f and r-f tubes.   C. J. Bakker[1] has published experimental results in confirmation of the formula of Eq. (15) in which β is taken equal to 5 for the type EF-50.   These results were approximately confirmed at the MIT Radiation Laboratory for the type 6AK5; but as was the case for the equivalent shot-noise resistor, agreement between experiment and theory was much less satisfactory for other tube types.   It is thus necessary to measure the equivalent induced grid noise resistor for any given tube either directly, as in the experiments of Bakker, or indirectly by measurement of F_1 after measurement of G_1 and R_{eq}.

For the type 6AK5, which has a very small induced grid noise current, G_r may be taken equal to 12 μmhos at 30 Mc/sec.   Because G_r varies as the square of frequency, G_r is equal to 48 μmhos at 60 Mc/sec for this tube type.   For the British type CV138, the value of G_r should be taken as 20 μmhos at 30 Mc/sec and 80 μmhos at 60 Mc/sec.   Most of the other tube types listed in Table 13·3 have values roughly the same as the CV138.

13·5. The First-stage Noise Figure.—The noise figure of the first stage is approximately the same for the three arrangements of Fig. 13·5.   A short discussion is given in this section of the conditions under which this approximation is valid.

From the expression for F_1 in Table 13·2 it can be seen that the terms involving R_{eq} are approximately the same if the following two inequalities are satisfied:

$$\left.\begin{array}{r} \mu \gg 1, \\ g_m \gg G_r. \end{array}\right\} \tag{48}$$

The first of these is valid for all pentodes and all high-gain triodes; the latter is true at all frequencies at which the tube can function as an amplifier.

A more obvious difference can be found in the contribution of the induced grid noise term.   For the grounded-plate tube this term is

[1] C. J. Bakker, "Fluctuations and Electron Inertia," *Physica*, **8**, No. 1, January 1941.

$$\frac{\beta G_r (g_m - G_s)^2}{G_s (g_m + G_r)^2}, \tag{49}$$

whereas for the grounded-cathode and grounded-grid cases it is

$$\frac{\beta G_r}{G_s}. \tag{50}$$

Expressions (49) and (50) are approximately equal only if

$$\left. \begin{array}{l} g_m \gg G_r, \\ g_m \gg G_s. \end{array} \right\} \tag{51}$$

The first inequality has been discussed above; but if the latter is not satisfied, the grounded plate appears to have an advantage over the other two arrangements. This advantage is not appreciable, however, unless $\beta G_r / G_s$ is comparable with $R_{eq} G_s$ when G_s is comparable with g_m. This is not usually true, so that the conventional approximation that the noise figure of a tube is independent of its configuration is, in most cases, quite justified. A further comparison of the three arrangements is deferred until expressions for the optimum source admittance have been obtained and other aspects of the noise figure discussed.

For the present it is assumed that the inequalities of Eq. (51) are satisfied, so that the noise figures for the three arrangements are all given by

$$F_1 = 1 + \frac{\alpha G_1}{G_s} + \frac{\beta G_r}{G_s} + \frac{R_{eq} |Y_s + G_r|^2}{G_s}, \tag{52}$$

where

$$Y_s = G_s + G_1 + jY_1. \tag{53}$$

Several conclusions can be drawn immediately. First, the noise figure is a minimum when

$$Y_1 = 0,$$

that is, when the total admittance presented to the input terminals of the tube is purely conductive.[1] Thus, quite apart from bandpass requirements, it is desirable to have the input circuit resonant at band center.

Second, the noise figure is increased by increasing G_1, not only because of the increase of the term $\alpha G_1 / G_s$ but also because Y_s is increased and hence $R_{eq} |Y_s + G_r|^2 / G_s$. Thus, the noise figure is always increased by increasing the loss in the input circuit, even if this loss conductance is effectively at zero temperature ($\alpha = 0$).

13·6. The Optimum Source Admittance.—The primary purpose of the input network is to present that admittance to the input terminals of the first tube which makes the noise figure a minimum. When the over-all

[1] This conclusion will have to be modified when the correlation between shot and induced grid noise is taken into account (Sec. 13·13).

noise figure is determined by the first stage alone, this value of G_s is the one which makes F_1 a minimum. If the second-stage noise contribution is small, the correct value of G_s differs from this only slightly.

From Eq. (53) this value of $G_{s,\text{opt}}$ is given by setting

$$0 = \frac{\partial F_1}{\partial G_s} = -\frac{\alpha G_1 + \beta G_r + R_{\text{eq}}[(G_1 + G_r)^2 + Y_1^2]}{G_s^2} + R_{\text{eq}}. \qquad (54)$$

Hence

$$G_{s,\text{opt}}^2 = \frac{\alpha G_1 + \beta G_r + R_{\text{eq}}[(G_1 + G_r)^2 + Y_1^2]}{R_{\text{eq}}}. \qquad (55)$$

To express this result in a more condensed form, the distinction between induced grid noise and thermal noise will now be dropped and two new quantities G_{B1} and ρ_1 will be introduced:

$$\left.\begin{array}{l} G_{B1} = G_1 + G_r, \\ \rho_1 G_{B1} = \alpha G_1 + \beta G_r. \end{array}\right\} \qquad (56)$$

The total damping across the input circuit is G_{B1}, and $\rho_1 T$ is its effective temperature. The expression for $G_{s,\text{opt}}^2$ now becomes

$$G_{s,\text{opt}}^2 = \frac{\rho_1 G_{B1} + R_{\text{eq}}(G_{B1}^2 + Y_1^2)}{R_{\text{eq}}}. \qquad (57)$$

In most practical cases, at least as far as the input circuit is concerned,

$$\rho_1 \gg R_{\text{eq}} G_{B1},$$

and in this special case the optimum source conductance at midband ($Y_1 = 0$) assumes the value

$$G_{s,\text{opt}} = \sqrt{\frac{\rho G_{B1}}{R_{\text{eq}}}}. \qquad (58)$$

The corresponding value for the optimum attainable midband noise figure $F_{1,\text{opt}}$ is

$$F_{1,\text{opt}} = 1 + 2\sqrt{\rho G_{B1} R_{\text{eq}}}. \qquad (59)$$

Equations (58) and (59) are of fundamental importance.

From Eq. (58) it is seen that at midband $G_{s,\text{opt}}$ is inversely proportional to $(R_{\text{eq}})^{1/2}$, so that the optimum conductance presented to the pentode is, other things being equal, always less than the optimum conductance presented to the same tube connected as a triode. Thus, even apart from its better noise figure, the triode is more suitable than the pentode tube type in applications requiring a large bandwidth, because the greater $G_{s,\text{opt}}$, the greater is the bandwidth of the optimum input circuit.

At frequencies different from midband, where Y_1 is no longer zero, Eq. (57) shows that $G_{s,\text{opt}}$ should be increased over its midband value

if optimum noise figure is to be obtained over the band.   In the majority of practical cases in which the input circuit is wide compared with the receiver pass band, the importance of this effect is negligible, especially since the variation of noise figure with G_s is small when G_s is in the neighborhood of its optimum value.

As the noise contribution of the second stage becomes more appreciable, $G_{s,\text{opt}}$ tends asymptotically toward the value that enables the greatest amount of power to be delivered to the input admittance of the tube.   This value is G_{B1} in the case of the grounded plate and grounded cathode and $\dfrac{g'_m Y_L}{Y_L + 1/r_p} + G_1$ in the case of the grounded grid.   This shift is small when the noise figure of the amplifier is good.

The above results will now be illustrated by a numerical example. With the type 6AK5,

$$\begin{aligned}
R_{\text{eq}} \text{ (triode)} &= 385 && \text{ohms,} \\
R_{\text{eq}} \text{ (pentode)} &= 1880 && \text{ohms,} \\
g_m \text{ (triode)} &= 6670 && \mu\text{mhos,} \\
g_m \text{ (pentode)} &= 5000 && \mu\text{mhos.}
\end{aligned}$$

If the frequency is taken to be 60 Mc/sec typical values for G_r and G_1 are

$$\begin{aligned}
G_r &= 48 && \mu\text{mhos,} \\
G_1 &= 20 && \mu\text{mhos.}
\end{aligned}$$

Taking $\beta = 5$, $\alpha = 1$ one gets for the type 6AK5, triode-connected, at midband,

$$\begin{aligned}
G_{s,\text{opt}} &= 730 && \mu\text{mhos,} \\
F_{1,\text{opt}} &= 2.14 && \text{db,}
\end{aligned}$$

while for the type 6AK5, pentode-connected, at midband

$$\begin{aligned}
G_{s,\text{opt}} &= 330 && \mu\text{mhos,} \\
F_{1,\text{opt}} &= 3.8 && \text{db.}
\end{aligned}$$

The relative magnitude of the contribution of induced grid noise in the grounded-plate and grounded-cathode tubes can now be discussed. For the type 6AK5 triode of 60 Mc/sec considered above,

$$G_{s,\text{opt}} = 730 \qquad \mu\text{mhos}$$

so that the theoretical noise figure for the grounded-plate tube is 1.92 db, an improvement over the grounded-cathode tube of 0.2 db.   This improvement is much less appreciable at lower frequencies and under most conditions is not significant.

13·7. Variation of Noise Figure with Source Conductance and with Frequency.—If the amplifier is driven from a crystal source, the source conductance may vary from crystal to crystal by a factor of 2 or 3 to 1 or

even more. Accordingly, the variation of noise figure and bandwidth with source conductance is important in the design of an input stage. Now the expression for first-stage midband noise figure given in Eq. (52) may be put in the form

$$F_1 = 1 + \frac{G_{B1}}{G_s} \left\{ \rho + R_{eq}G_{B1} \left[1 + \left(\frac{G_s}{G_{B1}} \right)^2 \right] \right\}. \tag{60}$$

This is plotted as a function of G_s/G_{B1} for various values of $R_{eq}G_{B1}$ and for ρ equal to unity, in Fig. 13·7. If the input network is chosen so that the
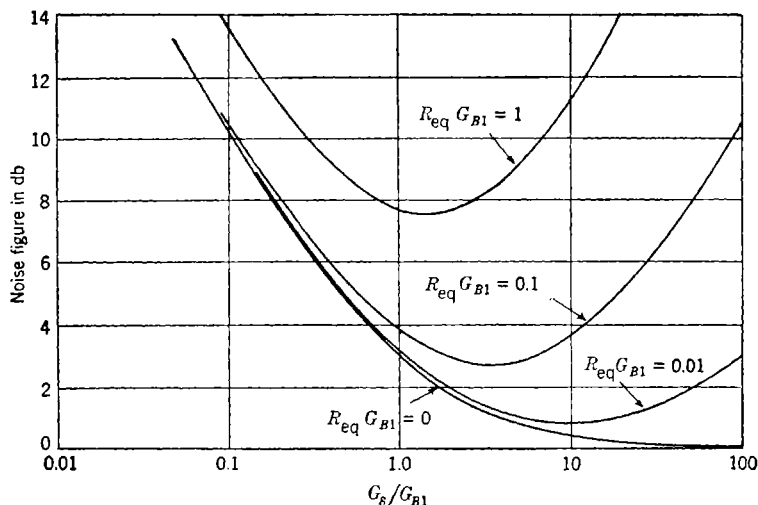
Fig. 13·7.—Variation of noise figure with source conductance various values of $R_{eq}G_{B1}$

optimum source conductance is presented to the tube when the crystal conductance assumes its geometric mean, the variation of noise figure with conductance is minimized. Thus, a four-to-one variation of source conductance produces a change in noise figure of

$$0.2 \text{ db when } R_{eq}G_{B1} = 0.01,$$
$$0.4 \text{ db when } R_{eq}G_{B1} = 0.1,$$
$$0.6 \text{ db when } R_{eq}G_{B1} = 1.0,$$

when the mean conductance is assumed to be optimum.

It will be noticed that the variation of noise figure becomes more pronounced as $R_{eq}G_{B1}$ increases, so that, in this feature as well, the triode is superior to the pentode.

In some cases, however, the source conductance may be very far from optimum. Under these circumstances a change of four to one in source conductance produces a change of nearly 6 db in noise figure.

When $G_s \ll G_{s,\text{opt}}$ the midband noise figure is given approximately by

$$F \approx \frac{G_{B1}}{G_s} + \frac{R_{\text{eq}}G_{B1}^2}{G_s}$$
$$\cong 1 + \frac{G_{B1}}{G_s} \text{ if } R_{\text{eq}}G_{B1} \ll 1;$$

(61)

when $G_1 \gg G_{s,\text{opt}}$ the midband noise figure is given by

$$F \cong 1 + R_{\text{eq}}G_s.$$

(62)

Equation (61) holds when the amplifier is driven from a pentode or grounded-grid stage and Eq. (62) holds when the amplifier is driven directly from the antenna or a grounded-plate stage.
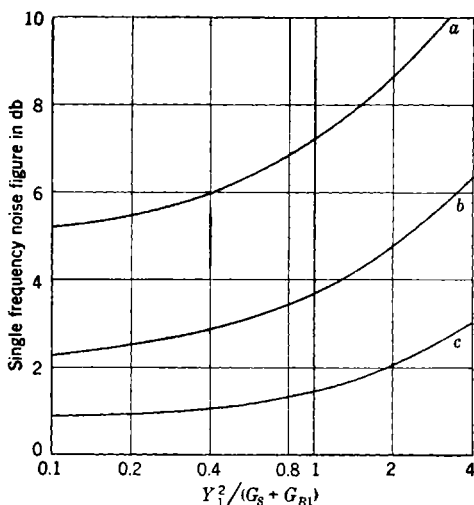
Fig. 13·8.—Single-frequency noise figure as a function of the ratios of input susceptance to input conductance. $R_{\text{eq}}G_{B1} = 1.0$ in (a), 0.1 in (b), 0.01 in (c).

In amplifiers in which the bandwidth of the input circuit is not large compared with the effective receiver bandwidth, the variation of noise figure with frequency is of interest. In Fig. 13·8 the noise figure is plotted as a function $Y_1^2/(G_s + G_{B1})^2$ for various values of $R_{\text{eq}}G_{B1}$ when G_s is chosen to give optimum midband noise figure, with ρ equal to unity.

If the input circuit is a grounded-cathode or grounded-plate tube, $Y_1^2/(G_s + G_{B1})^2$ rarely exceeds 1.25 over the effective amplifier bandwidth; but if the input circuit is a grounded-grid tube, so that the pass band of the input circuit is wide, then $Y_1^2/(G_s + G_{B1})^2$ might conceivably be large at the extremities of the effective receiver band and the variation of noise figure over the band might be appreciable.

13·8. Comparison of Alternative Tube Configurations.—It has been shown above that the noise figure of an isolated triode is appreciably

better than that of an isolated pentode. Nevertheless, because of its very low grid-plate capacity a pentode can be used in conventional grounded-cathode circuits with large stage gain and high stability. The simplicity of this design, which permits the employment of identical tubes and interstage couplings throughout the amplifier and does not require any special care in construction, justifies its use in cases in which optimum noise figure is not of paramount importance. A brief discussion of the over-all noise figure of a cascade of grounded-cathode pentodes is
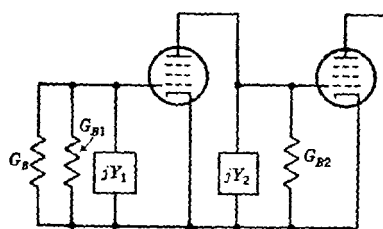
Fig. 13·9.—Grounded-cathode-pentode input circuit.

given in this section. For simplicity, it is assumed that the interstage couplings have two terminals (Fig. 13·9). The more general case is a simple extension of this.

Grounded-cathode Pentode Input Circuit.—Let G_{B2} be the total loss conductance of the second stage, including the transit time loading, $\rho_2 T$ the effective noise temperature of G_{B2}, and R_{eq_2} the equivalent noise resistor of the second tube. Now the output admittance, $1/r_p$, of the first pentode may be assumed to be high in comparison with G_{B2}, so that the midband noise figure F_2 of the second stage is of the approximate form of Eq. (59). It is given by

$$F_2 - 1 \approx r_p G_{B2}(\rho_2 + R_{eq_2} G_{B2}).$$

From Table 13·2, the available power gain of the grounded-cathode pentode is

$$\mathcal{W} = \frac{r_p g_m^2 G_s}{(G_s + G_{B1})^2},$$

where G_{B1} is given by Eq. (56).

Hence F, the noise figure at midband of the first two stages, is given by

$$F = F_1 + \frac{F_2 - 1}{\mathcal{W}} \approx 1 + \frac{\rho_1 G_{B1}}{G_s} + \frac{R_{eq_1}(G_{B1} + G_s)^2}{G_s} + \frac{G_{B2}(\rho_2 + R_{eq_2} G_{B2})(G_{B1} + G_s)^2}{g_m^2 G_s}.$$

The first-stage voltage gain $\mathcal{G}_1 \approx g_m(1/G_{B2})$, so that F may be written

$$F = 1 + \frac{\rho_1 G_{B1}}{G_s} + \frac{(G_{B1} + G_s)^2}{G_s}\left[R_{eq_1} + \frac{1}{\mathcal{G}_1^2}(R_{eq_2} + \rho_2 R_{B2})\right], \quad (63)$$

where $R_{B2} = 1/G_{B2}$. It is thus possible to neglect second-stage noise if

$$g_m^2 R_{B2}^2 = \mathcal{G}_1^2 \gg \frac{R_{eq_2} + \rho R_{B2}}{R_{eq_1}}. \quad (64)$$

When the inequality (64) is not satisfied, the effect of second-stage noise

can be taken into account by a suitable increase in R_{eq_1}, so that for optimum over-all noise figure, G_s must be decreased toward an asymptotic value of G_{B1} as second-stage noise becomes more important.

For stage voltage gains greater than 5, second-stage noise is virtually negligible, even if ρ is equal to unity, that is, if all the loading in the plate of the first stage is produced by a physical resistance. Under these circumstances, the noise figure of the amplifier is approximately equal to F_1, the noise figure of the first stage. Any appreciable improvement must, therefore, come by a reduction of F_1.

TABLE 13·4.—NOISE FIGURES OF AMPLIFIERS

Circuit	Tube type		Frequency, Mc/sec	3-db bandwidth Mc/sec		Noise figure, db*		
	First	Second		Input	Overall	Min.	Median	Max.
Pentode circuit:								
Grounded-cathode	6AC7	6AC7	30	10.0	1.5	3.4	3.9	4.4
pentode grounded-	6AK5	6AK5	30	12.0	6.0	2.6	3.3	4.5
cathode pentode	6AK5	6AK5	60	7.5	3.0	...	4.1	
Triode circuit:								
Grounded-cathode								
triode grounded-grid	6AK5	6J4	6	2	1	...	0.25	
triode†	6AK5	6J4	30	12	6	1.1	1.35	1.7
	6J4	6J4	180	30	2.5	...	5.5	
Grounded-plate triode grounded-grid triode	6J4	6J4	30	12	3.4	...	2.5	
Grounded-cathode	6AK5	6AK5	30	12	8	...	2.9	
triode grounded-cathode pentode	6AC7	6AC7	30	12	2	...	3.0	
Grounded-grid triode grounded-grid triode	6J4	6J4	180	100	2.5	...	7.0	

* Minimum and maximum noise figures are given in case of amplifiers of which at least 50 were measured. Maximum figure neglects worst 5 per cent.

† The optimum source resistance for the grounded cathode 6AK5 triode grounded-grid triode circuit was experimentally determined to be about

15,000 ohms at 6 Mc/sec,
2,500 ohms at 30 Mc/sec,
400 ohms at 180 Mc/sec.

The amplifiers in the 6- and 30-Mc/sec grounded-cathode grounded-grid cases above were built by J. L. Lawson and R. R. Nelson with every precaution to ensure optimum noise figure. The coil Q's in the input and neutralizing circuits were greater than 200, and the 6AK5 bias was selected for best average noise figure (70 ohm cathode bias resistor at 105 plate volts). The variability in noise figure in the 6-Mc/sec case was very small.

In Sec. 12·7 it was pointed out that it is theoretically possible to make the partition noise in the screen circuit nullify that in the plate by the proper use of feedback produced by a series reactance in the screen-grid

circuit. In individual experimental amplifiers at 30 Mc/sec it has been possible to improve the noise figure by about 1 db by selecting the proper value of screen bypass condenser or by connecting a small inductance between the screen-grid socket terminal and the bypass condenser. Such effects are very hard to control, and it has not been possible to manufacture amplifiers in quantity in which the noise figure was consistently improved by this kind of feedback. Accordingly, if better noise figures are to be obtained, a triode input circuit must be used.

In Table 13·4, the noise figures of typical pentode and triode amplifiers are given (also see Sec. 13·10). It should be emphasized that the first-stage noise figures of experimental pentode amplifiers increase more rapidly with frequency than the theory given in Sec. 14·3 would suggest. The reasons for this are not fully known; it is possible, however, that the larger number of uncontrolled feedback paths in the pentode tube may be responsible. This last fact makes the employment of a triode input circuit at high frequencies especially desirable when optimum noise figure is to be obtained.

Triode Input Circuits.—In Sec. 13·5, it was shown that under certain simplifying conditions, the noise figures of the three triode input circuits were the same. In this section, the analysis is developed further, and other relevant properties of the circuits are considered.

The most important of these properties are displayed in Table 13·5. They are discussed and compared in succession.

Instability and Input Capacity Due to Feedback.—The chief objection to the use of triodes in conventional high-frequency amplifiers is the presence of feedback capacity which tends to produce instability. In Row 1 of Table 13·5 the input conductance produced by feedback is given, where $G_L = 1/R_L$ is the load conductance, B_L, and B_{gp} are the load and feedback susceptances respectively, and the load reactance is assumed to have the value most conducive to oscillation.[1] In Row 2 the input capacity due to feedback is given for a case in which the load susceptance is zero.

Grounded Cathode.—The effects of feedback capacitance for a grounded-cathode triode are particularly undesirable. Not only is the effective input capacity materially increased and the bandwidth of the input circuit thereby decreased, but the negative input conductance may cause instability. Even if the source conductance is large enough to prevent oscillation, the rapid change of input conductance with frequency[2]

---

[1] For a derivation and discussion of these results see K. R. Sturley, *Radio Receiver Design,* Part I, Wiley, New York, 1943, pp. 37–56.

[2] Sturley, *loc. cit.,* shows that this input conductance is approximately equal to $(g_m B_{gp} B_L / G_L^2 + B_L^2)$, where B_L is the load susceptance and B_{gp} the feedback susceptance.

TABLE 13·5.—COMPARISON OF THE THREE TRIODE INPUT CONFIGURATIONS OF FIG. 13·5

Property	Grounded cathode	Grounded grid	Grounded plate
1. Value of input conductance due to feedback capacity when load reactance is assumed to have value most conducive to oscillation	$\dfrac{-g_m\|B_{gp}\|}{2\left(\dfrac{1}{r_p}+G_L\right)}$	$\dfrac{g_m'G_L}{\dfrac{1}{r_p}+G_L}-\dfrac{g_m\|B_{pk}\|}{2\left(\dfrac{1}{r_p}+G_L\right)}$	$\dfrac{-g_m\|B_{kg}\|}{2(g_m+G_L)}$
2. Input capacity when load is purely resistive, due to feedback	$C_{gp}\left(1+\dfrac{g_m}{G_L}\right)$	$\dfrac{C_{kg}}{\left(1+\dfrac{g_m}{G_L}\right)}$	$C_{pk}\left(1-\dfrac{g_m'}{G_L}\right)$
3. Output conductance when output coil loss is included in first stage	$\dfrac{1}{r_p}+G_D=\dfrac{1}{r_p'}$	$\dfrac{G_s}{g_m'r_p}+G_D\approx G_D$	$g_m'+G_D\approx g_m'$
4. Available power gain when output coil loss is included in first stage	$\dfrac{g_m^2 r_p'}{G_s}$	$\dfrac{G_s}{G_D+\left(\dfrac{g_m'}{G_s r_p}\right)}\approx\dfrac{G_s}{G_D}$	$\dfrac{g_m'}{G_s}$
5. First stage noise figure when output coil loss is included in first stage	$F_1+\dfrac{G_D G_s}{g_m^2}$	$F_1+\dfrac{G_D}{G_s}$	$F_1+\dfrac{G_D F_s}{g_m^2}$
6. Input coil loss G_1	$\dfrac{\omega_0}{Q}(C_1+C_{kg})$	$\dfrac{\omega_0}{Q}(C_1+C_{gp})$	$\dfrac{\omega_0}{Q}(C_1+C_{kg}+C_{kh})$
7. Neutralizing coil loss G_2	$\dfrac{\omega_0}{Q}C_{gp}$	$\dfrac{\omega_0}{Q}C_{kg}$	$\dfrac{\omega_0}{Q}C_{pk}\approx 0$
8. Output coil loss G_D	$\dfrac{\omega_0}{Q}C_2$	$\dfrac{\omega_0}{Q}C_2$	$\dfrac{\omega_0}{Q}(C_2+C_{gp})$
9. $G_{1,eq}$	$\dfrac{\omega_0}{Q}(C_1+C_{kg}+C_{gp})$	$\dfrac{\omega_0}{Q}(C_1+C_{kg}+C_{gp})$	$\dfrac{\omega_0}{Q}(C_1+C_2+C_{kh}+C_{kg}+C_{gp})$

produces asymmetry in the transfer characteristics of the input coupling network, especially if it is a double-tuned circuit. Ideally at least, these effects can be overcome either by neutralizing the feedback susceptance over the whole band[1] or, more simply, by resonating it out at midband frequency with an inductance. In the latter case, which is the only form of neutralization discussed in this chapter, the input conductance is small and positive and varies symmetrically about the midband frequency. As far as the bandwidth of the input circuit is concerned, however, the effective input capacity is still $C_{gp}(1 + g_m R_L)$. This form of neutralization becomes too critical for large values of $g_m R_L$ and is practical only in low-gain circuits.

Grounded Plate.—The effects of feedback capacitance are more complicated for the grounded-plate triode. Sturley[2] shows that the input conductance G_g due to feedback susceptance B_{kg} is given by

$$G_g = \frac{B_{kg}(G_L B_{kg} - g_m B_L)}{(G_L + g_m)^2 + (B_{kg} + B_L)^2}. \tag{65}$$

If $G_L B_{kg}$ is small compared with $g_m B_L$, G_g is negative when B_{kg} and B_L have the same sign. Hence tuning out the feedback capacity does not remove the danger of instability. It merely ensures that G_g is zero at midband and increases in magnitude symmetrically around midband to the maximum negative value given in Table 13·5. Although this maximum is much less than the corresponding quantity in the unneutralized grounded-cathode case, instability may still result.

In many practical cases,[3] however, $G_L B_{kg}$ is of the same order as $g_m B_L$, although in general it is smaller in magnitude, if B_{gk} is resonated out at midband. Under these circumstances it may be better not to tune out B_{kg} because if B_{kg} is a pure capacitance G_g does not become negative until frequencies appreciably higher than midband are reached, and the magnitude of G_g will be always less than when the capacity is tuned out. Another point to bear in mind is that it is very difficult to tune out the grid cathode capacitance at a given frequency owing to the variation of this capacitance with grid bias. In order, however, to provide an equal basis for comparison of the merits of the three alternative triode input circuits, it is assumed throughout this section that the feedback capacitance is resonated out at midband frequency.

Grounded Grid.—In conventional tube construction the grid wires lie between the plate and the cathode. By suitable arrangement of the electrode leads, therefore, it is possible to make the cathode-plate capacity

---

[1] See F. E. Terman, *Radio Engineer's Handbook*, 1st ed. McGraw-Hill, New York, 1943, pp. 467–474, for a discussion of possible neutralizing circuits.

