

TRANSISTORS I

RCA LABORATORIES

TRANSISTORS I

*A selection of papers describing
research and development work
of the
Radio Corporation of America
as it applies to the
theory, fabrication, and application
of transistors*

MARCH 1956

Published by
RCA LABORATORIES
Princeton, N. J.

© 1956 by Radio Corporation of America
All Rights Reserved

Printed in U. S. A.

PREFACE

The Transistor and related semiconductor devices are revolutionizing many aspects of electronics with amazing rapidity. Within the Radio Corporation of America, there has been such extensive research and development work on semiconductors, transistors, and their applications that scientific and engineering reports have accumulated in an unprecedented manner. Only a part of this work has appeared in the technical literature.

Because of the high quality and large quantity of the as-yet-unpublished material available within RCA, it was concluded that the most effective method of bringing the information to the attention of those who would profit by it would be the publication of this book. It was believed that the addition of a few published papers, and abstracts of others, would help to round out the over-all picture and would considerably enhance the usefulness of the book. Accordingly, the volume is made up of 31 complete papers (496 pages) which have not been published elsewhere, 10 papers (163 pages) which have appeared in periodicals, and 46 abstracts. The papers are grouped into six sections: General, Materials and Techniques, Devices, Fluctuation Noise, Test and Measurement Equipment, and Applications. It should be noted that the word "Transistors" as used in the title is intended to include semiconductor diodes.

The publishers are indebted to the Institute of Radio Engineers and to the Institute of Physics (Great Britain) for granting permission to reproduce papers which have appeared in the journals of these organizations.

TRANSISTORS I

CONTENTS

	Page
Preface	iii
General	
Basic Transistor Device Concepts	H. JOHNSON 1
New Advances in the Junction Transistor	E. W. HEROLD 27
Materials and Techniques	
A Switched-Zone Furnace for Germanium Purification	P. G. HERKART AND S. M. CHRISTIAN 59
Continuous-Process Apparatus for Growing Single-Crystal Germanium	H. NELSON 66
The Preparation of Single and Multiple P-N Junctions in Single Crystals of Germanium	A. R. MOORE 77
Transistor Fabrication by the Melt-Quench Process....	J. I. PANKOVE 82
Microscopic Examination of Germanium Crystals and Transistors... ..	S. G. ELLIS 97
Calculations of Alloying Depth of Indium in Germanium..	L. PENSAK 112
Uniform Planar Alloy Junctions for Germanium Transistors	C. W. MUELLER AND N. H. DITRICK 121
Some Aspects of Thermal Conversion in Germanium....	H. KROEMER 132
Surface Treatment of Silicon for Low Recombination Velocity.....	A. R. MOORE AND H. NELSON 136
Devices	
P-N-P Transistors Using High-Emitter-Efficiency Alloy Materials... ..	L. D. ARMSTRONG, C. L. CARLSON, AND M. BENTIVEGNA 144
Recent Advances in Power Junction Transistors.....	B. N. SLADE 153
A Silicon N-P-N Junction Transistor by the Alloy Process..	H. NELSON 172
Junction Transistor Designed for Iterative Operation at Low Frequencies	L. J. GIACOLETTO 182
A P-N-P Triode Alloy Junction Transistor for Radio-Frequency Amplification	C. W. MUELLER AND J. I. PANKOVE 189
The Drift Transistor	H. KROEMER 202
A Variable-Capacitance Germanium Junction Diode for UHF.....	L. J. GIACOLETTO AND J. O'CONNELL 221
Fluctuation Noise	
A Determination of 1/f Noise Sources in Semiconductor Diodes and Triodes	W. H. FONGER 239
The Noise Factor of Junction Transistors	L. J. GIACOLETTO 296
Noise Considerations of P-N-P Junction Transistors..	J. W. ENGLUND 309

	Page
Test and Measurement Equipment	
A Test Set for Transistor Performance Measurement at 455 KilocyclesD. D. HOLMES, L. A. FREEDMAN, AND T. M. SCOTT	322
A High-Frequency Measuring Equipment for TransistorsW. F. SANDS	336
Equipment for Measuring Junction Temperature of an Operating TransistorJ. OLLENDORF AND K. E. LOOFBOURROW	353
Evaluation of Transistors for Class-B Power AmplifiersW. HASENBERG	361
Applications	
Temperature Effects in Circuits Using Junction Transistors.....H. C. LIN AND A. A. BARCO	369
Stability Considerations in Transistor I-F AmplifiersD. D. HOLMES AND T. O. STANLEY	403
Cross-Modulation in Transistor R-F Amplifiers.....D. D. HOLMES	422
Performance of a Radio-Frequency Alloy Junction Transistor in Dif- ferent CircuitsL. J. GIACOLETTO	431
An Experimental Automobile Receiver Employing Transistors.....L. A. FREEDMAN, T. O. STANLEY, AND D. D. HOLMES	458
A Developmental Pocket-Size Broadcast Receiver Employing Trans- istors.....D. D. HOLMES, T. O. STANLEY, AND L. A. FREEDMAN	477
Design Considerations in Class-B Complementary Symmetry CircuitsT. O. STANLEY AND T. M. SCOTT	497
Transistor Audio AmplifiersA. I. ARONSON	515
A 20-Watt Transistor Audio AmplifierH. C. LIN	536
Modulated Transistor Oscillators and Their Applications...H. C. LIN	547
Transistorized Sync Separator for Television ReceiversH. C. GOODRICH	561
Transistorized Vertical Deflection for Television ReceiversM. B. FINKELSTEIN	579
Automatic-Frequency Control of Television Receivers Using Junction DiodesW. Y. PAN AND O. RAMANIS	598
Junction Transistor Switching Characteristics.....D. E. DEUTCH	609
A Novel Ring Counter Using Junction Transistors...D. E. DEUTCH	640
A Junction-Transistor Counter with High-Speed CarryT. P. BOTHWELL, G. W. BOOTH, AND E. P. ENGLISH	646
Abstracts	661

BASIC TRANSISTOR DEVICE CONCEPTS

BY

HARWICK JOHNSON

RCA Laboratories,
Princeton, N. J.

Summary—The properties of a p-n junction in a semiconductor are developed largely on the basis of physical arguments. From this conceptual point of view, the roles played by the junction barrier and by diffusion phenomena in minority carrier injection, diode conductance, transition capacitance, and diffusion capacitance are developed. This conceptual development is fortified by an elementary analysis. The utilization of these properties of a p-n junction in various transistor devices is discussed in a qualitative fashion.

INTRODUCTION

THIS paper considers some of the elementary device concepts which are essential to the physical understanding of the operation of transistor devices. The discussion is descriptive in nature with attempts to fortify the consequences of the physical reasoning with some heuristic analysis. While the results thus obtained have the correct first-order functional dependence, multiplying constants may differ from those obtained from a rigorous analysis. From such physical reasoning and simple analysis, one can develop a sense of understanding that many complicated analyses fail to convey.

Because most transistor devices are based on p-n junctions, a study of their properties constitutes the bulk of this paper; a brief description of how these properties are utilized in the operation of certain transistor devices is included.

PHYSICAL CONCEPTS OF A P-N JUNCTION

Consider a p-n junction to be formed by bringing together a p-type semiconductor and an n-type semiconductor as in Figure 1. This is not, today, a recommended method of forming a p-n junction. However, as an aid in visualizing the physical processes that go on, this model provides a simple initial state.

The p-type semiconductor contains impurity atoms (acceptors) having one less valence electron each than do the atoms of the host crystal. This robs the normal lattice of one electron for each impurity atom thus creating an electron vacancy or a hole. Since the hole repre-

sents an electron deficiency, it is represented as a positive charge. Since the impurity atom (originally electrically neutral) has acquired an additional electron, it is represented as a negative charge. The impurity atom is fixed in the crystal lattice but the hole is free to move and constitute a current.

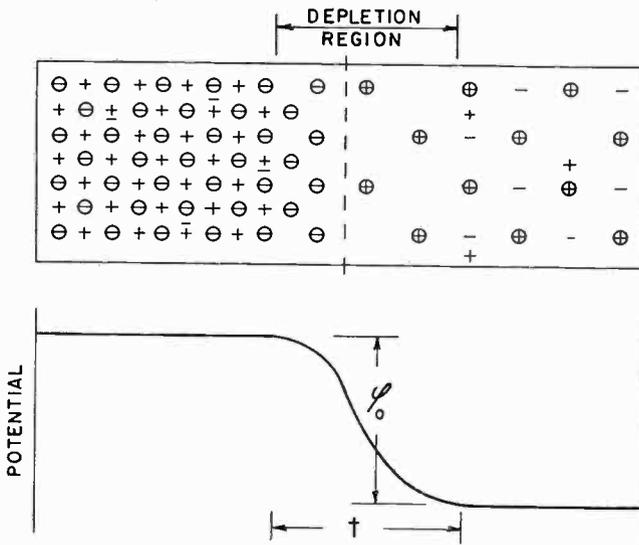
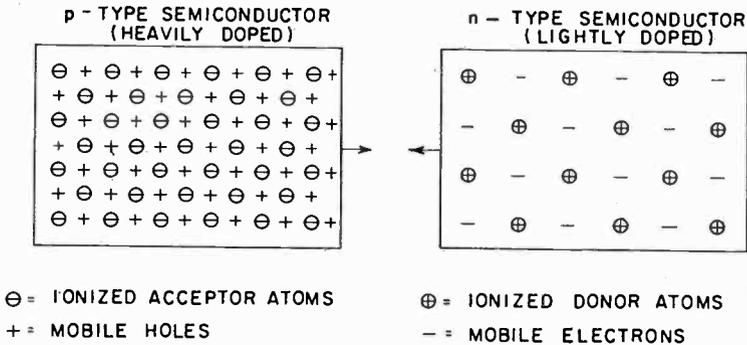


Fig. 1—Physical picture of a p-n junction.

The converse situation exists in the n-type semiconductor where the impurity atoms (donors) now each have one extra valence electron over those required by the crystal lattice. This extra electron is given up; whereupon the freed electron wanders through the crystal and is free to constitute a current. The fixed impurity atom having given up

an electron becomes charged positively. In the separated semiconductors, the mobile charges (electrons and holes) are uniformly distributed throughout the bulk material.

We now ask, what are the consequences of bringing together these two pieces of semiconductor? It may be expected that the mobile holes and electrons would tend to diffuse throughout the composite crystal. However, we shall see that restraining forces are set up which counteract this tendency so that the bulk of the holes remain in the p-type material and the bulk of the electrons remain in the n-type material. Thus, the charge densities in regions removed from the junction are essentially unaffected by this juncture.

It should be pointed out that the representation in Figure 1 is incomplete in that the holes and electrons are considered to be derived only from the ionization of the impurity atoms. In addition to the carriers thus derived, there are, in both types of semiconductor, additional electron-hole pairs formed by the release of electrons from atoms of the host crystal. The energy for ionization of these atoms comes from the thermal energy of the lattice. At normal temperatures and in material of the type ordinarily used for transistor devices, the number of electron-hole pairs thus generated is relatively small in comparison with the number of charge carriers derived from ionization of the impurity atoms.

Let us return now to a consideration of the mechanism whereby the holes and electrons are restrained from diffusing throughout the composite crystal. For simplicity consider one type of charge — the holes of the p-type region, for example. The holes tend to diffuse out of the p-type region where their density is high into the n-type region where there are but few holes. The fixed charges (negatively ionized acceptors) are then no longer electrically compensated and give rise to a negative charge density near the transition region between the p- and n-type materials. This negative charge being opposite to that of the holes will tend to retard the flow of holes out of the p-type region. Similarly, diffusion of electrons from the n-type region results in an uncompensated positive fixed charge density on the *n* side of the junction. This further restricts the loss of holes from the p-type region. Thus, under normal equilibrium conditions, the tendency for charges to diffuse across the junction sets up restraining forces in the form of uncompensated fixed charges in the transition region which act to keep the holes in the p-type region and the electrons in the n-type region.

To establish the uncompensated charge densities, the transition region is depleted of mobile charges and this region is often referred to as the depletion region or as the space-charge region. The uncom-

pensated charge densities set up an electric dipole resulting in an electrostatic potential difference, ϕ_0 , across the junction. Hence, reference is often made to this region as the barrier region, the thickness of the depletion layer being the barrier thickness, t , and the electrostatic potential, ϕ_0 , the barrier height. As in Figure 1, potential energy diagrams will be drawn, in this paper, so that electrons run down hill and holes run up hill.

CURRENT FLOW IN A P-N JUNCTION

The discussion thus far has referred to a p-n junction under conditions of thermal equilibrium with no applied voltage. Before considering a voltage applied to the junction, let us consider the nature of the currents that might flow.

First we have the picture of holes tending to diffuse across the junction from the p-type region but being hindered by a barrier of height ϕ_0 . The number that do succeed in crossing will depend on the density of holes in the p-type region, $p_p = N_a$ and on the barrier height, ϕ_0 . Statistical mechanics tells us that, in situations like this, the dependency on the barrier height is exponential and so we have for the hole density on the n-region side of the barrier,*

$$p_0 = p_p e^{-\frac{q}{kT} \phi_0} \quad (1)$$

We refer to this as an injected hole density in anticipation of hole injection in transistors. The current will be proportional to the injected charge density, so we write

$$I_i = \text{constant} \times e^{-\frac{q}{kT} \phi_0} \quad (2)$$

An analogous flow of electrons into the p region occurs and we may take Equation (2) as being the sum of the hole flow to the right and electron flow to the left. Although the particle flow is conceived of as being in opposite directions for electrons and for holes, the current flow is in the same direction since the holes and electrons are oppositely charged.

Recall that we are still discussing a junction to which no external potentials have been applied. Under equilibrium conditions no net current must flow, so there must be a counter flow to achieve this balance.

* A list of symbols used may be found in the appendix.

Let us see how this current arises. In the n region small numbers of holes are normally present as a result of electron-hole pair generation by thermal processes. There is also an internal potential established across the junction; this is in the direction to extract holes from the n region and sweep them into the p region. This, then, constitutes a reverse flow of holes which will balance those diffusing over the barrier into the n region. We ask how large is this current? The barrier potential can, of course, only extract those holes in the immediate vicinity, and in order for additional holes to be extracted they must diffuse to the barrier. Thus the magnitude of this current will be governed by the laws of current flow by diffusion. The "Ohms law," so to speak, for diffusion current flow is simply (for holes)

$$I_p = -qD_p \frac{\delta p}{\delta x}, \quad (3)$$

i.e., the current is proportional to the charge density gradient. This simply says that particles tend to move from regions of high concentration to regions of low concentration. Thus to determine this current we need to determine the density gradient.

A first-order approximation for the density gradient may be obtained by physical arguments concerning the density at the junction and at some distance away from the junction. Referring now to Figure 2; at the junction, it may be argued, the hole density is zero since the holes will be immediately swept away by the barrier potential. Now a charge carrier in a semiconductor leads a rather hazardous life since it was derived from an atom which normally would like to get it (or a similar charge) back. Thus, there is always a probability that an electron will drop back into a vacancy represented by a hole, thus ending its life (for the moment) as a free charge capable of carrying current. The barrier potential cannot collect those carriers which are created so far away that they are lost by recombination before diffusing to the barrier. If L_p is the average distance a hole in the n -type material can diffuse before recombination, then at a distance L_p from the junction, the normal hole density, p_n , will be unaffected by the presence of the junction. The density gradient then is p_n/L_p and the diffusion current flow is

$$I_p = -qD_p \frac{p_n}{L_p} = -\frac{kT}{q} \frac{b}{(1+b)^2} \frac{\sigma_i^2}{\sigma_n L_p}, \quad (4)$$

where the latter form may be obtained from simple manipulation using

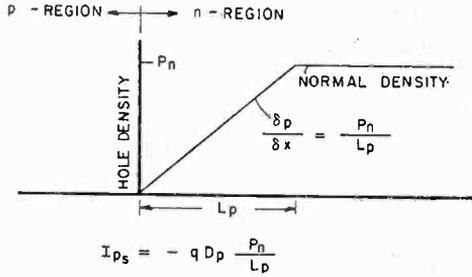


Fig. 2—Hole saturation current. Current flow is to left.

$$\sigma_i = q (\mu_p + \mu_n) n_i; \quad b = \frac{\mu_n}{\mu_p}; \quad p_n n_n = n_i^2; \quad \sigma_n = q \mu_n n_n; \quad \mu = \frac{q}{kT} D.$$

By analogy the electron flow from the *p* region to the *n* region is (Figure 3)

$$I_{n_s} = -q D_n \frac{n_p}{L_n} = -\frac{kT}{q} \frac{b}{(1+b)^2} \frac{\sigma_i^2}{\sigma_p L_n}. \tag{5}$$

The total counter flow of current under equilibrium conditions is then

$$I_s = I_{p_s} + I_{n_s} = -\frac{kT}{q} \frac{b}{(1+b)^2} \sigma_i^2 \left(\frac{1}{\sigma_n L_p} + \frac{1}{\sigma_p L_n} \right). \tag{6}$$

Note that these currents do not depend on the barrier height. This

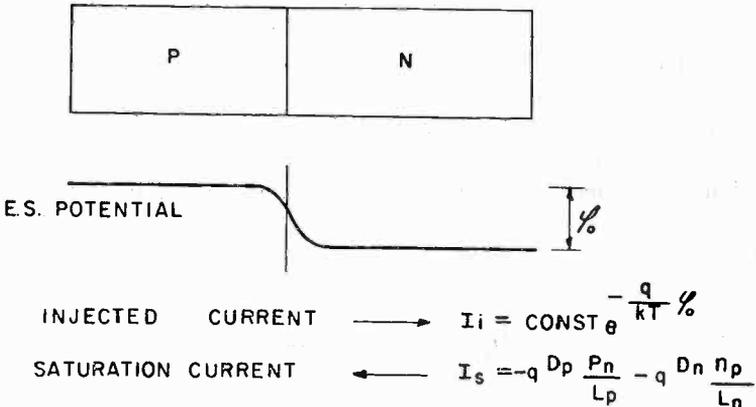


Fig. 3—Nature of current flow. Under equilibrium conditions $I_i = I_s$.

reverse flow then will *not* depend on the applied voltage; it is known as the saturation current.

Under equilibrium conditions, the saturation current just balances the injected current [Equation (2)]. The barrier height adjusts itself so that these currents are balanced for both hole and electron flow individually and no net current flows across the junction.

At this point we return to ask what happens to the injected carriers as in Equation (1) that succeed in overcoming the barrier. The injected holes tend to diffuse away from the barrier into the n-type material. As charge carriers, they are subject to the same hazardous life as the holes normally present, and on the average diffuse a distance L_p before being lost by recombination. Thus, the injected hole density distribution is approximately as shown in Figure 4. Recalling the form for current flow due to diffusion

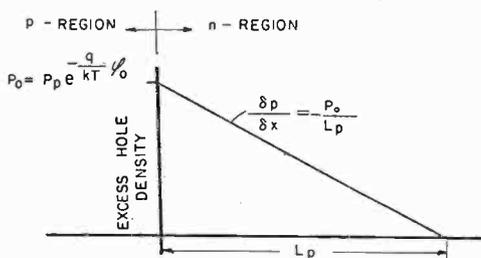


Fig. 4—Injected hole current.

$$I_p = -q D_p \frac{\delta p}{\delta x}$$

The hole current due to the injected charge density is simply

$$I_p = q D \frac{p_0}{L_p}$$

where p_0 is related to the barrier height by Equation (1). A similar relation holds for the electrons surmounting the barrier in the opposite direction.

If an external potential is applied to the junction, the effect is to alter the height of the barrier. As noted above, this does not affect the saturation component of current flow but will change the injection current flow. The injection current flow now becomes

$$I_i = \text{constant} \times e^{-\frac{q}{kT}(\phi_0 - V)}, \tag{7}$$

where V is taken positive when a positive potential is applied to the p side of the junction, as in Figure 5.

The constant in Equation (7) can readily be evaluated since we know that, when $V = 0$, the injection current must equal the saturation current. A little manipulation shows that the total flow is

$$I = I_s \left(e^{\frac{q}{kT} V} - 1 \right) \quad (8)$$

and the same form applies individually to both the electron and hole components of the current. When V is positive a very large current will flow because of the exponential relation. When V is negative the cur-

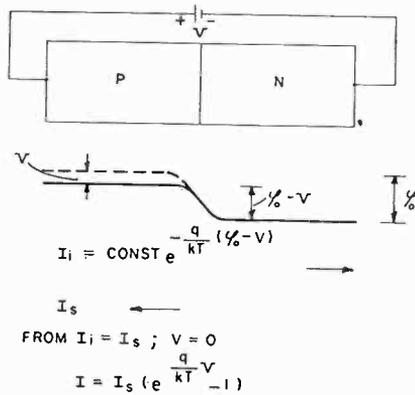


Fig. 5—Diode relation.

rent will be extremely small and is the saturation current. It is seen from this relation that the rectification properties of a p-n junction have no direct relation to the proportion of current carried by holes or electrons. The ratio of forward to reverse current being simply

$$\left(e^{\frac{q}{kT} V} - 1 \right).$$

On the other hand for many device applications, the property which enables a p-n junction to inject holes into n-type material or to inject electrons into a p-type material is essential. It will be of interest to discuss these properties further.

Let us first, however, indicate the a-c conductance obtained by differentiation of Equation (8).

$$g = \frac{\delta I}{\delta V} = \frac{q}{kT} I_s \frac{q}{e^{kT}} V. \quad (9)$$

$$\text{If } V \gg \frac{kT}{q}, g = \frac{q}{kT} I.$$

$$\text{If } V \ll 0, \quad g \rightarrow 0.$$

In particular, note the extremely small reverse conductance given by this simple theory. Subsequently we shall see how another phenomenon in a p-n junction leads to a finite a-c conductance.

HOLE AND ELECTRON CURRENTS — INJECTION EFFICIENCY

We note from Equation (8), that the current in a p-n junction is proportional to the saturation current. Hence, an inspection of the saturation currents for holes and electrons will give us information on the character of the diode current whether it be in the forward or reverse direction. From Equations (4) and (5) it is seen that the hole current depends only on the properties of the n-type region and the electron current only on the properties of the p-type region. The injection ratio of the hole and electron currents is, from Equations (4) and (5),

$$\frac{I_p}{I_n} = \frac{I_{p_s}}{I_{n_s}} = \frac{\frac{1}{\sigma_n L_p}}{\frac{1}{\sigma_p L_n}} = \frac{\sigma_p L_n}{\sigma_n L_p}. \quad (10)$$

For diffusion lengths of roughly equal magnitude, the predominant current across the junction corresponds to the majority carrier of the material having the greater conductivity. In this way a p-n junction may be utilized to inject a current of minority carriers into a semiconductor body and to suppress a flow of undesired majority carrier current from that body. This property of the p-n junction forms the basis for the common type of bipolar transistor.

It is common in discussing various aspects of transistor theory to refer to the injection efficiency, γ , rather than the ratio given by Equation (10). The hole injection efficiency is the fraction of the total current across the junction carried by holes, and is

$$\gamma = \frac{I_p}{I} = \frac{I_p}{I_p + I_n} = \frac{1}{1 + \frac{\sigma_n L_p}{\sigma_p L_n}} \approx 1 - \frac{\sigma_n L_p}{\sigma_p L_n}$$

where the approximation is valid for efficiencies near unity. An analogous relation for the electron injection efficiency can be written.

The above discussion of various aspects of the p-n junction has been given on the basis that the currents were small. However, as the currents are increased as would be desirable, for example, in a power transistor or even in the large-signal operation of a low-power transistor, other effects become important. This discussion of the injection efficiency is an opportune place to consider one of these effects, namely, the reduction of injection efficiency with increasing currents. This consideration will also bring to light a basic physical concept which, in the interests of simplicity, has not yet been pointed out.

For clarity, consider a specific type of junction — in particular one in which $\sigma_p \gg \sigma_n$, so that the current across the junction is carried principally by holes being injected into the n-type region. At low currents, the ratio of hole to electron current is given by Equation (10). We ask how this is altered at high currents.

When a hole is injected into the n-type region, an electron also enters through the external connection to preserve charge neutrality. Although we have considered principally the injection of one type of charge, this injection is accompanied by an equal flow of carriers of the opposite type. Reference is often made to this fact by stating that one should really speak and think of the injection of electron-hole pairs rather than confining one's attention to the injection of one type of carrier. Indeed, a consequence of this is the reduction of emitter efficiency at high currents.

If, then, p_0 is the injected hole density corresponding to an injected current, $I_p = qDp_0/L_p$, as in Figure 6, then an equal compensating charge density, $n_0 = p_0$ is added to the electron density normally present. Those electrons normally present arise from ionization of the donor atoms, N_d , so that the total electron density is $p_0 + N_d$. This, then, is the total electron density at the junction being held back from entering the p region by the barrier ($\phi_0 - V$). The number crossing the barrier is

$$(p_0 + N_d) e^{-\frac{q}{kT}(\phi_0 - V)}$$

From the earlier calculation of diode current where the increase in

electron density was neglected it will be recalled that for zero applied voltage the injected current was equal to the saturation current so that

$$N_d e^{-\frac{q}{kT} \phi_0} = I_{ns}$$

From this $e^{-\frac{q}{kT} \phi_0}$ can be evaluated so that I_n becomes approximately

$$I_n = \frac{p_0 + N_d}{N_d} I_{ns} e^{\frac{q}{kT} V} = \left(1 + \frac{L_p}{qD_p N_d} I_p \right) I_{ns} e^{\frac{q}{kT} V}, \quad (11)$$

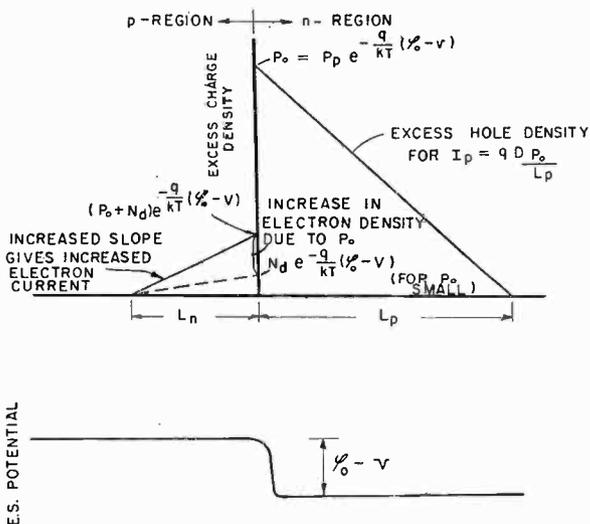


Fig. 6—Excess charge densities for p-n junction where $\sigma_p \gg \sigma_n$ and showing how electron current in p-region is increased by injection of p_0 into n region. The proportion of total current carried by holes is thereby reduced.

where, since large currents are assumed, the saturation current flow is omitted. The electron current flow across the junction has been increased by the factor $(1 + p_0/N_d)$ which is a function of the hole current. A similar argument may be applied to the hole current but for the case $\sigma_p \gg \sigma_n$ the factor is small and may be neglected for this first-order calculation. The injection ratio is then

$$\frac{I_p}{I_n} = \frac{I_{ps}}{\left(1 + \frac{p_0}{N_d} \right) I_{ns}} = \frac{\sigma_p L_n}{\sigma_n L_p} \left(\frac{N_d}{p_0 + N_d} \right). \quad (12)$$

In this case, the hole injection efficiency is reduced by a disproportionate *increase in the electron current* crossing the junction as the voltage is increased to increase the injected hole current.

BIPOLAR TRANSISTOR

Let us consider how the properties of the p-n junction may be applied to form the conventional bipolar transistor. Consider, as in Figure 7, a p-n-p transistor formed from two p-n junctions placed back to back with a base width, W , which is much less than the diffusion length for holes in this region and with voltages applied as indicated. With a positive voltage applied to the left-hand junction, a large current will flow into the base. If the conductivity of the emitter is much greater than that of the base, this current will be predominantly a hole flow. The injected holes are minority carriers in the base region

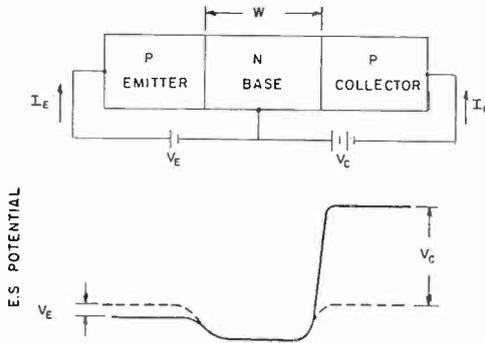


Fig. 7—p-n-p transistor.

and will diffuse through the base. Because the base region is thin, most of these will reach the right-hand junction and only a small fraction will be lost by recombination in the base. The right-hand junction is biased negatively, or in a direction to collect holes from the base region and transfer them to the right-hand p region. Thus the hole current injected by the emitter p-n junction diffuses through the base and is collected by the collector junction and the hole current is substantially constant through the device. Electron currents across the junctions are unwanted currents (in a p-n-p transistor) and one of the problems of transistor design concerns the minimization of these currents.

The forward conductance of a p-n junction is large, so that little power is required to inject hole current into the base region. On the other hand, the conductance of a junction biased in the reverse direction

is very small. Now, because the same current flows through the small conductance of the collector junction that was injected at the cost of very little power through the emitter junction, a considerably increased power may be developed in an external load.

Transistor action depends primarily on the diffusion of minority carriers through the base region. An analysis of the base region provides the essential features of transistor performance while the properties of the end regions are principally concerned with the flow of unwanted currents. A rigorous analysis then would solve the diffusion equation in the base subject to the boundary conditions imposed by the junctions. This becomes somewhat involved analytically, and therefore a somewhat heuristic approach on the basis of the current flow in p-n junctions discussed above will be followed.

The current flow across the emitter junction can be conceived of as

$$I_e = (\text{hole} + \text{electron flow due to } V_e) + (\text{hole} + \text{electron flow due to } V_c),$$

and that across the collector junction as

$$I_c = (\text{hole} + \text{electron flow due to } V_e) + (\text{hole} + \text{electron flow due to } V_c).$$

We might write the first term of I_c by considering this as a straightforward diode and using Equation (8); however, one modification must be made. The diode relation, Equation (8), was computed on the basis that the diffusion length, L_p , in the n region was smaller than the extent of the region. In the transistor, we have made W smaller than L_p . It will be recalled that L_p was the distance in which the injected hole density decreased to zero. In the transistor case, the collecting action of the collector junction reduces the hole density to zero at W . It is reasonable to replace L_p by W as the factor in determining the gradient giving rise to diffusion flow of current. Similarly, the second term of I_c can immediately be written down with the above modification. This is essentially the reverse current of the collector junction, i.e., the current for a junction biased negatively.

This leaves the transfer terms to be considered. Consider first the transfer term of I_c . From our physical explanation of the operation of the transistor this is simply the hole current injected by the emitter junction and collected at the collector (p-n-p transistor). It is given by the hole component of Equation (8) again with the modification of replacing L_p by W .

Now the structure is quite symmetrical and nothing in the device indicated that the left-hand junction should be the emitter. Thus we

must cause the transfer terms to have similar coefficients so that, if we choose, we can interchange the voltage polarities and obtain transistor action using the right-hand junction as the emitter. We therefore write the second term in I_e in a form similar to the first term in I_e . Thus we have

$$\left. \begin{aligned} I_e &= \frac{kT}{q} \frac{b}{(1+b)^2} \sigma_i^2 \left[\left(\frac{1}{\sigma_b W} + \frac{1}{\sigma_c L_e} \right) \left(e^{\frac{q}{kT} V_e} - 1 \right) \right. \\ &\quad \left. - \frac{1}{\sigma_b W} \left(e^{\frac{q}{kT} V_c} - 1 \right) \right] \\ I_c &= \frac{kT}{q} \frac{b}{(1+b)^2} \sigma_i^2 \left[\frac{1}{\sigma_b W} \left(e^{\frac{q}{kT} V_c} - 1 \right) \right. \\ &\quad \left. - \left(\frac{1}{\sigma_b W} + \frac{1}{\sigma_c L_c} \right) \left(e^{\frac{q}{kT} V_e} - 1 \right) \right] \end{aligned} \right\} \quad (14)$$

which is the first order approximation to the solution obtained from a rigorous mathematical development. In the rigorous solution the coefficients of $\left(e^{\frac{q}{kT} V_i} - 1 \right)$ are given in terms of hyperbolic functions and in practice it is the first-order approximation given in Equation (14) that is most often used in calculations.

We will return now to the simple p-n junction for a discussion of other basic properties. Before so doing, and for future use in some brief comments on the origin of temperature effects in transistors, note how temperature enters the transistor equations. While most of the parameters of Equation (14) have some temperature dependence, the most sensitive factor is the intrinsic conductivity, σ_i . This is discussed later.

CAPACITATIVE EFFECTS IN A P-N JUNCTION

Two phenomena give rise to the flow of capacitive currents in p-n junctions. The first of these, which we will term a diffusion capacitance, is the result of the nature of minority carrier flow in a semiconductor, i.e., a diffusion flow. The second, which we will refer to as a transition capacitance, is a consequence of the depletion of mobile charges near the junction as discussed earlier.