[2] See K. R. Sturley, *loc. cit.*

[3] See the grounded-plate grounded-grid circuit discussed in Sec. 13·10 below.

much lower than either of the other two interelectrode capacities. From Row 1 in Table 13·5 it may be seen that the grounded-grid configuration is less subject to oscillation than the other two configurations and would be so even if the feedback capacity were the same in all three cases. Since the feedback capacity is in general smaller, the grounded grid is unquestionably the most stable arrangement. If the feedback capacity is large, as is the case with the triode-connected 6AK5, neutralization may be desirable but will certainly not be critical. As may be seen from Table 13·5, the input capacity due to feedback is negative in the grounded-grid circuit so that the presence of this capacity actually increases the bandwidth of the input circuit. This effect is small, however, because C_{pk} is small and any possible advantage is outweighed by the fact that the input capacity of the grounded grid must include the cathode-heater capacity C_{kh}.

First-stage Noise Figures When Losses in Neutralizing and Output Coils Are Considered.—When the noise figures of the various triode input circuits were compared (in Sec. 13·5), it was assumed that the coil loss in the input circuit was the same in all three cases. Because input capacity varies with tube configuration, however, this assumption is clearly an oversimplification. A fairer comparison would allow for this difference and would also include the losses in the neutralizing and output coils, since the capacities with which those coils are resonant are, essentially, part of the first stage. This comparison is made here on the following assumptions:

1. All coils have the same magnification Q and are all midband resonant.
2. The output capacity of the input network, apart from interelectrode capacities, is C_1 and independent of the tube configuration.
3. The output capacity of the first stage, apart from interlectrode capacities, is C_2 and independent of the tube configuration.
4. The feedback capacity in the grounded-plate and grounded-cathode cases is resonated out at midband.
5. The plate-cathode capacity C_{pk} is negligible.
6. All the losses are assumed to be at temperature T.

1. *The Input Circuit Coil Loss.*—The effects of the loss G_1 in the input coil have already been discussed in Sec. 13·5. It contributes a term G_1/G_s to the noise figure.

2. *The Neutralizing Circuit Coil Loss.*—It can be shown that the thermal-noise constant-current generator associated with the loss conductance G_{12} in the neutralizing coil produces the same effect on the noise figure as if it were connected across the input terminals; that is, it contributes a term G_{12}/G_s to the noise figure. This result is proved for the

grounded-cathode triode in the discussion of the effect of feedback resistance on noise figure given in Sec. 13·12. For the grounded-plate triode the noise generator is in parallel with the induced grid noise equivalent generator. It was shown in Sec. 13·6 that it was permissible to assume that this generator is connected across the input terminals.

The Output-circuit Coil Loss.—The effects of the loss in the output coil G_D on the noise figure depend upon the configuration. Let F_1, G_p, and $\mathcal{W}$ be the noise figure, output conductance, and available power gain of the first circuit when the output coil loss is zero. Let F'_1, G'_p and $\mathcal{W}'$ be the corresponding quantities when this loss is G_D. Then

$$\left.\begin{array}{l} F'_1 = F_1 + \dfrac{G_D}{\mathcal{W}G_p}, \\[2mm] \mathcal{W}' = \dfrac{\mathcal{W}G_p}{G_p + G_D}, \\[2mm] G'_p = G_p + G_D. \end{array}\right\} \tag{66}$$

The values for G'_p, $\mathcal{W}'$, and F'_1 calculated from the expressions for $\mathcal{W}$, G_p, etc., given in Table 13·3 are given in Rows 3, 4, and 5, respectively, of Table 13·5 on the assumption that $G_s \ll g_m$ and

$$(\mu + 1)r_p \gg \frac{1}{G_D}.$$

Unless G_D is very large, it is clear that the output coil loss has no appreciable effect on the noise figure of the grounded-cathode or grounded-plate tubes. In the grounded-grid case, however, the noise figure is approximately the same as if the thermal-noise constant-current generator associated with the output coil loss were connected across the input terminals.

The values of G_1, G_{12}, G_D in terms of Q, the midband angular frequency ω_0, and the corresponding capacities are given in Rows 6, 7, and 8, respectively, of Table 13·5. The results of the comparison are summarized in Row 9 of the table, where the value of $G_{1,eq}$ is given. This last term is the equivalent conductance that, when placed across the input terminals, produces a thermal-noise power at the output of the amplifier equal to the sum of the noise produced by the losses in the input, neutralizing, and output coils. Its magnitude is given by the equation

$$G_{1,eq} = G_1 + G_{12} + \frac{G_D G_s}{\mathcal{W}G_p}. \tag{67}$$

This equivalent loss conductance is the same for the grounded-cathode and grounded-plate circuits but larger in the grounded-grid circuit, owing to the effects of plate-to-ground stray capacitance and cathode-heater capacitance.

Available Power Gains.—The most striking and important difference between the three alternative triode configurations lies in the available power gain, which is much greater for the grounded cathode than for the other two. If output coil losses are not taken into account, the grounded grid would have a greater available power gain than the grounded plate, since, in general, $\mu + 1 > g_m/G_s$. Because of its very low output conductance, however, it is very difficult to take full advantage of the available power gain of the grounded-grid triode. As shown above, output coil losses will change the available power gain from $\mu + 1$ to G_s/G_D at the same time as output conductance is increased from $1/(\mu + 1)r_p$ to G_D. If bandwidth considerations were unimportant, it would be possible to transform the output conductance of the first stage to present the conductance to the second stage which minimizes the second-stage noise figure. Under these conditions, the contributions of the second-stage noise at midband to the noise figure of the amplifier would be in direct ratio to the available power gain. Even when output-coil loss effects are allowed for, however, the comparison is still flattering to the grounded grid. Because the output coil loss conductance is much larger than the optimum source conductance of the second stage, it must be stepped up by a ratio[1] G_s/G_D, and the output susceptance of the first stage will be stepped up in the same ratio. The interstage bandwidth is thereby narrowed, and the off-resonance second-stage noise figure is increased.

With the grounded-plate tube, the reverse holds good. The output conductance is very large, $\approx 1/g_m$. This has to be stepped down, if the optimum source conductance is to be presented to the second tube, by a ratio equal to G_s/g_m, and the output susceptance of the first tube will be stepped down in the same ratio and hence will play a very small part in determining the noise figure and bandwidth of the second stage.

In the grounded-cathode case, the output conductance $1/r_p'$ of the first tube is usually less than G_{s,opt_2}, but of the same order of magnitude. The large available power gain ensures that the second-stage contribution is much less in this case than in either of the two alternatives, even if $1/r_p'$ is not transformed to be equal to G_{s,opt_2}.

The main results of this comparison are summarized in Table 13·6.

13·9. Noise Figures of Single-triode Input Circuits.—In this section, a discussion of the noise figure of a number of alternative input arrangements using a single triode to drive a pentode chain is given. This discussion is intended to form an introduction to Sec. 13·10, in which two-triode input stages—which represent the best that can be done in the present state of the art—are analyzed.

Historically, the first bandpass amplifiers with triode input stages

---

[1] This ratio is derived on the assumption that the optimum source conductance to drive the second stage is the same as that which drives the first stage.

used a grounded-grid tube because it was known that considerable voltage gains could thus be obtained without instability. The large input conductance of the grounded-grid circuit due to cathode feedback, which

TABLE 13·6.—COMPARISON OF VARIOUS SINGLE-TRIODE INPUT CIRCUITS

Input tube	Advantages	Disadvantages
Grounded cathode	1. Highest available power gain, hence maximum possible reduction of second-stage noise 2. Output conductance $1/r_p$ of same order of magnitude as optimum source conductance of second stage 3. Equivalent loss conductance equal to that of grounded plate 4. Highest voltage gain	1. Tendency to instability with large voltage gain, but easier to neutralize and more stable with small voltage gain than grounded plate
Grounded plate	1. High output conductance, hence easy to get wide-band interstage coupling 2. Induced grid noise contribution slightly less than in alternative configurations 3. Bandwidth of input circuit greater than in grounded-cathode case because of lower input capacity	1. Variation of grid-cathode capacity with grid bias makes neutralization difficult 2. Tendency to instability, particularly with small G_L, even if grid-cathode capacity is resonated out 3. Available power gain much lower than in grounded-cathode case
Grounded grid	1. High stability due to cathode feedback and low plate-cathode capacity 2. Large input conductance giving wide transfer band-pass characteristics for input network	1. Low available power gain 2. Critical dependence of first-stage noise figure on output circuit loss 3. Greatest equivalent loss conductance 4. Low output conductance, which combined with point 1 means second-stage noise contribution important 5. Dependence of input and output admittance on load and source admittance respectively

made the input network very wide and largely independent of variation in source admittance, also worked in its favor. Despite its low power gain, which often leads to rather poor over-all noise figures, these advantages are still important enough to warrant the use of such a triode as the first stage in special cases.

Grounded-grid Triode Input Circuit.—The interstage coupling may be either a single-tuned or a double-tuned circuit. The latter has been used in production amplifiers, all the additional damping being placed in the secondary. The effects on the noise figure of the loss in the output coil are far too serious to permit the use of additional primary damping. Although the power gain of the grounded-grid amplifier is so low as to cause an appreciable second-stage noise contribution to the noise figure, such amplifiers have been built with better noise figure than straight pentode amplifiers.[1] Unfortunately, double-tuned circuits with all the damping on one side are critical and liable to asymmetry. Despite the larger gain-bandwidth product of the unequal-Q coupled circuit, therefore, it may be better to use a single-tuned circuit as shown in Fig. 13·10.
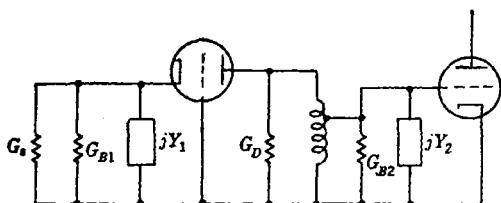
Fig. 13·10.—Grounded-grid triode driving a pentode with a single-tuned circuit interstage coupling.

To achieve the required interstage bandwidth, the input conductance of the pentode may be reduced, either by adding cathode lead inductance or by the use of resistive feedback between plate and grid, as discussed in Chap. 6. Although the second method is usually preferred for practical reasons, either of the two provides an input conductance that is at a very low effective noise temperature. The second stage in Fig. 13·10 is tapped down on the interstage coupling coil so that the coil losses are reduced.[2] For any given bandwidth there will be an optimum point at which to tap for minimum over-all noise figure. If the grounded grid is preceded by a grounded-cathode tube, as in the optimum circuit discussed in Sec. 13·10, this refinement is unnecessary.

Let G_D be the interstage coupling loss and n^2G_D the conductance presented to the input terminals of the second tube. The midband noise figure[3] F_2 of the second stage is given by

---

[1] To find the noise figure of such a combination the method detailed in Sec. 13·14 may be used, with suitable modifications. Because the analysis is quite lengthy, it will not be given here; the reader may refer to E. W. Herold, "An Analysis of the Signal to Noise Ratio of UHF Receivers," *RCA Rev.*, January 1942.

[2] The capacitance with which the coil is to resonate is thus reduced.

[3] The expression for the noise figure of the first two stages at frequencies other than midband is complicated and is not given here.

$$F_2 = 1 + \frac{\rho_2 G_{B2}}{n^2 G_D} + \frac{R_{eq_2}(n^2 G_{D2} + G_{B2})^2}{n^2 G_{D2}},$$

when the output coil loss is regarded as part of the first stage.

From Table 13·5, the noise figure and available power gain of the first stage are given by

$$F_1 = 1 + \frac{\rho_1 G_{B1} + G_D}{G_s} + \frac{R_{eq_1}(G_s + G_{B1})^2}{G_s},$$

$$\mathfrak{W}_1 \approx \frac{G_s}{G_D},$$

so that the over-all noise figure at midband is

$$F = F_1 + \frac{F_2 - 1}{W_1} = 1 + \frac{1}{G_s}\left[\rho_1 G_{B1} + G_D + \frac{\rho_2 G_{B2}}{n^2} + \frac{R_{eq_2}}{n^2}(n^2 G_{D2} + G_{B2})^2 \right]$$
$$+ \frac{R_{eq_1}}{G_s}(G_s + G_{B1})^2. \quad (68)$$

The form of Eq. (68) shows that when the first tube is a grounded grid, it is possible to take account of the noise contributions of the remainder of the amplifier by modifying ρ_1 to $\tilde\rho$, where

$$\tilde\rho G_{B1} = \rho_1 G_{B1} + G_D + \frac{\rho_2 G_{B2}}{n^2} + \frac{R_{eq_2}}{n^2}(n^2 G_{D2} + G_{B2})^2, \quad (69)$$

and the over-all noise figure becomes

$$F = 1 + \frac{\tilde\rho G_{B1}}{G_s} + \frac{R_{eq_1}(G_s + G_{B1})^2}{G_s}. \quad (70)$$

Thus, as second-stage noise becomes appreciable, G_s has to be increased for optimum over-all noise figure. The asymptotic value toward which $G_{s.opt}$ tends cannot be derived from Eq. (68) owing to the approximations made in deriving the power gain of the grounded-grid amplifier. It is possible to show, however, that this value is the one that enables the source to deliver the maximum power to the grounded-grid amplifier.

An alternative method of increasing bandwidth in the interstage coupling without adding physical resistance is to use a second grounded-grid stage to follow the first.

The interstage coupling network is of the same form as in Fig. 13·10, the tap position being determined as before by bandwidth and noise-figure requirements. A detailed discussion of this circuit is not given here, but it can be shown that the noise contribution of the second grounded-grid stage becomes appreciable at the edges of the pass band because of the large susceptance presented to its input terminals. Third-stage noise is also appreciable with this circuit except for very narrow bandwidths. The main use of such an input configuration is at fre-

quencies around 200 Mc/sec. The noise figure of a practical amplifier using this arrangement is given in Table 13·4.

Grounded-plate Triode Input.—If stability considerations are neglected, the grounded-plate triode input can give performance superior to those of the grounded-grid circuit as regards both bandwidth and noise figure. In the only application of this configuration yet used, a grounded-grid triode followed the input grounded-plate stage. Discussion of this case is therefore more appropriate in Sec. 13·10 and is given there.

Grounded-cathode Triode Input. Because of its high available-power gain, the grounded-cathode triode is the ideal choice for an input stage, provided it can be kept

FIG. 13·11.—High-stability interstage coupling with grounded-cathode-triode input stage.

stable. In the analysis of the stability of this configuration given in Sec. 13·8, it was pointed out that the negative input conductance was inversely proportional to the load conductance G_L presented to the output terminals of the tube. Accordingly, if neutralization is to be noncritical, G_L must be large.

If the over-all amplifier bandwidth is very large, the stage gain and hence $1/G_L$ are necessarily small. Under these circumstances, a straightforward coupling between the grounded-cathode input stage and the second stage might be practical, because second-stage noise will be no more important in this case than if a pentode input stage were used. The chief disadvantage would lie in the large capacitance across the input coupling network due to the Miller effect. At narrower bandwidths, however, the use of a wide-band, heavily damped interstage coupling network throws away much of the advantage inherent in the large available power gain of the first tube.

A possible compromise is effected by the circuit shown in Fig. 13·11. The plate of the grounded-cathode input tube is tapped down on the interstage coupling coil so that a large conductance[1] is presented to the plate of the triode.

If G_{B2} is the loading across the grid of the second stage with effective temperature $\rho_2 T$ and if n^2 is the impedance step-up ratio of the autotransformer, then F_2, the midband noise figure of the second stage, is given by

$$F_2 \approx 1 + n^2 r_p G_{B2}(\rho_2 + R_{eq_1} G_{B2})$$

[1] This conductance should lie between $g_m/\sqrt{3}$ and $g_m/4$, the minimum value depending upon the size of the grid-plate capacity of the grounded cathode, the ease with which the neutralization can be effected, and the midband frequency.

if $n^2 r_p \gg G_{B2}$ and the over-all noise figure at midband is

$$F = 1 + \frac{\rho_1 G_{B1}}{G_s} + \frac{(G_{B1} + G_s)^2}{G_s}\left[R_{eq1} + \frac{(\rho_2 R_{B2} + R_{eq2})}{n^2 \mathcal{G}^2} \right], \qquad (71)$$

where $\mathcal{G} = g_m/n^2 G_{B2}$ is the voltage gain to the plate of the first tube and $G_{B2} = 1/R_{B2}$. The quantity $\mathcal{G}$ is fixed by stability considerations and G_{B2} by bandwidth requirements. With narrow bandwidths, where $\mathcal{G}^2 n^2$ is large, the noise figure of the circuit is effectively that of the first stage. Because the plate-output susceptance of the grounded-cathode triode is stepped down by the coil, it has only a small effect on the bandwidth of the interstage coupling. Hence, if ρ_2 is appreciably less than unity,[1] good noise figures can be obtained even for quite appreciable bandwidths. This circuit therefore is preferable to the grounded-grid circuit. It is not the best that can be done, however, because the second-stage noise contribution is greater than if a pentode first stage had been used. The correct answer to the problem is to use a grounded-grid tube to follow the grounded-cathode input stage; this provides a large conductance across the output terminals of the first tube and yet gives a noise figure very close to the theoretical optimum. This configuration will receive detailed analysis in the next section.

It will be noticed from Eq. (71) that the second-stage noise contribution can be taken fully into account by suitable modification of the equivalent noise resistance of the first tube. Accordingly, as second-stage noise becomes predominant, the optimum source conductance tends asymptotically toward the value G_{B1} which causes the maximum power to be delivered to the first tube. This result also holds for the grounded-plate case and is in direct contrast with the grounded-grid case where second-stage noise is taken into account by suitable modification of ρ_1.

13·10. Double-triode Input Circuits.—The ideal triode input circuit should have the following features:

1. All the improvement in input stage noise figure over the pentode that is theoretically possible.
2. The contribution of second and later stages to the noise figure no greater than with pentode input stages of the same bandwidth.
3. A circuit that is stable and no more critical in adjustment than with pentode stages of the same bandwidth.

There are nine combinations possible with two triode input tubes because a choice between the grounded-cathode, grounded-grid, and grounded-plate connection may be made for each of the first and second stages. There is, however, only one circuit that meets all the above requirements.

---

[1] This is the case if the loading across the tuned circuit is provided largely by electronic means, as discussed above for the grounded-grid triode interstage coupling.

That is the combination of a grounded cathode followed by a grounded grid.[1] This circuit is considered below in detail, and a discussion of practical amplifiers using this input arrangement is given. A short discussion of alternative input circuits using a grounded-plate and a grounded-grid first stage illustrates the superiority of the grounded-cathode triode first stage.

Grounded-cathode-triode–Grounded-grid-triode Input Configuration.— The circuit described here has permitted noise figures as low as 0.25 db at 6 Mc/sec, 1.35 db at 30 Mc/sec, and 5.5 db at 180 Mc/sec, without critical adjustment of any sort (see Table 13·4).

When a grounded-grid stage succeeds the grounded cathode, as shown in schematic form in Fig. 13·12, a very large conductance $\approx g_{m2}$ is presented to the plate of the first tri-ode so that it is quite stable.[2] The bandwidth of the interstage coupling will be very large so that there is no need to add additional damping even in the most extreme cases. Finally it may be observed that the output conductance of the first tube is of the same order of magnitude as the optimum source conductance for the second tube, so that the full available power gain of the grounded-cathode

FIG. 13·12.—Grounded-cathode-tri-ode–grounded-grid-triode input con-figuration.

triode is utilized. Neglecting output coil loss, the output admittance of the grounded grid is low, because the admittance presented to its input is small compared with g_m. If coil loss is taken into account, the output admittance is equal to the conductance G_D of the coil loss, and the con-tribution of third-stage noise can be estimated as in Sec. 13·9.

The neutralizing circuit of Fig. 13·12 modifies both the output admittance and the available power gain of the first stage. For the present this effect is neglected, and it is assumed that the output con-ductance of the grounded-cathode tube is $1/r_p$. The noise figure of the second stage F_2, then, is given by

$$F_2 = 1 + \rho_2 G_{B2} r_p + r_p R_{eq_2}\left[\left(G_{B2} + \frac{1}{r_p}\right)^2 + Y_2^2\right].$$

[1] See Henry Wallman, A. B. Macnee, and C. P. Gadsden, "Low-Noise Amplifier," *Proc. I.R.E.* **36** (1948), 700–708.

[2] In Fig. 13·12 a neutralizing coil is provided to resonate out the grid-plate capacity, but this is included in order to improve the noise figure rather than for reasons of stability. Even at frequencies as high as 180 Mc/sec it is possible to omit the neutral-izing coil and still maintain stability; the noise figure in that case is degraded, however, from 5.5 db with neutralizing coil to 8.0 db without.

From Table 13·2 the available power gain of the first stage is

$$\frac{g_m^2 G_s r_p}{(G_s + G_{B1})^2 + Y_1^2};$$

and since the noise figure of the first stage is

$$F_1 = 1 + \frac{\rho_1 G_{B1}}{G_s} + \frac{R_{\text{eq}_1}[(G_s + G_{B1})^2 + Y_1^2]}{G_s},$$

the exact expression for the noise figure of the first two stages combined is

$$F = 1 + \frac{\rho_1 G_{B1}}{G_s} + \frac{(G_s + G_{B1})^2 + Y_1^2}{G_s} \left\{ R_{\text{eq}_1} + \frac{G_{B2}}{g_{m1}^2} \right. \\ \left. + \frac{R_{\text{eq}_2}}{g_{m1}^2} \left[\left(G_{B1} + \frac{1}{r_p} \right)^2 + Y_2^2 \right] \right\}. \quad (72)$$

In general, $1/r_p \gg G_{B1}$ so that, at midband, the shot noise of the second tube makes a contribution only $1/\mu_1^2$ that of the shot noise of the first tube.   It is not until the frequency is so different from midband that Y_2^2/g_{m1}^2 is comparable with unity that the second-tube shot noise becomes important.   The contribution of the second-stage thermal noise and induced grid noise will also be negligible in comparison with that of the first tube if $(G_s/g_m)^2 \ll 1$.

Thus, although it is necessary to make a careful choice of the first tube and its operating conditions, the choice of the second tube depends very little on its equivalent noise resistance.   It is usually made solely by the criterion of low cathode-plate capacity because if this capacity is large, neutralization of the second tube may be necessary, although in such cases it is usually quite uncritical.   The operating conditions of the second tube are also unimportant; if it is considered desirable to economize in power consumption, the tube can be run with high grid bias and low plate voltage.

The second stage should have a fairly high transconductance, however, because, as already noted, the stability of the first stage results from the heavy loading applied to the plate of the first stage by the input conductance ($\approx g_{m2}$) of the second stage.

If the grounded-grid stage is followed by a stage the midband noise figure of which is $F_3 \approx 1 + (\mu + 1)r_p G_{B3}(\rho_3 + R_{\text{eq}_3}G_{B3})$ when fed from a high output admittance,[1]

$$\frac{1}{r_p^2 g_{m2}'} = \frac{1}{(\mu + 1)r_p},$$

then the over-all noise figure of the combination, at midband,

[1] In deriving this expression it has been assumed that the dynamic resistance r_p of the first and the second tubes is the same.

$$F \approx 1 + \frac{\rho_1 G_{B1}}{G_s} + \frac{(G_{B1} + G_s)^2}{G_s} \left\{ R_{eq_1} + \frac{1}{g_{m1}^2} [\rho_2 G_{B2} + G_{B3}(\rho_3 + R_{eq} G_{B3})] \right\},$$

where the contribution of second-stage shot noise has been neglected. In general $\rho_2 G_{B2}/g_{m1}^2$ can also be neglected.   In this case if

$$\mathcal{G} = \frac{G_{B3}}{g_{m1}},$$

the expression for noise figure assumes the form

$$F \approx 1 + \frac{\rho_1 G_{B1}}{G_s} + \frac{(G_{B1} + G_s)^2}{G_s} \left[R_{eq_1} + \frac{1}{\mathcal{G}^2} (\rho_3 R_{B3} + R_{eq_3}) \right]. \quad (73)$$

But this is of exactly the same form as Eq. (63), which gives the noise figure of a pentode followed by a second stage.   The differences are that g_{m1} is greater for a pentode connected as a triode than for a pentode, so that $\mathcal{G}$ is larger for the triode connection, whereas R_{eq_1} is now the equivalent noise resistance for a triode instead of that for a pentode.   This double-triode circuit thus achieves the ideal standard laid down at the beginning of this section.

To conclude the theoretical discussion of the grounded-cathode grounded-grid circuit, the effects of the neutralizing circuit are analyzed below.   It is assumed for simplicity that the input coupling network is a single-tuned circuit whose magnification Q_1 is given by

$$Q_1 = \frac{\omega_0 C_s}{G_s + G_1},$$

where C_s is the input stray capacity (excluding Miller capacity) and G_{B1} is the input damping (excluding the neutralization coil loss).

If G_{12} is the conductance of the neutralization coil loss with magnification Q, then

$$G_{12} = \frac{\omega_0 C_{gp}}{Q},$$

and the output admittance Y_{22} of the first tube is given approximately by

$$Y_{22} = \frac{1}{r_p} + \frac{g_m G_{12}(1 + Q_1 Q \alpha^2)}{(G_s + G_{B1})(1 + Q_1^2 \alpha^2)} + j\omega_0 C_{gp} \alpha \frac{g_m}{(G_s + G_{B1})(1 + Q_1^2 \alpha^2)} \quad (74)$$

where $\alpha = (\omega/\omega_0) - (\omega_0/\omega)$.   This means that the midband output conductance is reduced to $1/r_p'$, where

$$\frac{1}{r_p'} = \frac{1}{r_p} + \frac{g_m G_{12}}{(G_s + G_{B1})}, \quad (75)$$

while it can be shown that the available midband power gain is

$$\frac{g_m^2 r_p'}{G_s}.$$

The effect on the contribution of the third stage is negligible, as can be seen from Eq. (73) in which $1/r_p$ does not appear explicitly. It is true that the noise contributions of the second-stage shot noise will be increased as $1/r_p$ is increased, as may be seen from Eq. (72), but this contribution is so small in any case that the over-all effect is insignificant.

Another, and more serious, effect of the grid-plate circuit of the first tube is that the output susceptance is increased by a factor $g_m/(G_s + G_{B1})$. This narrows the bandwidth of the interstage coupling and causes the shot noise of the second tube to contribute appreciably at frequencies far from the midband unless C_{gp} is small.[1]

The 3-db bandwidth of the single-tuned input circuit is equal to

$$\frac{G_s}{2\pi(C_1 + 2C_{gp} + C_{kg})},$$

where C_1 is the input capacity apart from that of the tube. It must be remembered, however, that as shown in Sec. 13·6, the equivalent thermal-noise constant-current generator, which gives a noise power equal to that produced by the losses in the neutralizing and input circuits, has a mean square value equal to

$$\frac{C_1 + C_{gp} + C_{kg}}{Q}$$

and not to

$$\frac{C_1 + 2C_{gp} + C_{kg}}{Q}.$$

Nothing has been said in this discussion about the effects of cathode-lead inductance. This might be expected to have an important effect on the performance of the grounded-cathode tube, because it is well known that feedback due to this inductance is responsible for the greater part of the input conductance of the tube. In Sec. 13·12, however, it is shown that this inductance has no effect upon the noise figure and only a small effect on the available power gain of the first triode. Its effect on the over-all noise figure is thus very small.

In Table 13·4, the noise figures of a number of practical amplifiers using this circuit are given. It can be seen that they represent a considerable improvement over any of the other arrangements listed.

A Practical Grounded-cathode-triode–Grounded-grid-triode Amplifier.— The circuit diagram is given in Fig. 13·13; photographs of the unit are shown in Figs. 13·14 to 13·16. This unit was designed to replace the first three i-f stages of a radar receiver that employed a type 6AC7 grounded-cathode pentode input stage. The output and power connections therefore are made through a built-in plug which plugs into the socket of the

---

[1] This is why it is necessary to tune out the grid-plate capacity of the first tube to get optimum noise figure (see footnote 2 on p. 657).

third tube of the original receiver. The input cable, which is attached to the unit, goes to a crystal mixer with an output resistance of 300 ohms. The capacity of the mixer plus that of the cable is about 30 $\mu\mu$f. Only minor alterations in the original receiver were necessary to accommodate this unit. Because of space limitations the input circuit coils had Q's of only about 120. The average noise figure in production was about 1.6 db. The input network is a degenerate "π" or "inverted L" network, as shown in Fig. 13·13. It is designed to operate into an input
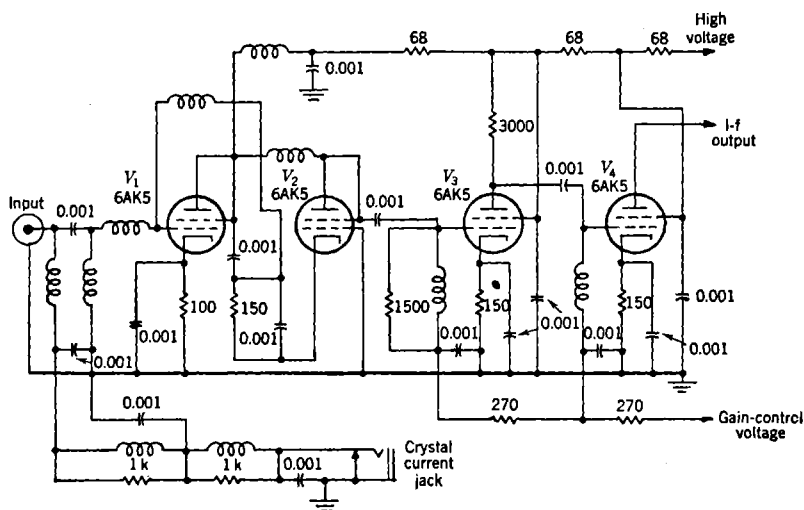
FIG. 13·13. Circuit diagram of practical grounded-cathode-triode–grounded-grid-triode input amplifier.

capacity of 7 $\mu\mu$f plus 2.4 $\mu\mu$f due to Miller effect, and it is transitionally coupled at a crystal resistance of 500 ohms. As a result, with the average crystal resistance of 300 ohms, it is somewhat less than transitionally coupled. The first tube is neutralized by L_7, which resonates the 1.2-$\mu\mu$f grid-plate capacity. Neutralization of the first stage causes an 0.25-db improvement in noise figure. The cathode-plate capacity of the grounded-grid 6AK5 tube, V_2, is 3.1 $\mu\mu$f, which is very high. It was found desirable to neutralize[1] the cathode-plate capacity by L_9.

The direct current flowing between the cathode of V_2 and ground follows a devious route: through R_3, which supplies grid bias, and then through L_7, L_5, and L_4 to ground.

---

[1] Use of a type 6J4 or 6J6 triode for the grounded-grid stage would be much better. The type 6J6 tube is best employed with only one of its triode sections functioning, pins 1, 3, 4, 5 being connected star-fashion to the tube socket center pin and grounded. Because of the large cathode-plate capacitance of a triode-connected grounded-grid 6AK5 (resulting from the internal connection of suppressor and cathode) the type 6AK5 makes a very inconvenient grounded-grid stage.
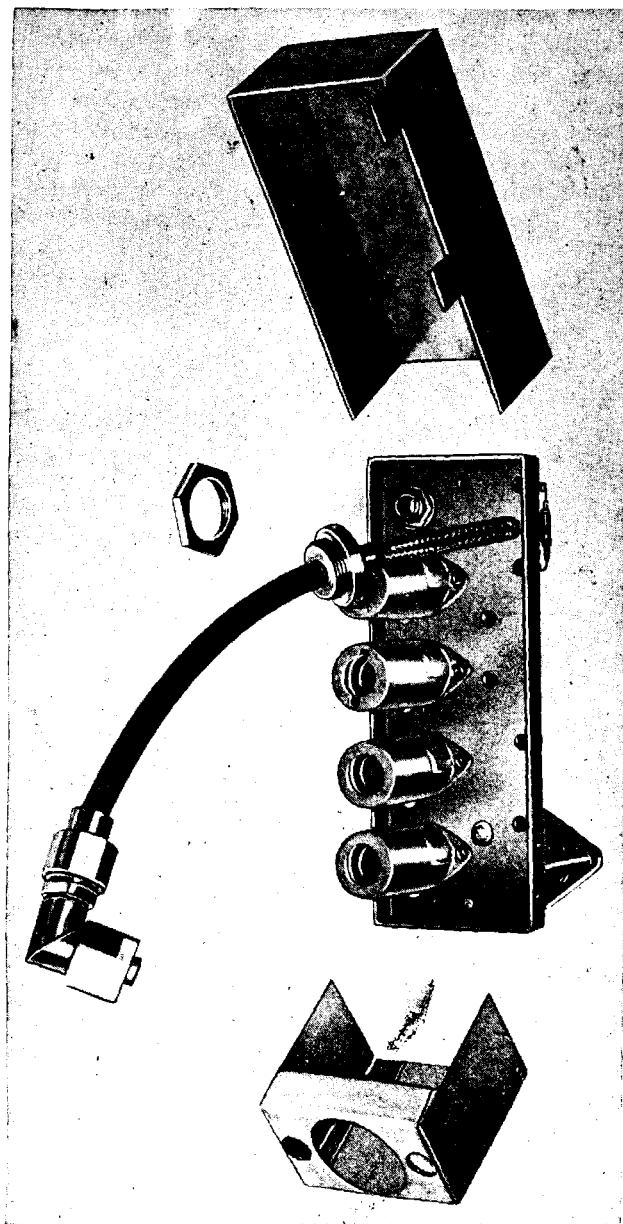
FIG. 13·14.—Exploded view of Fig. 13·13.

The type 6AK5 was selected as first tube because of its excellent noise figure. Its use as the second tube, in the grounded-grid stage, was purely a matter of convenience.

The neutralizing coils L_7 and L_9 were adjusted by connecting a signal generator to the input terminals, placing a tube with one heater pin broken off in sockets V_1 and V_2 in turn and so adjusting L_7 and L_9 that no output was obtained at the center frequency.

This amplifier was produced in considerable quantity. When the correct values of the neutralizing inductances were found for the prototype, the coils for other amplifiers were made by reproducing these inductances with a 5 per cent accuracy. These coils were then soldered into the amplifiers without further adjustment. The interstage cir-
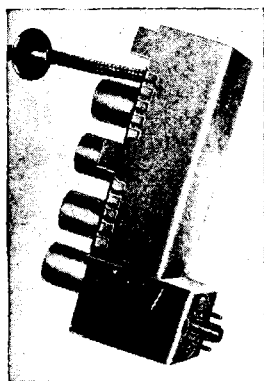
FIG. 13·15.—Side view of Fig. 13·13.
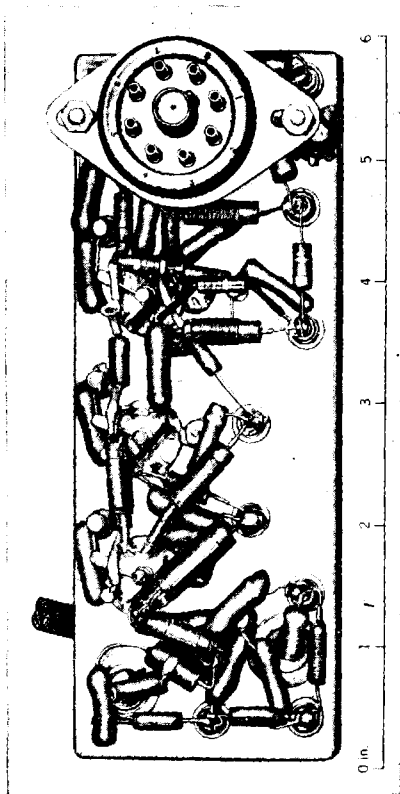
FIG. 13·16.— Bottom view of Fig. 13·13.

cuit between V_1 and V_2 is very broad, and therefore 5 per cent accuracy suffices for L_8 also. All the other coils are reproduced with an accuracy of 2 per cent.

Separate d-c returns are provided for the crystal mixer and the grid of V_1 by the use of a blocking condenser C_1. The choke coil L_1, in parallel with L_4, forms one of the members of the degenerate π-network. A decoupling filter, consisting of C_2, C_3, C_4, L_2, and L_3, is connected

between the end of L_1 and the telephone jack which is used for connecting in a meter for measuring crystal current.

Grounded-plate-triode–Grounded-grid-triode Input Configuration.— Because this amplifier configuration has been investigated at a number of laboratories, a short discussion[1] will be given here.

The basic circuit[2] is shown in Fig. 13·17. The cathode of the first tube is tapped down on the interstage coupling coil so as to present the optimum admittance to the grounded-grid tube for minimizing the second-stage noise figure. Since the input conductance of the grounded-grid stage is
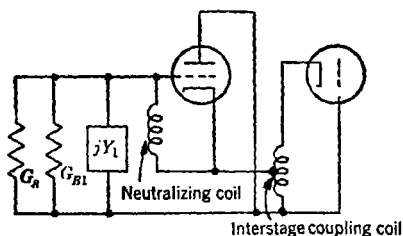
Fig. 13·17.—Grounded-plate-triode–grounded-grid-triode input configuration.

large, the bandwidth of this circuit is very large.

The available power gain of the first stage is g_{m1}/G_{s1}. If G_{s2} is the optimum conductance to present to the second stage, then the over-all noise figure of the first two stages, (assuming that the cathode-grid capacity is tuned out) is given by

$$F = 1 + \frac{\rho_1 G_{B1}}{G_{s1}} + \frac{(G_{B1} + G_{s1})^2 + Y_1^2}{G_{s1}} \left\{ R_{eq_1} + \frac{G_{s1}}{g_{m1}} \left[\frac{\rho_2 G_{B2}}{G_{s2}} + \frac{(G_{B2} + G_{s2})^2 + Y_2^2}{G_{s2}} R_{eq_2} \right] \right\}. \quad (76)$$

This expression is very similar[3] to that for the noise figure of the grounded-cathode-triode–grounded-grid-triode circuit, except that second-stage noise is reduced in the ratio G_{s1}/g_{m1} instead of $(G_{s1}/g_{m1})^2$.

To the first order, $G_{s1} \approx G_{s2}$ and $F_1 \approx F_2$, where F_1 and F_2 are the noise figures of the first and second stages respectively. Hence

$$F \approx 1 + (F_1 - 1)\left(1 + \frac{G_{s1}}{g_{m1}}\right) \approx 1 + 2\sqrt{G_{B1}R_{eq}}\left(1 + \sqrt{\frac{\rho G_B}{R_{eq}}}\frac{1}{g_m}\right)$$
$$= 1 + 2\sqrt{G_{B1}R_{eq}} + \frac{2G_{B1}}{g_m}, \quad (77)$$

so that the noise figure of the combination does not differ appreciably from that of the first stage above.

[1] For a more detailed analysis, see G. C. Sziklai and A. C. Schroeder, *Proc. I.R.E.*, **33**, 707–711, October 1945.

[2] This optimum adjustment of the grounded-plate–grounded-grid combination requires either two separate triodes or a dual triode with separate cathodes, as does the grounded-cathode–grounded-grid combination.

[3] The slight reduction in induced grid noise, which is a feature of the grounded-plate triode, has been neglected in deriving this expression.

The output admittance of the combination is that of the single grounded-grid stage, so that the presence of third-stage noise can be allowed for by suitable increase in G_{B2}, as for the grounded-cathode triode grounded-grid triode.   The lower available power gain of the grounded-plate triode means that the contribution of third-stage noise is correspondingly greater.

A more powerful objection to the use of this circuit is its tendency to instability.   This point has already been made in the general discussion of the grounded-plate circuit given in Sec. 13·8, and only a short analysis will be given here.   With the circuit of Fig. 13·17 the load conductance G_L presented to the output terminals of the grounded-plate tube is large compared with g_m.   In fact

$$G_L \approx \frac{g_m^2}{G_{s2}} \tag{78}$$

where G_{s2} is the conductance presented to the cathode of the grounded-grid triode.[1]

It was stated in Sec. 13·8 that the input conductance of the grounded-plate tube is given by

$$G_g = \frac{B_{kg}(G_L B_{kg} - g_m B_L)}{(G_L + g_m)^2 + (B_{kg} + B_L)^2} \tag{79}$$

where B_{kg} and B_L are the grid-cathode and load susceptances respectively.

If the grid-cathode and load capacitances are tuned out, G_g is symmetrical about midband frequency but is negative unless B_{kg} is increased by the addition of extra capacitance, since $g_m B_L$ is greater than $G_L B_{kg}$. To prove this it may be observed that the total cathode-ground capacitance load in the grounded-plate tube is

$$C_L = C_{kh} + C_2 + \frac{g_m}{G_{s2}} (C_{kh} + C_2 + C_{kg}), \tag{80}$$

when C_{kh}, C_{kg} are the cathode-heater and cathode-grid capacitances and C_2 is the wiring and tube socket stray capacitance.[2]   If C_0 is the self-capacity of the neutralizing coil, the total grid-cathode capacitance of the grounded-plate circuit is

$$C_{kg} + C_0$$

and

$$G_L(C_{kg} + C_0) - g_m C_L = \frac{-g_m^2}{G_{s2}} (C_2 + C_{kh} - C_0) - g_m(C_{kh} + C_2), \quad (81)$$

---

[1] In deriving the expression it has been assumed that the transconductances of the two triodes are the same and that $1/r_{p2} \gg G_{L2}$, the load conductance presented to the plate of the grounded-grid tube.

[2] It has been assumed that C_{kh}, C_{kg}, and C_2 are the same for both triode tubes.

which is negative unless

$$C_0 > (C_2 + C_{kh})\left(\frac{G_{s2}}{g_m} + 1\right).$$

If the grid-cathode capacity is not tuned out, the input conductance of the tube is asymmetrical about midband and becomes negative at high frequencies when C_0 is sufficiently large.

In either case the danger of instability at high frequencies is apparent. It may prove necessary, therefore, to introduce additional resistive loss in the input circuit[1] or to increase the grid-cathode capacitance to prevent oscillation. Either course of action increases the noise figure, and the second lowers the bandwidth of the input circuit.

To sum up, it may be said that this circuit is inferior to the grounded-cathode–grounded-grid circuit in noise figure and stability. It has a slight advantage in input capacity; and at frequencies so high that βG_r is comparable with $G_{s,\text{opt}}$, the lower induced grid noise contribution could possibly swing the scales in favor of the grounded-plate grounded-grid circuit. Usually, however, the grounded-cathode grounded-grid circuit is to be preferred.

In one particular simplification of the grounded-plate triode grounded-grid triode circuit, called the "cathode-coupled circuit," the cathode of the first triode is directly connected to that of the second, thus permitting the use of a double triode such as the type 6J6, having only one cathode. The noise figure of this simple combination is poorer than that of the grounded-plate grounded-grid circuit in which impedance stepup is employed between the first and second cathodes. The cathode-coupled combination is equivalent in noise figure to a single grounded-cathode triode with twice the value of R_{eq} and is equivalent in gain to a single grounded-cathode triode with half the transconductance. Somewhat better noise figures can be obtained, however, with the cathode-coupled circuit than with grounded-cathode pentode amplifiers.

13·11. General Considerations of the Effect of Feedback on Noise Figure.—In this section, it is shown that within certain limitations, the introduction of feedback does not influence the noise figure[2] but does affect the bandwidth of the input circuit by changing the input admittance of the first tube. Consider first the effect on noise figure. In Fig. 13·18 is shown a signal source A and amplifiers B and F. Feedback can be introduced by closing the switch S so that some of the output voltage of the amplifier B is transmitted through the network E back to the input

---

[1] This was, in fact, the means employed to obtain stability in a grounded-plate-grounded-grid amplifier built at another laboratory.

[2] This subject has been treated previously by W. A. Harris, "Fluctuations in Vacuum-tube Amplifiers and Input Systems," *RCA Rev.*, **5**, 505–524, April 1941; **6**, 115–124, July 1941.

terminals. Examples are feedback due to an impedance in series with the cathode of a tube (a class that includes the cathode follower) and feedback due to a capacitance or resistance connected between grid and plate of a single grounded-cathode stage. On the other hand, the grounded-grid amplifier, which is also a feedback amplifier, because its output current flows through its input circuit, utilizes feedback of another
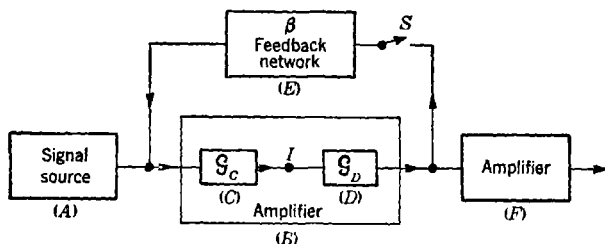
FIG. 13·18.—Block diagram of amplifier with feedback.

type. The point I denotes an independent source of noise within the amplifier B, and $\mathcal{G}_C$ and $\mathcal{G}_D$ denote the voltage gains between the input of B and I and between I and the output of B, respectively; β is the voltage gain of the network E and is assumed to be less than unity. In the argument that follows it is assumed that

1. The network E and the amplifier B transmit signals only in the directions denoted by the arrows.
2. The amplifier B has sufficient gain to make sources of noise within amplifier F negligible.
3. The network E contains no sources of noise.

The noise figure will be computed with no feedback (S open). The noise source I is replaced by a voltage E_i located in the signal source such that E_i gives the same voltage at the output of the amplifier B as the actual source of noise. This causes the available noise power of the signal source to increase by

$$\mathbb{W}_i = \tfrac{1}{4} G_s |E_i|^2. \tag{82}$$

According to the definition of noise figure given in Sec. 13·2

$$F - 1 = \sum_j \frac{G_s |E_i|^2}{G_s |E_s|^2}, \tag{83}$$

where the summation is taken over all sources of noise including the thermal noise of the signal sources, whose voltage is E_s.

The voltage at I is E_i. When feedback is introduced by closing S, the voltage at I becomes E_i', where

$$E_i' = E_i + \alpha\beta E_i',$$

and where $\alpha = \mathcal{G}_D \mathcal{G}_c$ the voltage gain without feedback.   This equation may be solved for E'_i to give

$$E'_i = \frac{E_i}{1 - \alpha\beta}. \tag{84}$$

It should be emphasized that $\alpha\beta$, the total gain of the networks around the loop, is independent of the location of the point I.   Therefore all independent noise voltages within the loop, including those at the input (such as the signal source) and at the output, are modified by the same factor[1] $1/(1 - \alpha\beta)$ so that the noise figure is independent of $\alpha\beta$.

In many wide-band amplifiers, however, the above assumptions (and therefore the previous argument) are not strictly valid.   In the first place, the feedback network E usually consists of passive elements only, and signals can be transmitted through it to the output of the amplifier B. If the amplifier B consists of a single wide-band stage with its inevitable low gain, the signal fed through E is not negligible and causes the equivalent voltage pertaining to the signal source I_s to be modified by a different factor from that which modifies other sources located within the amplifier B.   In such cases second-stage (amplifier F) noise must be considered also and becomes more important as the gain of amplifier B is reduced by introducing feedback.   In some cases, also, the feedback network contains resistors or may itself be a resistor from grid to plate of the amplifier tube.   Such resistors, of course, are sources of thermal noise.

When these conditions prevail, the general argument given previously is not completely valid, and the noise figure changes with introduction of feedback, although the change is usually small.

Another situation that is not covered by this argument is that in which a source of noise within the second amplifier F is coherent with a source within the first amplifier.   By proper choice of the value of $\mathcal{G}$, these sources can be made to nullify each other, and an improvement of noise results.   Sources of noise of this type are those due to partition of the electron stream between the screen grid and plate in a multigrid tube so that at one instant a fluctuation in the screen-grid current is accompanied by a fluctuation in the plate current of equal magnitude but opposite sign.   By introducing feedback between the screen grid and the input terminal of the tube, the fluctuation in the screen current can be made to cancel the effect of the fluctuation in the plate current.   This matter is discussed in Chap. 12 and briefly in Sec. 13·13.

Change of Input and Output Admittance Due to Feedback and Change in Available Power Gain Due to Feedback and Presence of Load.—Here the change of input admittance is calculated for a common situation shown

[1] See H. W. Bode *Network Analysis and Feedback Amplifier Design*, Van Nostrand, 1945, particularly pp. 33–34.

in Fig. 13·19. One input and one output terminal of the amplifier are grounded. Feedback is produced by an admittance Y_{12} connected between the ungrounded terminals. When Y_{12} is zero, Y_{11} and Y_{22} are the total input and output admittances, including the source and load admittances.
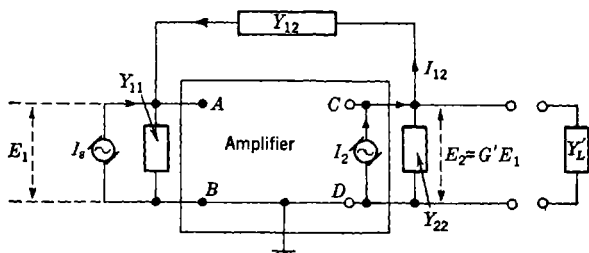
FIG. 13·19.—Amplifier with admittance connected between input and output terminals.

Take $Y_{11} = G_s + G_1 + jY_1 + Y_{10}$, where G_s is the transformed source conductance, Y_1 is the susceptance associated with the input circuit, G_1 is the input circuit loss, Y_{10} is the input admittance due to sources of feedback other than Y_{12} and includes the transit-time damping, unless, as in a grounded-plate circuit, this is part of Y_{12}.

It is assumed that Y_{11} is independent of Y_{22} (an assumption that is not valid if the amplifier consists of a grounded-grid amplifier whose load admittance Y_L is small compared with its dynamic plate conductance $1/r_p$). The amplifier may contain any number of stages. It is assumed here that at the center of the pass band, the phase angle between the input and output voltages is either zero, as in the case of a single-grounded grid or cathode-follower stage, or 180° as in the case of a single-grounded cathode amplifier. The total voltage across Y_{12} is then equal to $E_1(1 \pm \mathcal{G})$ where the upper sign pertains to a phase angle of 180° and the lower sign to a phase angle of zero. The current I_{12} flowing through Y_{12}, is then

$$I_{12} = -Y_{12}E_1(1 \pm \mathcal{G}). \tag{85}$$

The voltage gain of the amplifier from input to output is $\mathcal{G}$.

The input voltage E_1 is the same as though an admittance Y_F were connected in parallel with the input and the feedback through Y_{12} were disconnected, where

$$Y_F = -\frac{I_{12}}{E_1} = Y_{12}(1 \pm \mathcal{G}). \tag{86}$$

Let $\mathcal{G}$ be the voltage gain with the admittance Y_{12} present and α be the voltage gain with Y_{12} equal to zero. The connection between these two gains can be found by writing down the condition that the current I_2 flowing out of the amplifier should equal the current flowing in the load

plus the current flowing in the feedback network in both cases.   One gets

$$I_2 = \pm Y_{22} \mathcal{G} E_1 - E_1 Y_{12}(1 \pm \mathcal{G}) \tag{87}$$

and

$$I_2 = \pm Y_{22} \mathcal{G} E_1. \tag{88}$$

Equating these expressions, one gets

$$\mathcal{G} = \frac{\mathcal{G} \mp \dfrac{Y_{12}}{Y_{22}}}{1 + \dfrac{Y_{12}}{Y_{22}}}. \tag{89}$$

The first term in the numerator is due to signal coming through the tube, and the second term is due to signal transmitted from input to output terminals through Y_{12}.   Substituting the value of $\mathcal{G}$ into Eq. (86) yields

$$Y_F = \frac{Y_{12}(1 \pm \mathcal{G})}{1 + \dfrac{Y_{12}}{Y_{22}}}. \tag{90}$$

If Y_{12} is small compared with Y_{22}, the voltage gain is not altered by connecting the feedback admittance, which means $\mathcal{G}$ is very nearly equal to $\mathcal{G}$.

It can be therefore concluded that the output voltage is modified in two ways by the use of this type of feedback.   First, the loading on the signal source is changed; this changes the input voltage considerably and constitutes a relatively large effect.   The second effect—the change in voltage gain—is usually relatively small.   The change of loading changes the pass band of the input circuit.   If Y_F, averaged over the pass band, contains a positive conductance or a negative capacity, the bandwidth can be made to increase; but if it contains negative conductance or a positive capacitance, the bandwidth decreases.   If $Y_F = -Y_{11}$, the amplifier will oscillate.   The total input admittance Y'_{11} in the presence of feedback is given by

$$Y'_{11} = Y_{11} + Y_F = Y_{11} + \frac{Y_{12}(1 \pm \mathcal{G})}{1 + \dfrac{Y_{12}}{Y_{22}}}. \tag{91}$$

Output Admittance.—To find the output admittance, it is necessary to find the output voltage E_2 when a current source J is applied across the output terminals.   Let E_1 be the voltage across the input terminals of the amplifier and let I_2, the current flowing out of the amplifier, be related to the input and output voltages E_1 and E_2 by the linear equation

$$-I_2 = \pm G_m E_1 + g_p E_2,$$

where G_m is the generalized transconductance of the amplifier and g_p is the

generalized dynamic conductance of the amplifier.   It may be noted that

$$g_p + Y_L = Y_{22},$$

where Y_L is the load admittance.   If Kirchhoff's second law is applied
to the independent node pairs formed by the input and output terminals
respectively, one gets

$$J = Y_{12}(E_2 - E_1) + (Y_L + g_p)E_2 \pm G_m E_1,$$
$$0 = Y_{11}E_1 + Y_{12}(E_1 - E_2),$$

and solving for E_2 gives

$$J = \left[Y_{22} + \frac{Y_{12}(Y_{11} \pm G_m)}{Y_{11} + Y_{12}} \right] E_2,$$

so that the total output admittance Y'_{22} is given by

$$Y'_{22} = Y_{22} + \frac{Y_{12}(Y_{11} \pm G_m)}{Y_{11} + Y_{12}}. \tag{92}$$

Available Power Gain.—To find the available power gain when the
load Y_L is considered part of the feedback amplifier it is first necessary
to find the output voltage E_2 when a current I_s flows into the input
terminals of the amplifier and when an additional load Y'_L is connected.
If E_1 is the voltage across the input terminals, then application of Kirch-
hoff's second law to the two independent node pairs formed by the input
and output terminals gives

$$I_s = E_1 Y_{11} + (E_1 - E_2)Y_2,$$
$$0 = \pm G_m E_1 + g_p E_2 + (Y_L + Y_L)E_2 + (E_2 - E_1)Y_{12},$$

so that

$$I_s = \frac{-Y_{12}(\pm G_m - Y_{12}) - (Y_{11} + Y_{12})(g_p + Y_L + Y'_L + Y_{12})}{\pm G_m - Y_{12}} E_2$$
$$= \frac{Y_{12}(Y_{11} \pm G_m) + (Y_{22} + Y'_L)(Y_{11} + Y_{12})}{G_m \pm Y_{12}} E_2.$$

The power delivered to the load Y'_L is

$$E_2^2 G_L = \frac{G'_L |G_m \mp Y_{12})|^2 I_s^2}{|Y_{12}(Y_{11} \pm G_m) + (Y_{12} + Y'_L)(Y_{11} + Y_{12})|^2}.$$

If all the admittances in the circuit are conductances, this power is a
maximum when

$$Y'_L = G'_L = G_{22} + \frac{G_{12}(G_{11} \pm G_m)}{(G_{11} + G_{12})}. \tag{92a}$$

Because the available input power to the amplifier is $I_s^2/4G_s$, the
available power gain $\mathcal{W}$ is given by

$$\mathfrak{W} = \frac{G_s(G_m \mp G_{12})^2}{(G_{11} + G_{12})[G_{22}(G_{11} + G_{12}) + G_{12}(G_{11} \pm G_m)]}, \tag{93}$$

where G_{22}, G_{11}, G_{12}, G_s, are the conductive components of Y_{22}, Y_{11}, Y_{12}, Y_s respectively.