Consider first the diffusion capacitance arising from hole injection into the n region of a p-n junction. The hole flow in the n region corresponds, as we have seen, to a charge density gradient as shown in Figure 8. If the current is altered by changing the applied voltage,

the gradient, and hence the charge density distribution, must change thus changing the total charge. This change in total charge with the applied voltage corresponds to a capacitance (a diffusion capacitance).

The total hole charge within a diffusion length of the junction is

$$Q = q \frac{p_0}{2} L_p.$$

The diffusion current is

$$I_p = -qD_p \frac{\delta p}{\delta x} = qD_p \frac{p_0}{L_p}.$$

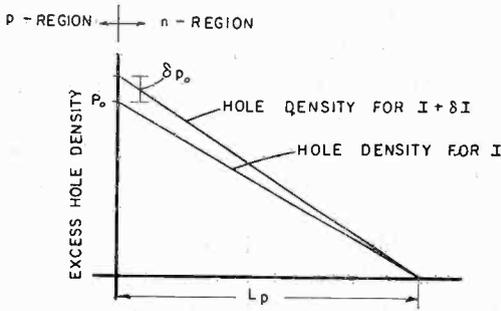


Fig. 8—Hole density distributions in *n* region for calculation of hole diffusion capacitance.

Then

$$C = \frac{\delta Q}{\delta V} = \frac{\delta Q}{\delta p_0} \frac{\delta p_0}{\delta I_p} \frac{\delta I_p}{\delta V} = \left(q \frac{L_p}{2} \right) \left(\frac{L_p}{q D_p} \right) \left(\frac{q}{kT} I_p \right) = \frac{q L_p^2}{kT 2D_p} I_p, \tag{15}$$

where the a-c conductance

$$\frac{\delta I_p}{\delta V} = \frac{q}{kT} I_p,$$

was obtained from Equation (9).

An analogous expression exists for the diffusion capacitance for the electron flow into the *p* region. The total diffusion capacitance is

ordinarily much larger than the transition capacitance and plays an important part in limiting the frequency response of the transistor. In the transistor, however, we must remember the minority carriers are stored in the distance W and not L_p , since ordinarily $W \ll L_p$. We can thus replace L_p by W in thinking of the diffusion capacitance of the injecting junction of a p-n-p transistor.

Consider next the transition capacitance which, in a transistor, is important for the collector junction. We have seen earlier how the potential barrier in a p-n junction was a consequence of the depletion of mobile charges in this region leaving uncompensated fixed charges. Further, if the potential drop is varied by an externally applied voltage, these densities must be altered to correspond to the new conditions. This process requires a flow of charge in response to the change in voltage, i.e., a capacitive flow. It is apparent that the exact nature will depend on the distribution of impurity atoms in the transition region. Two cases of practical importance have been much discussed: (1) an abrupt or step transition in which the impurity type changes discontinuously from n to p , and (2) a gradual transition in which the net impurity concentration changes linearly from n to p . We will consider the first case—a discontinuous transition—and, in particular, a junction in which $\sigma_n \gg \sigma_p$. This is shown in Figure 9, where the depletion layer is considered to exist only in the p region by virtue of its much lower conductivity. If we construct a pill-box of unit cross sectional area with one face at $x = 0$ at the edge of the depletion layer where the field is zero and the other face at $x = x$, the field at x is given by Gauss' Law as

$$\int_{\text{surface}} \mathcal{E}_n dA = \frac{4\pi Q}{\kappa}$$

here \mathcal{E}_n is the normal surface component of the field and Q is the charge enclosed in the pill-box. Upon integration,

$$\mathcal{E} = \frac{4\pi}{\kappa} qN_a x.$$

The potential can be found by integration over the barrier thickness, t :

$$V = \int \mathcal{E} dx = \frac{2\pi q}{\kappa} N_a t^2, \quad (16)$$

so the barrier thickness is

$$t^2 = \frac{\kappa\mu_p}{2\pi\sigma_p} V. \tag{17}$$

The capacitance can then be determined by taking the incremental change in charge with voltage, i.e.,

$$C = \frac{\delta Q}{\delta V} = \frac{qN_a \delta t}{\frac{2\pi q}{\kappa} N_a 2t \delta t} = \frac{\kappa}{4\pi t}, \tag{18}$$

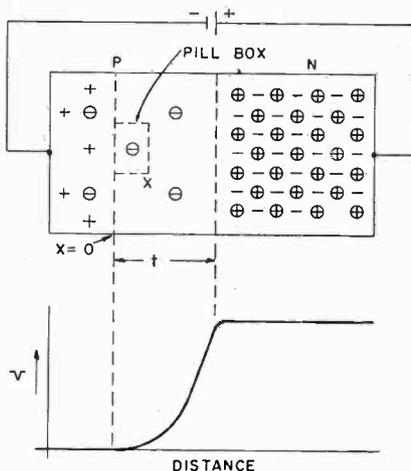


Fig. 9—Figure for calculation of transition capacitance when $\sigma_n \gg \sigma_p$.

which is the capacitance per unit area of a parallel plane capacitor with electrode spacing t filled with a material of dielectric constant, κ .

If one performs a similar analysis for the depletion region in the n region (which we neglected above in the interests of simplicity), the total depletion region thickness is given by

$$t^2 = \frac{\kappa\mu_p}{2\pi} \left(\frac{b}{\sigma_n} + \frac{1}{\sigma_p} \right) V, \tag{19}$$

where V is the total barrier potential including both the internal electrostatic potential and the applied voltage. In many practical transitions, such as those characteristic of alloyed junctions, one of the conductivities is much greater than the other and one term may be dropped, as in our example.

For a linear transition, the barrier thickness is given by

$$t^2 = \frac{3\kappa}{4\pi q\alpha} V, \quad (20)$$

where α is the net impurity density gradient.

SOME CONSEQUENCES OF VARIABLE BARRIER THICKNESS

Because the barrier thickness depends on the applied voltages, the simple picture of transistor operation we have given earlier must be modified in detail. Consider, for example, the effect of a variable voltage across the collector junction. Such a voltage is present when the transistor is operating into a load across which an a-c voltage is developed. The "electrical" thickness of the collector junction now varies with the instantaneous voltage and in so doing, alters the base width of the transistor (base width modulation). This effect modifies both the apparent conductance and susceptance of the collector junction. Let us consider a physical picture of these phenomena. Figure 10 shows the situation in the base region of the transistor as the effective position of the collector junction is varied.

Remembering that the diffusion current is proportional to the density gradient

$$I = -qD_p \frac{\delta p}{\delta x},$$

$$I_{\max} = qD_p p_0 \frac{1}{W - \frac{\delta W}{2}}, \quad I_{\min} = qD_p p_0 \frac{1}{W + \frac{\delta W}{2}}.$$

Then

$$g = \frac{\delta I}{\delta V} = \frac{I_{\max} - I_{\min}}{\delta V} = q \frac{D_p p_0}{W} \frac{1}{W} \frac{\delta W}{\delta V} = I \frac{1}{W} \frac{\delta W}{\delta V}, \quad (21)$$

giving an approximate expression for the a-c conductance due to modulation of the base width, W , by an applied signal. As we have seen earlier, the simple theory gave a conductance for a reversed bias

junction of the form $e^{-\frac{q}{kT} V}$. Even for relatively small negative values of V , this conductance is extremely small. The conductance developed

above turns out to be much greater and, for a practical unit with negligible leakage, does express the finite conductances found.

The situation of Figure 10 shows that the charge in the base region also varies as the effective position of the collector junction moves. This variation of charge with voltage may be interpreted as a capacitance which adds in parallel to the junction transition capacitance discussed above. This capacitance is derived as follows:

$$Q_{\max} = q \frac{p_0}{2} \left(W + \frac{\delta W}{2} \right), \quad Q_{\min} = q \frac{p_0}{2} \left(W - \frac{\delta W}{2} \right)$$

$$C = \frac{\delta Q}{\delta V} = \frac{Q_{\max} - Q_{\min}}{\delta V} = \frac{qp_0}{2} \frac{\delta W}{\delta V} = I \frac{W^2}{2D_p} \frac{1}{W} \frac{\delta W}{\delta V}. \quad (22)$$

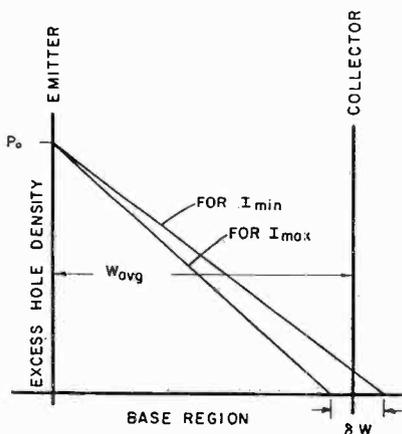


Fig. 10—Hole distributions in base of p-n-p transistor as effective base width is varied by variation in collector voltage.

This is similar in form to the diffusion capacitance associated with the emitter due to the flow of holes into the base. However, in this case the additional factor $(1/W) (\delta W/\delta V)$ makes this capacitance small and in practice it is but a fraction of the transition capacitance.

FIELD-EFFECT TRANSISTOR

This phenomenon of a variable barrier thickness has formed the basis for a different type of transistor device — the field-effect transistor. Such a device, which is shown in Figure 11, consists of a thin piece of semiconductor on the opposite sides of which are two p-n junctions. An ohmic contact is located at either end of the semicon-

ductor. It is seen that a conducting channel exists between the two ohmic contacts. This channel is defined by the two p-n junctions. If a reverse bias is applied to the two junctions, the conducting channel becomes still further limited by the depletion layers of the junctions. Thus the channel conductance may be varied by modulating its cross section through a variable voltage applied to the p-n junction. In this way the signal applied to the p-n junction controls the current flow between the ohmic contacts and through the load.

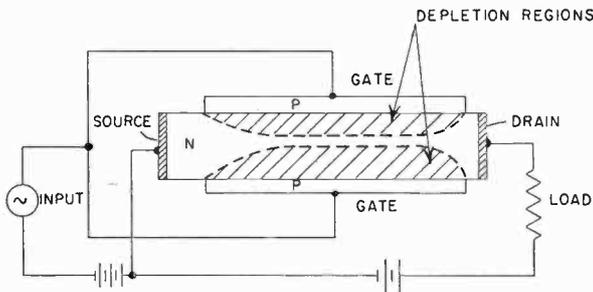


Fig. 11—Field-effect transistor.

ORIGIN OF TEMPERATURE EFFECTS

We will briefly examine the origin of the most important factor in determining temperature effects in transistors. In connection with the transistor equations, it was previously stated that the intrinsic conductivity was the most sensitive factor. Let us see how this comes about. Recalling that the diode currents were proportional to the saturation currents, we will look at the hole saturation current,

$$I_{ps} = -q D_p \frac{p_n}{L_p},$$

where p_n is the density of holes of thermal origin in the n-type material. These holes are created by the loss of an electron from an atom of the host crystal. Writing this in another fashion, using the relation $n_i^2 = n_n p_n$,

$$I_{ps} = -q \frac{D_p n_i^2}{L_p n_n},$$

where n_i is the density of electrons (or holes) in intrinsic material and

n_n is the density of electrons in the n-type material which come principally from the donor atoms.

Figure 12 shows, relative to thermal energy at room temperature (shaded) and 100°C (dotted), the energy required to release an electron from an impurity atom normally used in doping germanium and from a germanium atom and from a silicon atom. It is seen that thermal energy is somewhat larger than that required to release an electron from an impurity atom. It is for this reason that at normal temperatures we can assume that all of the impurity atoms are ionized and consequently that the normal electron density in n-type material, n_n , is equal to the donor density, N_d . Furthermore, having ionized all

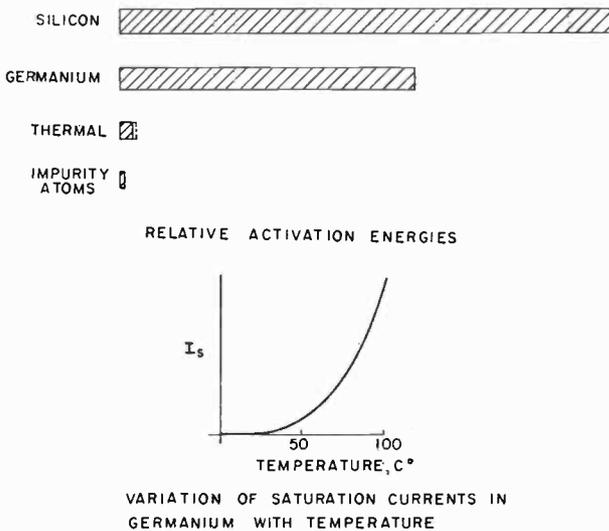


Fig. 12—Temperature effects.

of the donor atoms at normal temperatures, further increases in temperature do not change the number of electrons obtainable from this source. Then in I_{ps} above, n_n does not vary rapidly with temperature.

On the other hand, normal thermal energies are much smaller than the energy required to extract an electron from either a germanium or silicon atom. The number of electrons that will surmount a barrier as we have discussed in connection with the barrier in a p-n junction, is an exponential function of the barrier height in terms of thermal energy. Indeed,

$$n_i^2 \propto e^{-\frac{q}{kT} E_0}$$

where E_G is the "band gap" or activation energy shown relatively in Figure 12. Although the numbers are relatively small in comparison with n_n in practical transistors, the variation with temperature, due principally to the exponential dependence, is extremely rapid. The graph of Figure 12 illustrates the temperature dependence of the saturation current in germanium between zero and 100°C. This corresponds to the reverse collector current, I_{co} or COI_c . Because the activation energy in silicon is greater than that in germanium, the density of intrinsic electrons (and holes) at the same temperature is much less (again because of the exponential dependence on barrier height). However, the variation is similar in form and is about the same percentage-wise, but, because the density at normal temperatures is so small, the currents are unobjectionable until the temperature is increased considerably.

OTHER TRANSISTOR DEVICES

Let us consider briefly the mode of operation of a few other transistor devices made of an assemblage of p-n junctions.

The hook transistor shown in Figure 13 illustrates a different principle that provides a possible explanation for a certain behavior of point contact transistors. The hook transistor shown is a p-n-p-n structure with no external connection made to the internal p-region. The left hand junction is biased in the forward direction as an emitter. A negative potential is applied to the right-hand n-region. This biases the internal p-region negatively with respect to the base — in the direction to collect the holes emitted by the emitter. Under static conditions the internal p-region floats at a potential such that the currents entering and leaving it are equal. If holes are injected at the emitter and collected by the internal p-region, they are essentially trapped there — for we recall that holes tend to run uphill. The accumulation of holes lowers the potential of the internal p region biasing it in a forward direction with respect to the right hand n region. This permits an even greater number of electrons to be injected from the right which diffuse through the p-region and are collected by the n-type base. In this fashion the current injected by the left hand junction can control an even greater current flowing across the right-hand junction. Thus, a current-gain factor, alpha, which is greater than unity may be obtained. This is in contrast to the p-n-p junction transistor in which the collector current could at most be equal to the injected current, i.e., alpha is unity.

Figure 14 shows a point-contact transistor. Its operation can be at least qualitatively explained in terms of p-n junction theory. However, the details of its operation are not developed to the relatively refined

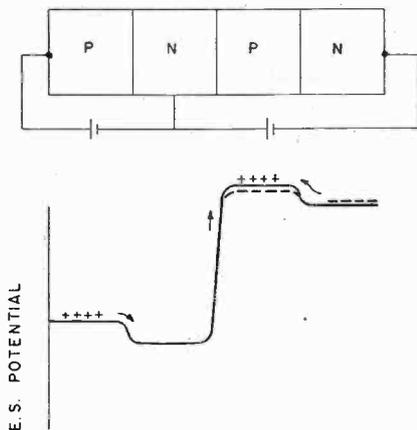


Fig. 13—Hook transistor.

state of the junction transistor. Here the emitter is a metal-to-semiconductor contact which, with the aid of surface states on the crystal, is able to create an electron-deficient or p-type region in the vicinity of the contact. Such a contact, similar to the p-n junction, can inject holes into the body of the germanium. The collector junction of a point contact is normally "formed" by pulsing it with an electric current. This forms a p-type region somewhat under the surface of the crystal. Biased negatively this acts as a collector for the holes injected by the emitter. Thus far we have something not much different than a p-n-p junction transistor except for geometry. However, point contact transistors, as it is well known, have current-gain factors (alphas) greater than unity. Just how this is achieved has not yet been completely determined. If one considers that between the formed p-region and the contact there exists an n-type region, then the structure is serially similar to the junction hook transistor discussed above. The reasoning

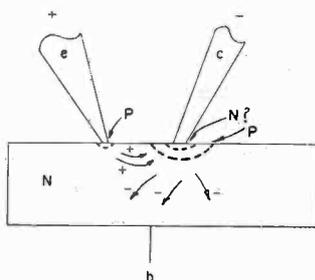


Fig. 14—Point contact transistor.

that gave the junction hook transistor a current-gain factor greater than unity can be invoked to "explain" this behavior of the point-contact transistor. This has not been entirely successful. An alternative proposal has been made that the collector forming process introduces traps in the region near the collector point. These traps are temporarily filled by holes injected by the emitter and in this state represent an accumulation of positive charge directly in front of the metallic collector point. The metallic collector point can supply an abundance of electrons and these are extracted from the metal by the collection of positive holes in the traps. This electron current flow may be much larger than the hole current injected by the emitter to give current-gain factors greater than unity.

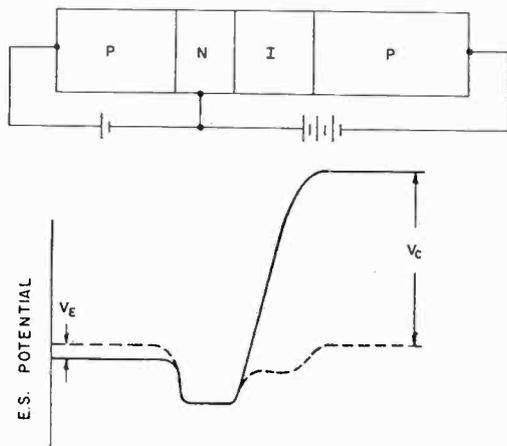


Fig. 15—p-n-i-p transistor.

A p-n-i-p transistor is shown in Figure 15. This may be considered much like a usual p-n-p junction transistor with a modified collector region. The modification in this case consists of the intrinsic region interposed between the n-type base and the p-type collector. In our discussion of depletion layers, we saw how the depletion layer extended principally into the low-conductivity material. The p-n-i-p transistor may be viewed as a p-n-p transistor in which the conductivity of the side of the base region near the collector has been made extremely low. When a potential is applied to the collector, the field extends through the intrinsic layer to the base. This provides a collector junction whose barrier thickness is very wide — the thickness of the intrinsic layer. Because the applied potential is distributed over a larger region, this

type of junction can withstand a high reverse voltage and has a lower capacitance. The low collector capacitance improves the high-frequency operation. The high-frequency operation of this type also depends, as in the conventional p-n-p transistor, upon making the n-type base region very thin. High-frequency operation also demands that the n-type region have a high conductivity which may seriously limit the voltage that may be applied to the collector junction of a conventional unit. In the p-n-i-p structure higher conductivity may be used in the n-type base with improved high-frequency performance without compromising the collector breakdown voltage.

CONCLUSIONS

This paper has discussed, largely on the basis of physical arguments, some of the basic transistor device concepts. The discussion has examined the various electrical properties of the p-n junction and pointed out the important roles played by the junction barrier and by diffusion phenomena. There has also been indicated in an elementary fashion how the properties of a p-n junction are utilized in a few transistor devices.

In other papers of the present volume, these concepts are further extended and applied to new and improved devices. Thus, E. W. Herold's "New Advances in the Junction Transistor" includes discussions of the alloy process of transistor fabrication, the effects of transistor geometry and an equivalent circuit for the transistor, among other topics. The limitations of diffusion phenomena on the frequency response of transistors are further developed by H. Kroemer. In his paper, "The Drift Transistor," a new transistor structure is described which uses a nonuniform impurity distribution in the base region to avoid the limitations of carrier diffusion. At the same time, this structure elegantly satisfies the other requirements for high-frequency operation. The capacitance of a p-n junction barrier is utilized by L. J. Giacoletto and J. J. O'Connell in "A Variable-Capacitance Germanium Junction Diode for UHF." Considerations of the injection efficiency of emitter junctions are invoked in two papers, "P-N-P Transistors Using High Emitter-Efficiency Alloy Materials," by L. D. Armstrong, C. L. Carlson, and M. Bentivegna, and "Recent Advances in Power Junction Transistors," by B. N. Slade; methods of improving injection efficiencies are described. Thus, the basic properties of a p-n junction discussed in the present paper are of general application throughout the transistor field, and it is hoped that the present discussion will provide a background for the important advances in the transistor art described in other papers in this book.

THE ALLOY JUNCTION†

The rectifying properties of the p-n junction were derived theoretically some years ago.⁷ Shockley's application of such junctions to an amplifying device came much later⁸; in the Shockley patent,¹ reference is made to diffusing an impurity into a semiconductor so as to produce a junction. A diffusion technique for making rectifying junctions was more fully described by Hall and Dunlap,⁹ but subsequent publications from the same laboratory indicate that diffusion in the solid state is extremely small.^{10, 11} Successful methods for making transistor junctions may utilize a different technique involving alloying. This has already been recognized^{2, 3} and is confirmed by experimental evidence to be discussed in the present paper.

Figure 1 shows the steps which take place in forming a junction by the alloy process, for example, using n-type, single-crystal germanium and indium as an impurity. In Figure 1(a) a piece of indium is placed on the germanium. The temperature is raised (Figure 1(b)) and the indium melts. Germanium is soluble in liquid indium, the solubility depending on the temperature. When the temperature is further raised, as in Figure 1(c), enough of the germanium is dissolved in the indium to cause a small depression in the crystal. In principle, the size of this depression does not depend on the time (provided equilibrium is reached in the germanium-indium solution) but only on the area, the volume, and the highest temperature reached by the molten indium. Subsequent lowering of the temperature reduces the solubility of germanium in the indium and the germanium crystallizes out, in indium-contaminated, p-type form. *The important factor at this stage is that the nucleation centers come from the original single-crystal base.* As a result, the p-type germanium recrystallizes *in single-crystal form* on top of the base material, and in crystallographic alinement with it. The p-n junction forms at approximately the original solid-liquid interface as a flat, abrupt transition of the Schottky type (Figure 1(d)). The

† Some of the research referred to in this section was done by J. Pan-kove, "Recrystallization of Germanium from Indium Solution," *RCA Review*, Vol. 15, pp. 78-85, March, 1954.

⁷ B. Davydov, "On the Rectification of Current at the Boundary between Two Semiconductors," *C. R. Acad. Sci. (U.S.S.R.)*, Vol. 21, p. 279, 1938.

⁸ W. Shockley, "The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors," *Bell Sys. Tech. Jour.*, Vol. 28, p. 435, July, 1949.

⁹ R. N. Hall and W. C. Dunlap, "p-n Junctions Prepared by Impurity Diffusion," *Phys. Rev.*, Vol. 80, p. 467, November 1, 1950.

¹⁰ W. C. Dunlap, "Measurement of Diffusion in Germanium by Means of p-n Junctions," *Phys. Rev.*, Vol. 86, p. 615, April-June, 1952.

¹¹ J. S. Saby and W. C. Dunlap, "Impurity Diffusion and Space Charge Layers in 'Fused-Impurity' p-n Junctions," *Phys. Rev.*, Vol. 90, p. 630, May 15, 1953.

resistivity of the recrystallized p-type germanium is about 0.001 ohm-cm, so that it forms an excellent emitter of holes into n-type material of 1 to 3 ohm-centimeter.

It is to be emphasized that, in principle, the position and shape of the junction are closely controllable by variation of the contact area, the volume of indium, and the maximum firing temperature. In practice, surface wetting is a possible variant which must also be controlled. Also of note is that any solid-state diffusion will play only a minor part in the junction position, changing it perhaps by a few tens of angstrom units.

The first evidence to indicate the nature of the alloy process was obtained by embedding an actual junction in plastic, cutting it in

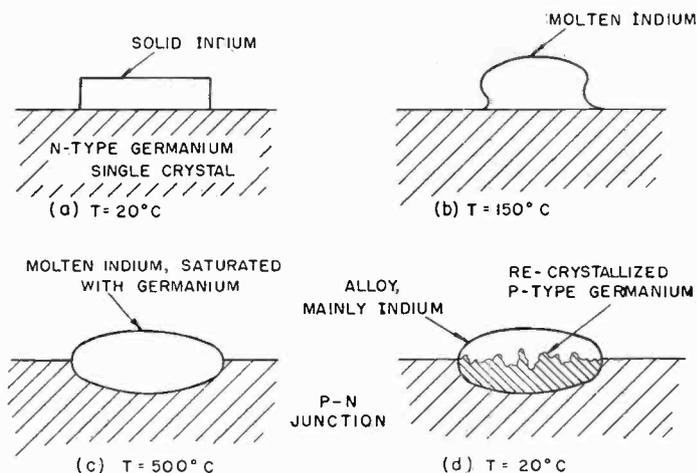


Fig. 1—Steps in the formation of an alloy junction.

cross section and studying it under a microscope. Better results were obtained by first dissolving out the indium with mercury followed by a final nitric acid wash. Some interesting pictures were also made of a rod-like, n-type, germanium crystal after it was dipped into molten indium, allowed time for equilibrium, and then cooled to room temperature. Correct choice of crystal direction in the rod permitted observation of recrystallized germanium on several crystal faces. After removing the indium, the end of the rod looked as in Figure 2. The typical crystal structure of the base material is clearly evident in the recrystallized germanium. A view from another angle, i.e., another crystal face, is shown in Figure 3. X-ray diffraction studies confirmed the basic single-crystal nature of the entire specimen.

After these photographs were taken, the end of the rod was cut in

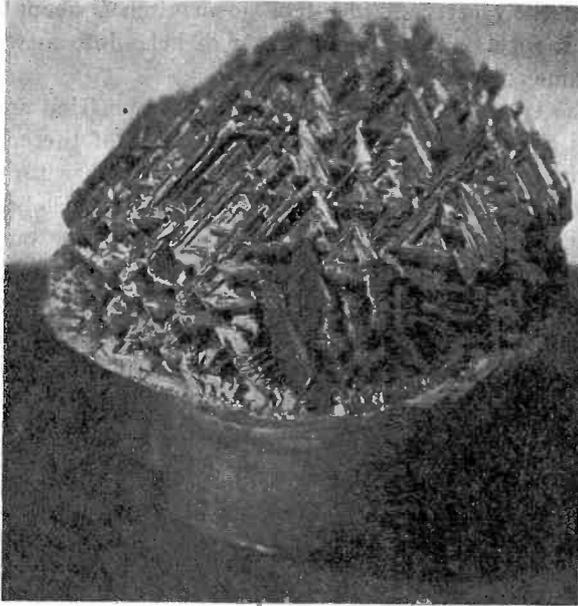


Fig. 2—View of end of single-crystal germanium rod after recrystallization from an indium melt and removal of the indium.

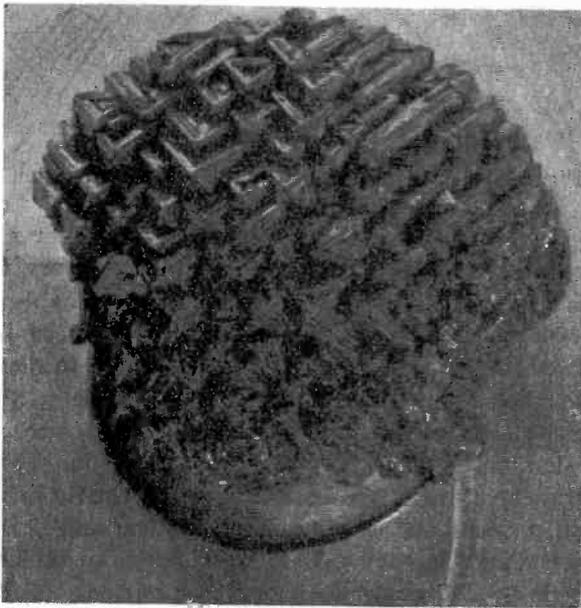


Fig. 3—Same rod as shown in Figure 2 viewed from another angle.

cross section, etched, and a series of low-power microscope pictures taken, to form the montage of Figure 4. The various crystal axes are identified by the arrows. By means of thermal probes it was found that a p-n junction had formed at a sharp line at about the boundary of the unmelted germanium. This junction line is visible in Figure 4 and is particularly clear at the (110) faces. The junction is more obscure at the (100) face. Although it may be only a coincidence, it is of interest that this crystal orientation is often avoided by those

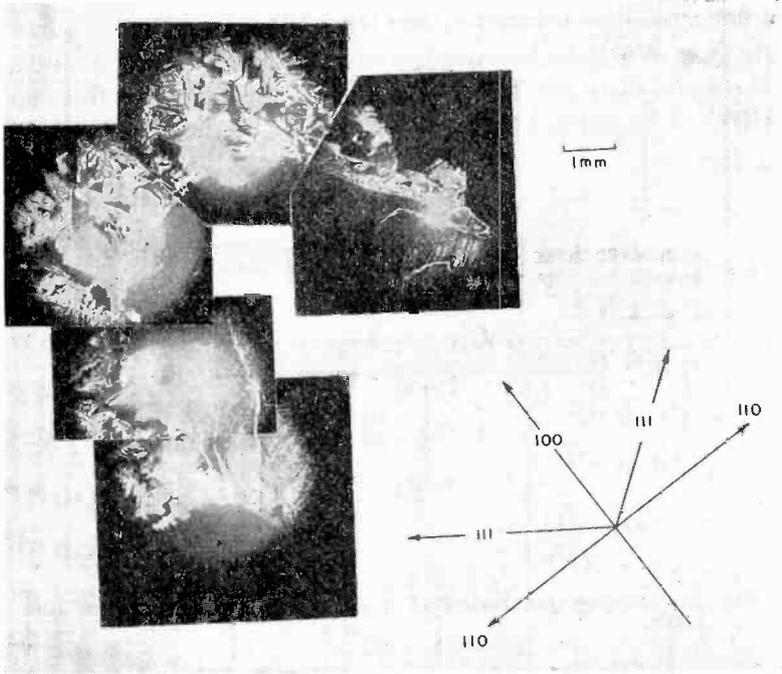


Fig. 4—Montage of cross-sectional views of end of rod shown in Figures 2 and 3. The p-n junction is seen as a fine white line inside the recrystallized portion.

making alloy transistors because it is said to yield less satisfactory junctions.

THE ALLOY JUNCTION TRANSISTOR

The p-n-p alloy junction transistor is made by using two such junctions on opposite sides of a thin germanium wafer. In cross section, a transistor is shown in Figure 5. Ordinarily, at least at low frequencies, an input signal is connected to the base lead, the output is taken from the collector, and the emitter, like the cathode of the electron

tube, is the common connection. The simplest single criterion of performance applicable to the junction transistor is its value of α , or current-gain factor. This is because any change in emitter current, which fails to reach the collector, contributes to a change in base current and this must be supplied in the form of input power. Ordinarily one thinks of two such current-gain factors, α_{ce} , the ratio of collector-current change to emitter-current change, and α_{cb} , the ratio of collector-current change to base-current change. We want α_{ce} to be unity and α_{cb} to be infinite. It is frequently most convenient to use α_{cb} because it is a more sensitive parameter than the other.

If α_{cb} is to be large, every source of base current must be a minimum. In the main, there are three sources: one is due to imperfect emitter efficiency, a second is bulk recombination of the injected holes in the

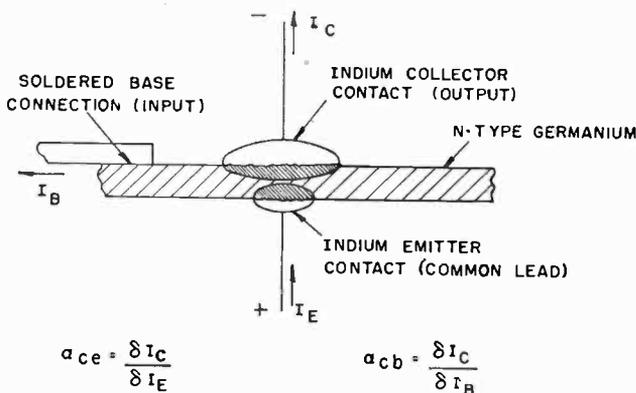


Fig. 5—Cross-sectional view of a p-n-p alloy junction transistor.

n-type base, and the third is surface recombination on the free surfaces of the base material. Let us examine these in turn.

Any part of the emitter current carried by electrons will not be able to arrive at the collector, and so represents base current. Primarily, this electron current is due to the free electrons in the n-type base. However, if the base conductivity is very small compared with the emitter conductivity, the emitter efficiency will be high. In the alloy transistor the emitter to base conductivity ratio is of the order of 10^3 . Except for the high-current-density case (e.g., power transistors), which will be examined later, a first approximation allows one to neglect emitter efficiency as a major source of base current.

Appreciable bulk recombination (i.e., short lifetime of minority carriers) in the base material prevents some of the injected holes from

arriving at the collector. In most transistor theory, this source of base current is emphasized. However, experience with the low-power alloy transistor indicated that bulk lifetimes of 5, 100, or 1000 microseconds alike gave equally good results. Since it is a poor sample of germanium indeed whose lifetime does not fall above 5 microseconds, one must assume that bulk recombination, again to a first approximation, is not a major source of base current in practical alloy transistors.