If there is no feedback of this type present, G_{12} is equal to zero and Expression (93) becomes

$$\mathfrak{W} = \frac{G_s G_m^2}{G_{22}(G_s + G_1)^2}. \tag{93a}$$

Equations (34a) and (34c), which give the available power gains of single-stage grounded-cathode and grounded-plate amplifiers with no load resistors connected, are special cases of Eq. (93a). In such cases, G_{22} consists of the output admittance of the tube.

.Equations (92a), (93), and (93a) do not apply accurately to a grounded-grid amplifier, because the input admittance of a grounded-grid amplifier, given by Eq. (22), depends upon the value of the load. At one extreme, when the dynamic plate conductance $1/r_p$ is large compared with the load admittance, the input admittance is approximately equal to $g_m' r_p Y_L$ or $(\mu + 1) Y_L$. At the other extreme (a situation that is frequently encountered in practice) $1/r_p$ is small compared with the load admittance Y_L, and the electronic component of the input admittance becomes constant and equal to g_m'. In this latter case the results of previous discussions can be made to apply. The output admittance is equal to the output admittance of the tube plus the admittance of the load; and because the available power of a constant-current generator is inversely proportional to its internal conductance, the available power gain $\mathfrak{W}$ with the load is equal to the available power gain without the load multiplied by the ratio of the tube's output conductance to the sum of the output conductance and the conductance of the load. The calculation of the output admittance and available power gain when the input admittance is not constant will not be given here but can be effected by evaluating the power output in terms of the input admittance and finding the value of auxiliary load that receives the maximum power.

13·12. Miscellaneous Types of Feedback and Their Effect on Noise Figure.—In this section two important cases of feedback are considered: (1) the grounded-cathode tube with resistance feedback from plate to grid and (2) the grounded-cathode tube with feedback provided by inductance in the cathode lead.

Feedback Produced by a Resistor between Input and Output.—A resistor between the grid and the plate of a grounded-cathode amplifier has been widely used as a means of increasing the over-all gain-bandwidth product. This subject was discussed at length in Chap. 6. It can be seen from Eq. (88) that if Y_{12} is a pure conductance G_{12}, the input admittance Y_F

due to feedback, consists of a conductance at the center frequency of the output circuit while at other frequencies it contains positive and negative susceptances when the load also contains positive and negative susceptive components.   This conductive component causes the bandwidth of the input circuit to increase and the voltage output of the amplifier to decrease because of the greater loading on the signal source.

Two possible applications of this feedback system are of interest.   It is possible to use resistor feedback to widen the pass band of the input network if the bandwidth of the latter is too narrow to give the required over-all amplifier bandwidth when the source conductance is chosen to give optimum over-all noise figure.   This possibility is discussed in greater detail in Sec. 13·14.   It is also possible to use resistor feedback in the first tube of the main amplifier chain so that the load into which the single- or double-triode input circuit works is at a very low effective temperature.   The distinction made

FIG. 13·20.—Pentode with plate-grid resistance feedback.

above between these two applications is not a fundamental one.   The theoretical results derived for one case apply also to the other.

The basic circuit with resistance feedback is given in Fig. 13·20.   It is assumed that a pentode stage is used so that grid-plate capacity feedback can be neglected.   If the circuit is driven from a source G_s, and if the effects of cathode-load inductance and finite transit angle are neglected, it has been shown[1] that the midband noise figure of the stage is given by

$$f = 1 + \frac{\rho_1 G_{B1}}{G_s} + \frac{G_{12}}{G_s} \frac{(g_m + G_s + G_{B1} + G_{12})^2}{(g_m - G_{12})^2} + \frac{\left(R_{eu} + \dfrac{G_L}{g_m^2}\right)}{G_s} \frac{(G_s + G_{B1} + G_{12})^2}{\left(1 - \dfrac{G_{12}}{g_m}\right)^2}, \quad (94)$$

where G_{B1} is the sum of the input circuit loss and the transit time conductance and G_L is the load conductance.

It can be seen from Eq. (94) that for values of G_{12} small compared with g_m, the noise figure at midband is the same as if G_{12} were connected across the input terminals.   When G_{12} approaches g_m, the noise figure approaches infinity, but by reference to Eq. (87) it may be seen that this

[1] A. B. Macnee, RL Internal Report 61–10/1/45.

corresponds to a value of G_{12} which gives zero gain, a situation of no practical interest.

All the formulas derived earlier in Sec. 13·11 can be applied to the amplifier of Fig. 13·20. From Eq. (90), G_F, the total input admittance at midband due to feedback, is given by

$$G_F = \frac{G_{12}(g_m + G'_L)}{G'_L + G_{12}}, \tag{95}$$

where $G'_L = G_L + (1/r_p)$. Because the tube is a pentode, $1/r_p$ can be neglected in comparison with G_L so that

$$G_F \approx \frac{G_{12}(g_m + G_L)}{G_L + G_{12}} \approx \frac{G_{12}g_m}{G_L}. \tag{95a}$$

The output admittance G'_{22} and available power gain $\mathcal{W}$, respectively, at midband are given, from Eqs. (92) and (93), by

$$G'_{22} = \frac{1}{r_p} + G_L + \frac{G_{12}(G_s + G_{B1} + g_m)}{G_s + G_{B1} + G_{12}} \approx G_L + \frac{(g_m + G_s)G_{12}}{G_s + G_{B1} + G_{12}}, \tag{96}$$

$$\mathcal{W} = \frac{G_s(g_m - G_{12})^2}{(G_s + G_{B1} + G_{12})\left(\dfrac{1}{r_p} + G_L\right)(G_s + G_{B1} + G_{12}) + G_{12}(G_s + G_{B1} + g_m)} \tag{97}$$

$$\approx \frac{G_s g_m^2}{(G_s + G_{B1} + G_{12})[G_L(G_s + G_{B1} + G_{12}) + G_{12}(g_m + G_s)]}$$
$$= \frac{G_s g_m^2}{(G_s + G_{B1} + G_{12})^2 G''_{22}}. \tag{97a}$$

The condition for transitional coupling is

$$2C_1C_2[g_m G_{12} + G_{11}G_{22} + G_{12}(G_{11} + G_{22})] = [G_{11}C_2 + G_{22}C_1 + G_{12}(C_1 + C_2)]^2, \tag{98}$$

while the 3-db bandwidth $\mathcal{B}$ of the negative feedback pair is given by

$$\mathcal{B} = \frac{[g_m G_{12} + G_{11}G_{22} + G_{12}(G_{11} + G_{22})]^{\frac{1}{2}}}{2\pi \sqrt{C_1 C_2}}, \tag{99}$$

where C_1 and C_2 are the total capacities to ground across the input and the output circuit respectively. In the important case when $C_1 = C_2$ and $G_L \gg G_{11}$, $1/r_p$, G_{12}, then the condition for transitional coupling becomes

$$2g_m G_{12} \approx G_L^2,$$
$$\mathcal{B} \approx \frac{1}{2\pi C} \sqrt{g_m G_{12}} = \frac{G_L}{2\pi C \sqrt{2}}. \tag{99a}$$

This last equation implies that the over-all 3-db bandwidth of the stage,

when negative feedback is applied, is the same as if there were no negative feedback but a conductance G_L connected across each tuned circuit. In this latter case the noise figure would have been

$$F = 1 + \frac{\rho_1 G_{B1}}{G_s} + \frac{G_L}{G_s} + \frac{R_{eq} + \dfrac{G_L}{g_m^2}}{G_s} (G_s + G_{B1} + G_L)^2;$$

and because $G_L = \sqrt{2 g_m G_{12}} \gg G_{12}$, the noise figure for the same 3-db bandwidth is considerably improved by the use of resistance feedback.

It can be shown that the noise figure of the stage, with resistance feedback, at frequencies other than midband is

$$F \approx 1 + \frac{\rho_1 G_{B1} + G_{12}}{G_s} + \frac{R_{eq} + \dfrac{G_L}{g_m^2}}{G_s} [(G_s + G_{B1})^2 + \omega_0^2 C_1^2 \alpha^2], \quad (94a)$$

where $\alpha = (\omega/\omega_0) - (\omega_0/\omega)$.

It may be seen from Eq. (94a) that although the pass band of the input circuit has been appreciably increased by feedback, the variation of noise figure with frequency is virtually unaltered by feedback. This result is in full agreement with the general arguments given in the center part of this section.

Feedback by Cathode-lead Inductance.—It has been shown that feedback due to cathode-lead inductance produces an input admittance, between grid and cathode terminals, that contains a conductive component.[1] If L_k is the cathode-lead inductance and if the reactance of this inductance is small compared with the other impedances in the circuit, particularly the grid-cathode reactance $1/\omega C_{kg}$, then this conductance G_c is given approximately by

$$G_c = \omega^2 L_k C_{kg} g_m,$$

where ω is the angular frequency. If grid-plate feedback is also present. it has been shown by Sturley[2] that the total input admittance due to the two effects is their sum.

This input conductance lowers the available power gain of the tube and hence increases the contribution of the second-stage noise to the noise figure in the same way as would a conductance G_c connected across the input terminals of the tube at effective noise temperature zero.

If the first tube is a triode, however, the cathode-lead inductance has no effect on the first-stage noise figure, although this is not true for a pentode. The equivalent noise representation for the triode case is shown in Fig. 13·21 where the equivalent shot-noise generator lies

[1] M. J. O. Strutt and A. van der Ziel, *Proc. I.R.E.*, **26**, 1011–1032 (1936).

[2] K. R. Sturley, *Radio Receiver Design*, Part I, Wiley, New York, 1943, pp. 50–53.

between plate and cathode. The cathode-lead inductance feeds back part of the shot-noise current in antiphase to the grid. It can be shown
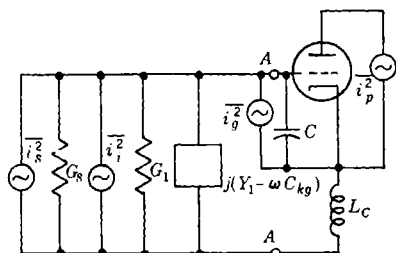
that the mean square current flowing from the shot-noise generator in the output lead that grounds the plate is reduced in the same ratio as the available power gain. The noise figure of the first stage is thus left unchanged. In the pentode case, however, the partition noise current that flows between screen and plate is unaffected by the presence of grid-cathode inductance; therefore, in a pentode, the noise figure is made worse by this feedback.

The mean square noise current flowing in an output short circuit, produced by the thermal-noise and induced-grid-noise generators in the input circuit is

$$\frac{g_m^2(G_s + \rho_1 G_{B1})}{(G_s + G_{B1} + G_c)^2 + Y_1^2} \, 4kTB,$$

where G_{B1} includes the transit-time loading but not G_c. In deriving this expression, it has been assumed that the grid-plate susceptance is zero.

It can be shown that the shot-noise constant-current generator produces a mean square output short-circuit current approximately equal to

$$\frac{g_m^2[(G_s + G_{B1})^2 + Y_1]}{(G_s + G_{B1} + G_c)^2 + Y_1^2} \, 4kE_{eq}TB.$$

Hence, the noise figure of the first stage above is

$$F_1 = 1 + \frac{\rho_1 G_{B1}}{G_s} + \frac{[(G_s + G_{B1})^2 + Y_1^2]R_{eq}}{G_s}, \tag{100}$$

which is independent of G_c.

The available power gain is now

$$\frac{g_m^2 r_p G_s}{(G_s + G_{B1} + G_c)^2 + Y_1^2}; \tag{101}$$

and since the output admittance is unchanged at $1/r_p$, the contribution of second-stage noise has been increased by the ratio

$$\frac{(G_s + G_{B1} + G_c)^2 + Y_1^2}{(G_s + G_{B1})^2 + Y_1^2}. \tag{102}$$

For some purposes it is convenient to regard G_c as part of G_{B1}, the

noise temperature of G_c being taken as zero. The formulas for the overall noise figure of an amplifier, derived earlier in this chapter, can then be used unchanged, provided that the equivalent shot-noise resistance of the triode is assumed to be

$$R_{eq} \frac{(G_s + G_{B1} - G_c)^2 + Y_1^2}{(G_s + G_{B1})^2 + Y_1^2}, \tag{103}$$

where G_{B1} is now assumed to contain G_c. The available power gain is

$$\frac{g_m^2 r_p G_s}{(G_s + G_{B1})^2 + Y_1^2}. \tag{104}$$

When the first tube is a pentode, the rigorous analysis is more complicated because the presence of cathode-lead inductance causes some of
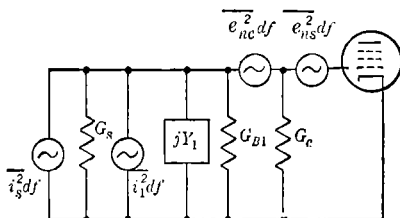
FIG. 13·22.—Equivalent noise representations of pentode with cathode-lead inductance

the screen shot-noise current to flow in the anode circuit. If this effect is neglected, however, as is justifiable for small values of cathode-lead inductance, the noise figure of the first stage assumes the form

$$F_1 = 1 + \frac{\rho G_{B1}}{G_s} + \frac{(G_s + G_{B1})^2 + Y_1^2}{G_s} \left\{ R_{ns} + \frac{R_{nc}[(G_s + G_{B1} - G_c)^2 + Y_1^2]}{(G_s + G_{B1})^2 + Y_1^2} \right\}, \tag{105}$$

where R_{nc}, R_{ns} are the equivalent shot- and partition-noise generators of the first tube. Their magnitudes are given by Eq. (17).

The effects of cathode lead inductance are most easily represented by the equivalent circuit of Fig. 13·22. The voltage generators $\overline{e_{nc}^2}$, $\overline{e_{ns}^2}$ are given by

$$\overline{e_{nc}^2} = 4kR_{nc}T,$$
$$\overline{e_{ns}^2} = 4kR_{ns}T. \tag{106}$$

If second-stage noise is important, its contribution can be allowed for by suitable increase of R_{ns}.

13·13. The Correlation between the Induced Grid-noise and the Shot-noise Currents.—Until now the correlation between induced grid-noise current and shot-noise current has been neglected, and it has been

assumed that these currents could be treated as if they were statistically independent. In this section the correlation is taken into account, and the possibility is considered of using the correlation to obtain improved noise figures at high frequencies, where induced grid noise is very important.

The discussion is based on the grounded-cathode triode circuit of Fig. 13·23. The analysis holds equally for the grounded-grid case as far as
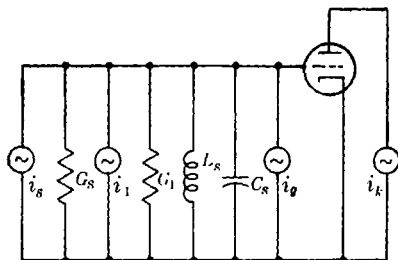
Fig. 13·23.—Equivalent noise representation for grounded-cathode triode with correlation between induced grid noise and shot noise.

first-stage noise figure is concerned, but not as regards available power gain.

In order to simplify the analysis it is assumed that the feedback susceptance between input and output terminals is neutralized over the effective amplifier pass band and that the input circuit is a single-tuned circuit. The effects of transit time on the magnitude of the equivalent shot-noise resistance are ignored as are the effects of cathode-lead inductance on the first-stage noise figure. The last assumption is justified by the discussion on this subject in Sec. 13·12.

Because of the approximate nature of the treatment and the still tentative theory on which it is based, it must be emphasized that this discussion can only be regarded as qualitative. The numerical results given are provided for illustration and have not been checked experimentally. Almost certainly they are too optimistic as to what can be done to cancel out induced grid noise at high frequencies.

The following symbols will be employed:

C_s = the total input capacity.
G_r = the transit-time damping conductance.
G_1 = the input coil losses.
G_s = the source conductance.
L_s = the input coil inductance.
Y_1 = the total input susceptance.
Y_s = the total input admittance equal to $G_1 + G_s + G_r + jY_1$.

For convenience G_r is regarded as noiseless, the induced grid current being associated directly with the shot current.

As stated in Sec. 13·3, the induced grid-noise current i_g is related to the cathode shot-noise current i_k by the equation

$$i_g = + j\omega \left(\frac{\tau}{3}\right) i_k$$

where ω is the angular frequency and τ the cathode-grid transit time. It has been assumed in deriving this formula that i_g and i_{pk} flow from the cathode to the grid or plate respectively. But the steady direct current in a tube flows from plate to cathode. Accordingly, if the grid-cathode voltage produced by i_g is i_g/Y_s, then this voltage produces a current $g_m i_g/Y_s$ from plate to cathode or a current $-g_m i_g/Y_s$ from cathode to plate.

Hence, the noise sources in the first stage produce a total noise current in the short-circuited output lead equal to

$$\frac{\left(i_s + i_1 - j\omega \left(\frac{\tau}{3}\right) i_k\right) g_m}{Y_s} + i_k.$$

The mean-square value of this current is

$$\frac{(\overline{i_s^2} + \overline{i_1^2}) g_m^2}{(Y_s)^2} + \frac{\overline{i_k^2}}{|Y_s|^2} \left| Y_s - g_m j\omega \frac{\tau}{3} \right|^2$$

and the noise figure F_1 of the first stage can be written

$$F_1 = 1 + \frac{\overline{i_1^2}}{\overline{i_s^2}} + \frac{\overline{i_k^2}}{i_s^2 g_m^2} \left| Y_s - g_m j\omega \frac{\tau}{3} \right|^2$$
$$= 1 + \frac{G_1}{G_s} + \frac{R_{eq}}{G_s} |(G_s + G_1 + G_r)^2| + \left(Y_1 - \sqrt{\frac{\beta G_r}{R_{eq}}} \right)^2, \quad (107)$$

from Eqs. (10), (15), (16), and (17). When $Y_1 = 0$, F_1 reduces to the familiar expression for the midband noise figure of the first stage:

$$F_1 = 1 + \frac{G_1}{G_s} + \frac{\beta G_r}{G_s} + \frac{R_{eq}}{G_s} (G_s + G_1 + G_r)^2; \quad (108)$$

when $Y_1 = \sqrt{\beta G_r/R_{eq}}$; however, F_1 is given by

$$F_1 = 1 + \frac{G_1}{G_s} + \frac{R_{eq}}{G_s} (G_s + G_1 + G_r)^2. \quad (109)$$

Therefore, by proper choice of the susceptance of the input network it is apparently possible to eliminate the induced grid-noise contribution, at least at a single frequency, and to obtain a noise figure determined by coil losses and shot noise alone.

This optimum value of Y_1 is directly proportional to the angular frequency ω since G_r/ω^2 is independent of frequency. Hence Y_1 should be the susceptance of a pure capacity C_r; that is, the input circuit should be capacitive at the midband frequency of the over-all amplifier. In the case of the type 6AK5, connected as a triode, where

$$R_{eq} = 385 \quad \text{ohms,}$$
$$G_r = 12 \quad \mu\text{ohms at 30 Mc/sec,}$$
$$\beta = 5,$$

this capacity comes out to be approximately 2.0 $\mu\mu$f.

In any practical case the input capacity C_s will be appreciably larger than 2.0 and the input circuit must then consist of a coil that resonates at midband frequency with $C_s - C_r$. Let $Y_i' = Y_1 - \sqrt{\beta G_r/R_{eq}}$; then F_1 can be written

$$F_1 = 1 + \frac{G_1}{G_s} + \frac{R_{eq}}{G_s}(G_s + G_1 + G_r)^2(1 + v^2) \tag{110}$$

where

$$v = \frac{\omega_0(C_s - C_r)}{G_s + G_1 + G_r}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right). \tag{111}$$

If $G_{s,\text{opt}} \gg G_1 + G_r$, the optimum source conductance when $v = 0$ is given approximately by

$$G_{s,\text{opt}} = \sqrt{\frac{G_1}{R_{eq}}}$$

and

$$F_1 \approx 1 + 2\sqrt{G_1 R_{eq}},$$

whereas in the case $Y_1 = 0$, the conventional midband case,

$$G_{s,\text{opt}} = \sqrt{\frac{G_1 + \beta G_r}{R_{eq}}}$$

and

$$F_1 \approx 1 + 2\sqrt{(G_1 + \beta G_r)R_{eq}}.$$

Thus in cases where βG_r is of the same order as or larger than G_1 an appreciable improvement in noise figure can, at least ideally, be obtained. To achieve this improvement a certain price has to be paid. The transfer characteristics of the input circuit are now asymmetrical; and if the amplifier is fed from a source such as a crystal, whose conductance can vary within appreciable limits, it is difficult to correct for this asymmetry later in the amplifier. It might be suspected that the mistuning would produce a loss in the available power gain of the stage. This is not the case, however, at least for the grounded cathode, as the decrease in $G_{s,\text{opt}}$ due to the improved noise figure more than compensates for the effect of mistuning, the ratio of the available power gains in the resonant

and mistuned cases respectively being

$$\frac{\left(G_c + \sqrt{\dfrac{G_1}{R_{eq}}}\right)^2 + \dfrac{\beta G_r}{R_{eq}}}{\left(G_c + \sqrt{\dfrac{G_1 + \beta G_r}{R_{eq}}}\right)^2},$$

which is less than 1, where G_c is the input conductance due to cathode-lead inductance.   The difference in the available power gains, however, is not large.

The mistuning discussed above is likely to be really worth while only at high frequencies where the induced grid noise is important and the asymmetry produced by mistuning relatively small.   To illustrate the magnitude of the improvement obtained by mistuning consider the case of the type 6AK5 triode-connected, when the midband frequency is 180 Mc/sec.

If the input coil magnification is 150 and the total input capacity 14 $\mu\mu$f, then

$$\begin{aligned} G_1 &= 90 \qquad \mu\text{mhos when the circuit is mistuned,} \\ &= 100 \qquad \mu\text{mhos when the circuit is midband resonant,} \end{aligned}$$

and

$$G_r = 400 \qquad \mu\text{mhos.}$$

When the circuit is midband resonant, the optimum source conductance is

$$G_{s,opt} = 2{,}340 \qquad \mu\text{mhos}$$

and the theoretical noise figure is

$$F_1 = 3.2 = 5.0 \qquad \text{db.}$$

When the circuit is mistuned, the optimum source conductance is

$$G_{s,opt} = 690 \qquad \mu\text{mhos}$$

and the theoretical noise figure is

$$F_1 = 1.9 = 2.75 \qquad \text{db.}$$

Thus it is apparently possible to improve the single-frequency noise figure of the first stage by 2.25 db by mistuning the input circuit 15 Mc/sec lower than band center.   As stated already in this section this result is probably too optimistic, but an appreciable improvement in noise figure should be obtainable, and rough experimental evidence exists of this trend.

The variation of noise figure with frequency may be obtained from Eq. (110).   In the present case the noise figure for the mistuned amplifier has deteriorated 0.5 db at frequencies ± 4 Mc/sec from the midband.

Throughout this chapter, it has been assumed that induced grid-noise and shot-noise currents can be regarded as statistically independent if the input circuit is midband resonant and the pass band of the amplifier geometrically symmetrical about midband frequency. The validity of this assumption will now be established.

Equation (107) can be written in the form

$$F_1 = 1 + \frac{G_1 + \beta G_r}{G_s} + \frac{R_{eq}}{G_s}\left[(G_s + G_1 + G_r)^2 + Y_1^2 - \frac{2Y_1}{G_s} G R_{eq}\right]. \quad (112)$$

If Y_1 is antisymmetrical about midband frequency, the average noise figure at frequencies geometrically disposed about the midband is equal to

$$1 + \frac{G_1 + \beta G_r}{G_s} + \frac{R_{eq}}{G_s}[(G_s + G_1 + G_r)^2 + Y_1^2]$$

if the variation of G_r over the frequency band is neglected. But this is just the result that would have been obtained if the correlation between the two noise currents had been ignored.

13·14. Input Coupling Networks.—The purpose of the input coupling network is to transform the conductance of the signal source to a value that gives the optimum noise figure consistent with the bandpass and pulse-response characteristics required of the amplifier.

The input networks discussed in this section are of two general types: (1) the single-tuned circuit, usually realized as autotransformers, shown in Fig. 13·24a, b; (2) the double-tuned inductively coupled circuit, which may be realized in any one of the equivalent forms of Fig. 13·25a, b, c.

In general the single-tuned circuit is used whenever it permits the bandwidth requirements to be met with the source impedance transformed to give optimum noise figure. This circuit employs the fewest number of parts, occupies the least space, and is easy to align. Changes of source conductance produce bandwidth changes no more serious than those produced in any alternative input network, and the effect on pulse response of errors in tuning is very small.

If the input network bandwidth is too narrow when the source conductance is transformed for optimum noise figure, two courses of action are open to the designer. Either he can retain the single-tuned circuit and increase its bandwidth by lowering the source conductance or adding damping by one of the methods discussed in Sec. 13·12, or he can employ a double-tuned input circuit. Such a circuit has approximately double the bandwidth for the same transformed source conductance as the single-tuned circuit,[1] and it has increased selectivity, which may prove valuable when image rejection is important. The double-tuned circuit

[1] When the double-tuned circuit is transitionally coupled. If the circuit is under-coupled, this improvement decreases rapidly.

has the disadvantages of increased size and complexity and greater dependence on circuit constants and tuning errors, particularly the latter.

The choice of input network to meet the design requirements of any particular wide-band amplifier thus necessarily involves some compromise and is based on a knowledge of the bandpass, output admittance, tuning, and distortion characteristics of the alternative networks. It is the purpose of this section to provide this information in a suitable form. Extensive use is made of the results of Chaps. 4 and 5, where the theory of single- and double-tuned circuits is discussed.

Before the properties of the networks of Fig. 13·24 and 13·25 are considered in detail, the statement made in Sec. 13·3 will be justified. There it was indicated that the circuit of Fig. 13·2 provides an equivalent network for the transfer and output driving-point admittances of any type of input coupling.

By Thévenin's theorem any two-terminal network containing both active and passive elements can be represented as a single-current generator J and an admittance Y_s in parallel with it, where Y_s is composed of a conductance G and a susceptance Y_1. In the particular case discussed in this chapter J equals the sum of the signal current I_s and the thermal-noise current whose mean square fluctuation i^2 is

$$i^2 = 4kT_1GB,$$

where T_1 is the effective noise temperature of G.

If the network has an available power gain $\mathfrak{W}$, which will be unity when the network is dissipationless, then it is possible to split G into two parts, G_s and G_1, where G_s represents the transformed source conductance and G_1 the network losses, such that

$$\mathfrak{W} = \frac{G_s}{G_s + G_1},$$
$$G = G_s + G_1.$$

The noise current i can be regarded as composed of two statistically independent noise currents i_s and i_1 whose mean square values are given by

$$(I_s + I_1)^2 = 4kT_1GB,$$
$$I^2 = 4kTG_sB,$$
$$I_1^2 = 4k\alpha TG_1B,$$

connected across the input terminals of the tube, where αT is the noise temperature of G_1, which is the network of Fig. 13·2, as required.