We are left with surface recombination to explain the base current in the alloy transistor. Early experiments showed that surface treatment of these units was one of the most crucial and critical parts of the processing technique, and confirm that this source of recombination is a major item in determining α .

Figure 5 shows why the surface may play such an important role. It is observed that the over-all transistor geometry is not parallel plane, as hitherto assumed in most transistor theory. Also one notes that the collector is larger than the emitter. The latter feature was incorporated because it was found to lead to high and more uniform values of α , and it is clear that it results in capturing much of the hole current at the top of the figure which might otherwise be surface recombination current. It is also seen that there is appreciable surface area surrounding both emitter and collector.

In the next section of the paper, the special geometry of the alloy transistor will be examined in the light of surface recombination, in order to explain observed values of α . Fortunately, with bulk recombination neglected, all conditions are on the boundaries and powerful analogue methods can be employed to effect a practical solution.

EFFECT OF GEOMETRY ON ALPHA*

The minority carriers (holes) flow in the base region primarily as a result of diffusion. The diffusion equation tells us that the current density in the base region is proportional to the gradient of the charge density. Thus

$$I = -qD_p \nabla P, \quad (1)$$

where q is the charge of a "hole," D_p is its diffusion constant, and P is the excess hole density, i.e., the hole density in excess of the thermal equilibrium value. Taking the divergence of each side

$$\text{div } I = -qD_p \nabla^2 P.$$

* See Moore and Pankove, Reference (4). For later work, see K. E. Stripp and A. R. Moore, "The Effects of Junction Shape—Part II," *Proc. I.R.E.*, Vol. 43, pp. 856-866, July, 1955.

In the steady state, if there is no bulk recombination, $\text{div } I = 0$ throughout the solid, so that

$$\nabla^2 P = 0, \quad (2)$$

a well-known equation for which powerful solution methods are available.

Let us compare this Laplace equation with one similarly derived for a conducting solid of bulk resistivity ρ . If ϕ is the potential.

$$i = -1/\rho \nabla \phi, \quad (3)$$

$$\text{div } i = -(1/\rho) \nabla^2 \phi = 0. \quad (4)$$

We see, then, an analogy in which $1/\rho$ corresponds with qD_p , and the potential corresponds with excess hole density.

To solve either the transistor or its conducting solid analogue, the boundary conditions must be set up. For the transistor, $P = P_0$, a constant over the emitter electrode surface and, for the usual highly efficient collector surface $P = 0$. In the analogue, corresponding surfaces are connected to a battery which places the emitter analogue at a potential ϕ_0 , the collector at potential zero. At the free semiconductor surface, if there were no surface recombination, there would be no current flow into this surface; the electrical analogue is an insulating surface. When surface recombination takes place, a current density flow proportional to the total charge, qP , takes place

$$I = qPs \quad (5)$$

where s is the proportionality constant, known as the surface recombination velocity. In the electrical analogue, such a surface current can be caused to flow by placing many small electrodes over the entire free surface, and connecting them through resistances back to the zero of potential. If A is the area of each small electrode, and R the value of each resistance, the surface current density is

$$i = \frac{1}{AR} \phi. \quad (6)$$

Simple division of the normal component of the general current Equations, (1) and (3) by the surface-current Equations (5) and (6) shows that

$$\frac{\nabla P_n}{P} = -\frac{s}{D_p} \quad (7)$$

and

$$\frac{\nabla \phi_n}{\phi} = -\frac{\rho}{AR} \quad (8)$$

By making ρ/AR equal to s/D_p , the analogy is preserved. We could solve the transistor equation, then, by building a solid conducting model, attaching a battery and measuring the ratio of collector current to emitter or base current.

Fortunately, the thin wafer geometry is such that the powerful tools of two-dimensional analogue solutions, such as the electrolytic tank or conducting sheet, can also be used. Figure 6 shows the cross section of a transistor, together with a conducting sheet analogue. If s/D_p is made equal to ρ_s/aR , where ρ_s is the surface resistance of the conducting sheet, and a is the length of each free boundary segment, the analogy is approximately applicable. The current-flow lines of the analogy are hole-flow paths in the transistor. To take into account the circular nature of the actual transistor surfaces, we may find the current density at each radius in the model and calculate the over-all emitter and collector current by weighting each current density value by $2\pi r$. Thus the corrected total emitter current is

$$i_e = \int_0^{r_e} 2\pi r i_e(r) dr,$$

and the total collector current

$$i_c = \int_0^{r_c} 2\pi r i_c(r) dr,$$

where r_e and r_c are maximum radii of emitter and collector electrodes. The ratio i_c/i_e is α_{ce} so that α_{cb} is readily computed as well.

Although not employed in the present investigation, it is clear that the resistance network analogy board¹² would permit a direct measurement of the cylindrical geometry; in addition, there is enough flexibility to allow the introduction of bulk recombination effects by use of bleeder resistors plugged in at intermediate points between emitter and collector.

¹² G. Liebmann, "Electrical Analogues," *British Journal of Applied Physics*, Vol. 4, p. 193, July, 1953.

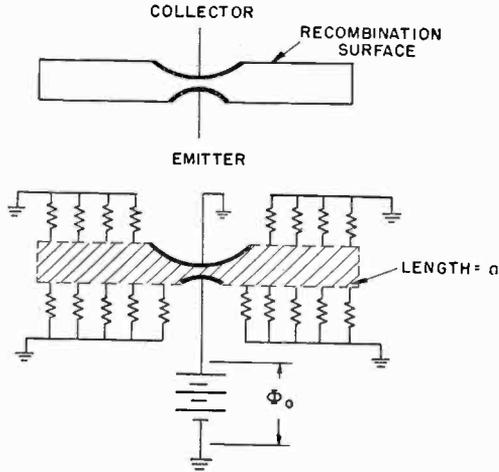


Fig. 6—Two-dimensional analogy to transistor using a conducting sheet. The transistor cross section is shown above, the analogy arrangement below.

Returning to the resistance-paper solution, Moore and Pankove took data on a variety of geometries and surface conditions.⁴ In Figure 7, a hole-flow map for a typical transistor geometry is shown. One junction is 0.045 inch in diameter, the other 0.015 inch, and the minimum

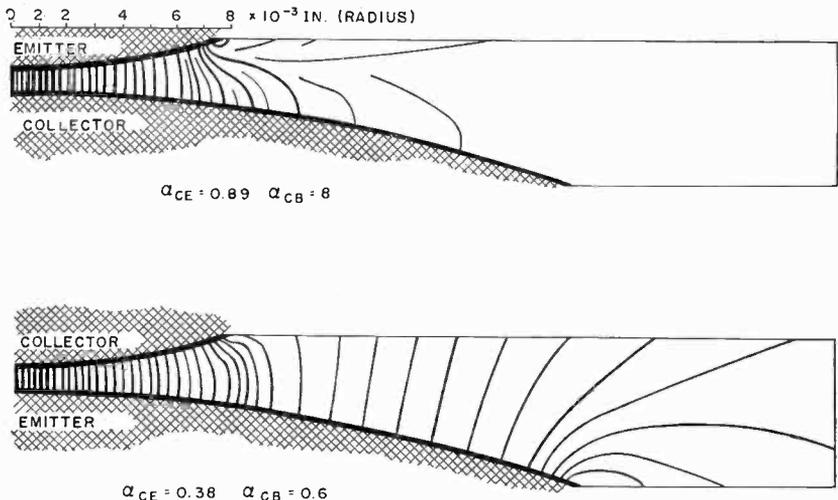


Fig. 7—Hole-flow maps with a surface recombination of 5000 centimeters per second. The upper figure shows normal connection of emitter (0.015-inch diameter) and collector (0.045-inch diameter); the lower figure shows these connections reversed, i.e., collector diameter 0.015 inch, emitter diameter 0.045 inch.

separation is 0.001 inch. Two plots are shown, corresponding to normal and reverse connections of emitter and collector. A surface recombination velocity of 5000 centimeters per second was chosen to correspond with some early transistor data. The surface currents are clearly delineated. The calculated collector-to-base α value was 8 in the normal connection, but only 0.6 in the reverse connection.

If there were no surface recombination, the normal and reverse α values would be the same. We arrive, therefore, at a very interesting suggestion. By measuring the ratio of forward-to-reverse α values in an actual transistor, we have a quantitative measure of the surface

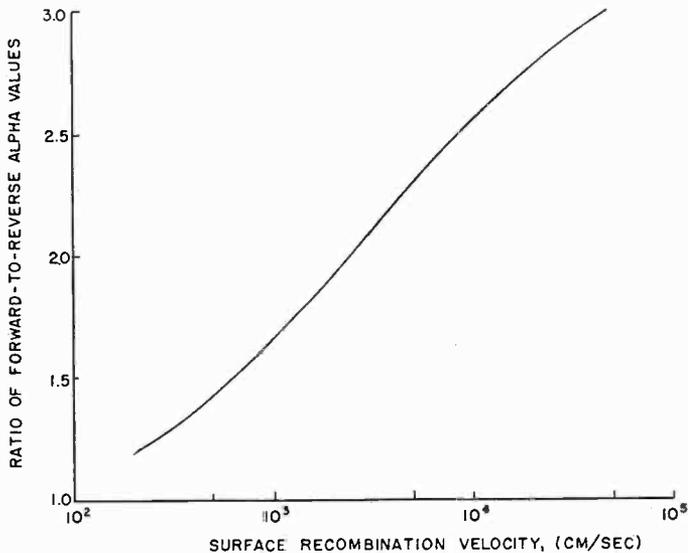


Fig. 8—Calibration of surface recombination for unsymmetrical transistors with 0.015-inch emitter, 0.045-inch collector, 0.001-inch minimum junction spacing and 0.0022-inch average junction spacing.

recombination. The electric analogue solution may be used to supply the calibration curve.

Figure 8 shows such a calibration curve, calculated for a minimum junction spacing of 0.001 inch (average spacing of 0.0022 inch), and for the typical geometry with one 0.045 inch and one 0.015 inch junction. In practice, transistors vary in both junction spacing as well as surface recombination, and both factors affect α . However, measurement of the emitter-to-base diffusion capacitance, to be discussed later in this paper, permits an independent evaluation of junction spacing. The forward-to-reverse α ratio, in conjunction with the diffusion

capacitance, provides a means of quantitatively determining junction spacing and surface recombination velocity, *even in a finished and already encapsulated transistor*.

Many actual transistors have been measured, and it is found that the surface recombination velocity varies from as little as a hundred to many thousands of centimeters per second. Very instructive experiments have been performed with a given transistor, open to the air, by observing forward-to-back α values while the surface treatment is varied.⁴

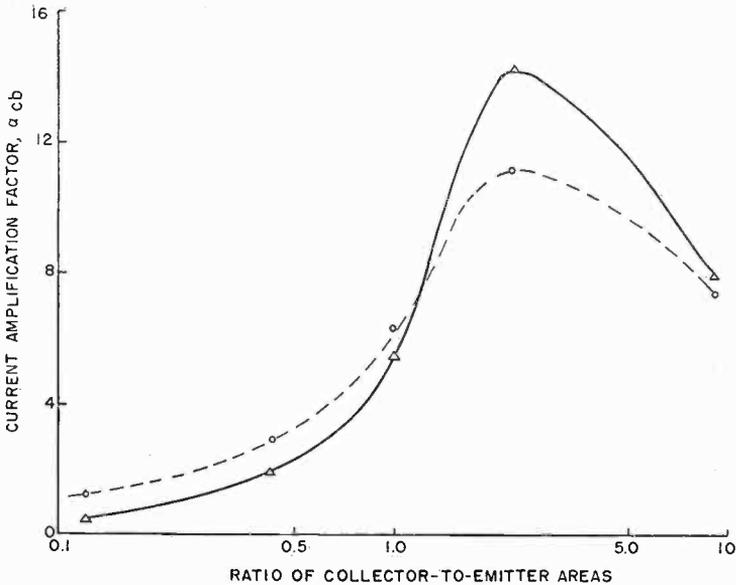


Fig. 9—Comparison of experimental and calculated values of α for p-n-p transistor with various area ratios. Largest junction diameter is always 0.045 inch. Surface recombination is 5000 centimeters per second. Average junction spacing is 0.0022 inch.

— Δ — Δ — calculated values
 - - \circ - - \circ - - experimental values

Hole-flow maps for different transistor geometries have explained the experimental results which showed the advantage of a large collector area. In Figure 9, the dotted curve shows a series of early experiments by Pankove in which alloy junction transistors were built with varying geometry. Each point is the average of a number of units. It is seen that highest α value is obtained when the collector area is about twice that of the emitter. The unbroken curve shows the computed values using the current-sheet analogue. Reasonable agreement is found.

It should be noted that lack of exact centering of the two junctions,

in practice, favors a larger area ratio than the optimum one shown in Figure 9. A 3:1 diameter ratio, therefore, has been widely used. Because modern alloy transistors with a 3:1 diameter ratio are found to have surface recombination velocities below 500 centimeters per second, the α values are in the range up to a hundred. The difference over early units is entirely due to improved etching and surface treatment.

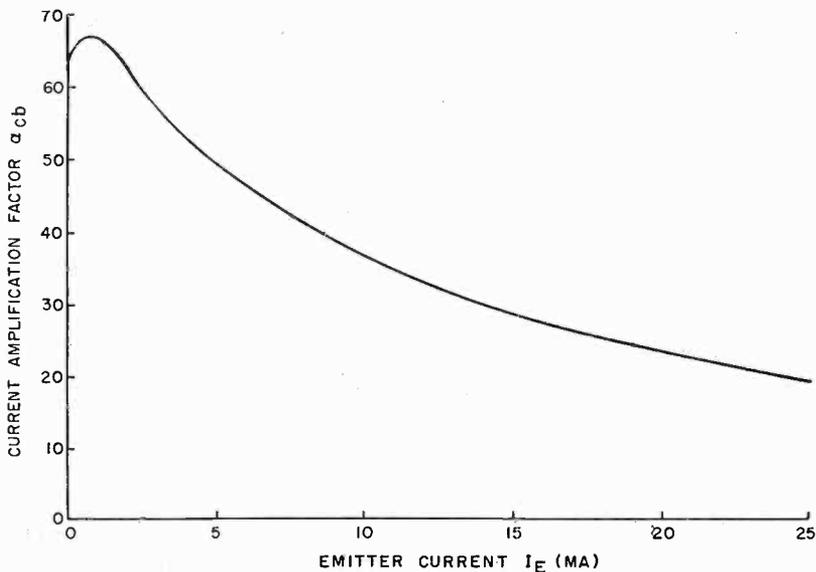


Fig. 10—Curve showing how α varies with emitter current in p-n-p alloy junction transistor. Collector diameter 0.045 inch, emitter diameter 0.015 inch.

EFFECT OF EMITTER CURRENT ON ALPHA*

Up to this point, transistor current gain, α , has been discussed under what might be called low-power conditions, i.e., the minority carrier charge density in the base region does not greatly exceed the donor concentration. In power transistors this is not always the case. If the collector-to-base current-amplification factor of a typical p-n-p alloy unit is measured, a curve similar to that shown in Figure 10 is obtained when the emitter current is varied. Low-power operation is below a few milliamperes and high α values are found. At high currents, the α value falls off substantially.

*The research referred to in this section was undertaken by W. M. Webster, Reference (4).

It is of interest to derive a relationship for α in terms of emitter efficiency, surface and bulk recombination and then to examine how these terms may vary when the emitter current density is increased. To make the problem amenable to solution, suitable approximations will be made, such as parallel-plane geometry; judicious interpretation makes the results applicable to the alloy transistor as well.

In Figure 5 a junction transistor is shown. An input variation on the base causes a collector current variation in the load, and we are interested in the ratio of the two current variations. The direct currents are I_E , I_B , and I_C . The emitter current is comprised of two parts, that due to holes, I_{Ep} , and that due to electrons, I_{Ec} ,

$$I_E = I_{Ep} + I_{Ec}$$

With high emitter-to-base conductivity ratio, however,

$$I_E \simeq I_{Ep}$$

The base current is the sum of the three factors discussed in an earlier section,

$$I_b = I_{Ec} + I_{VR} + I_{SR},$$

i.e., the emitter electron current, I_{Ec} , the volume recombination current, I_{VR} , and the surface recombination current, I_{SR} . Since α_{cb} is the derivative of collector current with base current, its reciprocal is

$$\begin{aligned} \frac{1}{\alpha_{cb}} &= \frac{\partial I_B}{\partial I_C} \simeq \frac{\partial I_B}{\partial I_{Ep}} \\ &\simeq \frac{\partial I_{Ec}}{\partial I_{Ep}} + \frac{\partial I_{VR}}{\partial I_{Ep}} + \frac{\partial I_{SR}}{\partial I_{Ep}}. \end{aligned} \quad (9)$$

The first of these terms is related to the ratio of electron-to-hole current, having to do with emitter efficiency. The second term is the change of volume recombination with emitter current and the third is affected by surface recombination.

Simple first-order theory, for very small density of minority carriers in the base region, gives us the three terms. Shockley and his co-workers¹³ show that

¹³ W. Shockley, M. Sparks, and G. K. Teal, "p-n Junction Transistors," *Phys. Rev.*, Vol. 83, p. 151, July 1, 1951.

$$\frac{\partial I_{Ee}}{\partial I_{Ep}} = \frac{I_{Ee}}{I_{Ep}} = \frac{\sigma_B W}{\sigma_E L_E}, \quad (10)$$

where σ_B and σ_E are the base and emitter conductivities, W is the base thickness, and L_E is the diffusion length in the emitter. Also from Shockley and co-workers' paper,¹³

$$\frac{\partial I_{VR}}{\partial I_{Ep}} = \frac{1}{2} \left(\frac{W}{L_B} \right)^2, \quad (11)$$

where L_B is the diffusion length in the base. By a simple integration of the diffusion equation, it can also be shown that⁵

$$\frac{\partial I_{SR}}{\partial I_{Ep}} = \frac{sW}{D_p} \frac{A_s}{A}, \quad (12)$$

where A_s/A is the ratio of "effective" recombination surface area to total emitter area and the other symbols are as used previously.

In this paper, it will only be possible to touch upon the highlights of the work and no attempt will be made to complete the derivations. However, we can readily understand the physics of each of the three terms. The first, Equation (10), represents simply the ratio of electron-charge carriers to hole-charge carriers available at the emitter junction. The second term, Equation (11), is an approximation to the solution of the one-dimensional diffusion equation and gives the loss in hole current because the base thickness is an appreciable fraction of the diffusion length. The final term, Equation (12), is proportional to s/D_p , as it should be, and also on the ratio of "effective" surface area, A_s , to emitter area, A , as might be expected.

In the alloy transistor, at low currents, the first term, Equation (10), is about 10^{-3} and the second, Equation (11), is perhaps a little smaller. The third term, due to surface recombination, is of the order of 10^{-2} and so predominates, as was already seen. However, the present purpose is to re-examine the situation when the direct currents are greatly increased.

Let us see how each term might be expected to vary when the injected hole charge increases to the point where it greatly exceeds the donor density in the base region. Charge neutrality requires an equal number of free electrons to be present, and these are then greatly in excess of the donors. As a result, the electron current through the emitter junction, I_{Ee} , is no longer proportional to the original base conductivity, σ_B , but is much larger. The effect is just the same as if

we had apparently increased the base conductivity, i.e., modulated its conductivity. Thus we see that the first term, Equation (10), increases with emitter current and gives a drop-off of α .

Looking at the second term, Equation (11), it is possible that the great excess of free electrons in the base, which accompanies our heavy hole-charge injection, will increase the bulk recombination. This would be the same as a decrease in L_B . If one assumes that "bimolecular recombination," i.e., recombination proportional not just to the hole density, but to the product of hole and electron density, sets in, the decrease in α due to the second term varies in exactly the same way as that of the first term; both may be thought of as due to an increase in base conductivity. In any event, even if Equation (11) is not a strong function of emitter current, the term of Equation (10) predominates.

The third term, Equation (12), requires a different interpretation. Up to this point in the discussion, no electric field in the base region has been included. All the current which flowed has been assumed to be due to diffusion because of a hole-density gradient. When the hole gradient becomes large, however, the electron gradient also becomes large. This gradient would, in the absence of an electric field, cause electron motion in the same direction as the holes. Since electrons cannot escape at the collector, such motion immediately sets up its own restoring force in the form of an electric field, which keeps the electron density high towards the emitter, where the hole density is also a maximum. Such an electric field aids the hole flow toward the collector and, for very high hole injection, gives exactly twice the expected current. It behaves just as if the hole diffusion constant, D_p had doubled.⁵

Naturally, if the hole flow toward the collector is augmented, that towards the surface is decreased, and the surface recombination is reduced. In fact, as indicated, D_p doubles and surface recombination is halved at high current densities. The term, Equation (12), then leads to an *increase* in α with current.

It is now seen how, in the alloy transistor, high emitter currents cause the first two terms of Equation (9) to increase in importance, and the last term, originally predominant, to decrease in importance. In place of the low-current expression, it can be found that,⁵

$$\frac{1}{\alpha_{cb}} = \left[\frac{\sigma_B W}{\sigma_E L_B} + \frac{1}{2} \left(\frac{W}{L_B} \right)^2 \right] (1 + kI_B) + \frac{sW}{D_p} \frac{As}{A} g(kI_B), \quad (13)$$

where

$$k = \frac{W\mu_e}{AD_p\sigma_B}, \quad (14)$$

and $g(kI_E)$ is a function which varies from unity to one half as kI_E is increased. Figure 11 shows curves of the function which multiplies the first two terms and that which modifies the last.

It is well to point out that Equation (13) interprets quantitatively each factor in transistor design. The critical nature of W , which enters into every term, even including the factor k , is to be noted. Since the equation is also applicable to n-p-n transistors by interchange of subscripts, a comparison of the two is in order. Examining k for germanium, it is clear that μ_e/D_p is about four times as large as μ_p/D_e which applies for the n-p-n type. The n-p-n transistor is, therefore, considerably less subject to variation of α values with emitter current. This had already been observed experimentally.

By substitution of appropriate numbers into the formula, and the adjustment of the unknown constants so as to obtain a reasonable

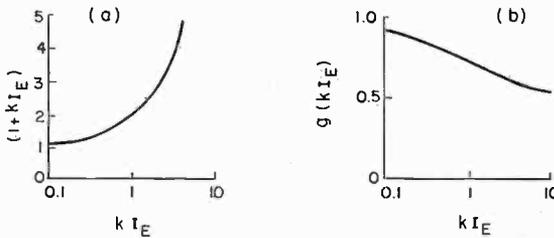


Fig. 11—(a) The function which changes the electron-to-hole current ratio and the bulk recombination current when emitter current is increased. (b) The function which changes surface recombination current when emitter current is increased.

check with surface recombination, Webster made a calculation of α against emitter current for a p-n-p alloy transistor.⁵ Figure 12 shows the result, and excellent agreement between the experimental values and the calculated values is found.

Figure 13 shows experimental and theoretical values for an n-p-n alloy transistor. Although agreement is satisfactory for the large-current region, the experiments show an even flatter curve than the theory predicts at low currents. The discrepancy is believed to be due to other factors not yet fully investigated. In this connection, it should be mentioned that less is known of the alloy mechanism in the n-p-n type which uses a binary impurity metal.³ There is also evidence that use of a binary impurity metal for the p-n-p type gives a flatter α curve than a single impurity. However, there is little doubt that, qualitatively, the analysis correctly interprets the observed phenomena of decreased α values with increase in current.

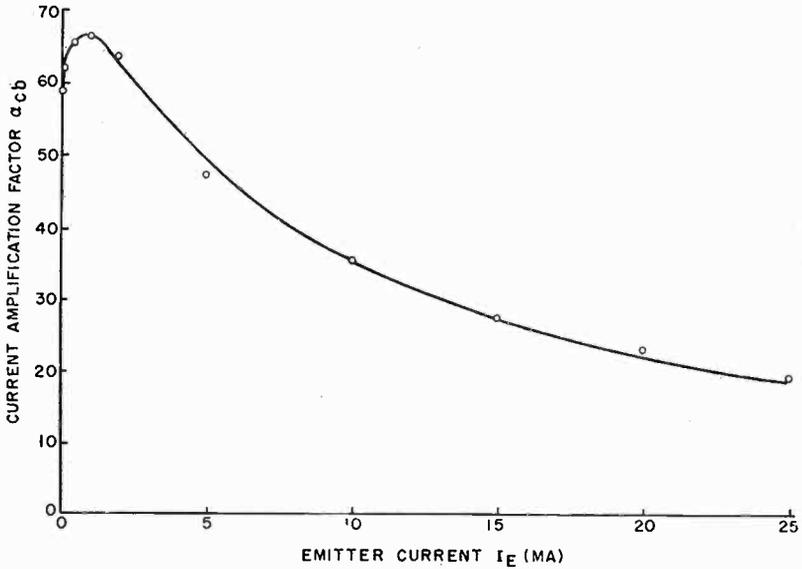


Fig. 12—Comparison of calculated α variation with experiment for typical p-n-p alloy junction transistor.

— calculated values
 ○ experimental values

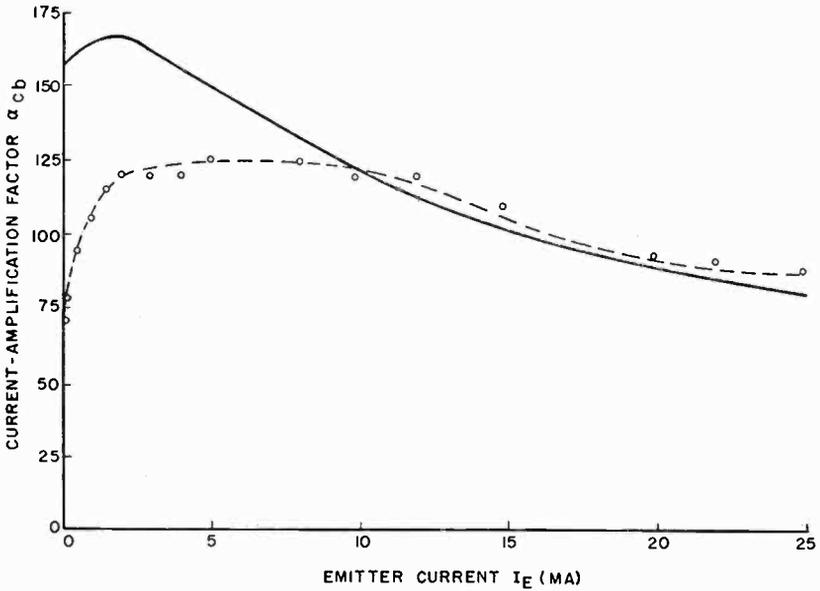


Fig. 13—Comparison of calculated α variation with experiment for typical n-p-n alloy junction transistor.

— calculated values
 --- ○ --- experimental values

HIGH-FREQUENCY PERFORMANCE AND THE EQUIVALENT CIRCUIT*

As with other electronic devices, there are high-frequency limitations of the alloy junction transistor. A possible limitation may be evident in the hole-flow map, Figure 7. There are appreciably different path lengths taken by the minority carriers in their travel from curved emitter to curved collector. Such a limitation is not basic, as we shall see, since junctions can be made reasonably parallel by variations in the alloy technique.

Much more basic is the fact that the carriers pass from emitter to collector by a diffusion process. Such current flow is relatively slow and, in addition to an expected phase shift, causes a decrease in α values as the frequency is raised. Much has been written about the so-called " α cutoff" and it is, indeed, a basic limitation in transistors having no electric field to assist flow through the base.

In the simplest, conventional alloy transistor, however, the decrease in magnitude of α is only partly responsible for the reduced gain at high frequencies. For this reason, it is better to consider an equivalent circuit and then find the various component values by a combination of intuitive and experimental approaches. A first clue was obtained because junction transistor theory failed to predict accurately the static characteristics of alloy junction transistors. In order to investigate the cause of the discrepancy, the transistor admittances were carefully measured.¹⁴ The first one chosen was that of the input, i.e., from the base to the grounded emitter. A bridge was set up and, to observe the degree of balance over a wide frequency range, a square-wave input signal was used as shown in Figure 14(a). A balanced-input oscilloscope was used as null detector, and the transistor was operated with normal bias supplies (collector by-passed to ground). For a typical p-n-p alloy transistor, a network which balanced quite well over the wide frequency range of the square wave is shown at Figure 14(b). The large capacitance and its shunting resistance could have been expected from junction theory as we shall see. The 350-ohm series resistance, however, was unexpectedly high and obviously cannot be ignored. Very little intuition was needed to find its source in the transistor construction: the base connection was made through a substantial length of germanium, r_{BB} , Figure 14(c).

One must recognize that in the equivalent circuit, between the in-

* The research referred to in this section has been performed by H. Johnson and L. J. Giacoletto.

¹⁴ L. J. Giacoletto, "Equipments for Measuring Junction Transistor Admittance Parameters for a Wide Range of Frequencies," *RCA Review*, Vol. 14, p. 269, June, 1953.

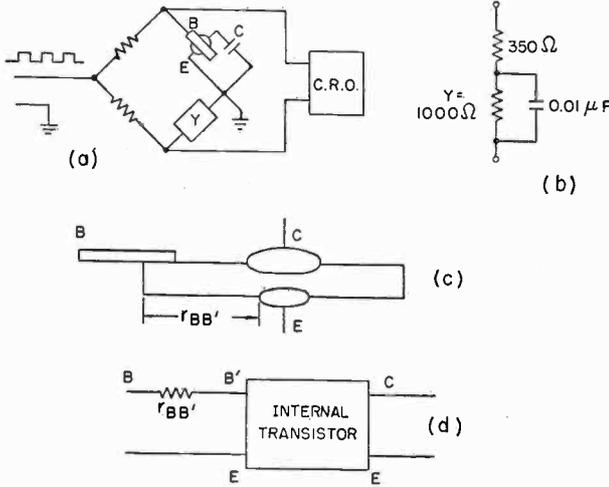


Fig. 14—Investigation of transistor input impedance: (a) Bridge for input impedance measurement; (b) Network which balances a typical p-n-p transistor; (c) Construction of transistor to show base-lead resistance; (d) Effect of base-lead resistance on equivalent circuit.

ternal intrinsic transistor and the available external base lead, there is a resistor of appreciable magnitude. Once this is recognized, and a correction made, it is found that the *internal* transistor agrees with predicted theoretical characteristics remarkably well, far better than is common with electron tubes. In Figure 14(d) is indicated the external base connection, *B*, and the internal point, *B'*, which is inaccessible, but which truly represents the base connection so far as theory is concerned.

We can now draw an equivalent circuit, as in Figure 15, and begin to evaluate its components from theory. Let us start by considering the emitter junction at such a low frequency that reactive effects are negligible. We know that the current through such a junction, with an applied d-c voltage V_B , is

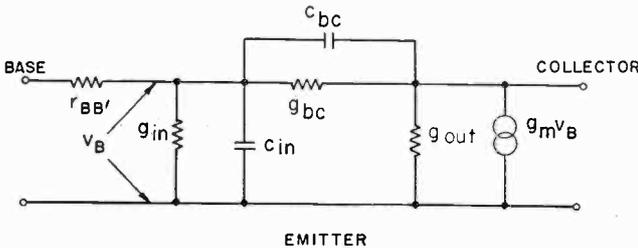


Fig. 15—Equivalent circuit of alloy junction transistor.

$$I_E = I_s \left[\exp \left(\frac{qV_B}{kT} \right) - 1 \right], \quad (15)$$

which is the conventional rectifier characteristic and q/kT is a constant of about 39 V^{-1} . For reasonably good transistors, most of this emitter current flows to the collector, so that a small input signal will produce a short-circuit output current and

$$g_m = \frac{\partial I_O}{\partial V_B} \simeq \frac{\partial I_E}{\partial V_B} \simeq \frac{q}{kT} I_E, \quad (16)$$

and we have then a value for the low-frequency part of our output generator. Intuitively, we recognize that the base-to-collector admittance is small, and we can then write down the input conductance, g_{in} , as

$$g_{in} = g_m / \alpha_{cb}, \quad (17)$$

where α_{cb} is our low-frequency current gain. This relation is almost obvious from the definition of α .

At a somewhat higher frequency, reactive effects set in. The output current generator has associated with it the time constant of the diffusion process W^2/D , which leads to both a time lag and an amplitude loss. In terms of a hypothetical short input pulse, it is delayed, spread out in time, and reduced in amplitude. The solution of the diffusion equation, and use of the approximation that the frequency be not too high, allows us to find an admittance, Y_m , in place of the conductance g_m , which is

$$Y_m = \frac{g_m}{1 + j\omega(W^2/4D)}. \quad (18)$$

When a small voltage δV_B is placed between base and ground, a small current flows which charges up the base region. If the voltage is removed, this charge must be swept out again. At low frequencies, the effect is like a capacitance, and quantitatively it is much larger than the normal Schottky-barrier capacitance. The total charge δQ included in the base region is $i\tau$, where τ is the "transit time" or diffusion time. Thus the equivalent capacitance is $\delta Q/\delta V_B$ and is

$$C_{in} = \frac{1}{\delta V_B} i\tau$$

$$= \frac{1}{\delta V_B} g_m (\delta V_B) \tau = \frac{q}{kT} I_E \tau.$$

However, the diffusion time τ taken to traverse the base thickness W is of the order W^2/D . The correct solution of the equation, at least for low emitter currents, is

$$C_{in} = \frac{q}{kT} I_E \frac{W^2}{2D}. \quad (19)$$

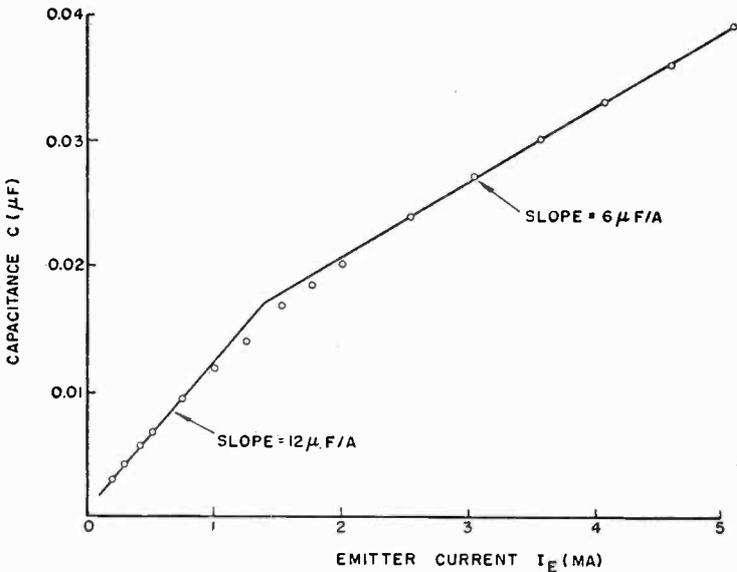


Fig. 16—Change of base-to-emitter capacitance with emitter current showing the two linear regions.

The formula for C_{in} is well checked by experiment, so far as variation with I_E is concerned, as may be seen from Figure 16. At low emitter currents, one finds a linear relationship with one slope, while at higher currents the diffusion constant effectively doubles due to large injected carrier density, as explained above. It is seen that the linear relation again holds, but with half the slope. The experimental evidence, therefore, confirms Webster's predicted effect of an electric field in the base.⁵ Because of the direct relation shown in Equation (19) between C_{in} and W^2 , a measurement of this capacitance is a means of calculating the base thickness W . This has been a standard procedure in this laboratory. We speak knowingly of the base thickness, as if there were

a direct measurement and, again, the data can be taken on a completed and encapsulated transistor.

For a base thickness $W = 0.002$ inch, and a p-n-p transistor, the diffusion time is

$$\tau = W^2/2D = 0.28 \text{ microsecond.}$$

This seems like a short enough time and, indeed, looking at Equation (18), whose time constant has only half this value, this is altered appreciably only when the frequency approaches 1 megacycle. But C_{in} , for $I_E = 1$ milliamper, and the same transistor, is calculated (and measured) to be 0.011 microfarad and, using the measured value of $r_{BB'}$ of Figure 14(b), the time constant of the $C_{in}r_{BB'}$ combination is about 4 microseconds. It is concluded that a major frequency effect is due to the input, since C_{in} cannot be compensated by tuning. It is also to be noted that C_{in} is *independent of junction area* and very critically dependent on the base thickness, W .

The equivalent circuit of Figure 15 uses α_{cb} as a basic low-frequency constant for the transistor, as is done with the μ of a triode electron tube. Frequency effects are entirely attributed to constant capacitances and conductances. This has many advantages over the use of a frequency-varying α and it is, perhaps, unfortunate that the latter concept is so prevalent. For the sake of completeness, however, it should be recognized that Figure 15 can be used to calculate a frequency-varying α_{ce} to be

$$\alpha_{ce} = \frac{(\alpha_{ce})_0}{1 + j\omega(W^2/2D)}$$

in agreement with the accepted formula.

To resume inspection of the equivalent circuit of Figure 15, a basic cause for g_{out} and g_{bc} is the change in position of the collector barrier when the base-to-collector voltage changes.¹⁵ It is simple to make such a calculation but not too important since surface leakage effects may contribute to these conductances to a significant extent. For the purposes of the present paper, these conductances are small enough to be neglected.

The capacitance C_{out} is that between collector and emitter and is also inherently small enough to be neglected. In fact, diffusion effects can be shown to lead to a small inductance in series with g_{out} , so that, at low frequencies, there is an inherent compensation of C_{out} .

¹⁵ J. M. Early, "Effects of Space-Charge Layer Widening in Junction Transistors," *Proc. I.R.E.*, Vol. 40, p. 1401, November, 1952.

The capacitance C_{bc} cannot be neglected. In the main, it is composed of the barrier capacitance of the collector junction, but there is a small contribution due to the change in barrier position with collector voltage. Of the two significant capacitances this is the only one which depends on junction area.

We shall now discuss the actual circuit values attained in a typical alloy p-n-p junction transistor, then show how changes in design can lead to a much improved high-frequency performance. Figure 17 shows the construction, dimensions and the equivalent circuit of a transistor which we shall here designate as the type TA 153 for convenience.² A typical unit uses a 0.006-inch base wafer of 3 ohm-centimeter germanium, with a round emitter of 0.015 inch diameter and an opposing

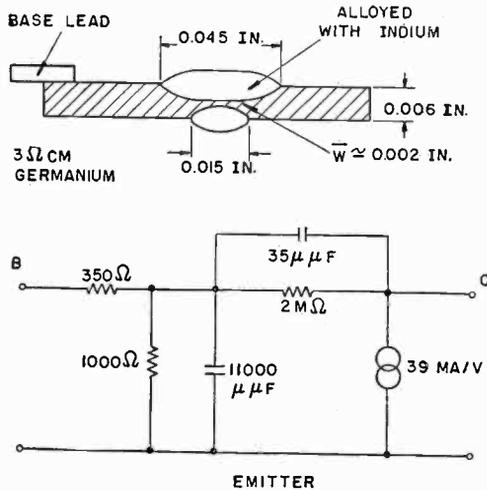


Fig. 17—Typical type TA 153 p-n-p alloy junction transistor dimensions and circuit constants at an emitter current of 1 milliamper.

collector three times as large. In practice, using pure indium as an alloying element, the junctions have an average spacing, \bar{W} , of about 0.0022 inch.

At an emitter current of 1 milliamper, collector voltage of 6 volts, a typical α value of 39 is obtained and the 4 kilocycles maximum power gain is about 40 decibels. Such a typical unit has the small-signal circuit constants shown in Figure 17, all of which can be measured readily. The base lead resistance is of the order of 350 ohms, the diffusion capacitance about 11,000 micromicrofarads and the collector junction barrier capacitance (plus its diffusion capacitance) is about 35 micromicrofarads. The resistance between collector and base is an extremely variable quantity but a typical value is 2 megohms. The

current generator has a transconductance of 39 milliamperes per volt but, it must be remembered, this is with respect to the internal base-to-emitter voltage, and *not* the applied voltage at the base terminal.

Let us examine the chief causes of poor high-frequency performance. First, we would like to reduce the base thickness, W , because this directly reduces the large diffusion capacitance. Then, since the germanium resistance largely determines the base lead resistance, we would like it reduced. Finally, the collector junction barrier capacitance is reduced by cutting down the junction area, so this is desirable. We shall now see how these principles have been applied to a recently developed radio-frequency amplifier transistor.

Before describing this work, perhaps some remarks about other high-frequency transistors are in order. The point-contact transistor has been operated as an oscillator up to 400 megacycles,¹⁶ but point-contact transistors are not well suited for amplifier use because of instability. Other work has been done on "tetrode" junction transistors, and these have been used successfully as amplifiers at frequencies up to the order of 50 megacycles.¹⁷ Their construction is inherently costly, however, in comparison to the alloy transistors hitherto described.^{2, 3} The research which we shall now discuss had as its objective the development of an improved radio-frequency transistor by a practical method which would retain the advantages of the alloy technique.

A NEW RADIO-FREQUENCY AMPLIFIER TRANSISTOR*

Figure 18 shows a cross-sectional view of the new transistor. First, it is noticed that, to reduce the base-lead resistance, a relatively thick wafer (0.020 inch) has been used, and the germanium resistivity of the base region is lowered to 0.7 ohm-centimeter. To allow a small base thickness, the collector junction is alloyed inside a small well which is cut into the wafer. This geometry has caused a five-fold decrease in the base-lead resistance. To reduce the collector barrier capacitance, the collector junction diameter has been reduced to one third. The emitter size was also reduced, to retain the favorable geometry which we have already discussed.

The most important improvement, however, lies in the means used for obtaining reduced base thickness. Because the wafer at the point of alloying is now extremely thin, and the junctions very small, accurate

¹⁶ F. L. Hunter and B. N. Slade, "High-Frequency Operations of p-Type Point-Contact Transistors," *RCA Review*, Vol. 15, p. 121, March, 1954.

¹⁷ R. L. Wallace, L. G. Schimpf, and E. Dickten, "A Junction Transistor Tetrode for High-Frequency Use," *Proc. I.R.E.*, Vol. 40, p. 1395, November, 1952.

* The research referred to in this section has been performed by C. W. Mueller and J. Pankove, Reference (6).

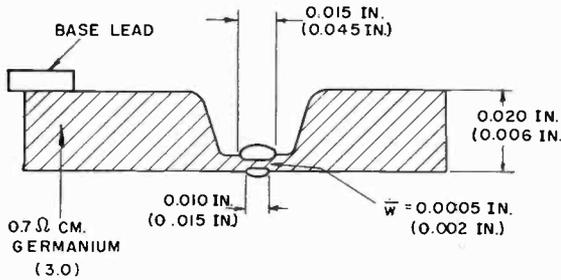


Fig. 18—Cross section of new p-n-p radio-frequency transistor. Dimensions in parentheses are those of the type TA 153, for comparison.

parallelism and control of the alloy penetration is essential. To prevent the relatively deep, hemispherical penetration which occurs when pure indium is used, the indium is first separately alloyed with about 5 atomic per cent of germanium. This partially-saturated alloy, then, can dissolve much less of the single-crystal base and penetrates less deeply than would pure indium. Simultaneously, it leads to a more parallel junction surface. The average junction spacing which can be attained in this construction is about 0.0005 inch or less. Since some of the important phenomena are associated with the *square* of this thickness, the improvement in these phenomena is a factor of about 20 or more over the transistor type TA 153 of Figure 17.

In Figure 19, we compare the equivalent circuit for the new transistor with values obtained on the earlier type TA 153. These figures are measured values and, although there are considerable variations among different units, may be considered typical. It is seen that the base-lead resistance is reduced by 5, the barrier capacitance by 3.5 and the diffusion capacitance by about 20 times. The resistance between collector and base, unfortunately, is not as high as that of the type TA 153 transistor. As a consequence, although the value of α remains

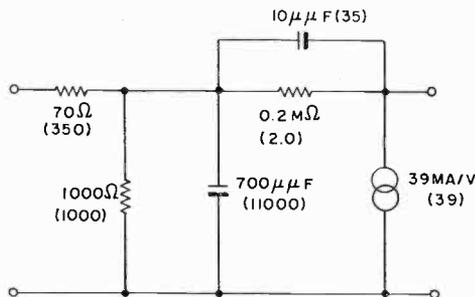


Fig. 19—Comparison of equivalent circuit of new radio-frequency transistor with type TA 153 both at 1 milliampere emitter current. Figures in parentheses are those of type TA 153.

about the same, the low-frequency gain is partially sacrificed. This is not a heavy price to pay for the very great improvement in high-frequency performance.

Figure 20 shows the measured power gain against frequency in a simple unneutralized gain test set, in which the output only was conjugate-matched by tuning at each test frequency. The input was a pure resistance. The unit tested was not quite so high in gain at low frequencies as the type TA 153 transistor, but it is seen that appreciable power gain remains (12 decibels) even at 10 megacycles. Both emitter-input and base-input curves are included. The former connection, in analogy to the grounded-grid triode, gives greater stability

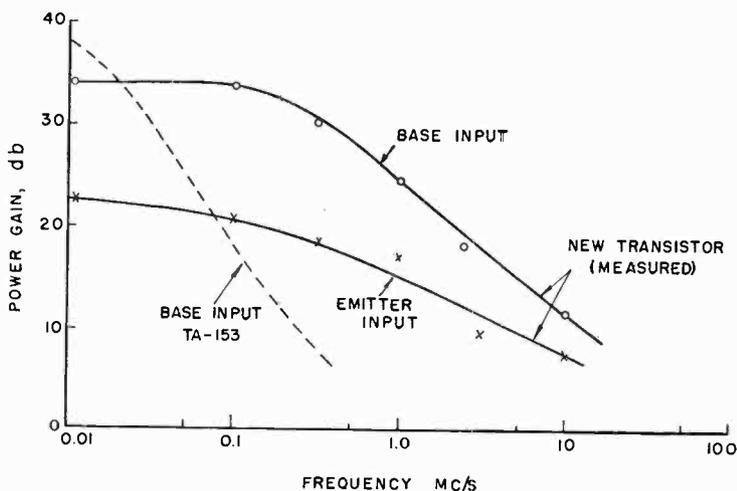


Fig. 20—Comparison of measured single-frequency power gain of a new unit with estimated gain of typical type TA 153 transistor at $I_E = 1$ milliampere, $V_C = -6$ volts. No neutralization.

at the expense of power gain. With base input, maximum gain requires neutralization of feedback and conjugate matching. When this is done, it has been found possible to obtain substantially higher gain than the values measured in Figure 20 for the frequencies up to 1 megacycle.

The noise factor against frequency of one of the new transistors is shown in Figure 21. It is there seen that the noise factor is quite low for intermediate-frequency or broadcast-band operation.

Although the upper oscillation limit of a transistor, as with a tube, is generally several times higher than its useful frequency range as an amplifier, it is one of the most convenient single numbers to specify performance. Transistors of the new construction described herein have an upper oscillation limit between 40 and 75 megacycles.

AN EXPERIMENTAL ALL-TRANSISTOR PERSONAL BROADCAST SET*

Although it has been possible to build an all-transistor broadcast set for some time, the performance attained with respect to sensitivity and signal-to-noise ratio was not as good as that of tube receivers. The new transistor just described, however, makes it possible to obtain performance comparable to an all-tube set in the radio-frequency characteristics, and exceed the performance of battery-operated all-tube receivers in nearly every other respect.

Such a receiver has been built and is shown in Figure 22. It is

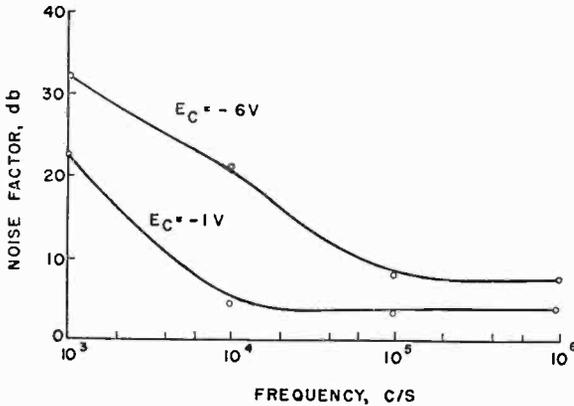


Fig. 21—Comparison of noise factor with frequency for new radio-frequency transistor. Base input. $I_B = 1$ milliampere.

smaller in size than the battery-operated all-tube receivers being sold in the United States under the designation "personal portable." It has, however, a much larger loudspeaker, about twice the audio output power and yet its low-cost battery of six ordinary flashlight cells will last 500 hours.

The receiver has nine transistors and one temperature-compensating junction diode. Six of the transistors are of the new radio-frequency (r-f) amplifier type we have just described. The remaining three are used for audio frequencies only. The signal is received by a ferrite-cored loop antenna. Miniature intermediate-frequency (i-f) transformers aid in keeping the chassis size small.

Figure 23 shows the circuit diagram of the receiver. There is a ferrite-cored, oscillator, capacitor tuned in conjunction with the ferrite loop. The mixer transistor has emitter injection of the oscillator, and

*The research work referred to in this section has been performed by L. E. Barton, "An Experimental Transistor Personal Broadcast Receiver," *Proc. I.R.E.*, Vol. 42, pp.1062-1066, July, 1954.

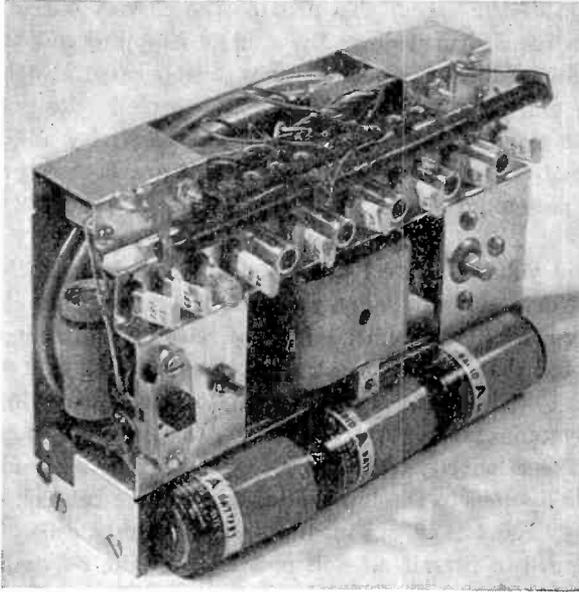


Fig. 22—Photograph of experimental all-transistor broadcast receiver using six of the new radio-frequency transistors, and three other transistors.

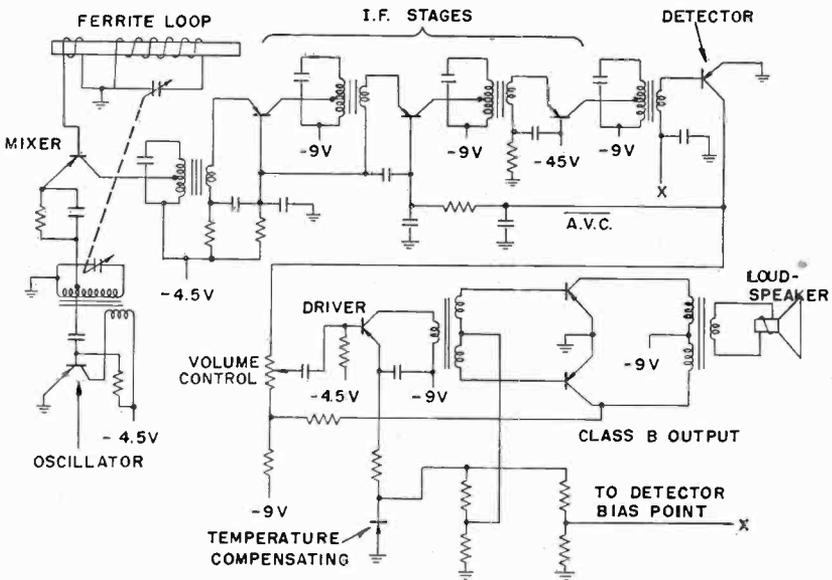


Fig. 23—Circuit diagram of receiver shown in Figure 22.

base injection of the signal. To provide the correct operating Q , the loop tuner is transformed down by a small coupling coil to the base. The three intermediate-frequency stages are operated at 455 kilocycles and use the emitter-input circuit. Small, specially designed, single-tuned, ferrite transformers provide both the selectivity and the proper impedance match for the transistors. There are both a primary tap and a small secondary winding. A stage gain of about 22 decibels is achieved.

Automatic volume control (a-v-c) is obtained from the collector circuit of the detector, which is another of the experimental transistors. Although a-v-c is shown connected to the base of the second i-f stage, it will be noted that variations of emitter current of this stage apply variable bias to the first i-f stage. Thus two stages are gain controlled. The audio system employs a driver stage, transformer coupled to two push-pull class B output transistors. Over-all inverse feedback from the output is applied to the low end of the volume control. The 9-volt battery supply consists of six medium-size flashlight cells. The supply is center tapped to permit 4.5-volt operation of the r-f amplifier transistors and to obtain the necessary d-c conditions for a-v-c and detection.

The circuits are arranged to reduce temperature variations in the transistors. This is achieved by current-limiting resistances in the emitters, which tend to maintain constant emitter current. The mixer has a self-bias feature which prevents excessive oscillator swing. A special means is provided to give temperature stabilization for the detector and the class B output units. This is done by means of a junction diode, which has a similar temperature characteristic to the collector-base circuit of the transistors. This diode is used in a potential-divider circuit and provides a bias which changes with temperature in the desired manner to maintain proper transistor operation.

The a-v-c characteristic compared with that of an all-tube receiver is shown in Figure 24. The audio response curve, similarly compared, is shown in Figure 25.

Comparison of Transistor Receiver with Battery-Operated Tube Receiver

	Tube receiver	Transistor receiver
Sensitivity (12.5 mW output), $\mu\text{V}/\text{m}$	125	155
Relative noise, db	0	+4
Maximum audio output, mW	75	150
Loudspeaker size, in.	2×3	4×6
Over-all size, cubic inches	140	73
Over-all weight, lb.	3.9	2.7
Total d-c power from batteries, mW	920	100
Battery life, hours	80-100	500
Battery cost per hour, cents	5.0	0.15

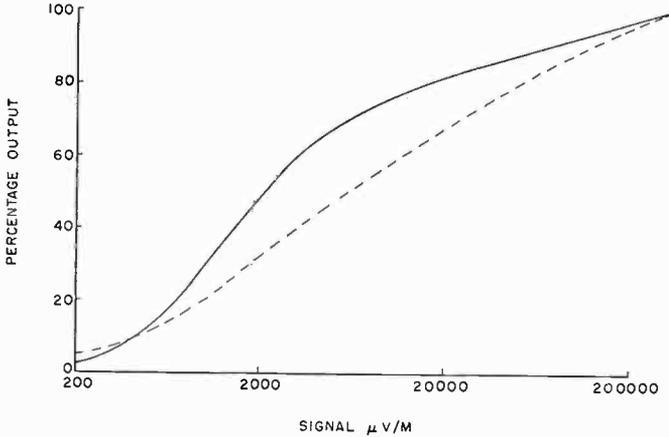


Fig. 24—Comparison of a-v-c characteristic of transistor receiver with that of a small battery-operated tube receiver.

— transistor receiver
 - - - tube receiver

Perhaps of greatest significance are the characteristics shown in the table. Of note are the sensitivity and signal-to-noise ratio, which differ insignificantly from the all-tube receiver. The higher audio output, improved quality of the larger loudspeaker, and reduced size and weight of the all-transistor receiver are marked. The really significant advantage, however, lies in the battery consumption, which is over 30:1 in favor of the all-transistor receiver.

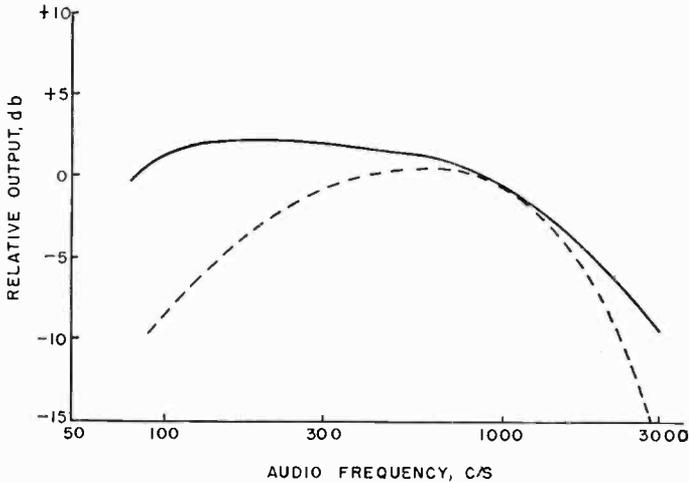


Fig. 25—Comparison of audio response of transistor receiver with small battery-operated tube receiver.

— transistor receiver (with feedback)
 - - - tube receiver

CONCLUSION

The research work of the writer's associates, as summarized herein, has led to a better understanding of the design parameters of the alloy-junction transistor, both for low- and high-frequency operation. Already this has led to an improved high-frequency transistor design which permits the frequency range to be extended by a factor of nearly a hundred over earlier units.

A SWITCHED-ZONE FURNACE FOR GERMANIUM PURIFICATION

BY

PAUL G. HERKART AND SCHUYLER M. CHRISTIAN

RCA Laboratories,
Princeton, N. J.

Summary—This paper describes a simple and improved apparatus for purifying germanium by passing a molten zone along the charge of germanium. In this method the heated zone is caused to travel along the charge without mechanical motion of crucible or heater. This is done by enclosing the charge in a series of separate resistance heating units which are individually turned on and off by a timing switch. To increase the speed, three traveling zones are used at one time.

This method is more compact than one using mechanically moved heaters or crucibles and is more economical than r-f induction heating. The entire unit, including power supply, occupies about 8 square feet, expends 4 kilowatts and produces 500 to 800 grams of intrinsic germanium overnight. The switching circuit provides for repeated passes of the molten zone, as many as desired, with no resetting or other attention.

THE ZONE-MELTING PROCESS

THE zone melting, as first described by Pfann¹ and applied to germanium, makes use of a segregation effect in which impurities tend to remain in the liquid portion of a freezing metal. As ordinarily used, a hot zone passes along a boat-shaped crucible, melting only a portion of the charge at any one time. Thus there is mechanical motion, either of the boat or of the heater, so that there is relative motion between the two. Such mechanical motion need not take place.

In the method described here this mechanical motion is replaced by the application of power to a succession of several heaters lying along the boat. This is done in sequence so that one or more molten zones in the germanium progress along the length of the charge.

CONSTRUCTION OF THE FURNACE

Figure 1 shows the external appearance of the furnace. It consists of 34 cells, of which one is sketched in detail in Figure 2. Each cell is one inch long. The charge is melted in the graphite boat, 40 inches

¹ W. G. Pfann, "Principles of Zone Melting," *Journal of Metals*, Vol. 4, p. 151, 1952.

long. This is kept in an inert or reducing atmosphere in the silica tube, extending through all the cells.

The cells are separated by $\frac{1}{4}$ -inch-thick firebrick in order to prevent the heat of one cell from melting the load in other cells at wrong times. These separators also prevent the silica tube from sagging. A hot zone occupies two cells, as described below.

Each cell contains a set of "GLOBAR" resistance heaters. The GLOBARS are supported by the side brick. Along the top of this lies a nichrome strip, from which spacer tabs extend down to hold the separators in place.

The brick furnace floor rests upon two transite sheets, separated by an air space and having their facing surfaces lined with aluminum sheets to reduce radiation. The covering bricks leave an open slit directly above the boat, serving three purposes: (a) to provide a view

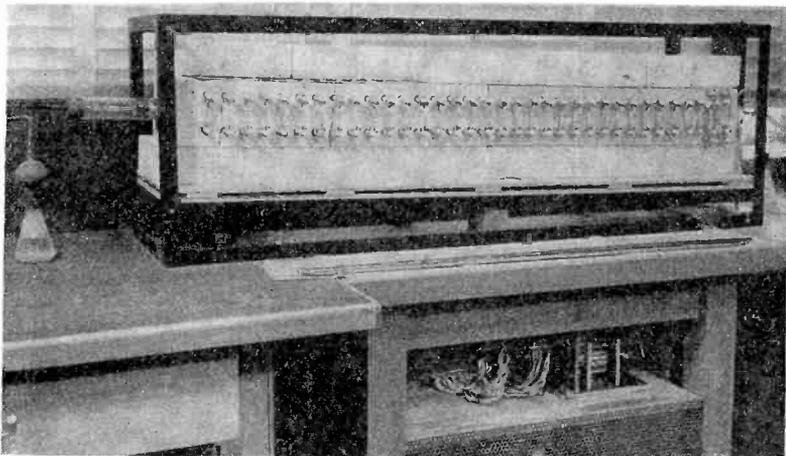


Fig. 1—Front view of furnace. A refined ingot may be seen on the table in front of the furnace.

of the molten material when necessary; (b) to hasten heat loss and freezing of germanium in cells not being heated; (c) to provide corrective heat loss so that the middle cells do not overheat. The opening is about $\frac{1}{2}$ inch wide near the middle of the furnace, tapering to $\frac{1}{4}$ inch near the ends.

The furnace is supported in an angle-iron frame and pivoted at one end to allow an adjustable slope. This is necessary because the expansion of germanium on freezing tends to push material ahead with the traveling zone. Tilting the forward end of the crucible about 2 degrees above the horizontal counteracts this effect and maintains a uniform cross section in the ingot.

It is necessary to heat two cells for each position of the molten zone to give overlap, so that the zone does not freeze before each new cell has heated to the melting point.

Each pair of GLOBARS is connected in series to form a cell resistor. To achieve the desired heating order, these resistors are in turn connected in series to form a ring of resistors as indicated in Figure 3. The junctions between resistors are attached each to two tap switches as shown. For instance, with the switches in position 2, resistor R_2 and R_3 are heated. One-fifth as much current also flows back through the rest of the ring, losing a negligible 4 per cent of the heating power.

A special switch construction is required. 12-tap "OHMITE" switches are mounted in tandem, with their stops removed to permit continuous rotation. To provide quick and positive movement of the switch from one position to the next, an auxiliary spring-action detent (Figure 4) is inserted between the switch and the reduction gear of the driving motor. The switch speed is one revolution in 45 minutes.

OPERATION

The boat charged with germanium is placed in the furnace so that the charge begins in the second cell. This preheats the end of the crucible and so aids in the initial melting of a zone. If power is applied at a constant level, the germanium may not melt during the first pass of the hot zone. This is permissible, since the furnace operates automatically overnight, making an ample number of effective passes.

On stopping the furnace, provision must be made for the last zone to travel to the impure end while no new zones start. Any unfinished zones would leave impure spots in the ingot, and might form troublesome "sprouts" of expanding germanium where the zones freeze. To prevent this difficulty, two 12-pole knife switches may be inserted in the leads going respectively to the first two groups of resistors. With these the power may be cut off from the first and second groups just in time to avoid starting new zones. Or more simply, the operator may progressively remove bricks from the top of the furnace, keeping the charge frozen behind the last zone.

The power supply is a large adjustable autotransformer. By adjusting the voltage of this source, the length of the molten zone is controllable within limits. About 35 amperes at 110 volts maintain proper zone length. This may vary between 3 and 4 inches, depending on changes of line voltage and of heat loss.

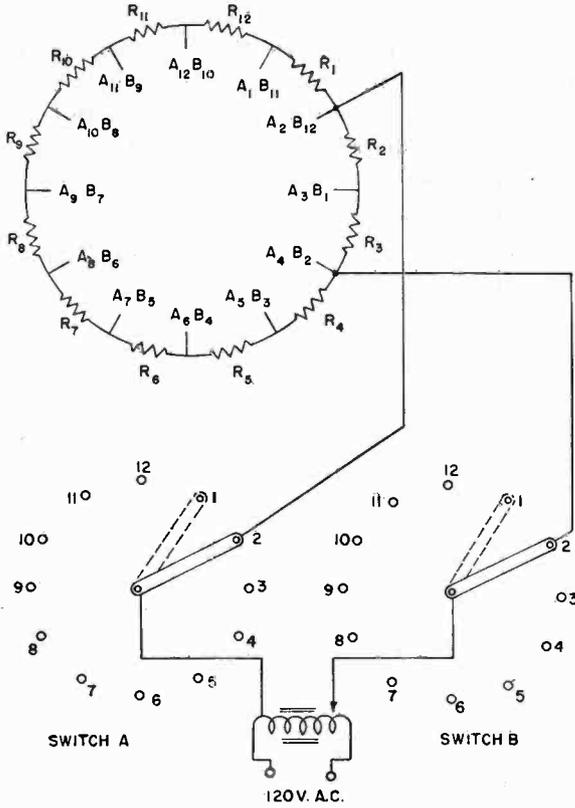


Fig. 3—Switching circuit.

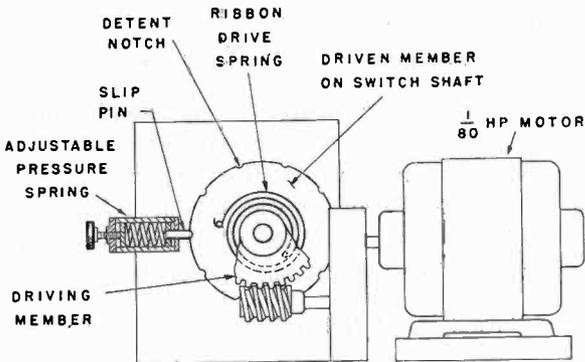


Fig. 4—Drive for tap switch.