One important point must be made here if confusion is to be avoided. In discussing the noise figure of the first stage, reference to Eq. (53) shows

that it is necessary to know the admittance presented to the terminals of the first tube by the external circuit. The input tube conductance due to feedback (whether produced by cathode-lead inductance or grid-plate resistance as in the grounded-cathode tube or by direct cathode feedback as in the grounded-grid tube) does not appear in the expression for first-stage noise figure, although this conductance affects the available power gain and hence the second-stage noise figure. In calculating the transfer admittance or pulse response of the input network, however, this conductance must be included. If the input tube is a grounded-grid triode,
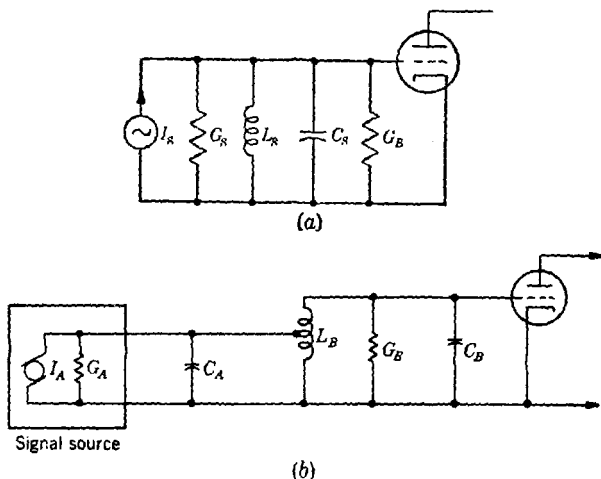
FIG. 13·24.—Single-tuned input network; (a) simple form, (b) autotransformer.

this conductance is so large that the bandwidth of the input network is wide enough for all practical amplifiers. Hence when the various means for increasing the pass band of the input circuit are discussed, attention is confined to the grounded-cathode case.

The two specific input networks with which this section is concerned will now be considered, the simpler case of the single-tuned circuit being dealt with first. In both cases it is assumed that the source admittance consists of a conductance G_A in parallel with a capacity C_A and that the input capacity of the tube, together with stray wiring and tube socket capacities, is C_B. Unless otherwise stated it is assumed that the input conductance of the tube due to feedback is negligible in comparison with the output conductance of the input network.

The Single-tuned Circuit.—The simplest network of this type is shown in Fig. 13·24a. Here C_s is equal to $C_A + C_B$, and this capacitance is resonated by L_s at midband.

It is only by rare coincidence that G_A is equal to the value of source

conductance that gives an optimum noise figure. If the source is an antenna or a crystal, the most important case, G_A is generally very much smaller than this value[1] so that some means of transforming G_A must be employed. The most commonly used method for achieving this is illustrated in Fig. 13·24b, where the source is tapped down the input coil so that L_s becomes an autotransformer. The conductance G_B is the sum of coil loss and transit-time damping, and L_A is the part of L_s between the tap and ground.

The only case that is considered is the one where the susceptance of L_A is much greater than the source admittance. Under these circumstances the equivalent circuit can be put in the form of Fig. 13·24a, where L_s resonates with C_s at the midband frequency and

$$G_s = \frac{K^2 L_A}{L_s} G_A,$$

$$C_s = C_B + C_A \frac{K^2 L_A}{L_s} = C_B + \frac{C_A G_s}{G_A}. \tag{113}$$

From Eq. (53) the single-frequency noise figure of the first stage is

$$F = 1 + \frac{G_B}{G_s} + \frac{R_{eq}}{G_s} \left[(G_B + G_s)^2 + \omega_0^2 C_s^2 \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right)^2 \right], \tag{114}$$

where G_s is independent of frequency.

In the discussion of optimum source conductance given in Sec. 13·5 attention was directed primarily to the midband noise figure. At frequencies other than midband it may be seen from Eq. (55) that the optimum single-frequency source conductance is greater than the midband value, so that the source conductance that gives the optimum noise figure over the entire pass band is greater than the conductance that gives optimum midband noise figure.

It has been pointed out in Sec. 13·5 that in many practical cases the source conductance G_A varies between wide limits—between mG_A and G_A/m—from amplifier to amplifier. If bandwidth considerations permit, it is desirable to choose the position of the tap on the input circuit so that the optimum source conductance $G_{s,opt}$ is presented to the input terminals when the source conductance assumes its mean value G_A, because by these means the variation of noise figure with source conductance is held to a minimum. The bandwidth of the input circuit varies directly with the source conductance, and its minimum value is

$$\mathcal{B}_{min} = \frac{G_{s,opt}/m + G_B + G_F}{2\pi C_s}, \tag{115}$$

[1] Occasions can arise when G_A is very much larger than the optimum value, for example, when the stage whose noise figure is to be minimized is driven from a pentode or grounded-grid stage. This case has already been considered in Sec. 13·6 and is not dealt with further here.

where $\mathfrak{G}$ is the 3-db bandwidth, and G_F is the input conductance of the tube due to cathode-lead inductance in the grounded-cathode case or to direct cathode feedback in the grounded-grid case.

If $\mathfrak{G}_{min}$ is such that the over-all amplifier bandwidth is adequate, the design of the input circuit is complete. If, however, the bandwidth is too narrow, a number of alternative procedures can be adopted, as follows.
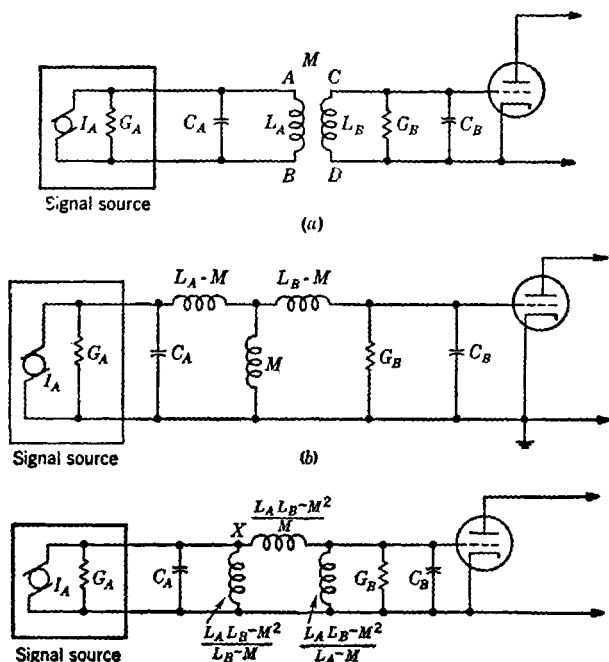
FIG. 13·25.—Double-tuned inductively coupled input network; (a) mutual-inductance-coupled circuit, (b) self-inductance-coupled T-circuit, (c) self-inductance-coupled π circuit.

Double-tuned-circuit Input Network.—The method most widely employed in the MIT Radiation Laboratory to obtain increased bandwidths in the input circuit was the use of a double-tuned inductance-coupled input network, which may be realized in any one of the equivalent forms of Fig. 13·25a, b, c. This network has a gain-bandwidth factor roughly double that of the single-tuned circuit. The capacity-coupled network is not used as a wide-band input network because it is as complicated as the inductance-coupled case while its gain-bandwidth factor is appreciably lower.[1]

Despite the fact that, in general, the primary Q is of the order of unity, the analysis given here will be based upon the "high-Q" theory of

[1] See Sec. 5·2 for a discussion of this point.

Sec. 5·2. The results are accordingly only approximate, but the general conclusions can be accepted as valid for all practical cases. This conclusion is justified by the fact that the high-Q theory gives correct values for gain-bandwidth facts and midband transfer impedance even when the fractional bandwidth is as high as 1.5, as is shown in Sec. 5·5.

It is convenient to assume that G_A, the primary conductance, is equal to the sum of the loss conductance of the primary coil L_A and the source conductance G_A'. In general the primary coil loss is negligible in comparison with G_A'. When this is not so, it can be allowed for by taking the over-all noise figure of the amplifier to be

$$\frac{G_A}{G_A'} F, \tag{116}$$

when F is the noise figure of the amplifier driven by a source conductance G_A.

Bandwidth and Transformed Source Conductance.—If G_s is the transformed source conductance, then from Eq. (5·42)

$$G_s = \frac{k^2 Q_A^2 \left(\dfrac{C_B}{C_A}\right)}{(1 + Q_A^2 \alpha^2)} G_A \tag{117}$$

where

$$\alpha = \left(\frac{\omega}{\omega_0}\right) - \left(\frac{\omega_0}{\omega}\right),$$
$$Q_A = \frac{2\pi f_0 C_A}{G_A}. \tag{118}$$

If $Q_B = 2\pi f_0 C_B / G_B \ll Q_A$, the transfer impedance of the network assumes the form of Eq. (5·36), the so-called one-side-loaded case,

$$|Z_{12}| = \frac{k}{2\pi f_0 \sqrt{C_A C_B} \left(k^4 + \alpha^2 \dfrac{1}{Q_A^2 - 2k} + \alpha^4 \right)^{\frac{1}{2}}}, \tag{119}$$

where $k = M/\sqrt{L_A L_B}$.

If the input network is transitionally coupled

$$k^2 Q_A^2 = \tfrac{1}{2}, \tag{120}$$

and so $\mathfrak{B}$, the bandwidth between half-power points, is

$$\mathfrak{B} = k f_0 \approx \frac{G_A}{2\pi C_A \sqrt{2}}. \tag{121}$$

But the transformed midband source conductance is

$$G_{s0} = \frac{k^2 Q_A^2 C_B G_A}{C_A} = \frac{C_B}{2C_A} G_A; \tag{122}$$

hence the 3-db bandwidth $\mathcal{B}$ can be written

$$\mathcal{B} = \frac{\sqrt{2}\,G_{s0}}{2\pi C_B}. \tag{123}$$

Referring back to Eq. (115) one sees that when $G_F + G_B \ll G_{s0}$ [a situation that has been assumed in deriving Eq. (120)] the 3-db bandwidth for the single-tuned circuit is

$$\mathcal{B} = \frac{G_{s0}}{2\pi C_B + \dfrac{G_s C_A}{G_A}}$$

when $m = 1$.

Even if $G_s C_A / G_A \ll C_B$ the double-tuned circuit has a 3-db bandwidth $\sqrt{2}$ times that of the single-tuned circuit. In general, however, it is not wise to let the input circuit bandwidth, at 1 db down, be narrower than the over-all bandwidth. If the bandwidths at 1 db down are compared, the transitionally coupled double-tuned circuit is twice as good as the single-tuned circuit, since the 1-db bandwidth of the double-tuned circuit is

$$\frac{G_{s0}}{2\pi C_B},$$

and the 1-db bandwidth of the single-tuned circuit is

$$\frac{G_{s0}}{2\pi 2\left(C_B + \dfrac{G_{s0} C_A}{G_A}\right)}.$$

This advantage becomes more pronounced as the ratio of input circuit bandwidth to over-all bandwidth increases.

In making this comparison it has been assumed that when the circuit is transitionally coupled, the transformed source conductance is optimum. It may be seen from Eq. (122) that this is true only if the ratio of secondary to primary capacity is correct. In general this is not the case, and it is necessary to increase either C_B or C_A and choose k so that the coupling remains transitional. The more important case, which will be considered here, is when C_A has to be increased. This procedure has two obvious disadvantages. An extra component has to be added, and the primary coil loss conductance is increased, thus increasing the noise figure of the first stage.

An alternative method for decreasing the transformed source conductance to its optimum value is to decrease k below transitional and leave the capacities unchanged. For fixed G_A and C_B, this optimum value of k is given by

$$k^2 = \frac{C_A G_{s0}}{Q_A^2 C_B G_A}. \tag{124}$$

The chief drawback to this method is that for given G_{s0}, the bandwidth is appreciably less than when C_A is increased and the coupling kept transitional.   In the limiting case, when the coupling is well below transitional, the shape of the response curve of the input circuit is approximately that of a single-tuned circuit with 3-db bandwidth

$$\mathfrak{B} = f_0 k^2 Q_A. \tag{125}$$

From Eq. (124), $\mathfrak{B}$ can be written

$$\mathfrak{B} = \frac{f_0 G_{s0} C_A}{C_B G_A Q_A} = \frac{G_{s0},}{2\pi f_0 C_B},$$

which is approximately equal to the 3-db bandwidth of the single-tuned circuit input when

$$C_B \gg \frac{G_{s0} C_A}{G_A},$$

so that most of the advantages of the double-tuned circuit is lost when coupling is below transitional.

Effects of Variation of Circuit Parameters.—The effect of variations in the coupling coefficient on the magnitude of the transformed source conductance can be deduced from Eq. (117); the conductance G_s varies as the square of the coefficient of coupling.   The variation in the shape and bandwidth of the amplitude characteristic is given in Fig. 5·4, where the absolute value of the transfer impedance is plotted as a function of k.

The coupling coefficient can be held to a 10 per cent tolerance fairly easily, but variations in source conductance G_A may be much greater, as much as 3 or 4 to 1, as pointed out in Sec. 13·5; hence the variation in source conductance is usually much more important than the variation in coupling coefficient.

The midband transformed source conductance can be written as

$$G_{s0} = \frac{k^2 (2\pi f_0)^2 C_A C_B}{G_A}$$

from Eq. (117).   In contrast to the single-tuned circuit case, G_{s0} varies *inversely* with G_A.   The variation, however, is still linear, so that a 4 to 1 change in G_A produces a 4 to 1 change in G_{s0}.   The variation in the shape and bandwidth of the amplitude characteristic of the input network with change of Q is given in Fig. 5·5.   It will be noticed that as G_A is *increased,* the bandwidth *decreases,* the shape of the response tending to that of a single-tuned circuit.

The variation in bandwidth is not linear, however, as was the case in the single-tuned circuit.   It now varies more rapidly than $1/G_A$.   In fact as G_A is increased, the bandwidth tends, in the limit, to the bandwidth of a single-tuned circuit input with the same value of transformed

source conductance. This result means that the advantage, in bandwidth, of the double-tuned circuit vanishes when the variation in source conductance gets very large.

Transient-response considerations normally impose a limit on the maximum height of the peaks of the amplitude characteristic of the input network. In one typical design the height of the peaks were limited to 1 db above response at midband; in another design it was required that the input network have 0 db peaks, that is, be transitionally coupled, when G_A assumed its minimum value.

It may thus be seen that there are three requirements on the input network for optimum performance:

1. When G_A assumes its minimum value, the ratio of peak to midband response must not exceed some arbitrary level, for example, 1 db.
2. When G_A assumes its maximum value, the bandwidth at some given level, say 1 db, must be greater than a certain amount.
3. When G_A assumes its geometric mean value, the optimum source conductance should be presented to the input terminals of the tube.

An example of the design of an input network on this basis is given in Sec. 13·15.

Noise Figure of First Stage with Double-tuned Input Circuit.—The noise figure of the first stage when a coupled-circuit input network is used is given by

$$F_1 = 1 + \frac{\rho G_{B1}(1 + Q_A^2 \alpha^2)}{G_{s0}} + \frac{R_{eq}}{G_{s0}} \frac{1}{|Z_{12}|^2} \tag{126}$$

where Z_{12} is the transfer impedance of the input network. This result follows from Eq. (117), which gives the transformer source conductance, and from Eq. (5·30), which relates the driving-point admittance Y_{11} of the network to the transfer impedance.

It may be seen from Eq. (126) that thermal noise associated with the secondary coil loss and the induced grid noise make contributions to the single-frequency noise figure that increase toward the edge of the pass band. If the edges of the pass band correspond to the 1-db points of the amplitude response of the input network, however, this effect is small.

Alternative Methods of Increasing the Bandwidth of the Input Circuit.— To get still wider bandwidths one of a number of methods can be used, and these will be considered very briefly.

1. *Increase in G_s.* One widely used method is to increase G_s, the transformed source conductance. In the single-tuned-circuit case this is done by moving the tap nearer to the top of the coil. The advantage of this method is that the curve of noise figure against transformed source conductance is very flat in the neighborhood of optimum source conductance and an increase of 2 to 1 above

optimum in G_s changes the noise figure only slightly, as may be seen from Fig. 13·7.

If G_A varies between wide limits, however, the noise figure when G_A is a maximum will now be appreciably worse than when G_A is a minimum. This wide variation of noise figure between one amplifier and another may prove undesirable. Another point is that the effects of source capacitance become more pronounced as the tap is moved further up the coil. From Eq. (115) it may be seen that the maximum bandwidth attainable by changing this tap position is given by

$$\mathbb{B} = \frac{G_A}{2\pi C_A},$$

and this sets a definite limit to what can be done by this procedure.

The bandwidth of the double-tuned circuit may also be increased by increasing G_s, either by increasing the coefficient of coupling or by decreasing C_A, if possible.

2. *Increase in the Secondary Loading by Adding Physical Resistance.*— This procedure should be followed only as an absolutely last resort, because the increase in noise figure is very large if a considerable increase in the bandwidth is to be obtained.

3. *Increase in the Cathode-lead Inductance of the First Tube.*—As pointed out in Sec. 13·12 the only effect of cathode-lead inductance is to reduce the available power gain of the first stage. The noise figure, if the tube is a triode, is unchanged. In Sec. 13·10 it was shown that the available power gain of the grounded-cathode-triode–grounded-grid-triode combination is so high that third-stage noise contribution is, in general, small. Accordingly this method of increasing bandwidth is probably the one that affects the noise figure least. The disadvantages are that the variation of input conductance, and hence of bandwidth, from tube to tube is likely to be large and that an extra and not easily controlled component has been added.

This method of increasing the bandwidth is also practicable when the input circuit is double-tuned, but it is not nearly so effective. As the secondary loading is increased, the Q's of primary and secondary tend to equality. The gain-bandwidth factor of an equal-Q coupled circuit is $1/\sqrt{2}$ that of an unequal-Q coupling when the coefficient of coupling is chosen to be transitional in each case, so that the proportional increase in bandwidth is less than in the single-tuned-circuit case.

4. *Increase in Loading of Input Circuit by Resistance Feedback.*—In Sec. 13·12 it was pointed out that resistance feedback could be

used to provide loading at a low effective noise temperature. In the discussion there given, the resistor was placed between grid and plate of a pentode tube. If the input configuration consists of a grounded-cathode triode followed by a grounded-grid triode the resistance feedback will be from the plate of the grounded-grid tube to the grid of the grounded-cathode tube, as shown in Fig. 13·26. The expressions for input conductance, bandwidth, etc., are the same as those given in Sec. 13·12 if g_m is taken to be the transconductance of the first tube.

The noise figure of the first stage is approximately that which would result if the feedback resistor were connected across the input terminals.
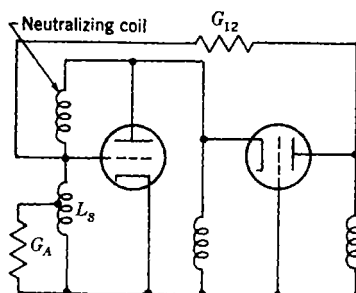
FIG. 13·26.—Feedback circuit for increasing bandwidth of input circuit with negligible deterioration of noise figure.

13·15. Example of Alternative Designs of Input Coupling Network.— The problem considered in this section is to design an input network to a 30 Mc/sec amplifier with a grounded-cathode triode (type 6AK5) as the first tube, when the source is a crystal the conductance of which varies between 5000 and 1330 μmhos. It will be assumed that the minimum stray capacity of the source is 5 $\mu\mu$f, that the total tube input capacity is 10 $\mu\mu$f, including Miller effect capacity resulting from the fact that $C_{gp} = 1.2$ $\mu\mu$f, and that all the coils have Q's of 150. The designer is at liberty to choose any input network that he pleases subject to the requirement that the amplitude response of the network have no peaks and that the bandwidth at 1 db be at least 5 Mc/sec.

For the triode-connected 6AK5 at 30 Mc/sec,

$$R_{eq} = 385 \quad \text{ohms,}$$
$$G_\tau = 12 \quad \mu\text{mhos,}$$
$$G_1 = \frac{2\pi f_0 C_B}{Q} = 12 \quad \mu\text{mhos.}$$

Under these circumstances the optimum transformed source conductance at midband is

$$G_{s,\text{opt}} = \frac{\rho G_r + G_1}{R_{\text{eq}}}$$
$$= 430 \quad \mu\text{mhos}$$

and the optimum value of the midband noise figure is

$$F_{1,\text{opt}} = 1 + 2\sqrt{(\rho G_r + G_1)R_{\text{eq}}} = 1.33 \qquad (1.3 \text{ db}).$$

It will be noticed that in this case, the induced grid-noise contribution is of far greater importance than the thermal noise of the coil loss. The effect of the source stray capacity on the noise figure is negligible. A number of alternative designs will now be given based on the discussion of Sec. 13·14.

Single-tuned Input Circuit.—As stated in Sec. 13·14 it is desirable that the optimum transformed conductance be presented to the tube when the source inductance assumes its mean value. But the optimum transformed conductance is 430 μmhos, and the mean value of the source conductance is 2540 μmhos. Hence the admittance stepdown of the input circuit must be

$$\frac{430}{2540} = \frac{1}{5.3}.$$

Accordingly the 1-db bandwidth of the input circuit will vary from 1.8 to 5.4 Mc/sec, while the noise figure at midband will have a maximum value of 1.4 db, an increase of 0.1 db over the optimum.

Before discussing the various alternative methods for increasing the bandwidth by electronic loading, the design of a double-tuned input circuit will be considered.

Double-tuned Input Circuit.—In this case the optimum transformed source conductance can be achieved either by increasing C_A or by decreasing k. These two alternatives will be considered in order.

If the input circuit is to be transitionally coupled when the source conductance assumes its minimum value and if the transformed source conductance is to be an optimum when the source conductance assumes its mean value, then the following two equations must be satisfied:

$$kQ_A = \frac{1}{\sqrt{2}\,G_{A,\text{min}}},$$
$$G_{s,\text{opt}} = k^2 Q_A^2 \left(\frac{C_B}{C_A}\right) G_{A,\text{mean}}.$$

If these equations are solved for C_A and k, the result is

$$C_A = 9 \quad \mu\mu\text{f},$$
$$k = 0.31.$$

The bandwidth at 1 db varies from

$$1.92 \text{ to } 10.8 \text{ Mc/sec},$$

a variation of $4\sqrt{2}$ to 1 as opposed to 3 to 1 in the single-tuned case. Because of the large variation in source conductance the advantage of this circuit over the single-tuned circuit is very small.

If no capacity is added to the primary, then the value of k that gives optimum transformed source conductance when G_A assumes its mean value is

$$k^2 = \frac{G_{s,\text{opt}} C_A}{C_B G_{A,\text{mean}}} \frac{1}{Q_A^2},$$

hence

$$k = 0.416.$$

In this case the bandwidth varies between

$$1.9 \text{ and } 5.0 \text{ Mc/sec},$$

so that the minimum bandwidth is the same as when C_A was increased; the bandwidth variation has, however, been cut in half.

Neither of these circuits provides adequate bandwidth, so that one of the other methods of Sec. 13·14 must be employed.

Increasing the Transformed Source Conductance.—If the tap position is so chosen that the bandwidth at 1 db is 5 Mc/sec when the source conductance assumes its minimum value of 1330 μmhos, then the admittance stepdown ratio is 1.59/1 and the transformed source conductance will vary from 3140 to 1040 μmhos.

The corresponding noise figure will vary from 3.46 to 1.65 db, and the bandwidth from 15 to 5 Mc/sec at 1 db.

Adding Cathode-lead Inductance.—If cathode-lead inductance is added to bring the minimum bandwidth up to 5 Mc/sec at 1 db, the noise figure of the first stage alone will be unchanged, while the bandwidth will vary from

$$8.6 \text{ to } 5 \text{ Mc/sec at 1 db}.$$

This greatly reduced variation is often of considerable advantage in designing wide-band amplifiers, as it enables compensating circuits to be used later in the amplifier.

Loading with Resistive Negative Feedback.—If it is assumed that the interstage capacity is the same as the input capacity, then the conductance of the feedback resistor turns out to be approximately equal to 100 μmhos. The optimum source conductance is now 620 μmhos, and the optimum noise figure is 1.8 db. The noise figure at the extremities of the band is 2.0 db. The variation of bandwidth is about the same as in the case where loading was provided by the cathode-lead inductance.

CHAPTER 14

MEASUREMENT OF NOISE FIGURE

By Yardley Beers

14·1. Introduction.—The measurement of noise figure affords an effective method of checking the ability of an amplifier to detect weak signals. The method of making such measurements is very simple and can be carried out quickly and precisely if the proper apparatus is available; such apparatus is described later in this chapter.

The measurement of noise figure is based upon one of its subsidiary definitions given in Chap. 12. In this definition it is assumed that all of the actual noise sources within the amplifier are replaced by a single source of noise connected to the input terminals and having an internal impedance equal to that of the signal source. The available power (a quantity that will be discussed in Sec. 14·2) of this fictitious noise source W_n is by definition such that this hypothetical amplifier has the same noise output power as the actual amplifier. If such an experiment were performed with an ideal amplifier, the only source of noise would be the thermal-agitation noise of the signal source, whose available power is kTB, where k is Boltzmann's constant, 1.38×10^{-23} joule/degree C, T is the temperature in degrees Kelvin, and B is the noise bandwidth defined below. Then the noise figure F is defined as

$$F = \frac{W_n}{kTB}. \tag{1}$$

Numerically $kTB = 4.00 \times 10^{-21}$ watt for $T = 290°K$ and $B = 1$ cps.

The definition of noise bandwidth B was given in Chap. 12 and is repeated here:

$$B = \int_0^\infty \frac{G_p \, dF}{G_{p0}},$$

where G_p is the actual power gain, defined as the power delivered to the indicating device (such as a meter, cathode-ray tube, or loud-speaker) divided by the available power of the signal source, and G_{p0} is the maximum value of G_p. This definition is tantamount to replacing the actual power-frequency curve of the amplifier by a rectangular curve having the same area and a height equal to G_{p0}. The width of this fictitious rectangle is equal to B. In almost all practical situations B is very nearly

695

equal to the bandwidth between half-power points.[1]   A notable exception occurs when the pass band is that of a single single-tuned circuit, in which case B is $\pi/2$, or 1.57, times the half-power bandwidth.

Noise figure is measured by the use of a signal generator that generates either noise or an unmodulated c-w signal.   If the amplifier is to be used with signals that consist of pulses or some other special form of modulation, it is valuable to supplement noise-figure measurements by determinations of minimum detectable signal making use of signals of the type for which the amplifier is intended.   Such measurements involve subjective factors that prevent them from being made as easily or with as high precision as measurements of noise figure.   However, they do give a valid check upon the amplifier, especially in regard to its pulse response.

In the following sections, it will be assumed that the signal-source impedance can be represented by a single resistance in parallel or in series with a single reactance.   In cases where this assumption does not hold, that is, where the signal-source impedance is a complicated function of the frequency, it is necessary to place in the amplifier a variable-frequency bandpass filter so narrow that this assumption is valid and then to measure what E. J. Schremp[2] calls the "single-frequency noise figure" as a function of frequency, and finally to average these values, weighted with respect to the gain, to give the complete noise figure.

The general procedure for measuring noise figure is (1) to measure W_n [see Eq. (1)] and (2) to determine the bandwidth B.   To measure W_n a signal generator of the same internal impedance as that of the amplifier source is connected to the input terminals of the amplifier. With the generator turned off the available noise-power output of the amplifier is observed, and then the generator is turned on and adjusted to double the output power.   Since the output power has been doubled, the available power that has been added must be equal to that previously present, that is, to W_n.   (It is assumed here that the amplifier is capable of handling this doubled power without overload.   Such an assumption is not always fulfilled, especially as noise peaks several times rms may be encountered.)

The signal generator may be of either the c-w or noise variety.   With a noise generator the procedure, although fundamentally the same, is simplified by the fact (to be shown in Sec. 14·5) that the available power of a noise generator is proportional to bandwidth.   Therefore when the expression for available power W_n is substituted into Eq. (1), the bandwidth cancels out, and the noise figure is obtained directly without the necessity of determining bandwidth.

[1] See footnote 1 on p. 169.

[2] E. J. Schremp, RL Internal Group Report 61-5/18/43. See also S. Roberts, "Some Considerations Governing Noise Measurements on Crystal Mixers," *Proc. I.R.E.*, **35**, 257–265 (1947).

14·2. Discussion of Available Power.—Since the definition of noise figure employs the concept of available power of generators, it is worth while to discuss this term and to prove some very elementary theorems that will be needed later in this chapter.

The most usual representation of an arbitrary electrical generator consists of a constant voltage in series with a constant internal impedance Z_a. Therefore, if these quantities and one of the following four quantities concerning the load are known, the other three can be calculated: the impedance of the load, the voltage across it, the current through it, or

(a)   (b)   (c)
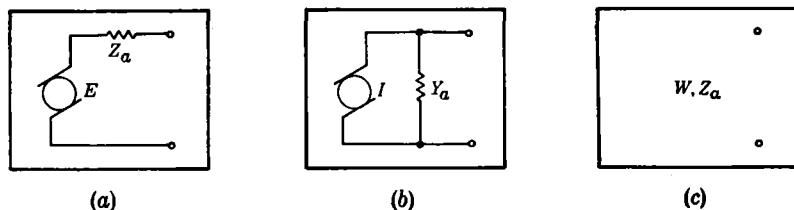
FIG. 14·1.—Three representations of a generator.

the power dissipated in it. Thévenin's theorem states that it is possible to represent in this way any network, containing generators and linear impedances, having two terminals to connect it to the outside world. According to this representation the electromotive force E, the internal impedance Z_a, and the external load Z are to be connected in series, as in Fig. 14·1a. Norton's theorem states that as far as the outside world is concerned, it is also possible to consider the same generator as a constant-current source generator in parallel with the internal admittance of the generator, $Y_a = 1/Z_a$, and with the load connected as indicated in Fig. 14·1b. It is possible to transform from one to the other of these representations by the relation

$$E = IZ_a = \frac{I}{Y_a}.$$ (2)

Finally, it is possible to consider the generator as a source of power and to describe it quantitatively in terms of "available power" W, defined as the maximum power that the generator can cause to be dissipated in an external load. This maximum power occurs when the external load impedance is the complex conjugate of Z_a and, therefore, when the voltage across the resistive component of the load is $E/2$ or the current through it is $I/2$. The relations by which one converts from the constant-voltage or constant-current representations are obvious:

$$W = \frac{E^2}{4R_a},$$ (3)

$$W = \frac{I^2}{4G_a},$$ (4)

where R_a is the resistive component of Z_a and G_a is the conductive component of Y_a.

Corresponding to each of these three representations is an experimental method for evaluating the generator. A voltmeter or potentiometer connected across the generator terminal measures E. The current strength I can be found by short-circuiting the terminals and measuring the current. The available power can be measured by means of a variable-impedance power-measuring device whose impedance can be adjusted to give maximum reading.

It is to be emphasized that these three representations are equivalent, and any of them can be used for any generator, only as far as the external
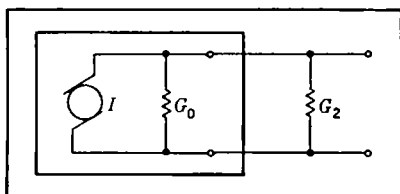
FIG. 14·2.—Generator with conductance in parallel with its terminals.

circuit is concerned; for calculating the power dissipated within the generator itself, detailed knowledge of the generator must be available.

Next consider what happens when various combinations of passive elements are connected to a generator. The combination can be considered as a new generator of different available power and internal impedance.

Theorem 1.—Connecting a network that consists of reactive elements changes only the internal impedance and not the available power.

Corollary.—Connecting a lossless transmission line whose characteristic impedance is R_0 or a four-terminal network containing only reactive elements and having both its iterative impedances equal to R_0 to a generator whose internal impedance is R_0 changes neither the internal impedance nor the available power of the generator.

The justification for these statements is self-evident and needs no further proof.

Theorem 2.—Connecting a conductance G_2 in parallel with a generator whose internal admittance has the conductive component G_0 changes the internal conductance to $G_0 + G_2$ and reduces the available power to $G_0/(G_0 + G_2)$ times the available power of the original generator.

This situation is represented by Fig. 14·2. The inner box contains the original generator, which is conveniently regarded as a current source I in parallel with an internal admittance G_0. The available power of the original generator is, from Eq. (4),

$$W_0 = \frac{I^2}{4G_0}.$$

It is clear that the internal conductance of the generator in the large box is $G_0 + G_2$. The available power of the generator in the large box is

$$W_2 = \frac{I^2}{4(G_0 + G_2)}.$$

Solving for W_2 in terms of W_0 yields

$$W_2 = \frac{W_0 G_0}{G_0 + G_2}, \tag{5}$$

which proves the theorem.

   Theorem 3.—Connecting a resistance R_1 in series with a generator whose internal impedance has the resistance component R_0 changes the
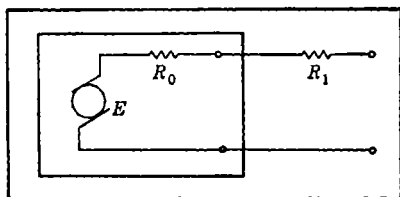
FIG. 14·3.—Generator with resistance in series with its terminals.

internal resistance to $R_0 + R_1$ and reduces the available power to $R_0/(R_0 + R_1)$ times the available power of the original generator.

   The proof of this theorem is similar to that of Theorem 2, except that it is convenient to consider the original generator as an emf E in series with an internal impedance, as indicated in Fig. 14·3.

   14·3. Measurement of Noise Figure with Unmodulated Signal Generators. The Relation of Noise Figure to Other Quantities That Express the "Noisiness" of an Amplifier.—The procedure for measuring the noise figure of any amplifier with an unmodulated c-w signal generator is similar to the one that has been used for many years for checking radio receivers. This method employs a signal generator containing a voltmeter, a calibrated attenuator, and a dummy antenna whose impedance is that of the actual antenna. The signal generator voltage is adjusted to cause the output power of the receiver to be twice that due to noise alone. When a standard dummy antenna is employed, this voltage itself is often used to describe the "noisiness" of the receiver. The quantity W_n can be calculated by substituting this value of the voltage and resistance of the dummy antenna into Eq. (3). The quantity W_n also is sometimes used to describe the "noisiness" of the receiver. Finally, the noise figure can be calculated by substituting W_n into Eq. (1).

The bandwidth B can be calculated by performing the integration indicated in Sec. 14·1 either numerically or analytically, or it may be approximated by the use of the half-power bandwidth as also noted in Sec. 14·1. In that section the fact was demonstrated that these three methods of describing the noisiness of a receiver are equivalent if it is assumed that the bandwidth and antenna impedance are specified. It is to be noted again that the voltage at the actual input terminals of the receiver is not known—only the voltage at the input of the dummy antenna is known.

There are certain difficulties connected with the use of c-w signal generators. The smallest power that can be directly measured with good laboratory apparatus to an accuracy of a few tenths of a decibel is about 10^{-6} watt. On the other hand, the noise figures of amplifiers of bandwidths of a few megacycles per second usually run from near one to about one hundred times (0 to 20 db), corresponding to equivalent noise powers of 10^{-14} to 10^{-12} watt. Hence, a signal generator must be equipped with a calibrated 80- or 60-db attenuator, which must be accurate to a few tenths of 1 db if the noise figure is to be known with similar accuracy. Such apparatus can be built but is difficult to maintain and is not practical for general use. Second, as has been pointed out earlier, the exact determination of bandwidth is a nuisance. Third, some devices that might be used to indicate receiver output power do not respond to noise and c-w signals in the same way. These difficulties can be avoided by the use of a noise generator.

NOISE GENERATORS

14·4. General Discussion.—There are a number of devices that have been used as noise generators for amplifier measurements. The outstanding feature of all of them is that their available power is proportional to an easily measured direct current. In the case of the temperature-limited diode, the constant of proportionality between the available power and the current is easily calculated, and it can be used as an absolute instrument up to about 300 Mc/sec. Above this frequency the measurement of impedance (which is required for determining the proportionality constant) becomes difficult, transit-time effects change the noise output, and the diode-noise generator can no longer be used as an absolute instrument, although it is still useful for relative measurements. Instruments of this type will be described in Secs. 14·5 and 14·6. For use at frequencies above about 1000 Mc/sec, a noise source has been developed[1] that utilizes a reflex velocity-modulated local-oscillator tube with its reflector voltage adjusted to prevent oscillation. This source is useful because large available powers are obtainable with it. At frequencies somewhat

---

[1] M. C. Waltz and J. B. H. Kuper, RL Report No. 443, Sept. 17, 1943.

above 3000 Mc/sec, transit-time effects in available tubes make the calibration vary rapidly with frequency, and this device is not satisfactory even for relative measurements. Third, silicon crystals (of the type used for rectifiers and mixers) generate considerable noise when direct currents are passed through them in the "reverse" direction. A practical crystal noise generator will be described in Sec. 14·7. Such a device is very compact and convenient for field measurements, but it must be calibrated in terms of signal generators of other types.

14·5. Theory of Noise Generators Using Temperature-limited Diodes. *High-impedance Noise Generators.*—A temperature-limited diode (that is, one whose plate voltage is high enough to saturate the plate current) acts like a generator of noise current having a certain rms value I_n, because of the "shot" effect—the fluctuations in the number of electrons emitted by the cathode. The mean squared current I_n^2 is given by

$$I_n^2 = 2eIB, \tag{6}$$

where e = charge of the electron = 1.60×10^{-19} coulomb,

$\quad I$ = direct current through the diode (in amperes),

$\quad B$ = bandwidth of the device being used to observe the noise.

Formula (6) does not apply rigorously to coated cathodes, which have another cause of noise (flicker effect), due to shifting of the active spots on the cathode.

A simple noise generator[1] consists of such a diode connected in parallel with a resistance R_a equal to the resistive component of the signal-source impedance and a reactance simulating the parallel reactance of the signal source. This combination is connected to the input of the amplifier and the value of I

FIG. 14·4.—High-impedance diode noise generator.

adjusted by varying the diode filament current to double the noise output of the amplifier (see Fig. 14·4). The available power of the generator W_n can be calculated by substituting Eq. (6) into Eq. (4), giving

$$W_n = \tfrac{1}{2}eIR_aB. \tag{7}$$

If this is substituted into Eq. (1), the noise figure F is

$$F = \frac{e}{2kT} IR_a \tag{8}$$

$$= 20IR_a \qquad \text{for } T = 290°K. \tag{8a}$$

[1] The use of temperature-limited diodes as signal generators for comparative measurements was reported by S. Ballantine, *Physics*, **4**, 294–306 (1933); and by D. O. North, *RCA Rev.*, **4**, 441–472, April 1940; and **5**, 106–124, July 1940. The high impedance diode noise source was developed at the Radiation Laboratory as a practical device for the measurement of noise figures by E. J. Schremp and C. P. Gadsden.

Equation (8) is a most important formula, for it gives the noise figure in terms of two easily measurable quantities, I and R_a.

This noise generator is very useful provided that (1) the noise generator can be connected in place of the signal source with very short leads and (2) the parallel reactance of the noise generator can be made entirely equivalent to that of the signal source over the entire range of frequencies of interest. There is always some capacity across the noise diode. In the cases where the desired shunt reactance corresponds to a capacity greater than the diode capacity, the desired capacity can be attained by adding a padding condenser. In other cases, which occur fairly often in
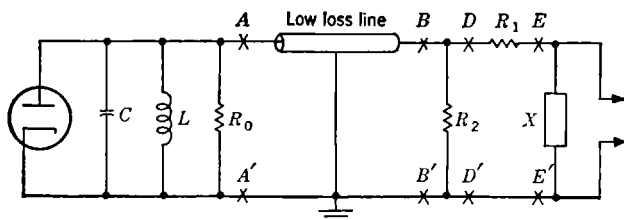
Fig. 14·5.—Matched-line diode noise generator.

practice, the desired reactance is that of a capacity smaller than the distributed diode capacity. These cases are somewhat embarrassing, because there is no completely satisfactory way of producing the desired reactance. The best that it is possible to do is to resonate out the undesired capacity by an inductance; this procedure is exact at only one frequency.

These difficulties can be largely avoided by the use of a "low-impedance" or "matched-line" noise generator.

The Matched-line Diode Noise Generator.[1]—This apparatus (see Fig. 14·5) consists of a noise diode feeding a lossless line of characteristic impedance R_0 with a termination at the generator end and a matching network consisting of R_1 and R_2 and X selected so that the internal impedance of the entire device equals that of the signal source.

The capacity C is the distributed capacity of the diode, and L resonates with C at the band center of the amplifier to be tested. More exactly, L and R_0 are selected so that the same impedance is measured within 1 per cent at AA' as at BB' (the other components to the right being disconnected in both cases). Since R_0 is a resistor of relatively low value, the bandwidth of the circuit C, L, R_0 is large compared with the usual amplifier bandwidths; therefore the diode capacity has much smaller effect than in the high-impedance noise generator. (In a typical

---

[1] This device was developed at the Radiation Laboratory by C. P. Gadsden and S. A. Wingate at the suggestion of S. N. VanVoorhis.

case, in which C is 10 $\mu\mu f$ and R_0 is 75 ohms, the bandwidth is about 200 Mc/sec.)

There is no fundamental reason why the line should be terminated in its characteristic impedance. Whatever the conditions of termination may be, it is possible by application of Thévenin's theorem to select matching networks giving the proper internal impedance and then to calculate the available noise power in terms of the direct current through the diode. However, for reasons that will become evident, the calculations are greatly simplified if the line is terminated in its characteristic impedance at the generator end (contrary perhaps to what one might first expect), the calibration then being independent of the length.[1]

The analysis of this generator is carried out in the following steps. First, the components to the left of terminals AA' form a simple noise generator whose available power is found by replacing R_a by R_0 in Eq. (7). According to Theorem 1 in Sec. 14·2, the addition of a lossless line of characteristic impedance R_0 means that the generator, viewed at the terminals BB', has internal impedance R_0 and the same available power as at AA'. The effects of R_2 and R_1 upon the internal impedance and available power can be found by applying Theorems 2 and 3 respectively. The reactance X is adjusted to make the parallel reactance of the generator equal to that of the actual signal source and does not change the available power.

If the available power of the generator is calculated in this fashion and the result substituted into Eq. (1), the noise figure is given by

$$F = \frac{R_0^2 I}{2kT \, R_a \left(1 + \dfrac{R_0}{R_2}\right)^2}, \tag{9}$$

$$= 20 \frac{R_0^2 I}{R_a \left(1 + \dfrac{R_0}{R_2}\right)^2} \qquad \text{for } T = 290°\text{K}, \tag{9a}$$

where R_a, the parallel output resistance of the generator, is made equal to the parallel output resistance of the actual signal source and R_0, R_1, and R_2 are chosen to give the desired value of R_a by the use of the following relation:

$$R_a = R_1 + \frac{R_0}{1 + \dfrac{R_0}{R_2}}. \tag{10}$$

Two special values of R_2 are of interest. If $R_2 = \infty$,

[1] Termination of the amplifier end of the line in its characteristic impedance does not yield any advantage. If both ends of the line are terminated, four times as much direct current is needed, and this sometimes exceeds the safe operating conditions of the tube.

$$F = \frac{e}{2kT}\frac{R_0^2 I}{R_a}, \tag{11}$$

$$= \frac{20R_0^2 I}{R_a} \qquad \text{for } T = 290°\text{K.} \tag{12}$$

By comparison with Eq. (8), it can be seen that for a given noise figure the ratio of the direct current required with the matched-line noise generator to that required with the high-impedance noise generator is $(R_a/R_0)^2$. To cite a typical example, if R_a is 300 ohms and R_0 is 75 ohms, sixteen times as much current is required with the matched-line noise generator as with the high-impedance noise generator.

If R_2 is made equal to R_0,

$$F = \frac{e}{2kT}\frac{R_0^2 I}{4R_a}, \tag{13}$$

$$= 5R_0^2 \frac{I}{R_a} \qquad \text{for } T = 290°\text{K.} \tag{13a}$$

$$R_a = R_1 + \tfrac{1}{2}R_0. \tag{14}$$

It can be seen by comparing Eq. (11) with Eq. (13) that four times as much diode direct current is required to measure a given noise figure if both ends of the line are terminated in the characteristic impedance of the line than if only the generator end is terminated.

14·6. Construction of Diode Noise Generators.—Diodes used as noise generators should have the following characteristics: pure tungsten or thoriated-tungsten filament, good saturation (the reciprocal of the slope of the plate-current vs. plate-voltage curve should be at least fifty times the desired generator resistance), short leads, and low capacity. Also, for matched-line use diode noise generators should have a moderately large emission and plate dissipation. Of all the tube types tested in the Radiation Laboratory the British type CV172 is the most satisfactory. The Western Electric type 708-A, the Eimac type 15E (the type 15R does not saturate very well), and the type 801A seem to be more or less satisfactory for matched-line generators. The type 01A and the Sylvania type X-6030 are satisfactory for high-impedance generators (except that the type X-6030 is not very rugged mechanically). At 30 Mc/sec comparisons among many of these types have yielded agreements within 0.05 db.

Oxide-coated filamentary tubes have proved unsatisfactory where highest accuracy is desired, and indirectly heated oxide-coated cathode-type tubes are doubly undesirable because of the difficulty and sluggishness with which they are controlled.

In Figs. 14·6 and 14·7 are shown an experimental noise diode under development at the time of writing.[1] The tube, designated the type

[1] Harwick Johnson, *RCA Review*, **8** (1947), 169–185.

R6212, is constructed in the form of a small section of coaxial line of 50-ohm characteristic impedance, and the filament is in the form of a loop of wire around the inner conductor. A section of high-loss 50-ohm cable, terminated at its far end by a 50-ohm resistor, is connected to one end of the tube and serves as the resistor R_0 in the circuit shown in Fig. 14·5. Another section of cable connected to the other end of the tube is used to
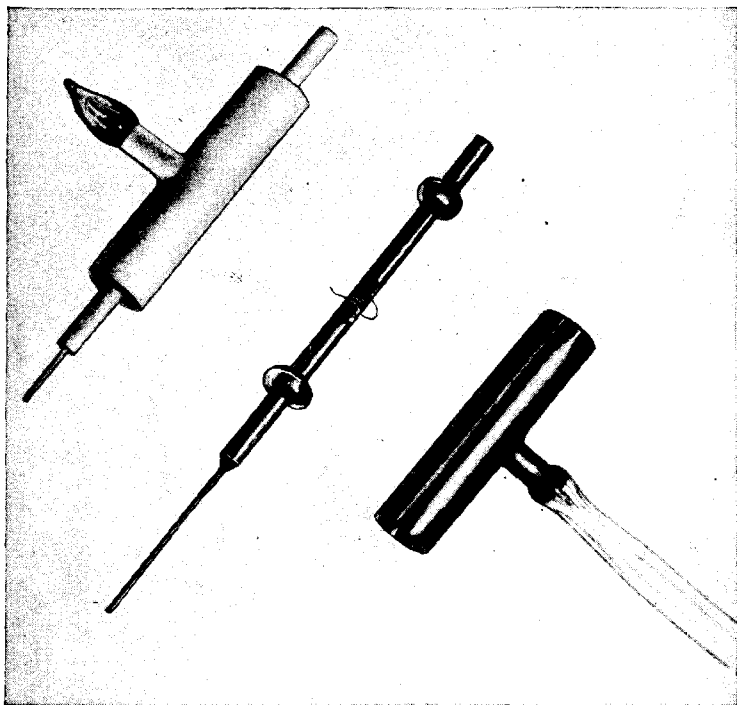
FIG. 14·6.—R6212 noise diode unassembled.   (*Courtesy of RCA Laboratories.*)

connect the tube to the input of the amplifier. Because of its unusual construction, this tube is capable of giving quantitative results at higher frequencies than conventional tubes.

The construction of noise generators is not difficult. The generator should be completely shielded. Its power leads should be filtered to prevent pickup of extraneous signals, particularly pickup from the output stages of the amplifier, since this pickup might cause regeneration. Because the diodes are operated under temperature-limited conditions, the plate-supply voltage need not have particularly good regulation or filtering.

Under temperature-limited conditions, the plate current varies very

rapidly with filament voltage.   Therefore, if the filament is to be operated
from the a-c mains, it will be necessary to provide a constant-voltage
regulator in the a-c line.   Alternatively the filament can be operated
from storage batteries.   Because of the rapid variation of emission with
filament voltage, two rheostats for the control of filament voltage should
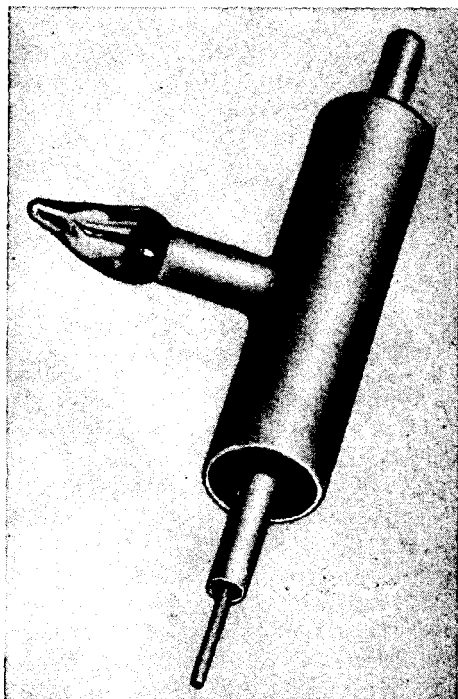
Fig. 14·7.—R6212 noise diode assembled.   *(Courtesy of RCA Laboratories.)*

be included in the apparatus: one for rough control and the other for fine
control.   It is important to have firm electrical connections in the filament
circuit.   Because of the large filament current the number of connectors
and toggle switches should be kept to a minimum; and if alternating
current is employed for the filaments, these connectors and toggle switches
and the control rheostats should be placed in the primary of the trans-
former, where slight changes of resistance will have very small effect.
Sometimes it is necessary to clean the filament pins of the tube and tube
socket.

A noise generator employing a type CV172 tube is depicted in Fig.
14·8, and the complete circuit diagram is shown in Fig. 14·9.   It was
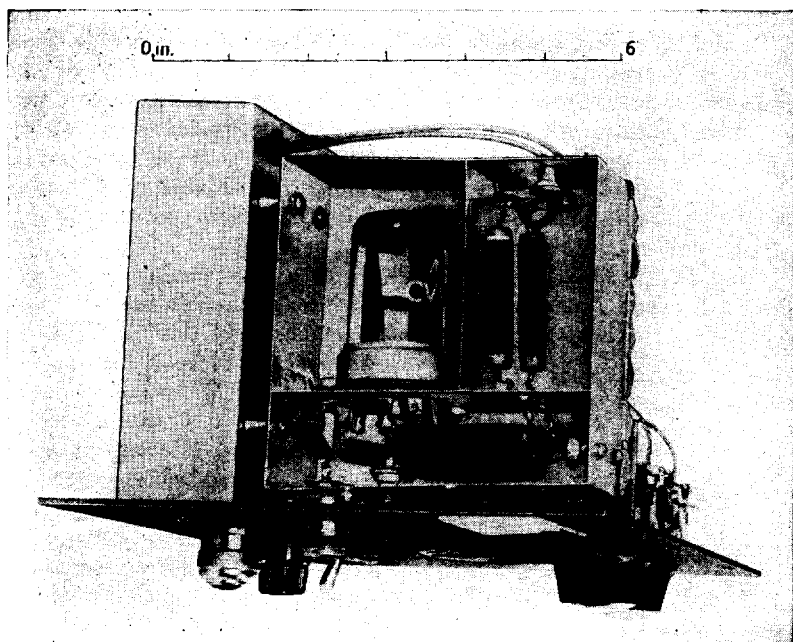
FIG. 14·8.—Interior view of CV172 noise source.
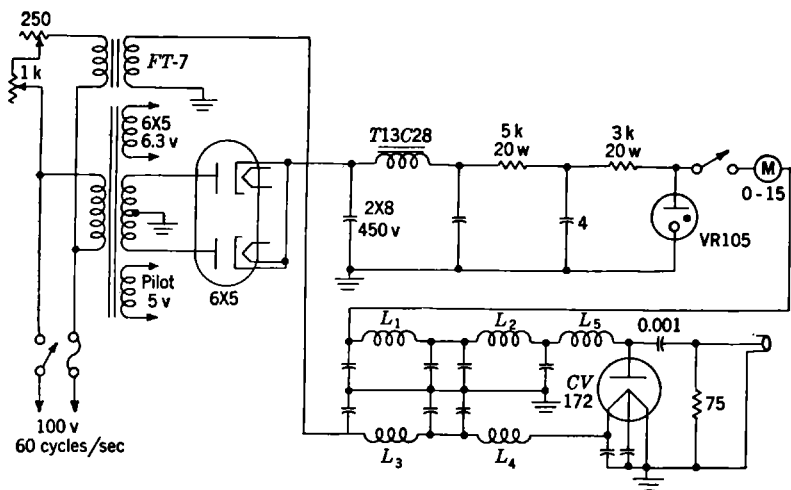
FIG. 14·9.—Circuit of noise generator.

designed primarily for use as a matched-line generator employing a section of 75-ohm cable (not shown in the diagram), and, therefore, a 75-ohm resistor is shown in parallel with the output terminal. However, this same design could be used as a simple high-impedance generator by raising the value of this resistor and adjusting the reactance in parallel to a suitable value.

In the light of the previous paragraph, the inclusion of the large amount of filtering and the VR-105 regulator tube in the power supply is, perhaps, not entirely necessary.

14·7. Crystal Noise Generators.—Silicon crystals produce considerable noise when a direct current is passed through them in the direction of lower conductivity.[1] Practical noise generators of this type, which
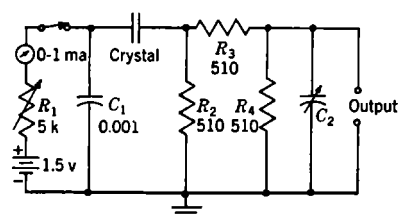
FIG. 14·10.—Crystal noise generator.

can be used at frequencies up to at least 3000 Mc/sec, have been developed at the Radiation Laboratory. The circuit diagram of a crystal noise generator designed for use at 30 Mc/sec is shown in Fig. 14·10. Because it was desired to use a meter with a sensitivity of 1 ma full scale,

it was necessary to provide a π-attenuator consisting of R_2, R_3, and R_4 to reduce the output power to a convenient value. The values were selected so that the parallel resistance at the output terminals was approximately 300 ohms, equal to that of the actual signal source used with the amplifier. The value of C_2 is adjusted to equal the capacity of the signal source. The crystal current is adjusted by variation of R_1 until the amplifier noise output power is doubled, and the noise figure is determined from this value of the crystal current and the calibration of the instrument, which must be made experimentally by comparison with a signal generator of another type. Such instruments are very compact and simple in construction but must be calibrated frequently if high accuracy is desired.