DISCUSSION

Mathematical analysis of zone refining^{1,2} shows that the impurity concentration should vary along an ingot approximately as shown by the solid lines in Figure 5. The present crucible length is about 10 zone lengths, and the effective distribution coefficient for native germanium

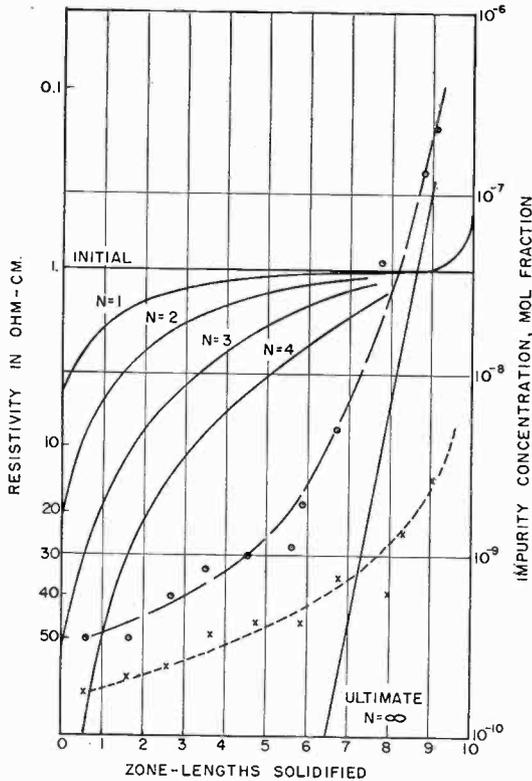


Fig. 5—Resistivity of two refined ingots. x, initial charge about 10 ohm-centimeters. \odot , initial charge about 1 ohm-centimeter. Solid lines show theoretical impurity distribution for this case.

impurities is about 0.2. The parameter N is the number of zone passes. The plotted points show measured resistivity values on typical ingots after 18 passes in an overnight run of the furnace.

The ingot represented by circles was produced from a 1,000-gram load of rejected ingot tails with resistivity about 1 ohm-cm. The impurity concentration is related to the resistivity.³ As the data indi-

² N. W. Lord, "Analysis of Molten-Zone Refining," *Journal of Metals*, Vol. 5, p. 1531, 1953.

³ P. G. Herkart and J. Kurshan, "Theoretical Resistivity and Hall Coefficient of Impure Germanium Near Room Temperature," *RCA Review*, Vol. XIV, pp. 427-440, September, 1953.

cates, about half of the charge is purified practically to the intrinsic range. A better charge yields about 75 per cent of intrinsic germanium.

The purity obtained is quite satisfactory for most purposes, so that more elaborate development of the furnace was not worthwhile. For higher purity: (a) the cell-length may be shortened to give zones down to about 2 inches; (b) the furnace may be operated for a longer time, though there is small advantage in more than 18 passes; (c) for the greatest improvement, the furnace may be reloaded with a charge of higher purity. Even with an infinite number of passes, there is a finite impurity concentration at every part of the ingot, because not all of the impurity can be restricted to the final zone to freeze. This ultimate impurity concentration is proportional to the total impurity in the charge.

The efficiency of each pass is somewhat low, probably because of the intermittent motion of the zones. Eighteen passes are more than are necessary for usual purposes, but are a convenient number for an overnight run.

Thus the switched-zone furnace is simple in construction and automatic in operation. It is shorter than systems using a moving crucible, and eliminates some problems involved in mechanical motion at high temperature. It is much smaller and less costly to build or operate than an induction-heated furnace of the same capacity. Several furnaces of this type have produced high-purity germanium with trouble-free operation over a period of many months.

CONTINUOUS-PROCESS APPARATUS FOR GROWING SINGLE-CRYSTAL GERMANIUM

By

HERBERT NELSON

RCA Laboratories,
Princeton, New Jersey

Summary—The development and testing of apparatus for growing continuous-process single-crystal germanium of controlled and uniform electrical resistivity are described. This continuous process, using the Czochralski technique, involves the simultaneous feed-in of polycrystalline material and the withdrawal of single-crystalline material. A mathematical expression relating pertinent parameters to the electrical resistivity of the single crystal is also given. With this expression as a guide, the technique described lends itself to the production of crystals which are of a predetermined and relatively uniform resistivity throughout the bulk.

GENERAL CONSIDERATIONS

WHEN in early transistor work with germanium and silicon it became evident that single-crystal material must be used for best results, the Czochralski¹ method became the popular technique for producing this material. By this method crystals are grown from a melt by touching a single-crystal seed to the liquid surface and slowly withdrawing it as new lattice layers are added to it by germanium freezing out from the melt. One type of apparatus for growing single crystals by this method has been described by L. Roth and W. E. Taylor.²

Single crystals produced in this and similar apparatus provide acceptable material for germanium transistors and rectifiers. These crystals do, however, suffer from the limitation that they vary in resistivity from one end to the other. Since, in general, the impurities have lower solubilities in solid than in liquid germanium, the concentration of these impurities in the melt and the rate at which they enter the crystal lattice increase as more and more of the melt is incorporated into the growing crystals. The end of the crystal first drawn from the melt will, therefore, possess a lower impurity concentration and a higher electrical resistivity than will the end drawn out last. A curve

¹ J. Czochralski, *Z. Physik. Chem.*, Vol. 92, p. 219, 1918; G. K. Teal and J. B. Little, "Growth of Germanium Single Crystals," *Phys. Rev.*, Vol. 78, p. 647, June, 1950.

² Louise Roth and W. E. Taylor, "Preparation of Germanium Single Crystals," *Proc. I.R.E.*, Vol. 40, p. 1338, November, 1952.

indicating the typical resistivity variation obtained is shown in Figure 1.

Since in the fabrication of semiconductor devices it is important that as little as possible variation in material exist from unit to unit, it follows that crystals more uniform in resistivity from one end to the other are desirable. An apparatus for growing single crystals of this type has been developed and is described in this paper.

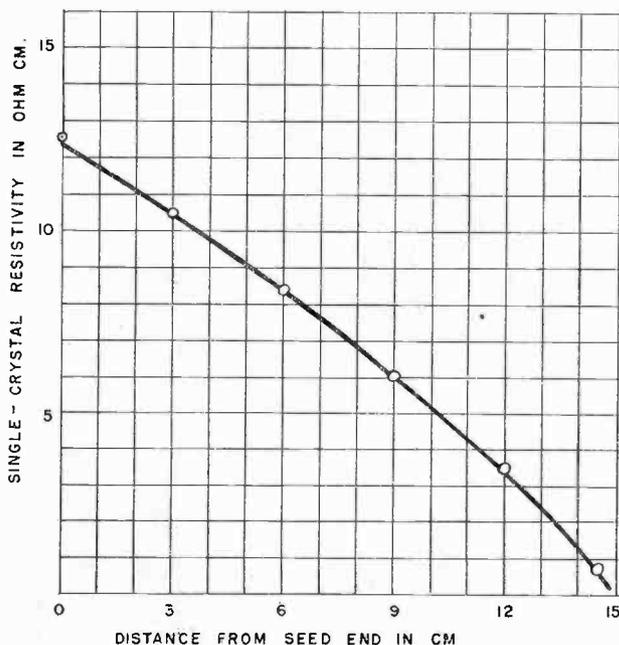


Fig. 1—Variation in resistivity along length of single crystal grown by original Czochralski method.

APPARATUS AND PROCEDURE

The apparatus which has been constructed and put into operation is of the continuous-process type and is shown in Figures 2 and 3. It consists of a horizontal vycor furnace tube 24 inches x 2½ inches, with two 18-inch x 1½-inch vertical side arms supported in a W & B No. 30 firebrick furnace and heated by "GLOBARS." In the furnace tube below the vertically placed side arms, a graphite boat is positioned so that one of the two channel-connected pot-holes is centered with the axis of each side arm. A germanium charge of 250 grams is placed

in the graphite boat. A properly oriented single-crystal seed is clamped in a stainless-steel holder which extends through a brass bushing closing the top of one of the vycor-tube side-arms. The threaded cylindrical rod which forms the upper part of the holder passes through a mechanism for lifting and turning the seed holder.

Suspended above the second pot-hole in the graphite boat is a rod of polycrystalline germanium which is clamped into a stainless-steel holder extending through a bushing at the top of the second vycor-tube

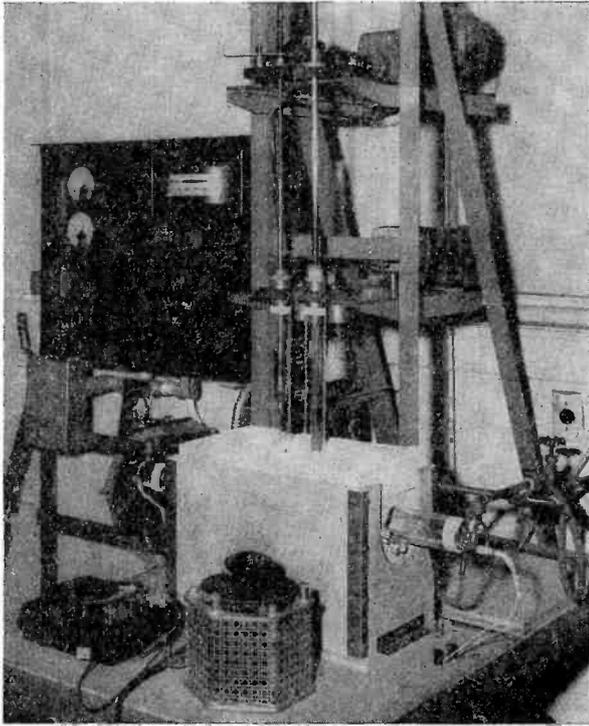


Fig. 2—Photograph of apparatus for continuous-process production of single-crystal germanium.

side-arm. This holder, after passing through a second guide bushing, is attached by means of a wire cable to a lowering mechanism. The open ends of the vycor furnace tube are closed by brass caps provided with quartz windows for observation purposes. The brass caps are also provided with gas inlets and holes for a thermocouple and an auxiliary heater.

With argon mixed with about 10 per cent hydrogen flowing through the system, the germanium is melted by heat supplied through the

Globars. The molten germanium is elevated to a temperature of about 1000°C and maintained at this temperature for about 5 minutes. The feed-in mechanism is then started to lower the polycrystalline rod of germanium into the melt at a rate of about one millimeter per minute.

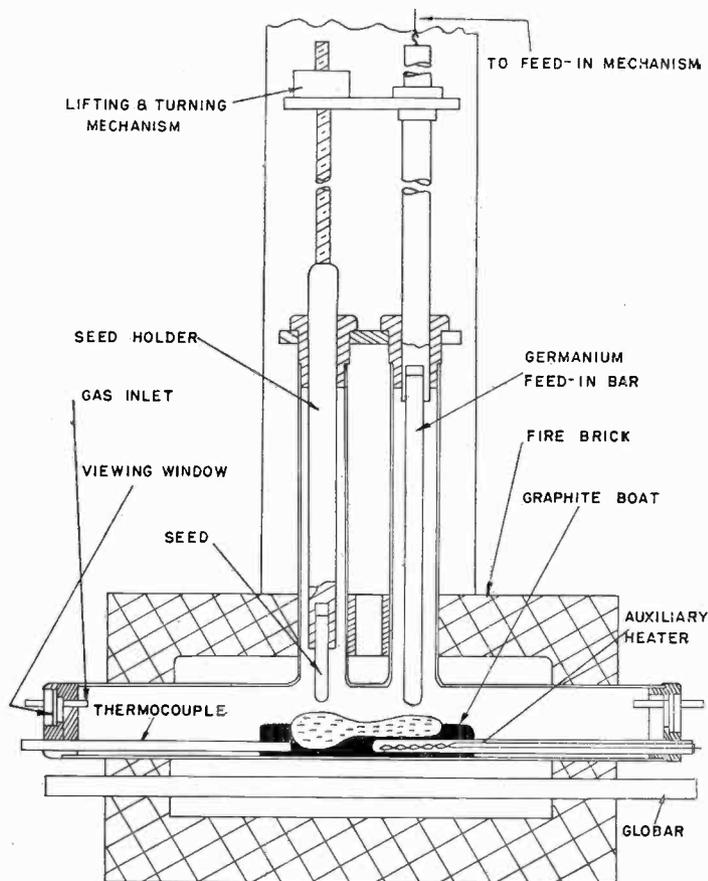


Fig. 3—Cross-section of furnace shown in Figure 2.

A Wheelco^{*} temperature controller, equipped with XacTline[†] anticipator and means for high-low rather than on-off regulation, is set to lower and maintain the temperature of the germanium at a value just above the melting point, 936°C . Temperature information is supplied to

* Trademark of the Wheelco Instrument Corp., Chicago, Ill.

† Trademark of the Claude S. Gordon Co., Chicago, Ill.

the regulator by a thermocouple enclosed in vycor tubing and inserted into the end of the graphite boat. To supply extra heat for melting germanium from the feed-in rod, an auxiliary nichrome heater, enclosed in vycor tubing, is inserted into the graphite boat below the feed-in pot-hole. The seed, which has been preheated by holding it about 2 millimeters above the surface of the melt, is brought into contact with the liquid germanium and withdrawal begun at a rate of 1 millimeter per minute and with the seed rotated at 1 revolution per second. Final adjustment of temperature is made to cause the single crystal to grow at a desired diameter (usually 2 centimeters when germanium is removed from the melt at the same rate as at which it is fed in).

CONTROL OF RESISTIVITY OF THE SINGLE-CRYSTAL MATERIAL

The apparatus and procedure described above can be used to grow crystals uniform in resistivity throughout the bulk. If, for instance, *n*-type single crystal is desired, the proper impurity concentration is introduced into the melt to cause donors to enter the crystal lattice under equilibrium growing conditions at a rate which produces the required donor concentration in the single crystal produced. To keep the impurity concentration in the melt constant, the feed-in material is chosen to cause the introduction into the melt of donors at the same rate as they enter the growing crystal.

To make possible control of the factors referred to above for the purpose of predetermining the resistivity of single-crystal material and for evaluating the effects of lack of perfect control under practical operating conditions, a mathematical expression relating pertinent parameters to the electrical resistivity of the single crystal must be derived.

For this purpose let us assume that in producing single crystals of predetermined resistivity we limit ourselves (as is for practical purpose desirable) to the use of intrinsic germanium for raw materials, and to the use of one impurity or "doping" element of the acceptor type for the preparation of *p*-type material and one doping element of the donor type for the preparation of *n*-type material. Let us further assume that the contribution of the thermally produced carrier to the conductivity of the germanium is negligible. It follows then, since each impurity atom supplies one current carrier to the crystal lattice, that the conductivity of the single-crystal material formed is directly proportional to the density, n_0 , of the impurity included in the lattice. Considering also that this density is directly proportional to the density, n , of impurity in the melt and that the factors of proportionality can be determined for any particular impurity used, it follows

that knowledge of n as a function of time will permit the calculation of the resistivity and the variation in resistivity from one end to the other of single crystals produced.

At the start of crystal formation and of feed-in, $n = N_0/V$ where N_0 is the number of impurity atoms originally introduced into the volume V of the melt. If, however, the rate at which the impurity atoms leave the melt at the point of crystal growth differs from the rate at which they are introduced at the point of feed-in, the total number of impurities in the melt, N , will change with time so that, after an increment of time Δt ,

$$N = N_0 + (A_f v_f n_f - A_c v_c n_c) \Delta t, \quad (1)$$

or

$$\Delta N = (A_f v_f n_f - A_c v_c n_c) \Delta t, \quad (2)$$

where A_f , v_f , n_f , and A_c , v_c , n_c are the cross-sectional area, the feed-in rate or growth rate, and the density of impurity atoms of the feed-in rod and of the single crystal grown, respectively.

If the operating conditions are arranged so that $A_c = A_f$ and $v_c = v_f$, and if the symbol V_r is substituted for both $A_f v_f$ and $A_c v_c$, then

$$\Delta N = (V_r n_f - V_r n_0) \Delta t, \quad (3)$$

or, since

$$\Delta n = \frac{\Delta N}{v},$$

$$\frac{\Delta n}{\Delta t} = \frac{V_r n_f}{V} - \frac{V_r n_0}{V}, \quad (4)$$

and since $n_c = kn$, where k is the segregation coefficient of the impurity used,

$$\frac{dn}{dt} + \frac{V_r kn}{V} = \frac{V_r n_f}{V}. \quad (5)$$

A solution of this differential equation with initial conditions taken into account gives

$$n = \left(n_0 - \frac{n_f}{k} \right) e^{-\frac{V_r k t}{V}} + \frac{n_f}{k}, \quad (6)$$

where n_0 = density of impurity atoms in melt at $t = 0$.

Since resistivity is inversely proportional to density of impurity atoms, the performance of the requisite algebraic operations gives

$$\rho_c = \frac{\rho_f}{1 + [(k\rho_f/\rho_0) - 1] e^{-\frac{V_r k t}{V}}}, \quad (7)$$

where ρ_c , ρ_f and ρ_0 are, respectively, the resistivity of the material added to the crystal lattice, of the feed-in material, and of the material used for loading the crucible.

An inspection of Equation (7) indicates directly that for all values of other constants, if $\rho_f = \rho_0/k$, ρ_c is independent of t and is equal to ρ_f , but that if $\rho_f \neq \rho_0/k$, the value of ρ_c approaches ρ_f only when t becomes very large.

Furthermore, Equation (7) makes possible the calculation of the value of ρ_0 required for any desired resistivity (below 10 ohm-centimeters) of the single-crystal material grown, and also the variation in this resistivity with length grown, for cases where the resistivity of the feed-in rod, ρ_f , differs from the resistivity desired for the single-crystal material.

At $t = 0$, Equation (7) reduces to $\rho_c = \rho_0/k$. Therefore, if a donor element for which $k = 0.005$, is used for preparation of n-type germanium, it follows that the value of ρ_0 to use for any particular ρ_c may be determined by the simple relation $\rho_0 = 0.005 \rho_c$. If single-crystal material having a resistivity of 2 ohm-centimeters is desired, for instance, intrinsic germanium doped with this donor element to reach a resistivity of 0.01 ohm-centimeter must be used for the crucible melt.

The change with time of the resistivity of the germanium added to the lattice of the growing single crystal may be calculated from Equation (7) by insertion of the numerical values for the constants involved. For a typical set of operating conditions, these are

$k = 0.005$ Distribution coefficient of donor element in germanium.

$V_r = 0.314 \text{ cm}^3/\text{min.}$ Corresponding to a single crystal 2 centimeters in diameter and of feed-in bar, and a rate of feed-in equal to the rate of crystal growth = 0.1 centimeter per minute.

$V = 70 \text{ cm}^3$ Volume of melt in boat.

$$\rho_0 = 0.01 \text{ ohm-cm.}$$

Equation (7) thus becomes

$$\rho_c = \frac{\rho_f}{1 + (0.5 \rho_f - 1) e^{-0.0000224t}} \text{ ohm-cm.} \quad (8)$$

It follows from an inspection of this equation that for values of $t < 1000$ minutes, $e^{-0.0000224t}$ is very nearly equal to 1, and that ρ_c is equal to 2 ohm-centimeters and is, for practical purposes independent of ρ_f for values of this parameter greater than 2 ohm-centimeters. For $\rho_f = 40$ ohm-centimeters, for instance, ρ_c calculated from Equation (8) is equal to 2.06 ohm-centimeters at $t = 1,000$ minutes. It is possible, therefore, by feeding in intrinsic or near-intrinsic germanium, to draw from doped melt a total length of 100 centimeters of single crystal 2 centimeters in diameter before an appreciable increase in resistivity occurs.

In Figure 4, a curve calculated from Equation (7) is compared with experimental data. In spite of the fact that the determinations of n_0 and k are subject to error, and that maintaining V and V_r constant is difficult during operating conditions, the correlation is good. Although near-intrinsic germanium was fed in during the growth of the crystal, its resistivity as shown by the data did not increase with time.

Results similar to those indicated above may be obtained when other doping agents are used, if the values of k for these impurities are of the same order of magnitude. As a consideration of Equation (8) shows, \bar{t} , the length of time of growth before a noticeable increase (about 3 per cent) in resistivity occurs as a result of feed-in of near-intrinsic germanium, varies with k , as indicated by the relation

$$\bar{t} = \frac{5}{k} \text{ minutes.}$$

If indium, $k = 0.001$, is used, for instance, for the preparation of p-type germanium, $\bar{t} = 5000$ minutes. With regard to the above relation as well as with regard to prior equations involving the distribution coefficient k , a word of caution is in order. If crystal growth occurs under conditions at which a pile-up of impurities occurs in the growth region of the melt, as a result of insufficient agitation or of rapid solidification, an effective rather than true value of k must be used. This effective value can, of course, be empirically determined for any particular condition of growth.

DISCUSSION

Single-crystal growing apparatus of the type described above have been operated under laboratory conditions for an extended period. Though the operation of these furnaces has not been trouble-free, it has led to the production of many kilograms of single-crystal germanium of excellent quality for transistors. By the use of single-crystal seeds of different orientation, crystals have been grown in different crystal directions, especially in the [100], the [110], and the [111] directions. Growth in the [111] direction appeared most trouble-free and, consequently a large majority of the crystals produced were grown in that direction.

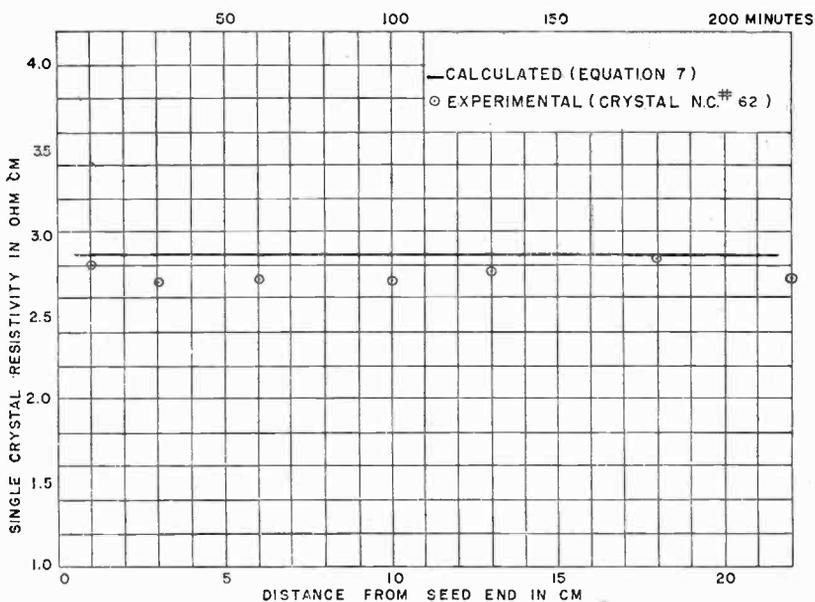


Fig. 4—Variation in resistivity along length of continuous-process crystal.

Some early difficulties were: melt solidification at edge of the boat crucible during crystal growth, twinning, and nonuniform distribution of impurities throughout the single-crystal lattice. The first of these difficulties, melt solidification, occurred as a result of the presence of "cold spots" at the ends of the graphite boat which "saw" the relatively cold end-caps of the vycor furnace tubes. To minimize these "cold spots" and to eliminate their adverse effects, the following measures were taken:

1. Introduction of heat shields between the ends of the graphite boat and the ends of the furnace tube.

2. Cutting of a slot in the furnace brick directly above the center of the graphite boat to eliminate or reduce temperature gradient from middle to ends of boat.

3. Use of GLOBARS having longer hot sections, and the placing of these closer to furnace tube.

By resorting to the above measures, it was found that the melt solidification difficulty could be eliminated but that a higher incidence of twinning in the growing crystal resulted. Apparently the increase in temperature gradient from the center of the melt to the hotter potrim, which led to the elimination of one difficulty, was a factor in bringing about the presence of another. The two upper GLOBARS were, therefore, placed slightly higher with respect to the melt surface, and provisions were made for greater conduction of heat away from the growth region of the melt through the single crystal. These changes resulted in a lower radial thermal gradient over the melt surfaces and a higher gradient in the very region of crystal growth, and brought about reduction in the incidence of twinning to the point where it was negligible.

In connection with work done to modify the thermal gradients in the furnace, the observation was made that the degree of inhomogeneity in the distribution of impurities in the crystal lattice was greatly affected by these modifications. When in early work with the furnaces a relatively high thermal gradient existed between center and ends of the graphite boat, and when the crystals were grown at a low rate of rotation, 1 revolution per minute, a cyclical variation in resistivity (striations) occurred along the length of the crystal.³ Since crystal growth is more rapid at the relatively cool side of the melt, rejection of impurities by the growing lattice tends to be diminished due to "pile-up" of rejected impurities in this region. Lower resistivity material is, therefore, added to the lattice on the "cool" rather than on the "hot" side where the crystal growth is lower and less increase in concentration of impurities occurs. To eliminate or minimize this non-uniformity in impurity distribution, the rate of crystal rotation was increased to 60 revolutions per minute to effect better mixing. As a consequence, the striation-type of inhomogeneity was greatly reduced.

The observation was also made that the thermal gradient conditions, which led to melt solidification when "cold spot" regions existed in the furnace, led to spurious resistivity variations along the length of the single crystal and also along its cross-sectional diameter, even at a high rate of rotation of the growing crystal. These types of nonuni-

³ P. R. Camp, "Resistivity Striations in Germanium Crystals," *Jour. Appl. Phys.*, Vol. 25, p. 459, April, 1954.

formities were found to be greatly minimized after the modifications to optimize thermal gradient conditions in the furnace were effected. Before-and-after results are shown in Figure 5.

The work done to optimize thermal gradients in the furnace was carried out on a qualitative rather than on a quantitative basis, and the above paragraphs should be interpreted accordingly.

With regard to the rate of single-crystal "pull-out," higher rates than 1 millimeter per minute have been used experimentally but sometimes led to exaggerated melt solidification and twinning difficulties. It appears, however, that optimization of thermal gradients in the furnace will make higher rates of crystal "pull-out" practical.

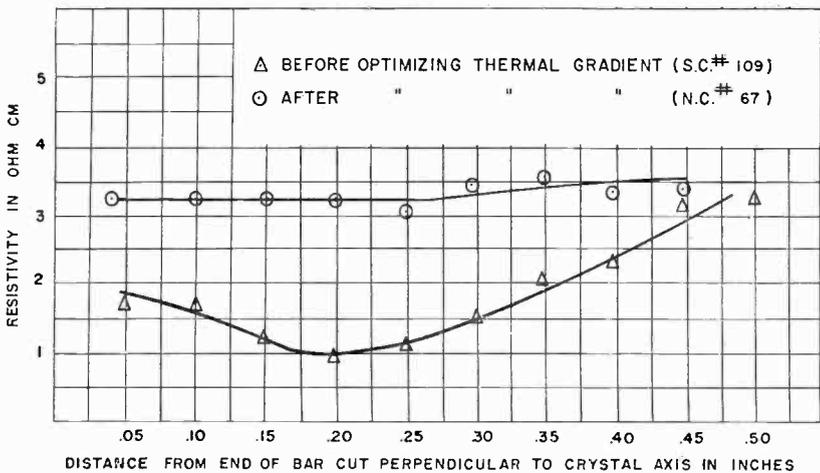


Fig. 5—Radial variation in resistivity of continuous-process crystal.

As has been suggested, the close control of the temperature of the germanium melt is important for the maintenance of optimum operating conditions. A very small change in the temperature (1 or 2°C) of the growing crystal surface will materially affect the rate of growth and upset the desired equilibrium conditions. For this reason manual control of the temperature was used to supplement automatic control furnished by a voltage regulator and a Wheelco temperature regulator. Since supplementary manual temperature control was needed only because of the change in the temperature loss which occurs as the growing crystal is gradually pulled away from the melt surface, it follows that proper programming of the temperature regulation can effectively replace supplementary manual control.

THE PREPARATION OF SINGLE AND MULTIPLE P-N JUNCTIONS IN SINGLE CRYSTALS OF GERMANIUM

BY

A. R. MOORE

RCA Laboratories,
Princeton, N. J.

Summary—A method and an apparatus are described for the preparation of junctions in crystals of Germanium. The composition of the melt from which a crystal grows is changed suddenly by rotation of a carbon block containing several melt compositions. A system of canals and valves keeps the crystal growing and minimizes mixing of the melts during the change-over. In principle any number of junctions between germanium of arbitrary composition can be formed.

INTRODUCTION

THE preparation of single crystals of germanium by the Czochralski technique has been discussed in the literature by Teal and Little.¹ With suitably designed furnaces it appears that this method permits the growth of crystals which are very nearly perfect and have minority carrier lifetimes in excess of 1000 microseconds. P-N junctions and junction transistors have also been described^{2,3} and their physical and circuit characteristics investigated. However, there have been few publications on the details of the technique of making such junctions in growing single crystals. This paper describes a furnace and a method for the laboratory preparation of junctions in a growing crystal, whereby the conductivity type (p or n), resistivity, and dimensions of the various sections can be controlled.

The principle of operation is that the composition of the melt from which the crystal grows is changed during growth in such a way as to incorporate n or p type impurities into the grown section. During the process, as the germanium is changed from n to p to n type, the crystal grows continually, and does not lose contact with the liquid germanium surface.

¹ G. K. Teal and J. B. Little (A), "Growth of Germanium Single Crystals," *Phys. Rev.*, Vol. 78, p. 647, June 1, 1950.

² W. Shockley, M. Sparks, and G. K. Teal, "p-n Junction Transistors," *Phys. Rev.*, Vol. 83, p. 151, July 1, 1951.

³ R. L. Wallace and W. J. Pietenpol, "Some Circuit Properties and Applications of n-p-n Transistors," *Proc. I.R.E.*, Vol. 39, p. 753, July, 1951.

DESCRIPTION OF THE APPARATUS

Figure 1 shows a furnace design which has been found suitable for junction preparation. The design illustrated is used mainly for n-p-n structures but it will be seen that single p-n junctions can also be made, and the method can be extended to any multiplicity of junctions. The furnace is composed of a 3-inch diameter quartz or vycor tube, (1), set vertically in a firebrick box. The heating elements, (2), are six vertical Globars* disposed symmetrically about the vertical quartz tube and supplied with power by means of a Variac† and some type

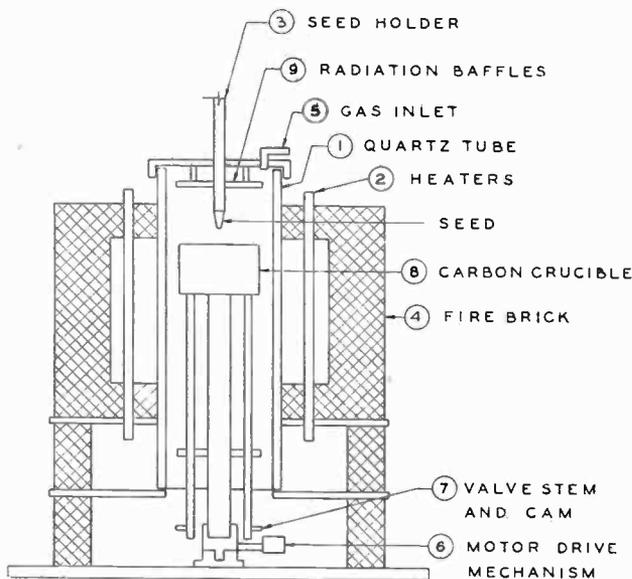


Fig. 1—Crystal-growing furnace.

of constant temperature control system. For short-length crystals, which are grown in a relatively short time, control by constant-voltage transformer is satisfactory; long crystals require thermocouple feedback-type control, inasmuch as the temperature must be held constant over long periods of time. During the growth period, the temperature is held within 1 degree C of the growth temperature. A rare gas atmosphere such as argon or helium is maintained in the tube to prevent oxidation. Radiation baffles at the top and bottom cut down heat loss.

The seed holder, (3), is a stainless-steel rod that goes through a

* Trade mark of the American Resistor Corp., Philadelphia, Pa.

† Trade mark of the General Radio Co., Cambridge, Mass.

hole in the upper plate and radiation baffles. It has a single crystal of germanium of known crystallographic orientation fastened to its lower end in such a position that it can be lowered until the seed just touches the liquid germanium in the crucible. The mechanism for withdrawing the seed at a controlled rate is not shown; it is a geared clock motor and pulley system winding a phosphor-bronze wire fastened to the upper end of the seed holder onto a drum.

The most important part of the furnace is the crucible, shown in

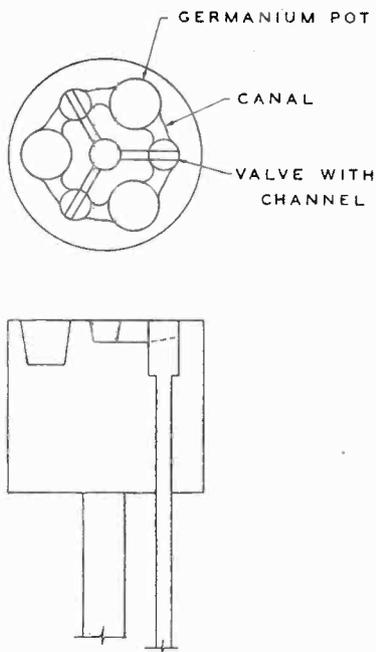


Fig. 2—Carbon crucible.

detail in Figure 2. It is a block of carbon $2\frac{1}{2}$ inches in diameter with three pots for germanium drilled into its surface 120 degrees apart. These pots contain germanium of the proper conductivity type and of such impurity concentration that, after allowance is made for the normal segregation of impurities⁴ during crystallization, the drawn crystal will have the required resistivity. The pots are connected by shallow canals. The liquid germanium from adjacent pots is prevented from flowing together through these canals by a set of rotatable carbon valves, one for each canal. These valves have channels cut into their

⁴C. H. Desch, *Metallography*, Longmans Green and Co., 1944; p. 146.

surface so that when the channels are rotated to be perpendicular to the canals the pots are effectively isolated; when they are rotated through 90 degrees the canals are connected by a stream of germanium. The rotation of the valves is done by means of the valve stems which extend through the crucible and protrude out of the bottom of the furnace as shown.