MEASUREMENT OF AMPLIFIER OUTPUT POWER

Experience at the Radiation Laboratory has indicated that most difficulties encountered in noise-figure measurements are due to factors external to the noise generator and are caused by improper methods of measuring relative power output of the amplifier.

The choice of an output-power measuring device depends upon

---

[1] J. E. Houldin, "The Crystal Capsule as a Generator of Noise," Report of the General Electric Company (Great Britain), July 9, 1943.

whether or not an unmodulated c-w generator is to be used and whether accuracy can be sacrificed for simplicity.

14·8. Attenuator and Postamplifier.—This combination represents the most accurate and generally satisfactory means of comparing amplifier output powers. Although it is more complex than the other schemes described here, it is very rugged and can be operated successfully by relatively untrained personnel. It is therefore ideally suited for use on production lines as well as in research laboratories.
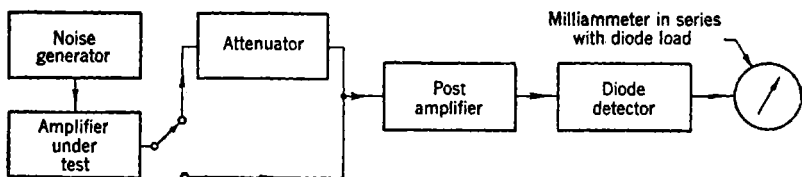
Fig. 14·11.—Attenuator and postamplifier method of measuring noise.

Method.—The method is illustrated by a block diagram in Fig. 14·11. The procedure is first to observe the reading of the output meter with the attenuator switched out and the noise generator turned off. Then the attenuator is switched in, and the noise generator adjusted to make the output meter have the same deflection as before. If the attenuator has a power ratio α equal to 2, the available power of the noise generator must have been that required to double the amplifier output power. In the event α is not exactly 2, a very simple correction can be made, as will be shown later. Because the postamplifier is operated at constant-signal level, no harm is caused by overloads within it, and this unit can be connected to the amplifier under test at a stage where the signal level is low enough to avoid overload difficulties. This connection can be made by means of an adapter that plugs in place of one of the tubes in the amplifier.

Construction.—The postamplifier should have the same center frequency as the amplifier under test and a bandwidth somewhat greater. Great care should be exercised in shielding the postamplifier, and it should be operated from its own power supply to avoid any chance of feedback to the amplifier. It is convenient to gang with the attenuator a switch connected to the noise generator plate supply, so that it can be turned on and the attenuator switched in by a single knob. However, this switch should be isolated from the attenuator to prevent signals from leaking out. The detector uses any of the common diode receiving tubes. A meter indicating 1 ma full scale in series with the diode-load resistor is adequate for many measurements, but the over-all sensitivity can be increased by using a meter of 30 or 100 μa full scale, with a dry cell and a variable resistance in series with it connected across

the meter to buck-out the "dark" current of the diode.   It is also useful
to have a jack connected to the high-voltage end of the load resistor
through a decoupling filter so that the apparatus may be connected to an
oscilloscope in order to observe the pass band of either the postamplifier
or the combination of the amplifier and postamplifier when a swept-
frequency signal generator is connected to their respective input terminals.

The attenuator and its immediate circuit are designed with these
objectives: (1) to make the calibration independent of the condition of
impedance mismatch at the output of the amplifier and (2) to have a
value of α as near to 2 as possible.   The diagram of the attenuator and
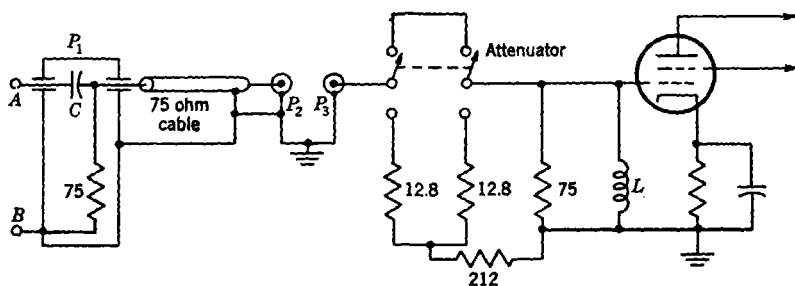
FIG. 14·12.—Details of attenuator and its immediate circuit.

its immediate circuit is shown in Fig. 14·12.   If the characteristic imped-
ance of the cable is other than 75 ohms, the values of the resistors are
to be multiplied by the ratio of the characteristic impedance to 75 ohms.
The resistors are $\frac{1}{2}$-watt carbon resistors, and those in the attenuator are
selected by a Wheatstone bridge to be within 2 per cent of the indicated
values.   If none are available within this tolerance, they can be produced
by filing away a portion of a resistor of lower value and covering the
filed·place with a coat of lacquer.

The terminating resistor R and the inductance L (which resonates
the input capacity of the tube) are selected so that the same impedance
(at the center frequency of the amplifier) is measured across the pins of
P_3 at both positions of the attenuator switch.   If this condition is met,
the calibration of the attenuator is independent of the impedance pre-
sented by the cable.   As a precaution against slight errors in the values
of L and R, the cable should present an impedance that approximately
matches the iterative impedance of the attenuator in order to absorb
any waves that might be reflected back from R and L.

In Fig. 14·12, the amplifier end of the cable is fitted with a plug P_1
that fits into a tube socket of the amplifier, here assumed to employ
single-tuned circuits.   This plug is designed to connect the cable in
parallel with the load of the tube in front of that displaced by the plug.

The plug contains a 75-ohm resistor terminating the cable in its characteristic impedance. The plug presents a net impedance of 37.5 ohms (75 ohms in the plug in parallel with a 75-ohm terminated line); this heavy damping leads to a very large bandwidth. An alternative procedure is to omit the resistor in the plug P_1 but to insert between P_2 and P_3 a T- or π-attenuator pad having a 75-ohm iterative impedance and at least 10-db attenuation. In fact if amplifiers with different gains or the same amplifier with various settings of the gain control are to be measured, the postamplifier must have sufficient gain to operate with
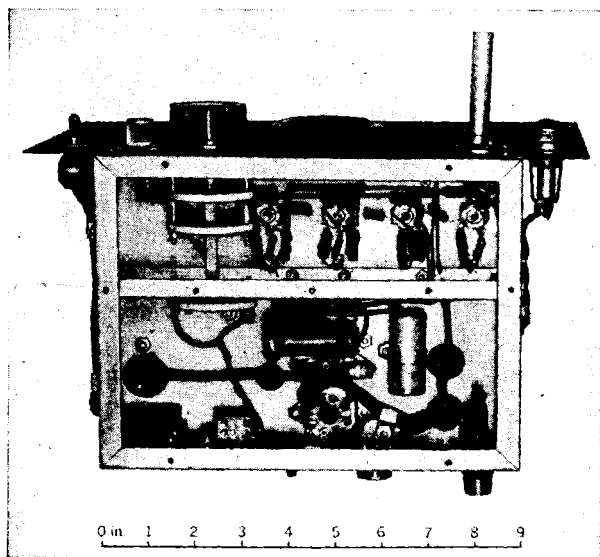
FIG. 14·13.—Bottom view of postamplifier.

an amplifier of minimum gain, and it is desirable to employ, between P_2 and P_3, a variable step attenuator of the type described in Sec. 8·5 to compensate for excess gain.

The method of calibrating the attenuator is to apply a c-w signal of variable power to it. The relative power level of the signal must be known in terms of a bolometer, vacuum-tube voltmeter, or waveguide attenuator. The attenuation factor α of the attenuator is given by the ratio of the two signal-power levels that give the same deflection on the output meter with the attenuator out and in respectively. It is necessary to use signal levels that are high compared to noise, and the precaution mentioned above of presenting to the attenuator an impedance equal to its iterative impedance should be observed. If the procedure previously outlined is followed, the attenuation will be within ± 0.2 db of 3 db at

30 Mc/sec.   Another method employing the calibration of a noise generator will be described at the end of this section.   This method usually can be used when the equipment mentioned above is not available.

Photographs of a unit of the type described in this section are shown in Figs. 14·13 and 14·14.   The circuit diagram is shown in Fig. 14·15.
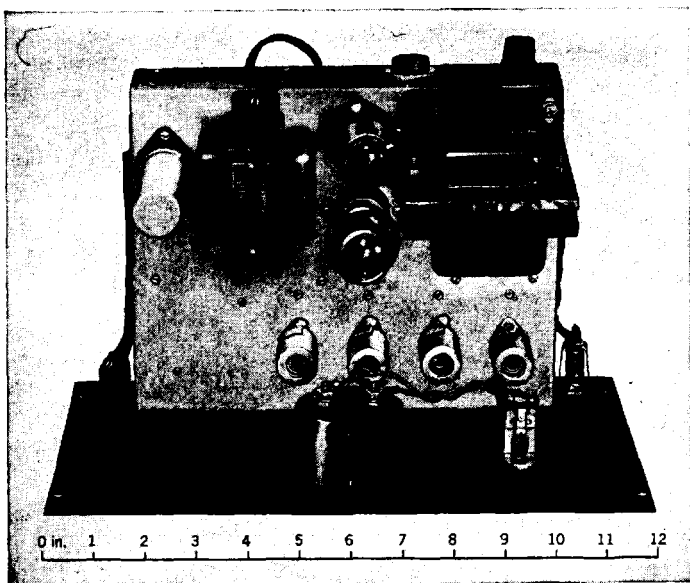
Fig. 14·14.—Top view of postamplifier.

The amplifier is shielded from the power supply and from the section of the selector switch that controls the plate supply of the noise generator. This unit has a bandwidth of 8 Mc/sec centered at 30 Mc/sec, and the voltage gain between the grid of the first tube and the plate of the third tube is approximately 350.

Correction to Be Applied When α Is Not Equal to 2.—The correction to be applied when α is not equal to 2 is very simple.   When a noise figure is measured, it is assumed that at the input terminal of the amplifier there is a single source of noise whose available power is W_n and which simulates all the sources of noise within the amplifier (see Sec. 14·1). With the attenuator in, the noise source is turned on and adjusted to give an available power W_1 such that the output power obtained is the same as that with the noise generator off and the attenuator out. Therefore

$$W_n + W_1 = \alpha W_n,$$ (15)

or

$$W_n = \frac{W_1}{\alpha - 1}.$$ (16)

The procedure is to (1) calculate the noise figure as though α were exactly 2 and (2) divide the result by $\alpha - 1$ if the noise figure is expressed as a ratio or subtract $10 \log_{10}(a - 1)$ if it is expressed in decibels.

Equation (16) shows the effect of an error in attenuator calibration on the accuracy of noise-figure determination. For example, if α is believed to be exactly 2 but is only approximately 2, then, because $\alpha - 1$ is approximately unity, half as large, the percentage error in noise figure
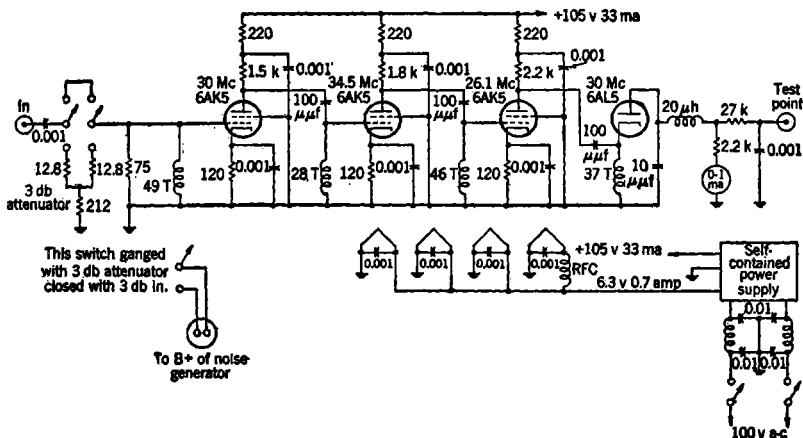
Fig. 14·15.—Auxiliary postamplifier for noise-figure measurement. Amplifier completely shielded.

is approximately twice as great as the percentage deviation between α and 2.

Attenuator Calibration by Means of a Noise Generator.—The following method permits calibrating an attenuator with extreme precision; the only requirement on the output power-measuring device is that it be able to indicate that power has not changed.

First, a direct current I_1 is passed through the noise generator so that the noise output power of the amplifier with the current I_1 flowing and the attenuator in is the same as the noise output power with the attenuator out and the noise generator off. If W_1 is the available noise-generator power corresponding to I_1, then

$$W_n + W_1 = \alpha W_n. \tag{15}$$

Next the noise-generator current is adjusted to a value I_2 such that with I_2 flowing and the attenuator in, the noise output power of the amplifier is the same as with the attenuator out and the current I_1 flowing. If W_2 is the available noise-generator power corresponding to I_2, then

$$W_n + W_2 = \alpha(W_n + W_1). \tag{17}$$

Subtracting Eq. (15) from Eq. (17) yields

$$\alpha = \frac{W_2 - W_1}{W_1}.$$

If W_1 and W_2 are proportional to I_1 and I_2, with the same proportionality constant (as is true for a given temperature-limited diode noise source),

$$\alpha = \frac{I_2 - I_1}{I_1} = \frac{I_2}{I_1} - 1. \tag{18}$$

It is thus possible to determine α with all the precision with which the direct currents I_1 and I_2 can be measured. Although Eqs. (15) and (17) are valid only if the amplifier preceding the attenuator is linear, this linearity is very easy to achieve simply by operating the part of the amplifier preceding the attenuator at low level.

14·9. Method Employing Gain-control or Uncalibrated Attenuator.— A method[1] has been developed for use with noise generators that requires only the noise generator and an output meter of any type whatever (usually a d-c voltmeter connected across the detector load resistor in a receiver) but that requires that the amplifier have a gain control and assumes that the noise figure is independent of gain-control setting. This latter assumption is not always valid and is discussed further in Sec. 14·13.

The method is as follows. First the deflection of the meter is noted with no noise-generator current. Then an arbitrary current I_1 is passed through the generator and the output meter is read a second time. Next with the current I_1 flowing, the amplifier gain control is set so that the output meter returns to its first deflection. Finally, the current is raised to a value I_2 so that the output meter has the second deflection. If W_1 and W_2 are the available powers of the generator corresponding to I_1 and I_2 respectively,

$$\frac{W_2 + W_n}{W_1 + W_n} = \frac{W_1 + W_n}{W_n}. \tag{19}$$

Solving Eq. (19) for W_n yields

$$W_n = \frac{W_1^2}{W_2 - 2W_1}. \tag{20}$$

Dividing by kTB, one has [see Eq. (1)],

$$F = \frac{W_n}{kTB} = \frac{\left(\dfrac{W_1}{kTB}\right)^2}{\left(\dfrac{W_2}{kTB}\right) - 2\left(\dfrac{W_1}{kTB}\right)}$$

[1] M. C. Waltz, RL Internal Group Report—61-9/15/43.

For a given temperature-limited diode noise source the ratios $W_1/(kTB)$ and $W_2/(kTB)$ have the form MI_1 and MI_2, where

$$M = 20R_a$$

for a high-impedance noise generator, and

$$M = \frac{R_0^2}{R_a\left(1 + \dfrac{R_0}{R_2}\right)}$$

for a matched-line noise generator, [see Eqs. (8) and (9)], at $T = 290°$K. Hence,

$$F = \frac{MI_1^2}{I_2 - 2I_1}. \tag{21}$$

This method is very useful for field measurements because it requires so little equipment, but the accuracy is often poor, mainly because $I_2 - 2I_1$ is usually small compared with I_1 or I_2, and therefore a small fractional error in either I_1 or I_2 leads to a relatively large error in $I_2 - 2I_1$. Often more accurate values are obtained if I_1 and I_2 are rather large (compared with the value of I used in the other methods). Unfortunately, however, many amplifiers overload at such high signal levels.

14·10. Crystal and Diode Rectifiers.—Crystal and diode rectifiers (which include the detectors incorporated in radio receivers) suffer disadvantages as amplifier power-output measuring devices except in regard to availability. Usually they do not respond to a combination of noise and c-w signals according to the same law as to noise alone and hence are not reliable for highest accuracy when c-w generators are employed. Even with noise generators, considerable effort must be expended, especially with wide-band amplifiers, to establish the output-deflection vs. input-power law of the detector. (Such detectors usually have a "dark current" with no signal applied. Depending upon individual conditions it is sometimes necessary to subtract the dark current and at other times to neglect it in establishing some simple law.) Also the law itself usually holds for only a small range of signal levels. Hence this method is usually not very accurate even with noise sources.

Crystal detectors used with indicating devices that have a deflection sensitivity of about $1\mu a$ full scale have been found to be very accurately linear in power, but with meters of lower sensitivity crystals may deviate considerably from such a law.

14·11. Bolometers.—A bolometer is the most satisfactory output-indicating device for use with unmodulated c-w signal generators because it responds to unmodulated c-w and noise powers according to the same law. It can be used with noise generators as well.

An unbalanced-bridge type of bolometer is sufficient for this purpose, since it is necessary to measure merely relative power and not absolute power in watts at the output terminals of the amplifier. Such a device is illustrated in Fig. 14·16, where R_2 is a $\frac{1}{100}$-amp Littelfuse or a Thermistor that (together with the decoupling resistors R_3 and R_4) forms one
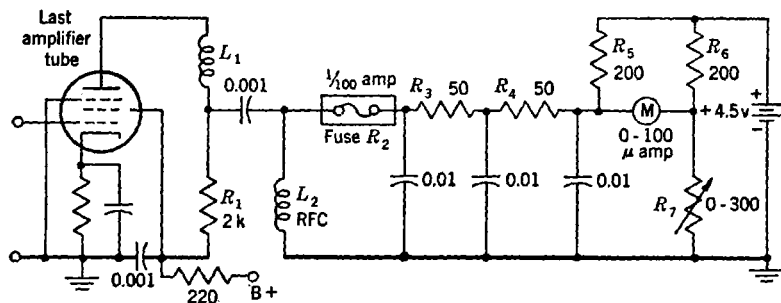
Fig. 14·16.—Unbalanced-bridge bolometer.

arm of a Wheatstone bridge, the other arms being R_5, R_6, and R_7. With the amplifier turned off, the bridge is balanced by the adjustment of R_1. With the amplifier operating, its output power, consisting of noise or signal or a combination of both, is dissipated in R_2. This changes its resistance, upsetting the balance of the bridge and causing the meter to deflect. If the amplifier is linear, the deflection is proportional to the input power to the bridge.

The value of L_1 is adjusted to resonate with C, the output capacity of the tube, at the center frequency of the amplifier. The bandwidth of this output circuit is

$$\mathcal{B} = 2\pi R_2 C f_0^2, \tag{22}$$

where f_0 is the center frequency. The ratio of the voltage across R_2 and the voltage at the input of the tube, if the dynamic plate resistance is assumed to be very high, is

$$\mathcal{G} = \frac{g_m}{C\omega_0 \sqrt{1 + R_2^2 C^2 \omega_0^2}} \tag{23}$$

where $\omega_0 = 2\pi f_0$. In order to obtain a large bandwidth it may be necessary to increase C by connecting a condenser in parallel with the output of the tube. According to Eq. (23) this action causes a reduction in gain. If an even greater bandwidth is desired (at the further expense of gain), L_1 may be made zero and L_2 made to resonate C, giving a parallel-tuned circuit loaded by R_2.

There is considerable danger of burning out R_2 or the meter, and care must therefore be taken in operating the apparatus. Furthermore, the

apparatus should be shut off while making connections or changes, because oscillation may result from improper connections.

Generally it is not very convenient to connect such a device to an already constructed amplifier. Therefore it is desirable to mount such a device in an auxiliary "postamplifier" whose input signal is obtained from a plug replacing one of the tubes of the amplifier under test.

14·12. Thermocouple Meters.—Experience obtained at 30 Mc/sec has indicated that depending upon the particular thermocouple in use, the deflection of such meters does not always vary linearly with power nor do they accurately obey the same law as at direct current. Therefore if they are used, they should be calibrated in terms of other devices such as attenuators or bolometers at the frequency at which they are to be used.

SPECIAL TOPICS

14·13. Effect of the Gain Control.—Usually the only important sources of noise in an amplifier lie within the first few stages, because the amplification of the first stages causes these noise sources to overshadow any noise in later stages. However, if a gain control is placed in one or more of the early stages, noise in later stages may predominate at low settings of the gain control, especially since these later stages are usually not carefully designed in regard to their noise. Hence, the noise figure may deteriorate at low-gain settings.[1]

A spurious effect sometimes appears at high-gain settings, where the noise figure measured by the attenuator or bolometer methods may appear to be larger than at intermediate-gain settings. This effect is due to overload of the stage or stages preceding the power-measuring device and causes the measurements to be in error, but it does not indicate that the actual ability of the amplifier to detect weak signals has deteriorated.

It is clear, therefore, that noise figure should be measured at several gain-control settings to be sure that no overload effects preceding the power-measuring device are taking place. Second, when it is desired to repeat the measurement of the noise figure of an amplifier with the highest possible precision, the gain-control setting must be reproduced.

14·14. Correction for Temperature.—At a temperature of 290°K, the numerical values of the constants in the various expressions for noise figure, Eq. (1) and especially Eqs. (8a) and (13a), are of a convenient value for calculations. Because 290°K is near room temperature, it is usually accurate enough to use these simple values of the constants without making individual determinations of the temperature. Situa-

---

[1] This is not usually a serious matter in the practical use of amplifiers, because, in general, an amplifier needs to have the best possible noise figure only when it is being used to amplify very weak signals, that is, when its gain is high.

tions have arisen, however, where the noise figure was so low (1.06 times = 0.25 db for amplifiers at 6 Mc/sec, 1.35 times = 1.3 db for amplifiers at 30 Mc/sec employing the grounded-cathode–grounded-grid-triode input circuit discussed in Sec. 13·10) and the precision of the measurements so high that the effect of variations in temperature had to be considered. Furthermore, because of heat produced within the apparatus it was necessary to place a thermometer on or near the resistor that simulated the resistance of the signal source.

In this connection it is interesting to consider a "radiometer" noise-figure determination carried out by J. L. Lawson. A brass cube about $1\frac{1}{2}$ in. on a side was provided with a well for the insertion of a thermometer and a cylindrical hole for insertion of a $\frac{1}{2}$-watt carbon resistor. The resistor was connected to the input terminals of a 30-Mc/sec amplifier so that the only source of noise, apart from that in the amplifier itself, was the thermal-agitation noise in the resistor. The amplifier noise output power was then recorded as a function of the temperature of the brass block over a range from 290° to about 390°K. (The brass block was presumably large enough to achieve a good approximation to temperature equilibrium; the source of heat was a soldering iron. The behavior of the resistance as a function of temperature had previously been allowed for.)

FIG. 14·17.—Radiometer determination of amplifier noise figure.

The points lay remarkably accurately on a straight line, as indicated in Fig. 14·17. Extending this line back to 0°K gave the noise originating in the amplifier alone (the excess noise of the amplifier); the noise figure at 290°K was in fact

$$F = \frac{\text{output noise power at } 290°K}{\text{output noise power at } 290°K - \text{output noise power at } 0°K}.$$

For the particular amplifier under test, this value came out to be 1.50, whereas the noise–diode determination of noise figure had yielded 1.52.

The success of these simple procedures requires an amplifier with sufficiently good noise figure to show significant variation in total output noise power for a small change in source-resistance temperature; difficulties in attaining temperature equilibrium would arise at temperatures far above room temperature.

The value of the noise figure F at given temperature T can be calcu-

lated by substituting the exact value of T into the various formulas, but for the comparison of different measurements it is necessary to calculate F_0, the value of the noise figure that would have been obtained if the temperature were some standard value T_0. According to the definition of noise figure contained in Sec. 14.1, it is possible to replace all of the sources of noise within the amplifier and signal source by a generator of available power W_n at the input of the amplifier. From Eq. (1)

$$W_n = FkTB.$$

Part of this power, kTB, is due to the thermal-agitation noise of the signal source, whereas the rest, $(F - 1)kTB$, is due to sources within the amplifier. If the measurement had been performed at T_0, the first part would have had the value kT_0B. For the lack of information to the contrary, it is generally assumed that the second part would have been unchanged. Therefore, the total available power would have been

$$W_{n_0} = kT_0B + (F - 1)kTB, \tag{24}$$

and the noise figure F_0 would have been found by dividing W_{n_0} by kT_0B:

$$F_0 = 1 - \frac{T}{T_0} + \frac{FT}{T_0}. \tag{25}$$

The last term on the right is the value of the noise figure that would have been obtained if the temperature T_0, instead of the actual temperature, had been used in the formulas for noise figure [Eqs. (1), (8), and (13)].

Obviously the most convenient value for T_0 is 290°K, because $e/(2kT)$ is then equal to 20. In this case

$$F_0 = 1 - \frac{T}{290} + F', \tag{26}$$

where F' is the noise figure calculated by using 290° for T in the formulas for noise figure, or, in other words, the value obtained using the simple value of $e/(2kT) = 20$, as in Eqs. (8a) and (13a).

14.15. Noise Figure of an Amplifier with Push-pull Input Connections.—Although it is possible to construct signal generators with push-pull output connections, they are not necessary for noise-figure measurements of amplifiers with push-pull input connections.

Ordinary generators can be used in the following manner: The normal signal source has an impedance consisting of a resistance R_a in parallel with a reactance X; therefore, the signal generator is adjusted to have an output impedance equal to that of R_a and X in parallel and is connected across one input terminal and ground. A dummy impedance equal to that of R_a and X in parallel is connected between the other terminal and ground. Some of the power supplied by the signal source is transmitted through the input circuit and is dissipated in the dummy

impedance. This effect of the dummy impedance modifies the voltage across the terminals of the signal source in the same way as would a resistance R_a' and a reactance X' connected in parallel with its terminals. According to Theorem 2 in Sec. 14·2, the available power is effectively reduced by the factor $R_a/(R_a + R_a')$. Usually it can be assumed that the input circuit acts as transformer with a unity turns ratio, causing R_a' to be equal to R_a and the ratio to be equal to one-half. In other situations, R_a' can be measured by an impedance-measuring device, and the ratio can be calculated.

The procedure then is to perform the measurements and calculations as though the amplifier were single sided and the parallel resistance of the signal source were R_a. Then the noise figure determined in this way is multiplied by the ratio $R_a/(R_a + R_a')$. In order to eliminate the effects of unbalance, the signal generator and dummy impedance should be interchanged, the measurements repeated, and the two results averaged.

14·16. Measurement of Noise Figure of Superheterodyne Radio Receiver with Image Response.—In a superheterodyne radio receiver there is always some response at the image frequency. Consequently there is a contribution to the output noise power from noise sources at the image frequency; these include thermal-agitation noise, grid noise of the first tube, and possibly other effects. Hence an equivalent noise *generator at the amplifier input terminals, at signal frequency, has to* have larger power than if there were no image response. A perfect receiver has no image response, and the bandwidth B in Eq. (1) is evaluated for the signal band only. Therefore the noise figure of the receiver deteriorates because of the image response. If the gain were the same at the image frequency as at the signal frequency, and if the noise sources enumerated above were the only ones present, the noise figure would be increased by a factor of 2. Since the gain at image frequency is usually less than at signal frequency and since other sources of noise are present, such as plate noise, the factor by which noise figure is multiplied is usually considerably less than 2 but, of course, never so small as 1.

When unmodulated c-w signal generators are employed to measure noise figure, no difficulty is encountered, since these generators do not produce a signal at the image frequency. However, noise generators produce noise at the image frequency as well as at the signal frequency.

For calculating the noise figure, it is necessary to know the value of the d-c noise-generator current needed to double the noise output power if there had been no image response. This value is equal to the observed current multiplied by the ratio $(G_p + G_p')/G_p$, where G_p and G_p' are the average actual power gains at the signal and image frequencies respectively. Consequently, noise figures calculated by substituting the actual observed current into Eq. (8) or (13) should be corrected by multiplication by this factor.

APPENDIX A

REALIZABILITY OF FILTERS

By Henry Wallman

Given a certain amplitude characteristic, the question arises as to whether or not it can be realized by means of a physical filter. For example, is there any physical network whose absolute value is a gaussian-error curve? The only restriction on a "physical network" is that it show no response to an applied signal until the input switch is closed; i.e., that the Fourier transform be zero for time $t < 0$.

A complete and very simple answer to the question of realizability is provided by a theorem of Paley and Wiener, drawn from the theory of Fourier transforms in the complex domain.

The result for the gaussian-error curve is that it is not realizable.
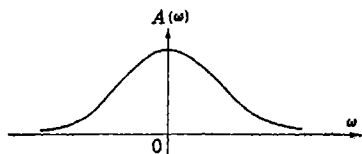
FIG. A·1.—Gaussian-error-curve filter, $A(\omega) = e^{-\omega^2}$.
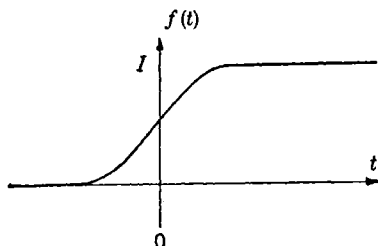
FIG. A·2.—Step-function response of Gaussian-error-curve filter.

A·1. The Paley-Wiener Criterion.—If it were possible for a filter to have a gaussian-error curve $A(\omega) = e^{-\omega^2}$ as amplitude characteristic (Fig. A·1), and zero phase lag, so that the complex system function is $H(\omega) = e^{-\omega^2}$, then the step-function response

$$f(t) = \frac{1}{2\pi} \int_{-\infty-jc}^{\infty-jc} \frac{H(\omega)}{j\omega} e^{j\omega t} d\omega$$

of such a filter is the error function (Fig. A·2)

$$f(t) = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{t} e^{-t^2/4} dt.$$

Now consider an idealized low-pass filter, having for amplitude characteristic the function $A(\omega) = 1$ for $-1 < \omega < 1$ and $A(\omega) = 0$ other-

wise (Fig. A·3), and zero phase lag, so that the complex system function is $H(\omega) = 1$ for $-1 < \omega < 1$, and $H(\omega) = 0$ otherwise. Then the step-function response is the sine-integral function (Fig. A·4)

$$f(t) = \frac{1}{\pi} \int_{-\infty}^{t} \frac{\sin t}{t} dt.$$

There is something very disquieting in Figs. A·2 and A·4, in that they show response for time $t < 0$ although a physical system obviously cannot react to a step function before the step function has even been applied.
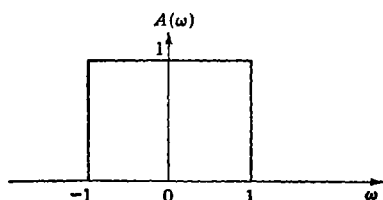
FIG. A·3.—Idealized low-pass filter $A(\omega) = 1$ for $-1 < \omega < 1$, $= 0$ otherwise.
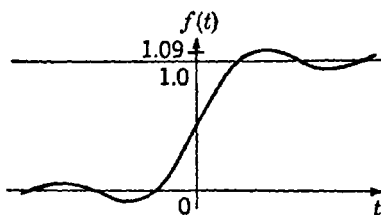
FIG. A·4.—Step-function response of idealized low-pass filter.

The reason for this difficulty lies in the false assumption that it is possible for a filter to have a gaussian-error, or idealized low-pass amplitude characteristic, and zero phase lag.

Moreover, the difficulty cannot be removed by associating a linear phase lag, no matter how great, with the gaussian-error or idealized low-pass amplitude characteristics. Although a linear phase lag has the effect of shifting the time origin in Figs. A·2 and A·4 to the left, no matter how great this shift to the left, it would still remain true that there would be some response even before the input switch were closed.

As a practical matter, a response of, for example, one-hundred-millionth per cent for time < 0 can be neglected, in the sense that it has no experimental meaning; hence these filters are "practically" realizable, with large enough phase lags (see Sec. A·3 for a more precise discussion of this point).

Nevertheless, it is of considerable mathematical interest to determine whether or not there is *any* phase function *whatever*, linear or otherwise, that can be associated with either the exact gaussian-error amplitude curve or the exact idealized low-pass amplitude curve so as to give *strictly zero* transient response for $t < 0$. The answer is no, and this negative answer is a consequence of an extraordinarily elegant theorem of Paley and Wiener,[1] which states, in the language of engineering:

[1] R. E. A. C. Paley and Norbert Wiener, "Fourier Transforms in the Complex

If $A(\omega)$ is an arbitrary amplitude characteristic, i.e. an even non-negative function of frequency, having a Fourier transform, $A(\omega)$ is said to be "realizable" if it is possible to associate with the amplitude function $A(\omega)$ a phase-lag function $\phi(\omega)$ (not necessarily linear) such that the combined frequency function $A(\omega)e^{-i\phi(\omega)}$ yields zero transient response for $t < 0$. (This is clearly a very general and nonrestricted conception of realizability.) Then

Theorem of Paley and Wiener.—A necessary and sufficient condition for an amplitude function $A(\omega)$ to be realizable is that

$$\int_{-\infty}^{\infty} \frac{|\log A(\omega)|}{1 + \omega^2}\, d\omega \tag{1}$$

be finite.[1]

In words: A realizable amplitude characteristic cannot have too great a total attenuation.

A realizable characteristic may have infinite rejection for a discrete set of frequencies, but it cannot have infinite rejection over a *band* of frequencies.

The theorem of Paley and Wiener has the great merit that although it is of a complex-variable nature in content and proof, Criterion (1) itself is entirely expressed in the domain of real variables.

Both the gaussian-error curve and the idealized bandpass characteristic attenuate too much to satisfy Criterion (1). Hence neither is exactly realizable. *This is not just a practical difficulty but a theoretical impossibility.*

The mathematical meaning of this nonrealizability is as follows: For every sequence of filters whose amplitude curves approximate more and more closely to the gaussian-error curve or to the idealized low-pass amplitude curve, it will be found that the successive phase characteristics diverge.

A·2. Examples. *Gaussian-error Curve.*—To illustrate this divergence of phase characteristics one may examine a familiar method of approxi-

---

Domain," *Am. Math. Soc. Colloq. Pub.*, **19** (1934), Chap. I, "Quasi-analytic Functions," Theorem XII, pp. 16 and 17.

The exact quotation is as follows: "Let $\phi(x)$ be a real nonnegative function not equivalent to zero, defined for $-\infty < x < \infty$, and of integrable square in this range. A necessary and sufficient condition that there should exist a real- or complex-valued function $F(x)$ defined in the same range, vanishing for $x \geqq x_0$ for some number x_0, and such that the Fourier transform $G(x)$ of $F(x)$ should satisfy $|G(x)| = \phi(x)$ is that

$$\int_{-\infty}^{\infty} \frac{|\log \phi(x)|}{1 + x^2}\, dx < \infty."$$

[1] Integral (1), if it exists, is in fact an evaluation for a special case of the expression giving the logarithm of a complex impedance function in terms of the logarithm of its absolute value.

mating in amplitude to a gaussian-error curve to see how it leads to divergent phase functions.
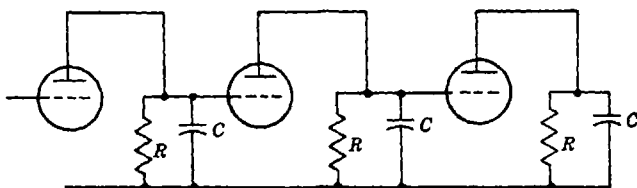
FIG. A·5.—Cascade of n identical RC-coupled pulse-amplifier stages.

Consider a cascade of n identical RC-coupled pulse-amplifier stages (Fig. A·5). The amplifier has the complex selectivity function

$$(1 + j\omega RC)^{-n}, \qquad (2)$$

whose absolute value is $[1 + (\omega RC)^2]^{-\frac{n}{2}}$. If the over-all 3-db bandwidth is held constant as n increases, e.g., if the 0.707 voltage point is maintained

FIG. A·6.—Absolute value vs. frequency curves of (a) cascade of n identical RC-coupled stages, (b) gaussian-error curve; both curves adjusted to have their 3-db points at $\omega = 1$.

at $\omega = 1$, then $RC = (2^{1/n} - 1)^{\frac{1}{2}}$. From the Taylor's series approximation $2^{1/n} - 1 \approx (\ln 2)/n$, Expression (2) becomes

$$\left(1 + \frac{j\omega \sqrt{\ln 2}}{\sqrt{n}}\right)^{-n}. \qquad (3)$$

The absolute value of Expression (3) is

$$A_n(\omega) = \left(1 + \frac{\omega^2 \ln 2}{n}\right)^{-\frac{n}{2}}, \qquad (4)$$

and the phase lag of Expression (3) is

$$\phi_n(\omega) = n \tan^{-1} \frac{\omega \sqrt{\ln 2}}{\sqrt{n}}. \qquad (5)$$

Compare Expression (4) with a gaussian-error curve having $\omega = 1$ as its 0.707 point (Fig. A·6), the formula for which is

$$e^{-\omega^2 \frac{\ln 2}{2}}.$$

It is plain from Expression (4) that $A_n(\omega)$ converges to $e^{-\omega^2 \frac{\ln 2}{2}}$, as follows from the definition of e as $\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$; on the other hand $\phi_n(\omega)$ diverges, by becoming indefinitely large for all ω. That is, the sequence of cascaded RC-coupled stages tends to a gaussian-error curve in amplitude, *but* (and this is the whole point of the example) the associated phase functions (5) diverge.

Idealized Low-pass Filter.—A similar state of affairs prevails for the idealized low-pass filter. This filter is *not* realizable (not because of the *steepness* with which the amplitude curve cuts off but rather because the amplitude curve cuts off to *zero*). The approximate low-pass filter
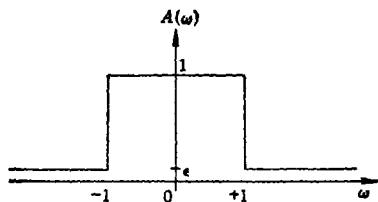
FIG. A·7.—Approximate low-pass filter, $A(\omega) = 1$ for $-1 < \omega < 1$, $= \epsilon$ otherwise.
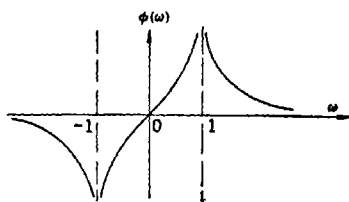
FIG. A·8.—Phase vs. frequency curve, $\phi(\omega) = |\ln \epsilon| \cdot \ln \left| \dfrac{1 + \omega}{1 - \omega} \right|$, corresponding to absolute-value curve of Fig. A·7.

of Fig. A·7, which cuts off to ϵ, is realizable, however, no matter how small ϵ may be. The corresponding phase-lag function is[1] (Fig. A·8) proportional to

$$\phi(\omega) = |\ln \epsilon| \ln \left| \frac{1 + \omega}{1 - \omega} \right|.$$

Filters Having $|\sin \omega/\omega|$ *and* $\sin^2 \omega/\omega^2$ *as Absolute Values.*—These two amplitude characteristics (Fig. A·9) do satisfy the Paley-Wiener criterion
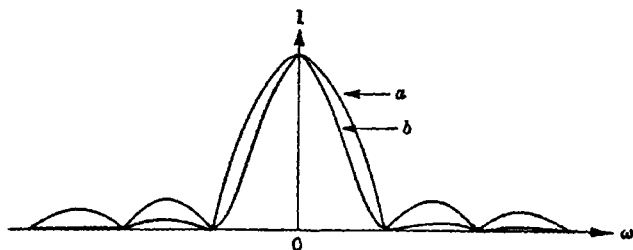
FIG. A·9.—Absolute value vs. frequency curve of (*a*) $(\sin \omega/\omega)$, and (*b*) $(\sin^2 \omega/\omega^2)$.

(1) and, as it happens, are very simply realizable, with linear phase lags, in the forms $(\sin \omega/\omega)e^{-j\omega}$ and $(\sin^2 \omega/\omega^2)e^{-2j\omega}$. Routine integration shows the step-function responses to be as shown in Fig. A·10. (The step-function response of the $(\sin^2 \omega/\omega^2)e^{-2j\omega}$ filter is made up of two parabolic arcs.)

The fact that $|\sin \omega/\omega|$ and $\sin^2 \omega/\omega^2$ satisfy Criterion (1) does not mean that these two amplitude curves, with their infinity of arches, are exactly realizable by means of a finite number of resistances, inductances,

[1] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, Oxford, New York, 1937, p. 121.

and capacitances, but it does mean that it is possible to have a sequence of filters, each made up of a finite number of lumped circuit elements, that tend to $|\sin \omega/\omega|$ or $\sin^2 \omega/\omega^2$ in amplitude and *whose phase functions converge* (as a matter of fact, converge to linear phase lags of ω and 2ω radians respectively).
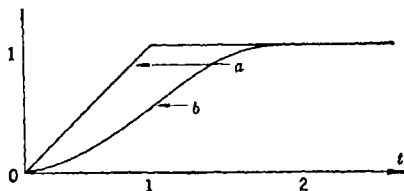
Fig. A·10.—Step-function response of (*a*) $(\sin \omega/\omega)e^{-j\omega}$ filter, and (*b*) $(\sin^2 \omega/\omega^2)e^{-2j\omega}$ filter.

Indeed, an example of such a sequence of finite filters has been suggested by E. A. Guillemin. In pulse-forming networks of the sort shown in Fig. A·11, as the number of elements increases, the current I_2 tends
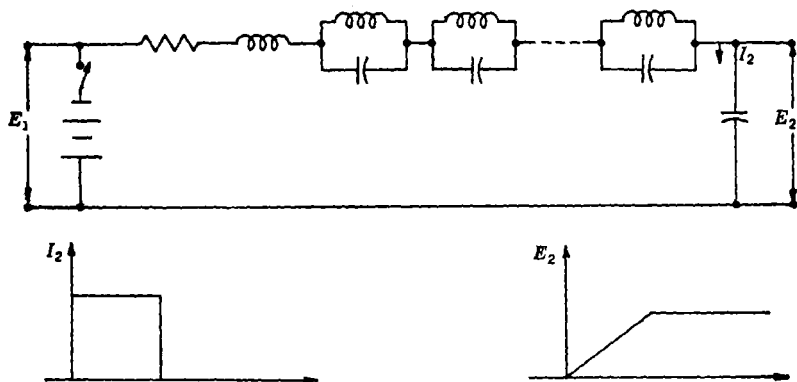
Fig. A·11.—Idealized pulse-forming network. The system function for E_2/E_1 is $(\sin \omega/\omega)e^{-j\omega}$.

more and more to a square pulse, and hence the voltage E_2 to an inclined step. Therefore the ratio E_2/E_1 (see Fig. A·10) tends to $(\sin \omega/\omega)e^{-j\omega}$.

It is even possible that the complex functions $(\sin \omega/\omega)e^{-j\omega}$ and $(\sin^2 \omega/\omega^2)e^{-2j\omega}$ can be exactly realized by a finite number of lines with distributed constants.

A·3. The Practical Meaning of the Paley-Wiener Criterion.—For sufficiently negative values of t, the Fourier transform of an arbitrary amplitude characteristic $A(\omega)$ [subject only to the (physically inevitable) condition of having a Fourier transform] is as small as is desired, whether or not the Paley-Wiener Criterion (1) is satisfied. Consequently

it is possible to approximate to the given amplitude characteristic, as closely as desired over any finite frequency interval,[1] by a real filter with large phase lag, i.e., large time delay. If it happens that Criterion (1) is satisfied, then the theorem of Paley and Wiener asserts that there is, in fact, a way of carrying out this approximation so that the phase functions converge to some finite $\phi(\omega)$, thereby yielding a *finite* time delay.

The practical significance of the Paley-Wiener criterion is then as follows:

An amplitude characteristic having a Fourier transform can be approximated arbitrarily closely by finite filters whether it satisfies the Paley-Wiener criterion or not; but if it does satisfy the Paley-Wiener criterion, then the entire approximation process can be carried out within the bounds of a finite time delay, whereas if the Paley-Wiener criterion is not satisfied, the approximation process necessitates an infinite time delay.

The following may be said in summary: Amplitude characteristics having a Fourier transform can be divided into three types:

1. Those exactly realizable by a finite number of R, L, C elements, for example, $1/\sqrt{1 + \omega^2}$. Amplitude characteristics of this type satisfy the Paley-Wiener criterion, and in addition the corresponding complex frequency characteristic $[1/(1 + j\omega)$ for the above example] is rational.

2. Those exactly realizable by an infinite number of R, L, C elements (or by a finite number of lines with distributed constants), e.g., $|\sin \omega/\omega|$. These are exactly realizable in the sense that as the approximation to the given amplitude characteristic becomes closer and closer, the associated phase functions converge also. Amplitude characteristics of this type satisfy the Paley-Wiener criterion.

3. Those not exactly realizable at all but *approximable* arbitrarily closely over any finite frequency interval, although only at the expense of increasingly large time delay, e.g., $e^{-\omega^2}$. Amplitude characteristics of this type do not satisfy the Paley-Wiener criterion.

[1] But the approximating filter would then behave differently at $\omega = \infty$ and would therefore have an entirely different abstract-mathematical character.

APPENDIX B

CALCULATION OF LOAD-TUNING CONDENSER

By Duncan Mac Rae, Jr.

A more refined treatment of the method of selecting the tuning condenser for the output stage of the two-stage amplifier (Sec. 9·6) can be considered if extremely small phase shift is desired.   For a circuit with cathode- and plate-circuit impedances, the equivalent circuit can be
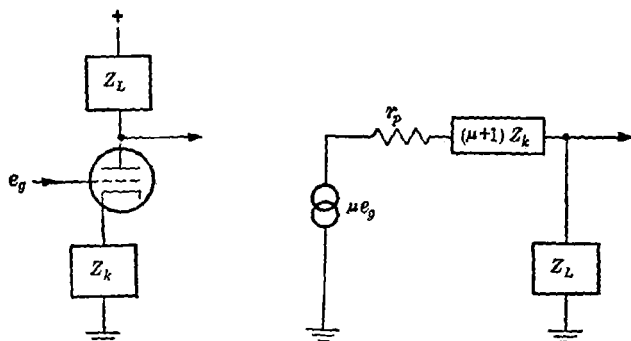
Fig. B·1.—Equivalent circuit of amplifier with impedance in cathode circuit.
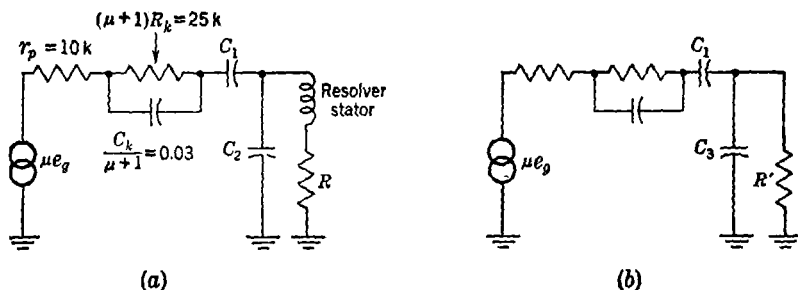
Fig. B·2.—Equivalent circuits of output stage. R' = apparent resistance of parallel-resonant circuit. C_3 = excess of C_2 over value required for resonant tuning.   (a) Equivalent circuit of output stage; (b) simplified equivalent circuit if stator is tuned capacitatively.

drawn as is shown in Fig. B·1.   For the operating point chosen, the 6C4 has $\mu \approx 16$, $r_p \approx 10,000$ ohms.

Therefore with the constants of Fig. 9·15b in the output stage, the equivalent circuit is as shown in Fig. B·2.   The stage can be adjusted to

728

zero phase shift if the load is tuned slightly capacitively. The magnitude of apparent capacitance can be determined as follows: The *LRC*-circuit may be considered as composed of a pure resistive tuned circuit (20*k*) in parallel with an additional capacitor C_3; if the cathode impedance is neglected (or if it is replaced by the corresponding series *RC*-circuit), the equivalent circuit of Fig. B·1 becomes part of a Wien bridge, and the condition for zero phase shift (in the case $Z_k \approx 0$) is $\omega^2 r_p R^1 C_1 C_3 = 1$; then,

$$C_3 = \frac{1}{(2\pi 500)^2 \times 17,000 \times 7500 \times 10^{-7}} = 0.008 \ \mu\text{f}.$$

This circuit requires an extremely accurate selection of condensers, which may be feasible for experimental equipment but is impractical when wide production tolerances and temperature coefficients are involved.

APPENDIX C

DRIFT OF VACUUM-TUBE CHARACTERISTICS UNDER CONSTANT APPLIED POTENTIALS

By John W. Gray

The severest limitation is imposed upon direct-coupled amplifiers by the fact that the characteristics of any vacuum tube will gradually shift while the tube is in operation even though the conditions of operation are all held constant. For example, if the plate and grid voltages and filament or heater voltage are held constant, the plate current will drift slightly even after temperature equilibrium has been reached. This drift (except in the case of "microphonics" or dimensional changes due to shock or vibration) is attributable to cathode change resulting in variation of the average initial electron velocity of emission. This premise is substantiated by the nature of the shift of tube characteristics; it is found to be of the same type as that caused by variation of cathode temperature (see Fig. 11·7). Generally, as a tube ages, the drift is in the direction of decreasing emission; but short-time drifts are erratic as to sense and rapidity.

As in the case of cathode temperature effect (Sec. 11·6), the most convenient way to express the drift is in terms of the variation of control-grid bias that is required to hold the plate current constant with a given value of plate voltage (and screen-grid voltage). In addition to being more directly descriptive of the effect upon amplifiers, this quantity is more independent of plate current and voltage than is the drift of current with constant grid bias.

The first stage of an amplifier is usually the only stage wherein the tube drift is of much importance, since the drift of the second stage is less effective by a factor equal to the voltage gain of the first, etc. (Exceptions occur when the first stage is a cathode follower or an electrometer tube designed for low grid current but having very low gain.) This fact is fortunate, because the drift of a power stage is greater than that of a voltage amplifier operated at low power. Minimum drift in most ordinary receiving tubes occurs at plate currents of between 0.1 and 1 ma (10 to 100 μa for the small, filament types) and at plate and screen voltages as low as permissible from the standpoint of control-grid current, although neither current nor voltage is at all critical. The drift takes place as a general shift of the characteristics, as for cathode temperature change.

Rate of drift is relatively great in new tubes. Most of the drift during the life of a tube usually occurs during the first hundred hours of cathode operation. It is therefore advisable to age a tube, with filament voltage applied, for several days before use where stability is essential.

Figures C·1 and C·2 show the long-term drift of a group of 6SL7 double triodes. Readings of grid bias required, for a certain plate current at a certain voltage, were made at the times indicated. Heater voltage was held constant only during readings; the rest of the time it varied with the a-c line and was turned off several times. Because of the much more rapid variation at first, a logarithmic time scale is used.

Differential amplifiers are often used as input stages to help neutralize the effect of heater-voltage variation. The question arises as to whether any cancellation of drift is also accomplished. For short-time drift, where a zero adjustment can be made every day or so, the drift of each tube is erratic, and the use of two tubes instead of one in the first stage will multiply the probable drift by $\sqrt{2}$. For intervals of weeks or months, however, Figs. C·1 and C·2 seem to indicate that there is some drift cancellation between double triodes.

Figures C·1 and C·2 show that the amount of drift over periods of weeks is measured in tens of millivolts. (Some tubes are considerably worse than the group illustrated.) Thus, unless a zero adjustment can be made at frequent intervals, it is rather pointless to make an amplifier with a gain such that its output range is traversed as the input voltage changes by only 1 mv. If the zero set can be made just before the application, if a selected tube is used, and if heater voltage is constant, it is possible for input variations of a fraction of a millivolt to have significance.
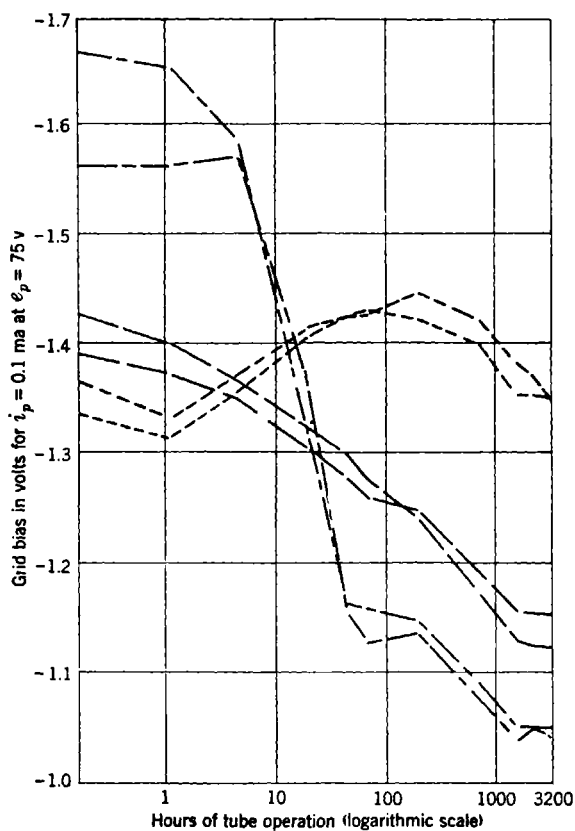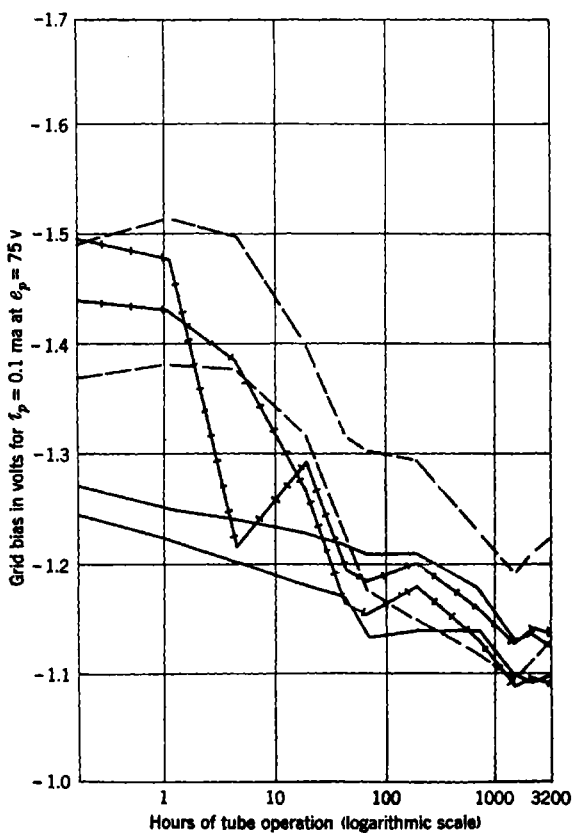
FIGS. C·1 and C·2.—Drift of three 6SL7's (six triodes) dur-

ing 3200 hr.   Triode pairs are indicated by similar lineation.

Index

A