The three radial channels connecting each valve to the center hole, as spokes of a wheel to the hub, are for the purpose of preventing contamination of adjacent pots during the time interval that the valves are turned so that the canals connect. The bottom of each valve channel has a slight pitch, so that when it is turned perpendicular to the canal, all germanium in the valve, which includes the intermix zone in the liquid germanium, is dumped into the center hole. This hole collects all contaminated or intermixed material, which is later returned to a germanium purification unit, not part of the equipment under discussion. If the operation of rotating the valves is done rapidly enough the pots of germanium remain essentially unchanged in composition.

OPERATION OF THE APPARATUS

In typical operation, pot I is filled with 1-2 ohm-cm n-type material, pot II with 0.05-0.1 ohm-cm p-type material, and pot III with 0.001 ohm-cm n-type germanium. To prevent the possibility of contamination due to melt-off from the grown crystal, the junctions are always grown from the purest germanium first. Then a small quantity of melt-off has little effect on the next pot which has a much higher concentration of impurities. The crystal is started in pot I with all valves closed. After a suitable length of material has been grown by raising the seed, the crucible is rotated by a motor-driven mechanism to the next pot position, while a cam opens the valve as the crystal passes through and immediately closes it again after passage. The germanium left in the valve is emptied into the center hole. The entire operation of transfer from pot to pot takes about one second. The width of the central p-type region is adjusted by controlling the time the crystal is allowed to remain in pot II. Then the operation is repeated and a length of crystal is grown from pot III. With melts of the type specified above, single-crystal junctions can be made with the resistivities of the respective sections: 5-6 ohm-cm, 0.2-0.3 ohm-cm, 0.005-0.01 ohm-cm. The central p-type region has been made as thin as several mils, with a normal pulling rate of 1 millimeter per minute. For the thinnest p-type regions (about 1 mil) the pulling rate can be reduced to .2 millimeter per minute during growth in pot II.

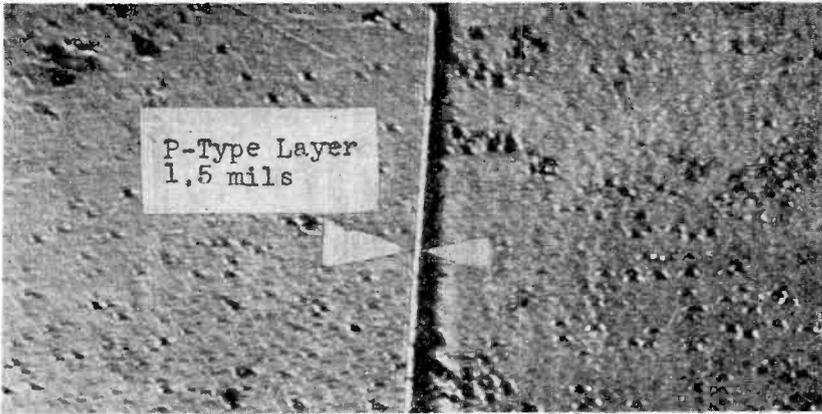


Fig. 3—Photograph of an n-type crystal containing a p-type layer.

Figure 3 is a photograph of one such crystal in which the p-type region has been made visible by cutting the crystal longitudinally and by differential etching of the surface. In this crystal the p-type region is about 1.5 mils thick. Figure 4 is a potential plot along another crystal when the ends are held at a fixed difference of potential of 0.5 volt. The location of the junctions can be clearly seen. In this case the central p-type region was made about 15 mils thick so that a potential probe plot could be obtained.

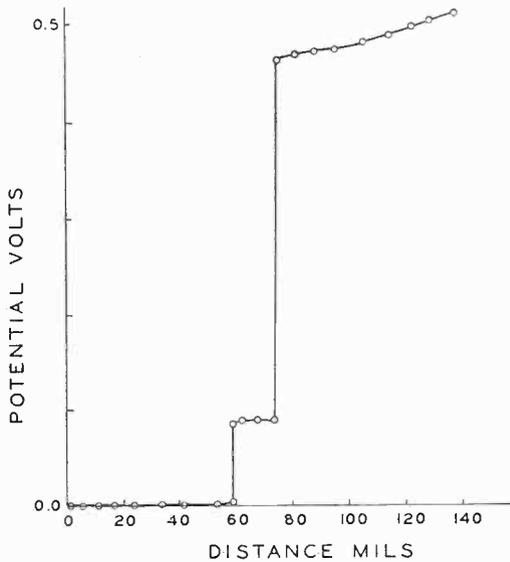


Fig. 4—Potential plot along a crystal showing the location of the junctions.

TRANSISTOR FABRICATION BY THE MELT-QUENCH PROCESS*†

By

JACQUES I. PANKOVE

RCA Laboratories,
Princeton, N. J.

Summary—A small cylindrical germanium crystal doped with donor and acceptor impurities of different segregation coefficients is partly melted, then caused to freeze rapidly. Due to the impurity segregation effect a p-n junction is formed at the stopping level of the liquid-solid interface. As the freezing process is accelerated, quenching occurs and impurity segregation can no longer take place. This produces a second p-n junction very close to the first junction. It is shown that most of the heat is dissipated by conduction through the crystal. Neglecting convection and radiation losses, the freezing rate in a typical structure can be greater than 0.85 centimeter per second. It is found that the whole structure remains a single crystal after such a treatment.

INTRODUCTION

THIS paper discusses a novel method of making in a germanium bar junctions suitable for use in transistors. In this "melt-quench" process of transistor fabrication, a small bar of suitably doped germanium is partially melted, then caused to freeze very rapidly. Two closely spaced, plane-parallel p-n junctions can thus be formed in a matter of a few seconds. Since these junctions may be designed to have a variety of electrical properties, the melt-quench process promises to be applicable to the fabrication of a variety of devices. The simplicity and rapidity of this process appear attractive for exploitation in low-cost mass production of transistor devices. However, this possibility has not been fully evaluated.

A method of making n-p-n junctions by melting back and then slowly refreezing has been described by Pfann.¹ In the present paper, however, refreezing is done very rapidly. As a result, greater control of junction parameters is possible.

PRINCIPLE OF JUNCTION FORMATION

As shown in Figure 1, a bar of germanium containing a concentra-

* Parts of this paper appeared in *Proc. I.R.E.*, February, 1956.

† This material was presented at the *IRE-AIEE Conference on Semiconductor Device Research*, Philadelphia, Pa., June 20, 1955.

¹ W. G. Pfann, "Redistribution of Solutes by Formation and Solidification of a Molten Zone," *Journal of Metals*, Vol. 6, p. 294, February, 1954.

tion of N_d donors and N_a acceptors is melted from the right, so that the liquid-solid interface is at X_0 . At this stage in the process the impurity distribution is not altered. The detail of choosing the impurities will be described later. Let K_d and K_a be respectively the donor and acceptor segregation coefficients. For the present, consider that $N_d > N_a$ (the crystal is n-type) and that $K_d < K_a$ (N_a/N_d).

The melt is caused to freeze at first slowly, then very rapidly.

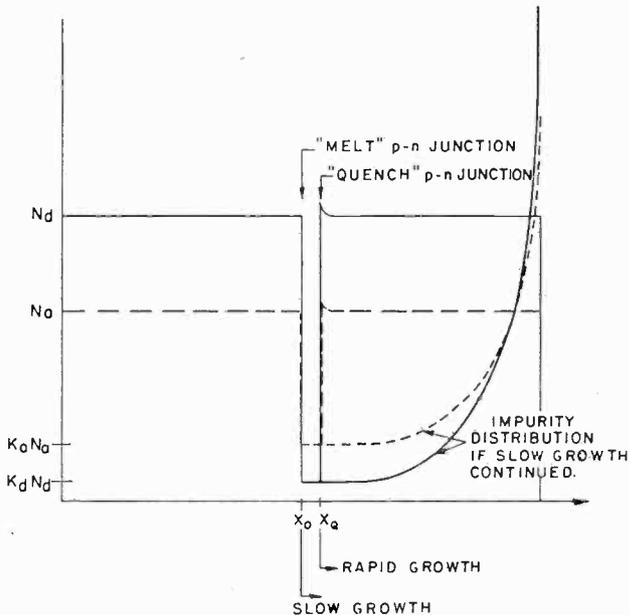


Fig. 1—Impurity distribution in the melt-quench process.

During the period of slow growth, the impurities are trapped according to the rules of impurity segregation. Consequently the initial impurity concentration is such that $K_a N_a > K_d N_d$ and the recrystallized material is p-type, thus forming a p-n junction. This junction will be referred to as the "melt junction".

If the crystal were allowed to continue to grow slowly as described by Pfann,² the concentration of donors in the liquid would increase more rapidly than the concentration of acceptors. Eventually the donors would predominate and a second junction form. The impurity concentrations and their segregation coefficients determine the spacing

² W. G. Pfann, "Segregation of Two Solute, With Particular Reference to Semiconductors," *Journal of Metals*, Vol. 4, p. 861, August, 1952.

between the junctions as well as the conductivity of the emitter and base regions, thus also influencing the emitter injection efficiency. While theoretically any two of these parameters may be chosen arbitrarily, in practice it is difficult to adjust them independently especially in a structure requiring high base conductivity — such as would be needed for high-frequency operation.

In the present process after a short period of slow regrowth, the crystal is caused to grow rapidly (quenched). This occurs at X_0 in Figure 1 thus determining the spacing between junctions. Upon quenching, the impurities cease to segregate and the donors again dominate. The p-n junction produced upon quenching will be called the "quench junction." Since the thermal treatment (quenching) determines the spacing between junctions the impurity concentration can be chosen independently to satisfy the requirements of good injection efficiency and low base resistivity. This separation of parameters controlling the junction spacing and injection efficiency allows a freedom of design not attainable in a slow growth process alone. If the slowly grown region is small compared to the quenched region, the impurity concentration in the quenched region is almost identical to the initial concentration in the crystal.

Quenching to form the second p-n junction is done very rapidly, and it might be expected that the quenched region would be polycrystalline. This does not turn out to be the case. X-ray examination shows the quenched region to be a continuation of the single crystal.

Design Criteria

The following reasoning is applied to the design of an n-p-n unit. Analogous reasoning can be used for a p-n-p unit.

Since the unmelted region, i.e., the starting material, is n-type,

$$N_d > N_a, \quad (1)$$

and the conductivity of this region is

$$\sigma_n = (N_d - N_a) q\mu_n, \quad (2)$$

where q and μ_n are the electronic charge and mobility, respectively.

If g denotes the fraction of the liquid that has solidified, then

$$g = \frac{x - x_0}{x_e - x_0},$$

where x_e is the length of the bar.

In the slowly freezing region the impurity distributions (n_a , n_d) are then given by

$$\begin{aligned} n_a(x) &= K_a N_a (1-g)^{K_a-1} \\ n_d(x) &= K_d N_d (1-g)^{K_d-1}. \end{aligned}$$

But if the slowly grown region is small compared to the volume of liquid, i.e., $g \ll 1$:

$$\left. \begin{aligned} n_a(x) &\approx K_a N_a \\ n_d(x) &\approx K_d N_d \end{aligned} \right\} \quad (3)$$

Since the slowly grown region is required to be p-type it is required that

$$n_a(x) > n_d(x),$$

and therefore from Equation (3),

$$K_a N_a > K_d N_d. \quad (4)$$

The conductivity in the p-type region is

$$\sigma_p = (K_a N_a - K_d N_d) q \mu_p, \quad (5)$$

where μ_p is the hole mobility.

To summarize (1) and (4), the important condition to be satisfied is:

$$N_d > N_a > \frac{K_d N_d}{K_a}. \quad (6)$$

This result has also been presented by Pfann,¹ but is reproduced here for completeness.

For closely spaced junctions and neglecting impurity pile-up due to segregation, the conductivity of both n-type regions is the same. Thus the units should be nearly symmetrical. For an efficient injection σ_n/σ_p should be made high. That is,

$$\frac{\sigma_n}{\sigma_p} = \frac{(N_d - N_a) \mu_n}{(K_a N_a - K_d N_d) \mu_p} \gg 1. \quad (7)$$

Substituting (4) into (7), it is found that

$$\frac{K_d}{K_a} \ll 1. \quad (8)$$

Although (8) is satisfied by (6), the greater the inequality (8), the better the injection efficiency of the emitter will be.

The melt junction is always abrupt. The quench junction is abrupt also, but may be gradual if the growth rate is suitably controlled during the junction formation.

THERMAL PROCESS

Thermal Characteristics

The thermal properties of germanium which are of importance here are given in Table I where they are also converted into units convenient for the case of a cylinder 30 mils in diameter.

Table I—Thermal Characteristics of Germanium

Property	Basic Values	Values for a cylinder 30 mils in diameter
Heat of fusion ³	8,000 cal/At gr	2.68 cal/cm
Specific heat ⁴	0.074 cal/gr °C	1.8×10^{-3} cal/cm °C
Thermal Conductivity ⁵	0.1 cal/sec cm °C	$(1/2.2) \times 10^{-3}$ cal cm/°C
Emissivity ⁶	0.1	

The above data shows that at the melting point there is somewhat more heat stored in heat of fusion than as specific heat.

Temperature Distribution

Since the present thermal treatment is a dynamic process (as contrasted with an equilibrium process) and involves a varying boundary condition (motion of the liquid-solid interface), the exact heat-flow problem is very difficult to solve. However, much has been learned by an approximate approach. The model considered here is one-dimen-

³ S. E. Bradshaw, "Phenomena Observed in the Melting and Solidification of Germanium," *Journal of the Electrochemical Society*, Vol. 101, p. 293, June, 1954.

⁴ *Handbook of Physics and Chemistry*, Chemical Rubber Publishing Co., 33rd Edition, p. 1874.

⁵ A. Grieco and H. C. Montgomery, "Thermal Conductivity of Germanium," *Phys. Rev.*, Vol. 86, p. 570, May, 1952. Estimated also to be 0.1 cal/sec cm °C in liquid by applying the Wiedemann-Franz law.

⁶ Estimated from comparison with values for other elements.

sional, has a perfect heat sink and for further simplicity its thermal conductivity is independent of temperature.

Melting

In the melting process, heat is fed at one end of the germanium filament from a high-temperature source while the other end is kept at a nearly constant low temperature, T_0 . The time constant for the heat-diffusion process in the solid is small compared to the duration of temperature rise at the heater. On the basis of a lumped RC representation, the diffusion time constant comes out as 0.04 second; a distributed representation, which is more accurate, would give a shorter time constant. The time constant for the actual heating process, which in addition to supplying heat for the diffusion process also supplies heat of fusion, measures in the order of 0.3 second. Here it can consequently be assumed that heat flow is instantaneous. Therefore, as the temperature at the heated end rises up to the melting point, T_{MP} , the temperature is linearly distributed along the length of the crystal if the thermal conductivity is constant. As the heated end reaches the melting point, heat of fusion is supplied to the liquid solid interface in addition to the heat flow necessary to sustain the thermal gradient. In other words, more heat enters the heated end during the melting of the bar than goes out to the heat sink during this period. Consequently a larger temperature gradient exists in the liquid than in the solid. This is shown in Figure 2. At equilibrium, the liquid-solid interface is stationary, no heat of fusion is supplied and the temperature gradient is uniform. The rate of heat flow in the solid is

$$\left(\frac{dQ}{dt} \right)_S = \sigma A \left(\frac{dT}{dx} \right)_S, \quad (9)$$

where σ = thermal conductivity, A = cross-sectional area through which heat flows. In the liquid, the rate of heat flow is

$$\left(\frac{dQ}{dt} \right)_L = \left(\frac{dQ}{dt} \right)_S + HA \left(\frac{dx}{dt} \right) = \sigma A \left(\frac{dT}{dx} \right)_L, \quad (10)$$

where H = heat of fusion; dx/dt = velocity of the liquid-solid interface. Since in practice the equilibrium condition (liquid-solid interface at $x_0 = 0.1$ centimeter) is reached roughly 0.5 second after melting begins, the following evaluation can be made:

$$\left(\frac{dx}{dt} \right)_{\text{average}} = 0.2 \text{ cm/sec,}$$

$$\left(\frac{dT}{dx} \right)_s = \frac{1000}{0.1} = 10^4 \text{ }^\circ\text{C/cm,} \quad (11)$$

where the liquid-solid interface is taken to be 1000 C° above the heat sink.

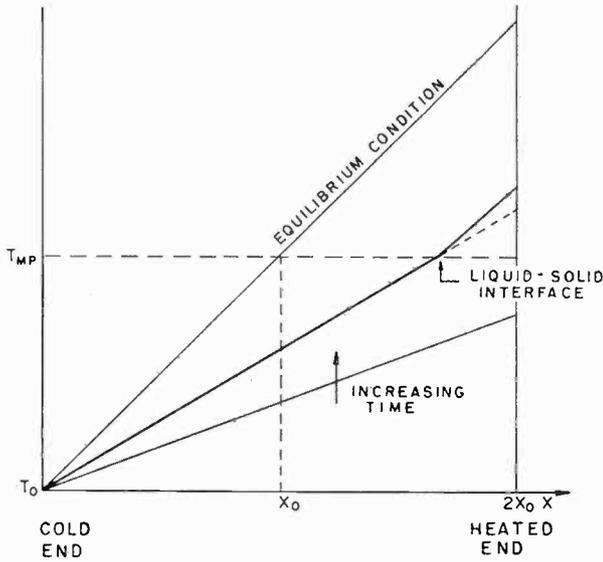


Fig. 2—Temperature distribution in cylinder during melting process.

Substituting (11) into (9) and using the converted data of Table I,

$$\left(\frac{dQ}{dt} \right)_s = \frac{10^{-3}}{2.2} \times 10^4 = 4.55 \text{ cal/sec} = 19 \text{ watts.} \quad (12)$$

$$AH \frac{dx}{dt} = 2.68 \times 0.2 = 0.536 \text{ cal/sec.}$$

Substituting into (10),

$$\left(\frac{dQ}{dt} \right)_{L_{\text{ave}}} = 4.55 + 0.54 = 5.09 \text{ cal/sec,}$$

$$\left(\frac{dT}{dx} \right)_{L_{\text{ave}}} = \frac{1}{\sigma A} \left(\frac{dQ}{dt} \right)_{L_{\text{ave}}} = 1.12 \times 10^4 \text{ } ^\circ\text{C/cm.}$$

Comparing this with (11) shows that the average increase in temperature gradient to produce fusion is only about 10 per cent.

The above description neglects radiation and convection losses which require that the temperature gradient be a function of temperature and largest at the hottest region. A rough estimate of the heat lost by radiation, W_R , can be made as follows:

$$W_R = \epsilon k A (T^4 - T_0^4), \quad (13)$$

where ϵ is the emissivity, k is the Stephan-Boltzmann constant and A the radiating area (about 3×10^{-2} square centimeters for a 0.1-centimeter length of liquid). If the temperature is assumed uniform in the liquid, the calculated radiation heat losses are as given in Table II. These are negligible compared to the 19 watts lost by conduction according to (12).

The convection losses are very difficult to evaluate but there is experimental evidence indicating that they are not more than 1.1 watts.

Table II—Calculated Radiation Losses

T	W_R
1200° K	0.035 watts
1800° K	0.18 watts
2400° K	0.57 watts

Therefore, at the end of the melting process the present model exhibits a uniform temperature gradient—the equilibrium condition of Figure 2. Then

$$\left(\frac{dQ}{dt} \right)_s = \left(\frac{dQ}{dt} \right)_L. \quad (14)$$

Freezing

Here it is assumed that the time constant of the heat source is negligible.

During freezing, latent heat of fusion is liberated. Assuming all the heat is lost by conduction through the solid, the process can be formulated as

$$\left(\frac{dQ}{dt} \right)_s = \left(\frac{dQ}{dt} \right)_L + HA \frac{dx}{dt}, \quad (15)$$

where dx/dt is the velocity of the liquid-solid interface or the growth rate (see Figure 3). Because of the initial condition (14), the initial growth rate is zero. Physically this means that the initial heat flow comes from the specific heat stored in the liquid.

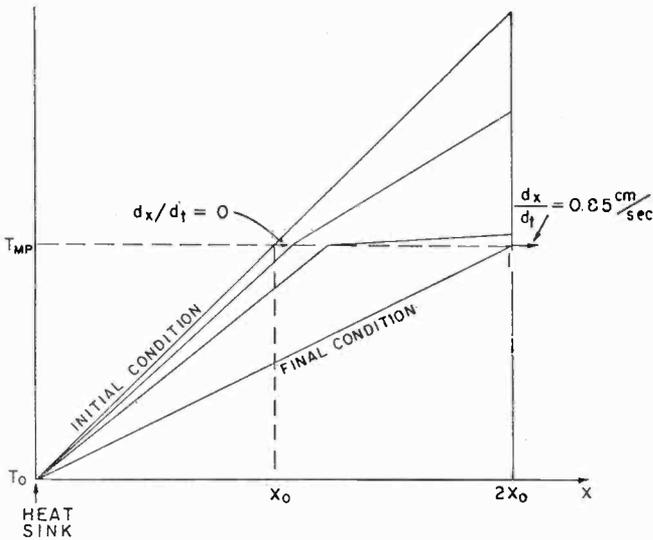


Fig. 3—Temperature distribution in cylinder during freezing process (heat loss only by conduction to heat sink).

As mentioned previously, the time constant for the diffusion process is less than 0.04 second. Hence the temperature in the liquid drops rapidly down to T_{MP} . During this time the growth rate is very low, the impurities are segregated and the p-type layer is formed.

Obviously all the liquid is at freezing temperature, T_{MP} , when the liquid-solid interface is an infinitesimal distance from $x = 2x_0$, the end of the cylinder. Then

$$\left(\frac{dQ}{dt} \right)_L = 0,$$

and the growth rate is:

$$\left(\frac{dx}{dt} \right)_{x=2x_0} = \frac{1}{HA} \left(\frac{dQ}{dt} \right)_s = \frac{1}{HA} \sigma A \frac{dT}{dx}$$

$$\left(\frac{dx}{dt} \right)_{x=2x_0} = \frac{1}{HA} \sigma A \frac{T_{MP}}{2x_0} \quad (16)$$

$$= \frac{1}{2.68} \frac{10^{-3}}{2.2} \frac{10^3}{0.2} = 0.85 \text{ cm/sec,}$$

which is a very rapid growth rate. Should the temperature of the liquid be T_{MP} before the liquid-solid interface reaches the end of the crystal, the growth rate would be even higher. But it could not be higher than by a factor of 2, (1.7 centimeters per second) which corresponds to all the liquid being at temperature T_{MP} from x_0 to the end of the crystal—except in the following case.

If the heat losses at the hot end of the cylinder are appreciably greater than the losses by conduction to the heat sink alone, then the temperature at $2x_0$ will drop so rapidly that the liquid will eventually comprise two regions as shown in Figure 4. One region will be supercooled and separated from the solid by the other region composed of hot liquid. When the non-supercooled liquid reaches the freezing temperature, T_{MP} , the supercooled region suddenly sees the liquid-solid interface and freezes at an even greater rate than before.

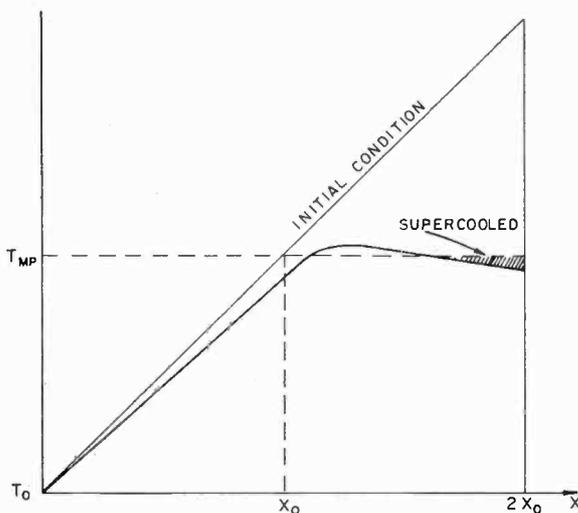


Fig. 4—Temperature distribution in cylinder during freezing process (case of supercooled liquid).

Crystallographic Considerations

Since the growth is nucleated by a relatively large-area single

crystal it is reasonable to expect that the processed unit remains a single crystal in the region of slow growth. In the quenched region or region of rapid growth, because of the very high growth rate discussed above, it is not obvious that the crystal will remain monocrystalline. This growth rate is about 500 times greater than the rates used to pull single crystals of germanium by standard techniques. X-ray back-reflection analysis of the interior of two samples showed these to be single crystals throughout, one with stresses in the recrystallized region, the other without stress. The physical appearance and electrical behavior of good samples not X-rayed lead to no suspicion that they are other than single crystals.

It is desirable to cut the cylinders along the [111] axis, for then the (111) equilibrium face of the freezing germanium will more probably form a flat quench junction normal to the cylinder's axis. The melt junction is almost always flat and normal to the cylinder axis as determined by the thermal gradient. The [111] orientation is also desirable for obtaining a thin p region or base layer, since crystal growth is slowest along the [111] direction.

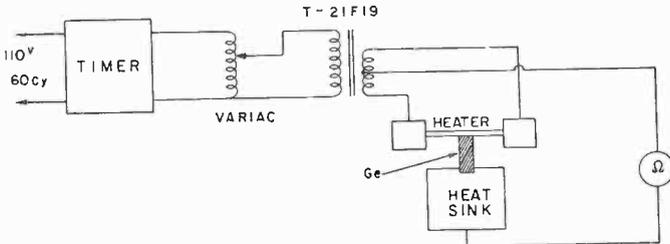


Fig. 5—Experimental setup for melt-quench process.

EXPERIMENTAL TECHNIQUES

A germanium crystal doped with donors and acceptors as discussed under Design Criteria is cut up into 100-mil-thick slices which are examined for their resistivity and uniformity. For this experimental work, resistivities up to 0.1 ohm-centimeter were used. The slices are cut up into 30-mil-diameter cylinders with an ultrasonic machine tool. It has been found that etching the crystal prior to the melt-quench process is not necessary.

The cylinders are mounted in a heat sink and placed in a helium atmosphere in contact with a heating element which may be a carbon filament or a machined graphite strip (see Figure 5). The position of the crystal is controlled by a micrometric screw until contact is indicated by the ohmmeter labelled "Ω" in Figure 5.

A timer is operated to pass a one-second pulse of current through the heating element. At the heater reaches a temperature of 2300° to 2900°C depending on the Variac* setting, the crystal melts about half-way down its length. When the heating current is turned off, the crystal cools (mostly by conduction to the heat sink) at an initial rate of about 1000°C per second at the liquid-solid interface.

Actually the freezing process is slowed down by the fact that the heater remains hotter than the germanium while the latter freezes. Yet the process is still sufficiently rapid to give satisfactory results. A furnace has been built in which the heater is supported by the armature of a relay so that the heater can be removed from the liquid germanium at the end of the melting process. This arrangement has the advantage of decoupling the thermal capacity of the heater from the germanium, and of greatly reducing the possibility of strains in the growing crystal.

It is remarkable that in most cases the surface tension is able to restrain the liquid to the initial cylindrical shape.

Graphite, copper, and forcibly cooled copper have been used for the heat sink with the crystal stuck in a well in the case of graphite and soldered in the case of copper. The use of a heat sink is imperative in order to establish a strong longitudinal temperature gradient which causes the melting front to advance as a flat surface. Crystals supported by a thermal insulator show a lopsided melt junction.

It is possible to accelerate the cooling process by increasing the radiation losses if the crystal surface is blackened to increase its emissivity. Colloidal carbon condensed from smoke or deposited by the heater itself serves this purpose; but a coating of aquadag forms a nucleating surface and hence yields a polycrystalline region. Although radiation losses are usually negligible as shown in Table II, they are multiplied by a factor of 10 when the crystal is blackened.

TRANSISTORS BY THE MELT-QUENCH PROCESS

The most suitable structure for the melt-quench process is a cylinder. Hence, the resulting transistor will have axial symmetry and will comprise a pair of plane parallel junctions (Figure 6) separated by a distance w . Considering the use of such a transistor at high frequency it is reasonable to design it for operation where its power gain varies at the rate of 6 decibels per octave so that the approximate power gain can be calculated from⁷

* Trade Mark of the General Radio Company, Cambridge, Mass.

⁷ L. J. Giacoletto, "Study of P-N-P Alloy Junction Transistor From D-C Through Medium Frequencies," *RCA Review*, Vol. XV, p. 506, December, 1954.

$$\text{P.G.} = \frac{g_m}{4\omega^2 r_{bb'} C_{b'e} C_c} \quad (17)$$

The base-lead resistance $r_{bb'}$ can be minimized by making the connection completely surround the base layer. It is then⁸

$$r_{bb'} = \frac{\rho}{8\pi w}$$

The other terms to be evaluated are

$$C_{b'e} = 1.33 \times 10^{-9} w^2 I_e \quad (\mu\text{f, mils, ma})$$

$$C_c = \frac{0.071 d^2}{\sqrt{\rho V_c}} \quad (\mu\mu\text{f, mils})$$

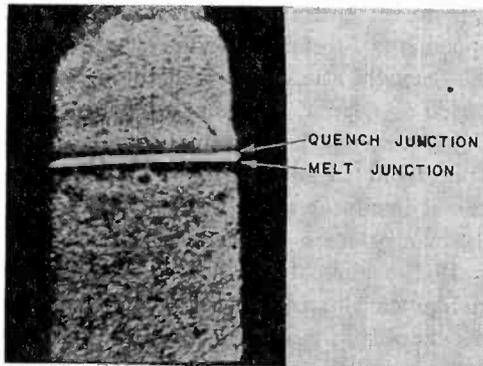


Fig. 6—Etched cross section of germanium cylinder processed by the melt-quench technique. The spacing between the junctions is 1.2 mils.

Hence, a 30-mil-diameter unit with $w = 0.5$ mils, $\rho = 1$ ohm-centimeter, $V_c = 6$ volts, $I_e = 1$ milliamper, gives

$$r_{bb'} = 31 \text{ ohms,}$$

$$C_{b'e} = 334 \mu\mu\text{f,}$$

$$C_c = 26 \mu\mu\text{f,}$$

$$\text{P.G. at 1 mc} = 29.6 \text{ db.}$$

The present paper is intended to describe the principles and techniques of the melt-quench process. Much of the evaluation has been

⁸ J. M. Early, "PNIP and NPIN Junction Transistor Triodes," *Bell Sys. Tech. Jour.*, Vol. 33, p. 526, May, 1954.

done by direct physical observation (junction spacing, etc.) rather than by more indirect electrical measurements on completed transistors. A wide variety of structures have been made with base width from about 0.1 mil to over 1 mil. To make the principles and techniques of this new process known at this early time no attempt has been made to accumulate extensive transistor data on any one structure.

To give a rough estimate of the possibilities of this technique some individual measurements will be cited. An early unit of about 0.8 mil junction spacing gave the amplifier performance of Table III. This is considered promising in view of the fact that this unit had a collector-to-base current-gain factor of but 5. Its base-lead resistance was 36 ohms. Other units have had current-gain factors up to 260. Measurements of junction capacitance show a dependence on voltage of $V^{-1/2}$ indicating that the transitions between the p and n type regions are abrupt.

Table III—Characteristics of Early Experimental Melt-Quench Transistor

	Audio Frequency	455 kc
Power Gain ($I_c = 1$ ma, $V_c = 3v$)	31.5 db	33.4 db
Power Gain ($I_c = 2$ ma, $V_c = 6v$)		37.1 db
Input impedance	200 Ω	200 Ω
Output impedance	30,000 Ω	> 52,000 Ω
Neutralizing Capacitance		7-10 $\mu\mu\text{f}$

ACKNOWLEDGMENT

The assistance of F. H. Corregan and the advice of S. G. Ellis are gratefully acknowledged.

APPENDIX—RADIATION BY GERMANIUM THROUGH ITS OWN SOLID

Since germanium is transparent to much of the infrared, and since heat is carried by infrared radiation, it is interesting to evaluate this process. Since radiation is an instantaneous process compared to diffusion the question arises as to the relative importance of these two heat losses.

Wien's law for the spectral distribution of radiation from a black-body gives the maximum energy at

$$\lambda_{\text{MAX}} = \frac{2884}{T} \text{ microns.}$$

Then the radiation from the liquid-solid interface ($T \approx 1200^\circ\text{K}$) gives

$$\lambda_{\max} = 2.4 \text{ microns.}$$

This is well in the transparent region of room-temperature germanium. According to V. A. Johnson and H. Y. Fan⁹ the energy gap depends on temperature as

$$E_g = 0.73 - 1.1 \times 10^{-4} T \text{ i.e., } 0.60 \text{ ev at } 1200^\circ\text{K}$$

which corresponds to a transmission threshold at $\lambda = 2.05$ microns. Hence the hot crystal is transparent to most of the radiation.

The number of thermally excited electrons at the solid near its melting point is calculated as

$$\frac{(2\pi mkT)^{3/2}}{h^3} e^{-E_g/2kT} = 5.25 \times 10^{18}.$$

This is such a small concentration that it should not contribute to the absorption. Therefore solid germanium is transparent to radiation from its liquid-solid interface.

The power radiated by the liquid-solid interface is

$$W_R = \epsilon k (T^4 - T_0^4) = 1.14 \text{ watts/cm}^2.$$

The power lost by thermal conduction through a slab of germanium of 0.1 centimeter thickness is

$$W_c = \sigma A \frac{dT}{dx} = 3820 \text{ watts/cm}^2.$$

Hence, in the present case the radiation loss to the heat sink supporting the cylinder is negligible.

⁹ V. A. Johnson and H. Y. Fan, "Transistor Dependence of the Energy Gap in Germanium from Conductivity and Hall Data," *Phys. Rev.*, Vol. 79, p. 899, September, 1950.

MICROSCOPIC EXAMINATION OF GERMANIUM CRYSTALS AND TRANSISTORS

BY

S. G. ELLIS

RCA Laboratories,
Princeton, N. J.

Summary—This paper discusses the techniques which may be employed in the microscopy of germanium crystals, and the information which may be obtained by the examination of such crystals and of transistors which are made from them. Microscopy is valuable as a supplementary method of examination since it permits a point-by-point study whereas most other methods, such as electrical, mechanical, or thermal methods yield volume-average results.

Useful information may be obtained regarding crystal orientation, twinning, junctions, defects and dislocations, small angle grain boundaries, and the external and internal structure of transistors.

MICROSCOPY

TO study crystal orientation, twinning, the shape of junctions and certain crystal defects in germanium by means of the microscope it is necessary to cut, grind (and sometimes polish) and etch the specimen. Germanium, as it is used in semiconductor devices, is much easier to examine under the microscope than is steel, for example. This is because it is a single element, in a very pure state, and the crystals, usually single, have a high degree of perfection. As a consequence, the etching techniques are relatively simple and the crystal defects are often far enough apart to be resolved by the light microscope.

It is too often assumed that on looking through a microscope one sees the specimen — magnified. Actually one sees a magnified *image* produced under very special conditions of illumination. The appearance of this image can be changed widely by changing the illumination. If the illumination is not correctly chosen important information may be completely absent. One consequence of the variation in the image with changes in the illumination is that a photomicrograph of a germanium surface can convey very little information unless the conditions of illumination are very precisely described. A photomicrograph is usually meaningful only to the man who takes it. It serves as a memorandum on what he has seen. This is one reason why no photomicrographs are shown in this paper. Instead a plan drawing of the observed structure is given. Beneath this is a sectional elevation of the structure as

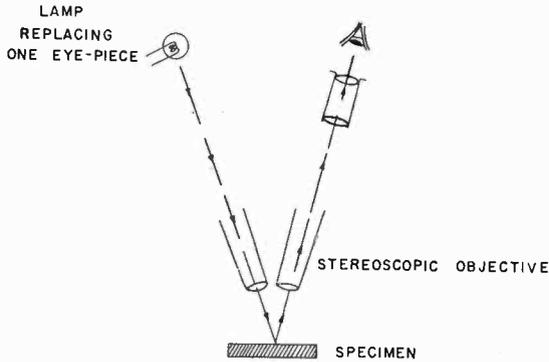


Fig. 1—Illuminating the specimen through one tube of a stereoscopic microscope.

deduced from the behavior of the image as the specimen and the illumination are manipulated.

Germanium surfaces are most readily examined when they are nearly flat, when the normal to the surface passes close to the center of the objective, and when the illumination is normal to the surface. This may be called the bright-field condition. When using a low-power ($\times 15$ to $\times 90$) stereoscopic microscope this condition can be achieved, at least approximately, by removing one eye piece and directing light down this tube of the instrument, as shown in Figure 1. The specimen is then positioned to reflect light into the other objective. Vertical illumination is standard on the metallurgical microscope ($\times 50$ to $\times 600$) and can be achieved at low power on the Leitz Panphot microscope by the use of an auxiliary mirror as shown in Figure 2.

If the normal to the germanium surface passes near the center of the objective but if the illumination is oblique so that the specularly

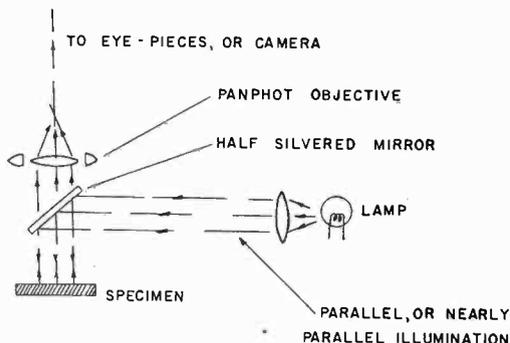


Fig. 2—Leitz Panphot microscope with an auxiliary mirror to illuminate the specimen,

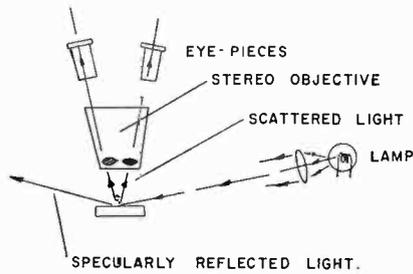


Fig. 3—Auxiliary illuminator for dark-field studies with stereoscopic microscope.

reflected light does not enter the objective, we have the dark-field condition. This is achieved with the stereoscopic and metallurgical microscopes by the use of an auxiliary illuminator as shown in Figure 3. It is standard on the Leitz Panphot microscope, where it is achieved in the manner indicated in Figure 4. Under dark-field conditions the surface is seen, if it is seen at all, because of light scattering. This method of illumination provides a sensitive method for detecting and examining dirt, contamination, and corrosion on the germanium surface. It is a very poor method of examining the germanium itself, however. When a stereoscopic microscope is used with oblique illumination, some parts of the specimen surface may reflect light into one objective and others may not, so that mixed light and dark field conditions are obtained. At times it may be preferable to use only one side of the microscope to simplify the interpretation of the image.

Oblique illumination is valuable in studying surface shape. Con-

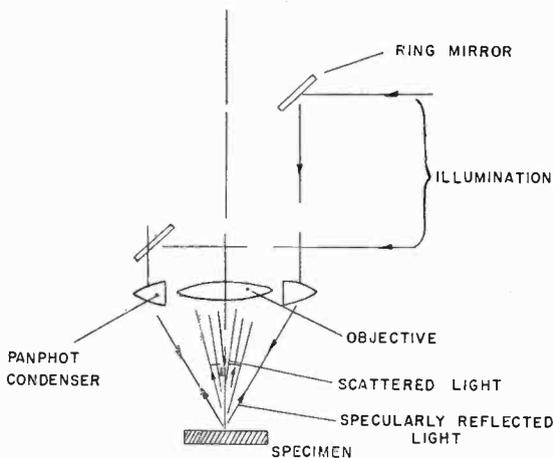


Fig. 4—Dark-field illumination, Leitz Panphot microscope.

sider the surface ABCD shown in section in Figure 5. When in the position shown, the whole surface is seen under dark field conditions. If now the surface is rotated counterclockwise about B, the regions AB and CD will first appear bright because they reflect light into the objective while BC will still appear dark. It can be deduced therefore that the surface AB is higher than the surface CD. This is a very sensitive method. It is particularly useful when the length of BC is large enough, or its inclination to AB small enough, that the difference in focus between AB and CD is difficult to determine.

Under vertical illumination the following factors help in determining the surface structure. With wide aperture illumination and consequently small depth of focus one can focus on the higher and lower

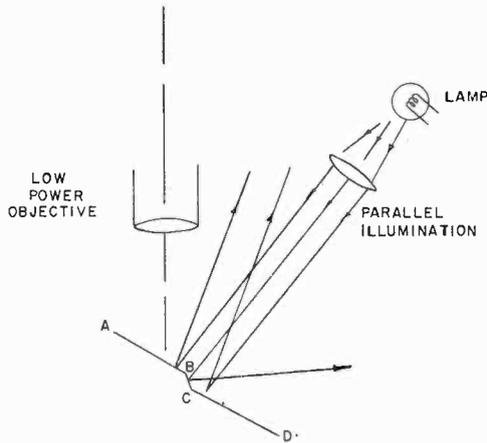


Fig. 5—Oblique illumination.

levels of the surface separately, provided that the difference in level is greater than approximately 5×10^{-5} centimeter. For small structures it may help to remember that concavities turn from dark to light on racking up *through* focus. Convexities give the opposite behavior. On racking up from tilted regions a bright reflection will move across the field. Consider the surface shown in section in Figure 6. On racking up so that the plane on which the microscope is focused moves from PQ to RS, the reflection from AD moves to CE, and of course goes somewhat out of focus. If the height AB is measured by the fine-focus knob, and BC with an eye piece scale, θ can be determined roughly from the relation

$$\theta = \frac{1}{2} \tan^{-1} (BC/BA).$$

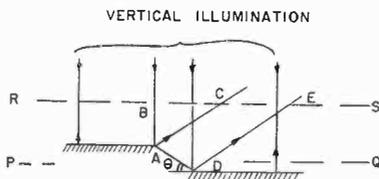


Fig. 6—Surface structure determination.

Since many examinations require that the germanium crystal be tilted until a certain plane is perpendicular to the axis of the microscope, it is convenient to have a stage which will permit this manipulation. Such a stage is shown in Figure 7. The test for correct orientation is first brilliance of reflection from the surface under study, and then more critically that details in the surface go out of focus symmetrically on racking up, i.e., do not show a sideways displacement.

The following techniques and instruments, while not widely available, may be of value in special cases.

Interference Microscopy: Using monochromatic light at vertical incidence one observes the interference fringes formed between the germanium surface and a silvered flat placed close to it. This is a powerful method of detecting and measuring small differences in height on the surface.

Polarizing Microscope: This has proved of little value in studying germanium, though it has in special cases shown some optically active solids on germanium surfaces. The solids could not be identified.

Electron Microscope: One looks not at germanium but at a replica of the surface. The technique being less direct is therefore slower but it is the only one which will give higher resolution than the light microscope. It is wise to precede electron microscope studies by careful light microscope studies.

SPECIMEN PREPARATION

The crystal is first cut approximately parallel to one of the simpler

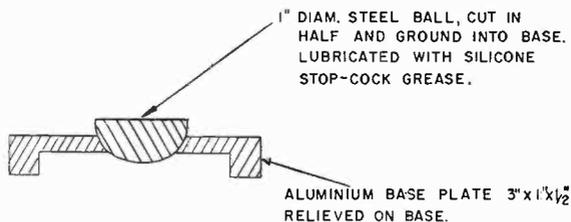


Fig. 7—Tilting stage for light microscope.

sets of planes: (100), (110), or (111). It may then be ground on a glass plate using a water slurry of American Optical Co. Nos. 303½, and 305 polishing powders in succession, the plate and crystal being carefully washed before proceeding to the finer powder. In general a high polish is not required.

If much work of this character is anticipated it is worthwhile to use a glass polishing wheel. If the surfaces to be ground are small, the crystal should either be held in a jig or embedded in plastic, so that the ground surface is plane.

The etches in general use in device work, while not necessarily the best for microscopy, are at least suitable and are in ready supply. Their compositions are given in Table I. The mode of attack of some of these etches has also been studied.¹

Detailed recommendations on the choice of etches are given in the following sections. For the present we content ourselves with some generalizations. A ground surface will have had many dislocation loops created in it. These dislocations begin and end at the surface. Number 2 etch, used for a short time, attacks the crystal at these dislocations and etches downwards until the larger ones are consumed. It then continues to attack laterally forming a more or less flat-bottomed pit. These pits often overlap. They are useful, as described below, in determining the orientation of crystal surfaces.

When the bulk properties of the crystal have to be studied it is important to remove the worked, or dislocated surface produced by grinding.* This is most conveniently done with a fast etch such as number 1 or number 4. On a (111) surface number 4 is particularly convenient since it will show the edge dislocations. Etching for two or three minutes is usually adequate.

In general it is desirable to use a volume of etch that is large compared to the volume of the crystal, and to maintain good circulation so that the etch does not become either hot or depleted. It is also important, particularly in dislocation studies, that the surface to be examined shall *at no time during or subsequent to the etching be touched with a solid body*, either the beaker or filter paper. After etching it should be quickly rinsed in running water and blown dry.

¹ P. R. Camp, "A Study of the Etching Rate of Single-Crystal Germanium," *Jour. Electrochem. Soc.*, Vol. 102, pp. 586-593, October, 1955.

* X-ray studies by S. Weissmann at Rutgers University show that the disordered layer, for purposes of the present report, can be somewhat thicker than the 2 to 10 microns found by etch-rate techniques in Reference (1). It is suggested that from 25 to 50 microns (i.e., one to two mils) be removed to give a margin of safety.

Table I—Composition of Commonly Used Germanium Etches

No. 1:	2 cc HNO ₃ 4 cc HF 4 cc H ₂ O 200 mg CuNO ₃	No. 3:	56 cc HF 56 cc HNO ₃ 12.5 cc H ₂ O
No. 2:	10 cc HF 10 cc H ₂ O ₂ 40 cc H ₂ O	No. 4:	50 cc HNO ₃ 30 cc CH ₃ COOH 30 cc HF 0.6 cc Br

Silver Etch:

40 cc HF
20 cc HNO₃
40 cc H₂O
2 gm AgNO₃

H₂O₂ Technical Quality (Allied Chemical Co.) Assay 30-35%

HNO₃ Reagent Quality (Baker) Minimum Assay 69.2%

HF Reagent Quality (Allied Chemical Co.) Minimum Assay 48%

CH₃COOH Reagent Quality (Merck) Minimum Assay 99.8%

Number 2 etch, when old, may leave a visible film on the germanium surface. This can be avoided by increasing the H₂O₂ concentration. It is preferable to do this just before the etch is used since the H₂O₂ is unstable at room temperature. 30 cc of H₂O₂ added to the above formula is sufficient. A sufficient volume of the etch should be used so that it does not weaken appreciably during the reaction. The test for this condition is that the pit shape does not change on a further short etch with fresh solution.

The number 1 etch is preferable when studying dislocations that terminate on the (100) surface. Number 3 etch when used on a polished surface will show p-n junctions.

OBSERVATION OF CRYSTALS

Orientation

To check the orientation of a surface it is convenient to start with it in a fine-ground state. It should then be etched in number 2 etch with agitation for approximately 45 seconds. It should be examined at $\times 90$ or higher magnification; $\times 300$ to $\times 600$ is preferable, using vertical illumination.

If the surface is close to one of the major planes it will show one of the pit structures sketched in Table II. In the early stages of etching, the pits will be closely packed together. Structures which are

ambiguous by the criteria of Table II indicate inaccurate orientations. Where the orientation is important such crystals should be recut using an X-ray diffraction examination to determine the plane of the cut.

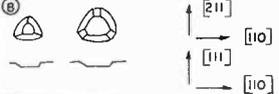
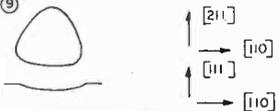
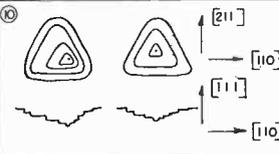
A common problem is the following: From the way in which a crystal has been grown and cut it can be assumed that the surface is close to the (111) planes. The problem is to determine whether the cut is within a few degrees of the (111) planes. If the surface is fine ground the worked layers should be removed with number 4 etch. The

Table II

THE PIT SHAPES ARE SHOWN FIRST AS THEY WOULD APPEAR UNDER VERTICAL ILLUMINATION. BENEATH THIS IS SHOWN A VERTICAL CENTER SECTION OF THE STRUCTURE. THESE PITS ARE BEST VIEWED AT X100 OR HIGHER MAGNIFICATION.

SURFACE	ETCH	PIT SHAPE AND ORIENTATION	INTERPRETATION AND NOTES
(100)	I	①	THE CHARACTERISTIC FEATURE IS THE CENTER DOT WHICH APPEARS BRIGHT ON RACKING UP FROM FOCUS. THIS PIT FORMS AT THE INTERSECTION OF AN EDGE DISLOCATION WITH THE SURFACE. THE TERRACES PROBABLY INDICATE JOGS IN THE DISLOCATION. SIZE OF PITS VARIABLE UP TO ABOUT 20 μ ALONG AN EDGE. THESE CAN AGGREGATE TO FORM SMALL ANGLE GRAIN BOUNDARIES.
		②	THE CHARACTERISTIC FEATURE IS THE ABSENCE OF A CENTER DOT. GROUND SURFACES ETCH WITH THIS PIT FORM. THE SIZE IS VERY VARIABLE RANGING UP TO 20-40 μ FOR FORM ②. FORM ③ IS GENERALLY SMALLER.
		③	
(110)	II	④	THE CHARACTERISTIC FEATURE IS THAT THE CENTER ETCHES HIGHER THAN THE SURROUNDING SURFACE. BOTH ④ AND THE TERRACED FORM ⑤ ARE PROBABLY THE SITE AT WHICH AN EDGE DISLOCATION MEETS THE SURFACE. THE SIZES RUN UP TO 10x20 μ FOR FORMS ④ AND ⑤.
		⑤	⑥ IS SMALLER AND IS PERHAPS THE BEGINNING FORM OF A STRUCTURE SUCH AS ⑤
		⑥	⑦ IS A TERRACED FORM SEEN WHEN THE ORIENTATION OF THE SURFACE ETCHED IS NOT ACCURATELY (110). THE DOT IS RAISED WITH RESPECT TO THE GENERAL SURFACE.
		⑦	

Table II (Continued)

SURFACE	ETCH	PIT SHAPE AND ORIENTATION	INTERPRETATION AND NOTES
(111)	II		THERE IS NO CENTER SPOT IN THESE PITS. SIZES UP TO 15μ - VERY VARIABLE.
			PROBABLY A TERMINAL FORM OF (B) USUALLY LARGE - UP TO 40μ ON A SIDE.
			THESE TERRACED FORMS INDICATE EDGE DISLOCATIONS, THE TERRACES BEING DUE TO JOGS IN THE DISLOCATIONS. THE CHARACTERISTIC FEATURE IS THE CENTER DOT - WHICH NEED NOT HOWEVER BE CONCENTRIC WITH THE OUTER TERRACES.
	IV		DUE TO EDGE DISLOCATIONS

crystal is then etched for two minutes in number 2 etch. If the orientation is exactly (111) the pits will be as shown in Table II. As the orientation departs from (111) the pits become asymmetric. At about 3° misorientation the pits become unrecognizable. An experienced observer can often judge the orientation to better than 3° . If no pits or only faint streaks are seen it can be concluded that either the orientation is more than 3° from the (111) planes or that there are no edge dislocations meeting the surface. At the time of writing the latter condition has not proved an impediment to the method.

In this, as in all light microscope observations, the information obtained pertains only to the observed surface of the crystal. Only in the absence of twinning and polycrystallinity can it be assumed to apply to the bulk of the crystal.

Twinning

The unambiguous proof of twinning by the use of the light microscope is difficult and laborious. A better method involves taking a back reflection Laue pattern with the X-ray beam bisected by the suspected twin plane.

Twinning is nevertheless a sufficiently common occurrence that the observer should be familiar with its appearance. Germanium twins on the (111) plane, one twin being rotated 60 degrees about the [111] direction with respect to the other. The simplest case, an A-B twin is shown in Figure 8. Twinning is often multiple. The next simplest

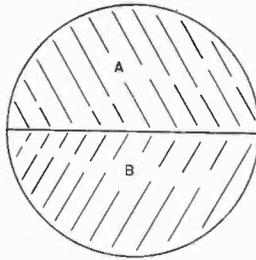


Fig. 8—A-B twin.

case an A-B-A twin is shown in Figure 9. It can be shown from the geometry of the germanium lattice that the part A can form a continuous single crystal, but that if the part B terminates as at PQ, then PQ cannot be a twin plane and will in general be the site of defects. During crystal growth the defects at PQ may nucleate other crystals, making the specimen polycrystalline.

With the crystal prepared as for orientation studies, an AB twin appears as a straight line, or trough, separating two regions of different orientation. If the $[111]$ twin plane is not perpendicular to the surface under examination the surfaces A and B will usually etch to different depths. An A-B-A twin may cause confusion if the B region is so narrow that it cannot be resolved in the microscope. The identification of twins is usually aided by their very specific orientation and by the fact that the twin plane often (though not always) runs for a long distance through the crystal.

Since twinning and slip both occur on (111) planes there is some danger of confusing the two conditions. Slip lines are only seen on *unetched* surfaces, they are moreover softer in appearance and usually less extensive in length than twin lines on a surface. They are discussed further below.

Impurity Effects

Resistivity striations, such as those which have already been described,² also cause irregular etching on longitudinal cuts on crystals.

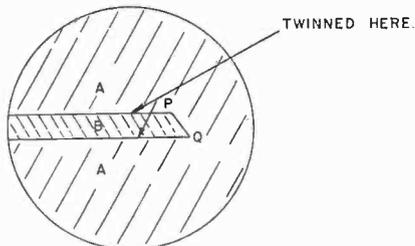


Fig. 9—A-B-A twin ending at PQ.

This is mentioned not as a method for studying such striations (the pulse plating method of Reference (1) is superior) but because it is a possible source of ambiguity in studying the irregular etching figures produced by dislocation arrays.

The position is further complicated by the fact that dislocations and impurities may be expected to interact in the crystal, locking each other in place. It is probable that this interaction is relatively weaker in valence crystals than in metals.

Junctions

One method of delineating junctions, involving differential electroplating, has been described in the literature.³ In another method,⁴ the crystal is cut, ground, and polished. The polishing can be done on a glass plate using a water slurry of Linde A powder, or on 4/0 paper using cutting oil, or if the crystal is small and is embedded in plastic the early polishing may be done on polishing paper, the operation being finished with a pad of paper towel and wet Linde A powder. Next, using a camel-hair brush the polished surface is vigorously swabbed with number 3 etch for about 30 seconds. Near the junction the n-type region etches faster than the p-type region.

The examination is best carried out with a low-power stereoscopic microscope using parallel illumination. The mixed dark- and light-field condition described in connection with Figure 5 is employed. In that Figure, AB would correspond to the p-type region.

Defects and Dislocations

The subject of defects and dislocations in single crystals has been reviewed by Seitz⁵ and Cottrell.⁶ It is probable that, in those etching reactions which are not self-catalytic, each pit nucleates at a defect. Some pits are known to nucleate at dislocations. It does not follow that an observation of the pits, however, will lead to an unambiguous picture of the distribution of dislocations.

The dislocations in germanium, and the techniques for detecting

² P. R. Camp, "Resistivity Striations in Single-Crystal Germanium," *Jour. Appl. Phys.*, Vol. 25, pp. 459-463, April, 1954.

³ E. Billig and J. J. Dowd, "P-N Junction Revealed by Electrolytic Etching," *Nature*, Vol. 172, p. 115, 1953.

⁴ A. R. Moore, "The Preparation of Single and Multiple P-N Junctions in Single Crystals of Germanium," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 77-81. A photograph of an etched junction is shown in this paper, but the etching procedure is not described in detail.

⁵ F. Seitz, "Imperfections in Nearly-Perfect Crystals," (W. Shockley, Editor), John Wiley & Sons, N. Y., p. 3, 1952.

⁶ A. H. Cottrell, "Progress in Metal Physics," (B. Chalmers, Editor), Interscience Publishers, N. Y., Vol. 1, p. 77, 1949; Vol. 4, p. 205, 1953.

them have been described elsewhere.* For present purposes, and at the risk of some oversimplification the dislocations may be divided into two groups: edge dislocations, and composite dislocations. The edge dislocations provide acceptor levels in the crystal and also act as recombination regions. The more numerous composite dislocations are not at present known to have any electrical effects. We therefore restrict the following discussion to the detection of the edge dislocations.

Since a mechanically worked surface has had many dislocations formed in it, it is necessary to etch deeply before attempting to study the distribution of dislocations in the undisturbed bulk. When studying surfaces within 3° of the (111) planes this may be done by a five minute etch with number 4. Conical pits (Table II) now show where the edge dislocations meet the surface.

Number 2 etch may be used on (110) and (111) surfaces. Its attack is slower than that of number 4 etch so that it is convenient to remove the worked surface with a faster etch. Number 1 etch is preferred on (100) surfaces. The interpretations are given in Table II. As a general rule it should be noted that the edge dislocations give pits on (111) and (100) surfaces which have a central depression. Pits with smooth bottoms can form where a U-shaped dislocation loop has been etched out of the crystal, or where an edge dislocation changes to the composite orientation and ceases to etch in the same way.

It is not uncommon to find the dislocations arrayed in lines. The interpretation is discussed under slip lines, below. When the pits are very close, i.e., less than 10^{-4} centimeter apart, they become difficult if not impossible to resolve. One sees only an etched line, often curved, and usually terminating within the crystal.

On surfaces other than (100) and (111), dislocation arrays may etch proud, i.e., be left above the surrounding surface. This behaviour is seen on (110) surfaces. Proud etching is not an unambiguous sign of dislocations since it may be caused by impurity concentration (see above).

Dislocations may be caused by slip, by excessive impurity concentration where the solute atoms are misfits, by the termination of twinned regions within the crystal, and may be formed during crystal growth.

Slip Lines

At temperatures above 600° C, germanium slips on the (111)

* S. G. Ellis, "Dislocations in Germanium," *Jour. Appl. Phys.*, Vol. 26, pp. 1140-1146, September, 1955.

planes. The intersection of these regions of slip with the free surface of the crystal produces very small steps which are called slip lines. With practice they can be distinguished from twin lines by their appearance. A safer procedure is to etch the crystal. Twin lines are intensified while slip lines disappear, being replaced by dislocation pits or small angle grain boundaries the corresponding substructures having been formed as a result of the slipping.

If a crystal is cut or etched after deformation the slip lines cannot be seen. There are cases in which it may be important to decide if deformation has taken place. The dislocation distributions which result from deformation will therefore be described.

When a crystal slips the edge dislocations increase in number and move along the slip planes. The first indication of slip is therefore a local increase in the dislocation density, the corresponding pits being lined up on the trace of the slip plane, $(11\bar{1})$ for example, in the plane, (111) , under study Figure 10.

As the slip proceeds the continued motion of these dislocations may be prevented, either by another dislocation or by impurity atoms. The dislocations then come to rest in a stable array, known as a small-angle grain boundary. Its trace, in the plane under study, is usually perpendicular to the trace of the slip planes Figure 10.

Slip lines are best seen under vertical illumination.

Wafers and Bars

Wafers and bars cut preparatory to making devices may be examined by the above techniques for orientation, twinning, junctions, and small-angle grain boundaries.

In addition to the appearances described above there will often be found irregular mounds and hollows. Some of these are due to the fact that the germanium was not flat when the last layers of worked material were etched off. These irregularities persist in reduced form on further etching. There are other irregularities which cannot be explained on this basis. Those which persist with little change of shape on further etching are probably due to structural faults in the crystal—arrays of dislocations, inhomogeneities in impurity distribution, or both together.

OBSERVATIONS OF TRANSISTORS

This section will be concerned with the examination of unpotted, alloy junction, p-n-p transistors, using indium or indium-germanium as the alloying material. A cross section of such a transistor, not typical, is shown in Figure 11. We first consider chemically etched transistors.

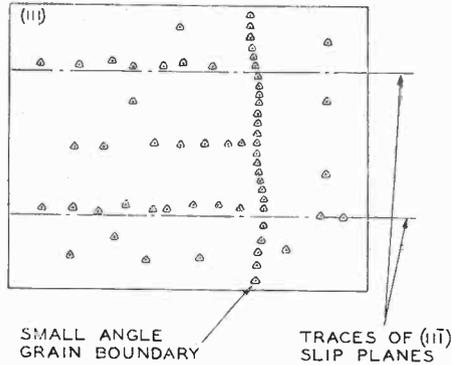


Fig. 10—Appearance of (111) plane in slipped crystal after etching with number 2 etch.

An examination of the transistor as received will show the spread of the dots, their alignment (using a stereoscopic microscope to look at the transistor in profile), and whether or not the final etching has been done. This last can be judged by the extent to which the isolated germanium crystals in the dots can be seen. A dark-field examination will also indicate the cleanliness of the surfaces of the wafer.

In the next step the indium dots are removed using an etch which does not attack germanium. This may be concentrated hydrochloric acid or mercury. Mercury is faster but must be cleaned from the transistor by a dip in nitric acid. The nitric acid dip should be a short one since nitric acid slowly attacks germanium. The germanium crystals which had previously been isolated in the indium dot may now be washed away and the wafer dried.

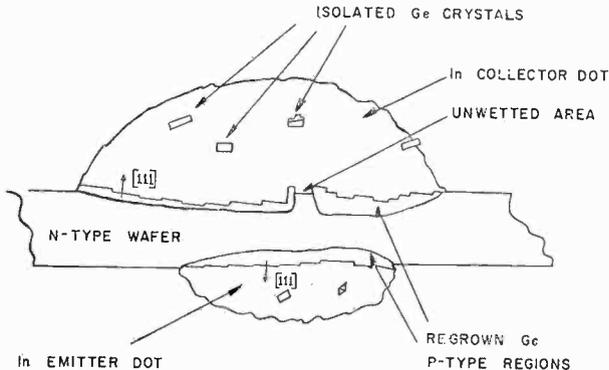


Fig. 11—Cross section of Indium-on-Germanium p-n-p transistor.

The region of recrystallization should now be observed under vertical illumination (if the wafer surfaces are nearly (111) planes). With the wafer tilted to give the maximum reflection from the regrowth crystals the [111] direction coincides with the axis of the microscope. Hence the departure of the wafer from the horizontal measures its misalignment from the [111] direction. Unwetted regions can usually be detected by the misalignment of the surface with the [111] direction, but even when the orientation is good, it will be seen that the unwetted germanium has a different surface texture from that of the regrowth crystals. This surface texture is the same as that found outside the region of the dots. Sometimes the regrowth crystals will form a cap over an unwetted region. In such cases the regrowth crystals must extend above the general surface of the wafer. Judicious probing with a needle will test such cases.

We now consider the case in which the final etching of the transistor was electrolytic, the etching current having been injected into the base wafer from the emitter and collector. The examination for dot spread, and alignment, and the nature of the recrystallization and wetting may proceed as before. One difference will be noticed in the base wafer. The etching occurs where injected holes reach the surface of the base.* The emitter and collector are surrounded by troughs which will usually be smooth and clean if the etching-current density was high (of the order of 1 ampere per square centimeter). The curved walls of these troughs make them difficult to examine in the light microscope. Raised regions within the troughs suggest that the hole current reaching this part of the surface was reduced, as would be the case with local hole-electron recombination within the base. Recombination at the dislocations in small angle grain boundaries has been observed to cause this effect.*

A crude examination of the junction can sometimes be made by breaking the wafer so that the fracture surface crosses the dot regions. Such fracture surfaces are seldom flat enough to permit careful study. A better method is to embed the wafer in Bioplastic or Araldite, being careful to exclude air bubbles so that the wafer is well supported. It may then be cut, or ground down to a suitable cross section, polished on 4/0 paper with oil, and etched to show the junctions.

Examination under the microscope will now show the junction shapes, the variation of base thickness between the junctions, and the thickness of the regrowth regions. This information, of course, pertains to only one cross section of the transistor.

*S. G. Ellis, "Electrolytic Etching of Small-Angle Grain Boundaries in Germanium," *Phys. Rev.*, Vol. 100, p. 1140, November 15, 1955.

CALCULATIONS OF ALLOYING DEPTH OF INDIUM IN GERMANIUM

BY

LOUIS PENSAK

RCA Laboratories,
Princeton, N. J.

Summary—A method is described for calculating the depth of alloying of indium in germanium for two simple cases of indium shapes. The results are presented in a graphical form convenient for use. Some limitations on the general applicability of the calculations are described briefly.

INTRODUCTION

IN the fabrication of semiconductor devices such as transistors or rectifiers, the formation of junctions by the alloy process is commonly used. It is often necessary to be able to calculate the depth of alloying under a variety of conditions. The special case of indium on germanium has been selected for treatment here because it is, at present, the most frequently used combination of materials. It is also a convenient example of how calculations can be made for other combinations for which the alloy phase diagram is known.

PHASE DIAGRAMS

The phase diagram is one way of describing the solubility of one component in another as a function of temperature. The case of indium and germanium is shown in Figure 1 where the curve shows the atomic percentage of germanium which will dissolve into a melt of indium at various temperatures. However, atomic percentages are not directly measurable and must be reduced to volumes for practical use. To do this, let the ratio of germanium atoms to the total number of atoms in a melt at some temperature be "S," and the number of atoms per unit volume for germanium and indium be N_{Ge} and N_{In} , respectively. Then

$$S = \frac{N_{Ge}}{N_{Ge} + N_{In}}.$$

This can be transposed to the form

$$N_{Ge} = \frac{S}{1 - S} N_{In}. \quad (1)$$

This expression can be converted to volumes by substituting in it the term for the number of atoms (N) in a body,

$$N = \frac{\text{grams of material}}{\text{grams per atom}} = \frac{\text{Volume} \times \text{Density}}{\text{Atomic weight} \times \text{Weight of hydrogen atom}},$$

and Equation (1) becomes

$$V_{Ge} = \frac{S}{1-S} \cdot V_{In} \cdot \frac{\text{Density (In) Atomic weight (Ge)}}{\text{Density (Ge) Atomic weight (In)}}. \quad (2)$$

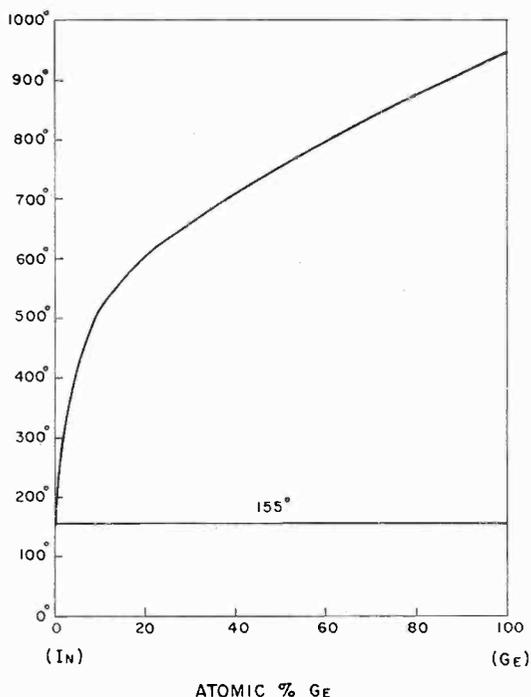


Fig. 1—The indium-germanium system.

Table I

	Indium	Germanium
Density	7.31 gms/cc	5.32 gms/cc
Atomic weight	114.8	72.6

Using the values shown in Table I for the constants, Equation (2) reduces to

$$V_{Ge} = .870 S V_{In} / (1 - S).$$

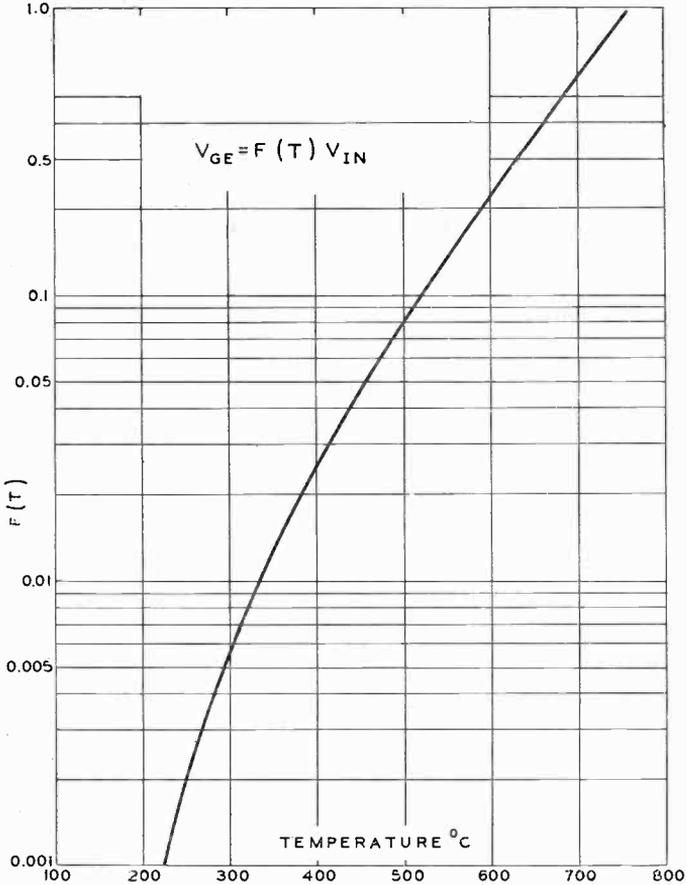


Fig. 2—Volume solubility of germanium in indium versus temperature.

Since S is a function of temperature as given by the phase diagram, this can be rewritten as

$$V_{Ge} = F(T) \cdot V_{In}. \quad (3)$$

Figure 2 is a graph of $F(T)$ plotted against temperature. While $F(T)$ may be obtained in principle from Figure 1, the plotted curve was obtained from the data of Keck and Broder¹ as a more accurate source.

¹ P. H. Keck and J. Broder, "The Solubility of Silicon and Germanium in Gallium and Indium," *Phys. Rev.*, Vol. 90, p. 521, May, 1953.

ALLOY DEPTH CALCULATIONS

When one puts a small piece of indium on a flat surface of germanium, the depth to which it will alloy depends not only on temperature, but also on the area it covers and to a lesser extent on the shapes that the molten solution may take during the heating cycle. Not all cases can be calculated but two simplified cases of practical value will be treated.

Case I (see Figure 3); The indium is assumed to be confined to a cylindrical shape of diameter " d " and height " h " with one plane end in contact with the germanium. The dissolved germanium is also assumed to be a cylinder of the same diameter, but its height x is to be determined.

The volume of indium is

$$V_{In} = \frac{\pi d^2}{4} \cdot h.$$

The volume of germanium dissolved is

$$V_{Ge} = \frac{\pi d^2}{4} \cdot x.$$

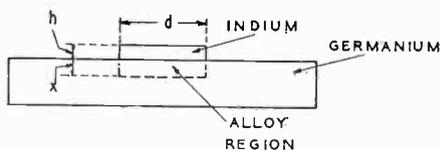


Fig. 3—Geometry for the case of the confined dot.

Substituting in Equation (3) we get

$$\frac{\pi d^2}{4} \cdot x = F(T) \frac{\pi d^2}{4} \cdot h$$

and

$$x = F(T) \cdot h.$$

The curve of Figure 2 can now be used directly to find x at any temperature given only the value of h .

Case II: The indium is now assumed to be not confined in any way. On heating up it will wet the germanium and spread out to the shape of a section of a sphere. This will be true for small volumes of liquid

with high surface tensions such that the force of gravity can be neglected. It is also assumed that the spreading occurs at low enough temperatures so that the depth of alloying during spreading can be neglected. The germanium dissolved is assumed to be in the shape of a cylinder whose diameter is that of the contact area of the indium and whose height x is to be determined.

There are two ways of specifying the diameter of spreading of the indium. One can refer to a spreading diameter, D_s , in terms of a ratio K with the diameter, D , of an indium sphere (as a measure of its volume) or one can refer to the contact angle θ of the liquid indium with the germanium surface. The relationship between K and θ can be shown as follows:

The volume of indium is

$$V_{In} = \frac{\pi D^3}{6}.$$

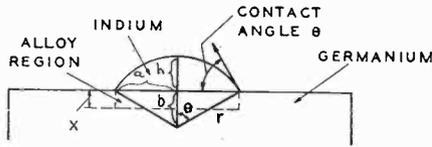


Fig. 4—Geometry for the case of the unconfined dot.

This volume, as a section of a sphere, (see Figure 4) is

$$V_{In} = \frac{\pi}{6} (h^3 + 3a^2h), \quad (4)$$

where h = maximum height of the liquid and

$$a = \text{radius of the contact area} = \frac{D_s}{2}.$$

In terms of the contact angle θ ,

$$V = \frac{\pi}{6} \frac{D_s^3}{8} \left[\left(\frac{1 - \cos \theta}{\sin \theta} \right)^3 + 3 \left(\frac{1 - \cos \theta}{\sin \theta} \right) \right].$$

Since this is equal to the volume of the initial sphere, we can find the ratio of D_s to D as a function θ :

$$\frac{D_s}{D} = 2 \left[\left(\frac{1 - \cos \theta}{\sin \theta} \right)^3 + 3 \left(\frac{1 - \cos \theta}{\sin \theta} \right) \right]^{-1/3}$$

The range of values for θ from 60° to 110° was chosen as adequate for most spreading conditions. These values give the curve of Figure 5 in which D_s/D is labeled K (called the spreading ratio).

The calculation of alloying depth can now be made in terms of D (diameter of sphere of indium), K , and temperature. The volume of germanium dissolved is

$$V_{Ge} = \frac{\pi D_s^2}{4} \cdot x \quad (\text{where } x \text{ is depth of alloying}).$$

Substituting the germanium and indium volumes in Equation (3),

$$\begin{aligned} \frac{\pi D_s^2}{4} \cdot x &= F(T) \frac{\pi D^3}{6} \\ x &= F(T) \cdot \frac{2}{3} \frac{D^3}{D_s^2}, \end{aligned}$$

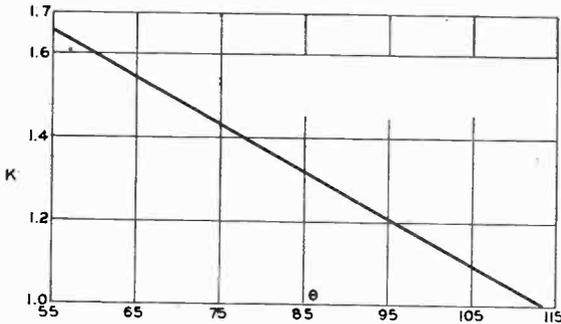


Fig. 5—Spreading ratio (K) versus contact angle (θ).

but

$$\frac{D_s}{D} = \frac{1}{K},$$

therefore
$$x = F(T) \cdot \frac{2}{3} \frac{D}{K^2},$$

which is the general equation for depth of alloying, x , as a function of spreading ratio, K , temperature, $F(T)$, and indium volume of sphere diameter, D .

As a result of experience, it has been found that $K = 1.4$ is an average value for pure indium on clean germanium in a dry hydrogen atmosphere. This value was therefore used in preparing the curves of Figure 6 which have been found useful in the design of transistors. However, the curve of Figure 7 was also prepared to simplify the use of Figure 6 when different values of K are necessary. For any value

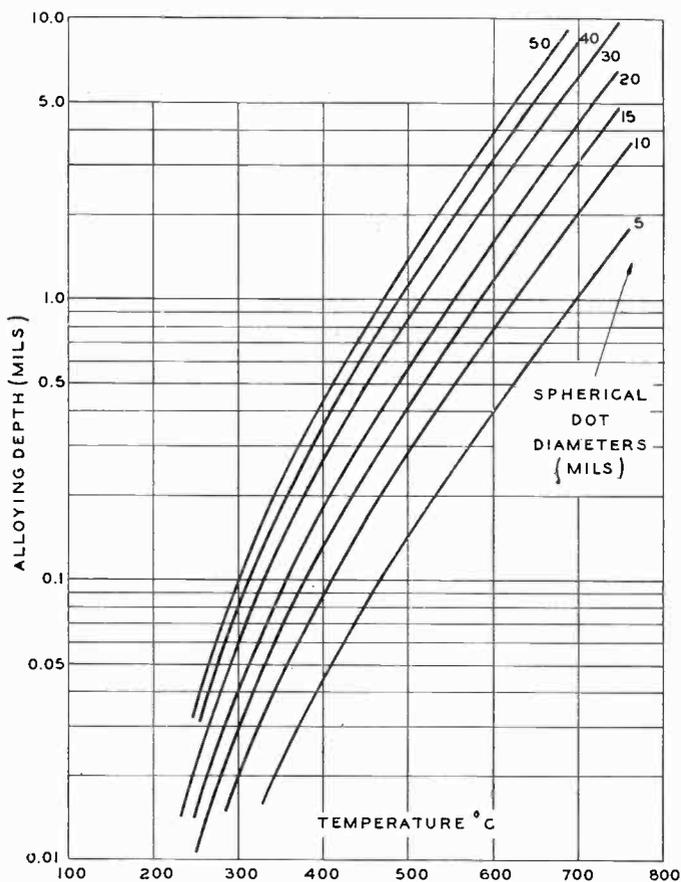


Fig. 6—Alloying depth versus temperature and dot size ($K = 1.40$).

K , a value of R is shown which is the multiplying factor for correcting the alloying depth values shown in Figure 6. The curve shows how sensitive the alloying depth is to spreading diameter. This is because the volume of germanium dissolved is a constant for a given temperature and as the spreading diameter changes, the contact area changes with the square of the diameter and the depth varies inversely with the area.

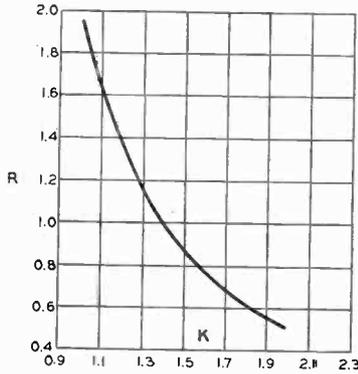


Fig. 7—Correction factor R for modifying curves of Figure 6 when K is different from the value 1.40.

The curves of Figure 8 are given as a convenience in finding equivalent sphere diameters from the dimensions of indium dots when they are fabricated as punchings from sheet material in the form of cylinders.

LIMITATIONS ON THE GENERALITY OF THE CALCULATION

These calculations were checked experimentally and found to hold within experimental error over the range of 400° to 600°C. The value of spreading ratio of 1.4 is not always to be expected. For reasons not yet clear, large deviations are occasionally encountered. Also, the spreading may vary with temperature. Above 600°C there is a tendency of the indium to spread beyond the value of $K = 1.4$ and then return to 1.4 on cooling. Measurements on the alloyed indium would indicate a K of 1.4 and give a misleading result because this effect reduces the alloying depth from the values predicted by the curve. The magnitude

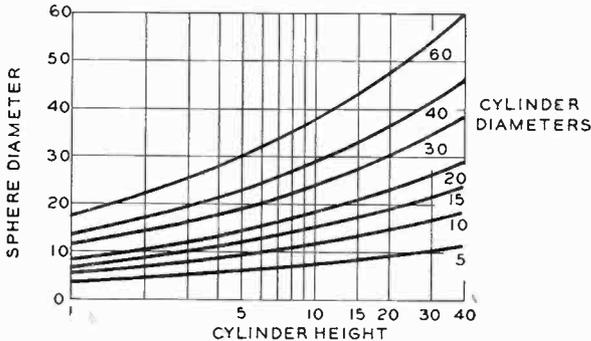


Fig. 8—Conversion of cylinders to spheres of equal volume.

of reduction depends on the values of the spreading diameter as a function of temperature and will not be treated in this paper. Other effects that cause departure from the ideal are described by Mueller and Ditrick.² They are not always large enough to seriously invalidate the usefulness of the above calculations but should be kept in mind to prevent misleading conclusions.

² C. W. Mueller and N. H. Ditrick, "Uniform Planar Alloy Junctions for Germanium Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 121-131.

UNIFORM PLANAR ALLOY JUNCTIONS FOR GERMANIUM TRANSISTORS*†

BY

C. W. MUELLER AND N. H. DITRICK

RCA Laboratories,
Princeton, N. J.

Summary—In the alloying process previously used, for example in germanium p-n-p junction transistors, wetting and alloying proceed simultaneously. This often produces a rounded junction whose shape, area and penetration are not adequately controlled for close-spaced transistor devices. Techniques are described which, by separating the wetting from the alloying steps, achieve a control over junction geometry not heretofore realized. These techniques permit the alloy transistor to be economically exploited to considerably higher frequencies and constitute a major advance in junction preparation.

Uniform wetted areas are obtained in a low-temperature soldering operation using a small amount of zinc in the indium dot material and a liquid flux. Control of the dot material volume, alloying temperature, and crystal orientations provide for uniform penetration during the alloying process in an atmosphere chosen to restrict further spreading. The dense (111) plane of the germanium wafer is used to provide a planar junction front. By this means an inherent structural property of the crystal is utilized to provide a flat junction front independent of minor variations in the physical environment. With these techniques junctions can be made that are flat to within $\frac{1}{2}$ micron (0.02 mil) over 90 per cent of their diameter.

A slow cooling rate provides for uniform recrystallization of the dissolved germanium. The uniform recrystallized region requires less etching and gives improved electrical performance.

INTRODUCTION

THE alloy junction has been widely used in audio transistors where a rounded junction shape can be accepted without serious undesirable effects. However, for many reasons, flat parallel junctions would be preferred. Flat parallel junctions are especially important to improve the uniformity of characteristics and to facilitate construction of transistors with small base widths. This paper describes new techniques which are believed to constitute a major advance in the alloy junction art and which achieve a control over the junction geometry not heretofore realized by the alloy process. By the use of these new techniques, junctions can be made that are flat within $\frac{1}{2}$ micron (0.02 mil) over 90 per cent of their diameter.

At present, the method most generally used for making alloy junc-

* Reprinted from *RCA Review*, March, 1956.

† Presented at the I.R.E.-A.I.E.E. Conference on Semiconductor Research, Philadelphia, Pa., June 22, 1955.

tions¹ is to position an impurity dot on the base wafer and fire the assembly in a reducing atmosphere. In this process, the dot material first wets at one point from which point alloying and wetting progress simultaneously. Thus, the dot penetrates and spreads out at the same time and a curved junction front as shown in cross section in Figure 1 results. The final shape of the junction front can be altered somewhat by changing the temperature during the process cycle or applying weights to the dots. However, as long as alloying and wetting take place at the same time, it is extremely difficult, if not impossible, to make a truly flat junction. The important feature of the new technique is that the present one-step process is separated into a three-step process: (1) the dot material is caused to wet a certain area of the

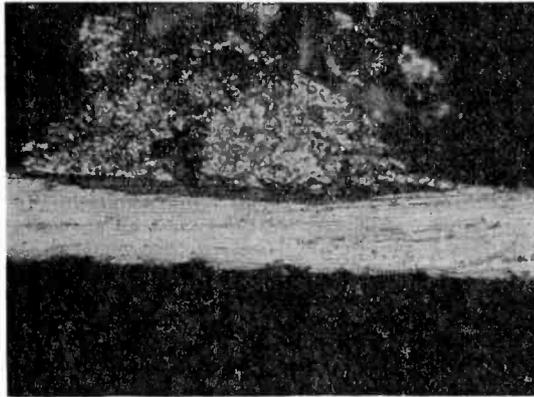


Fig. 1—Cross-section of p-n alloy junction diode.

base wafer; (2) the dot material is alloyed into the base wafer without further wetting; (3) the dissolved germanium with a small amount of dot material is recrystallized upon the base wafer. Each of these steps is important and will be discussed in order. The new technique has been applied to p (indium-rich) dot material on n germanium with about 10,000 edge dislocations per square centimeter, and the description that follows is applicable thereto. Similar techniques to those to be described can be employed for other types of dot materials and semiconductors.

WETTING AND SOLDERING

Good wetting over the entire dot area is a prerequisite for obtaining uniform planar junctions. To prevent appreciable alloy penetration

¹ R. R. Law, C. W. Mueller, J. I. Pankove, and L. D. Armstrong, "A Developmental Germanium P-N-P Junction Transistor," *Proc. I.R.E.*, Vol. 40, pp. 1352-1357, November, 1952.

from taking place during the wetting process, this operation must be carried out at a low temperature, say 300° to 350° C, for an indium-rich material. At these low temperatures, hydrogen is no longer useful as a flux and other fluxes must be used. The flux must clean the surface to permit wetting and also reduce the molten liquid surface tension. The dot material then flows out over a well-defined reproducible area. Any residual products of the flux must be either removed by subsequent cleaning or must produce no deleterious effects in the final junction. Wetting is also dependent upon the dot material, and alloys of indium with certain others metals have been found to be superior to pure indium. Although the art of soldering is old, adequate theories or data are not available to explain the various wetting properties. This subject has been surveyed by Bondi.²

A liquid flux containing zinc chloride, similar to that used by tin-smiths for decades, was employed with some success in wetting indium to germanium. It has good wetting properties, but dissolves a considerable amount of indium when heated. The dissolved indium is then deposited as a thin spotty layer around the dot periphery. This condition, aptly called "measles," can cause death or crippling illness to the junction and in any event causes penetration over an uncontrollable area during the subsequent alloying step. One possible solution to the problem is to use a different flux, but no flux other than those containing $ZnCl_2$ has been found which will cause pure indium to wet germanium satisfactorily at about 300° C. However, if 1 per cent zinc is added to the indium, the wetting properties are improved. Zinc is a p-type impurity and does not change the electrical characteristics of the junction. However, the better wetting properties permit the use of a weak flux to obtain satisfactory wetting at 300° C. A suitable nonmetallic flux is the following:

Ammonium chloride	— NH_4Cl	2 grams
Hydrazine monohydrobromide	— $N_2H_4 \cdot HBr$	2 grams
Methanol	— CH_3OH	10 cc
Water	— H_2O	5 cc
Glycerin	— $C_3H_8(OH)_3$	1 cc

This flux dissolves less indium and avoids the difficulties mentioned above when zinc chloride is used. Some of the commercial fluxes as for instance Divco 521* also work reasonably well in this respect.

² A. Bondi, "The Spreading of Liquid Metals on Solid Surfaces," *Chemical Review*, Vol. 52, pp. 417-458, 1953.

* Product of Division Lead Company, Chicago, Illinois.

The details of the wetting or soldering operation are as follows:

(1) The 1 per cent zinc-indium alloy is made into balls of the desired volume.

(2) The balls are etched in a 1 per cent HCl solution for three minutes, then washed in deionized water. Rinsed in methanol and dried.

(3) The balls are placed in the previously described nonmetallic flux, the excess flux drained off and then placed on the germanium wafer in an anodized aluminum jig. (Jigs are used only during the soldering operation.)

(4) The assembly is placed in a furnace with a pure dry hydrogen atmosphere, heated rapidly to 340° C and held there for 3 minutes. (This firing may also be done in air with reasonable success.)

(5) The germanium wafer is now washed in a 10 per cent HCl solution and then rinsed in a deionized water solution containing a few drops of Triton X-102* wetting agent.

With the above processing, the dots are firmly soldered to the germanium, and by a process of alloying, a small amount of germanium is taken into solution by the zinc-indium. The initial penetration is about 0.1 mil.

It is important to have some means of evaluating the wetting-soldering operation. The standard technique for doing this is to remove the dot material from the germanium by dissolving the dot with hydrochloric acid. (This does not attack the germanium.) The exposed surface structure is then examined microscopically for unwetted areas. This technique is tedious and difficult to interpret exactly. A refinement of this technique using preferential crystallization of germanium makes the interpretation and absolute evaluation more definite. A germanium wafer with its face about 10 degrees off the (111) crystal plane is used for these tests. The recrystallized germanium surfaces follow the (111) faces so there will be a 10-degree difference between the surface planes of the recrystallized area and the unwetted area. Thus, if vertical illumination is used for the examination and if the surface is adjusted so light is reflected into the lens from the germanium wafer surface, all unwetted areas will be bright and easily detected. Figure 2 shows the application of this technique to accentuate an unwetted area.

ALLOYING

In the alloying operation, the dot material penetrates the germanium wafer. It is generally desirable to arrange the alloying

* Product of Rohm and Haas Resinous Products Division, Phila., Pa.

condition so that it is an equilibrium process in which case the depth of penetration is independent of time of alloying. The depth of penetration is determined by alloying temperature, dot volume, and wetted area with the aid of standard phase diagrams. The dot volume and wetted area are so chosen that the wetting angle between the dot and germanium surfaces is such that further spreading due to liquid forces is minimized. To prevent further wetting during alloying, a neutral or slightly oxidizing atmosphere is desirable.

Indium alloying into germanium is a dissolving process similar in many respects to etching of germanium by acid. The rate of penetration is different for the various crystal planes so that the crystallographic orientation of the germanium wafer is important. The (111) crystallographic plane (see Figure 3) is most densely populated and

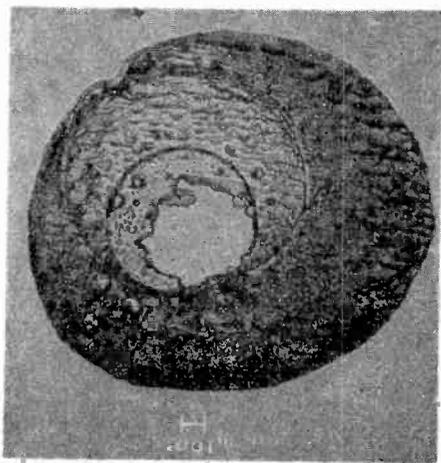


Fig. 2—Recrystallized region on an off-axis crystal.

the atoms in these planes are tied together by 3 bonds. The single bond (pointing upward in Figure 3) is most easily removed, thus giving penetration principally by the peeling off of layers. The (111) plane, therefore, acts as a natural leveling means to flatten out the alloying front. Better leveling action is obtained if near-equilibrium conditions prevail. For this reason, a gradual increase in temperature is desirable; 20° C per minute from 300° C to the final temperature is satisfactory.

In a germanium crystal there are other (111) planes that can act to resist alloy penetration. These planes can be located by remembering that the 4 bonds of each atom of Figure 3 point in a [111] direction and the location of the planes are then perpendicular to these directions. A set of planes will then intersect to form a truncated pyramid

as sketched in Figure 4a. Figure 4b shows the top view of a junction in which the triangular shape is evident. If a junction is sectioned along the place indicated in Figure 4a, it should intersect the bottom plane and be perpendicular to the right (111) plane. The line of intersection should include a 110-degree angle. Figure 4c shows such a cross-section with the correct angle of 110 degrees. The left intersection is rounded because the original round liquid front and the sharp angle of the (111) planes are incompatible.

The highly restrictive action of the (111) plane is shown in Figure 5a with a corresponding pictorial sketch in Figure 5b. Here the alloy-

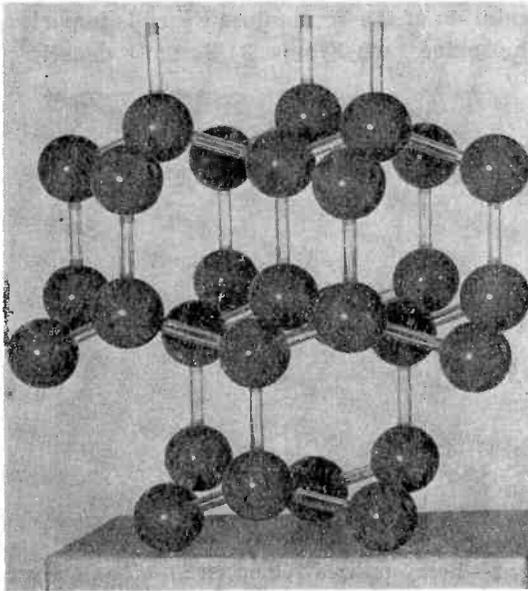
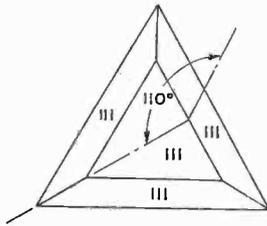


Fig. 3—Model of germanium crystal lattice.

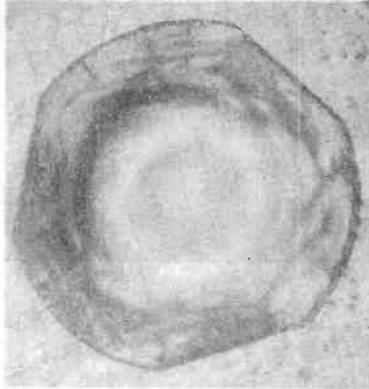
ing was done on a crystal 9 degrees off axis. The resulting junction was flat but at the corresponding 9-degree angle with the surface.

The restrictive influences of the side (111) planes have a deleterious effect if there are unwetted spots. Where the (111) planes are well defined, they tend to prevent the undercutting of an unwetted section on the triangular sides and cause the unwetted region to propagate itself. Such an unwetted spot is shown in Figure 6. Note that the junction is flat on both sides of the unwetted area.

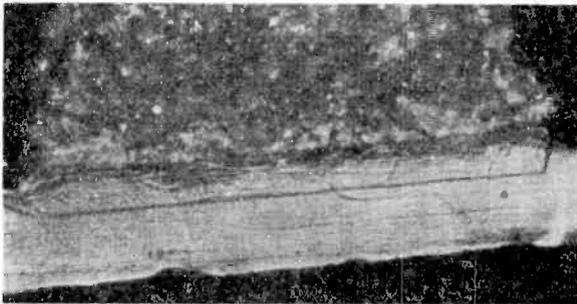
The success of this new technique in obtaining flat junctions can be seen in Figure 4c. No deviations can be detected in the bottom of the junction when compared with the microscope cross hair at a mag-



a. Sketch of (111) planes in solid.



b. Top view of indium dot showing triangular shape.



c. Cross section of p-n junction.

Fig. 4—Influence of (111) planes on wetting.

nification of 480 times. This degree of flatness is convincing evidence that the junction surface is determined by a crystal plane. In experimental lots of six units, flat junctions have been made that showed a variation of ± 0.025 mil in the location of the junction. Thus, not only do these new techniques provide for microscopically flat junctions, but also for precise control of their location.

Junction sectioning as shown in Figure 4 was very useful in evaluating the new alloying process. The major requirement of a cross-

sectioning method in addition to showing a true high-definition picture is that the method be fast so that a large number of samples can be readily examined. An improved method for rapidly exposing junctions was developed and proved to be an important tool. Details are given in the Appendix.

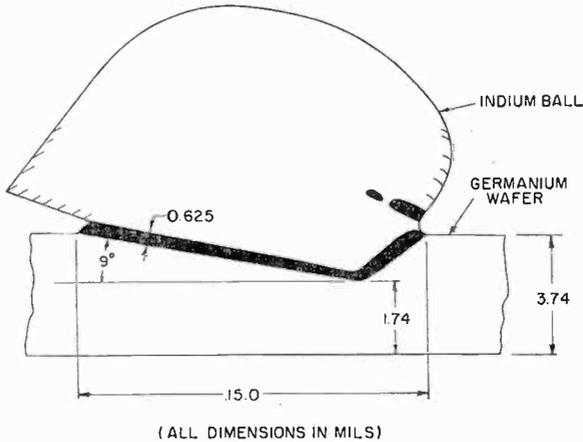
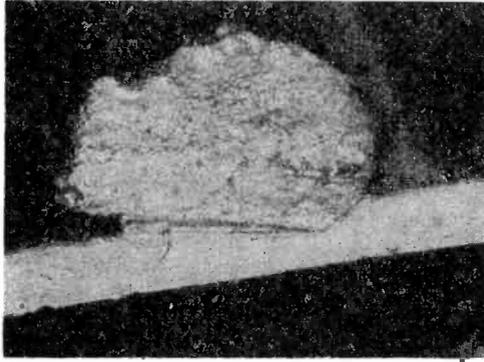


Fig. 5—Effect of crystal orientation.

RECRYSTALLIZATION

Recrystallization of the dissolved germanium back upon the base wafer as a good single crystal is the final step in the formation of a junction. For greater uniformity and crystal perfection, recrystallization should proceed on a near equilibrium basis. A gradual decrease in temperature is required; a 20°C reduction per minute has been satisfactory.

When the temperature is decreased gradually, substantially all the dissolved germanium is redeposited on the base and a uniformly thick recrystallized region as shown in Figure 4c is obtained. Dendritic growth, a potential source of noise, is also eliminated. The thin recrystallized end regions obtained in the single step alloying process (see Figure 1) are not present. The usual heavy etching used to remove these thin regions is no longer required. Another result of the gradual decrease in temperature is the absence of germanium crystallites scattered through the resolidified indium. This makes the connection of wire leads to the dots considerably easier.



Fig. 6—p-n junction showing an unwetted area.

CONCLUSIONS

Junctions made by the three-step process outlined have the following desirable characteristics:

- (1) Flat bottom surface.
- (2) Exactly defined depth and area.
- (3) Uniformly thick recrystallized regions.
- (4) Absence of thin end regions and dendritic growth.
- (5) Absence of germanium crystallites in the dots.

The net result of these characteristics is a uniform well-defined junction with considerably improved electrical properties. Thus, the upper frequency limit to which the alloy transistor can be economically exploited is considerably extended.

ACKNOWLEDGMENT

Contributions by L. Pensak, especially to the soldering technique, are gratefully acknowledged. Discussions with H. Kettering on chemical problems were very valuable. The skill of Mrs. E. Moonan in preparing and sectioning a great many samples expedited the progress of the work.

Appendix — JUNCTION SECTIONING TECHNIQUE

To prepare a germanium junction for sectioning, it is firmly mounted on a suitable stem and lead wires are attached to the indium dots. No encapsulant or potting material is necessary, thus speeding up the process greatly.

The mounted pellet assembly is then clamped in the sectioning jig, Figure 7, with the free end of the germanium crystal down through the slot in the jig. A piece of No. 1 metallographic paper is placed on a smooth flat surface with the abrasive side up. Two paper strips are placed on the paper with their edges parallel and about $\frac{1}{2}$ inch apart. The jig is then placed on the paper strips so that the jig body itself is on the paper strips, but the slot is over the abrasive between the

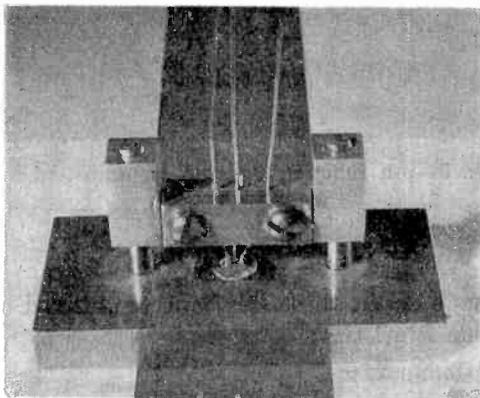


Fig. 7—Sectioning jig.

paper strips. The jig is then moved along the paper strips with the weight of the jig slide holding the germanium down against the abrasive in one direction of travel and with the slide raised as the jig is moved in the other direction. The cutting is done in one direction only in order to obtain a plane surface over the germanium wafer. When the germanium wafer is ground down to within 5 or 6 mils of the point where the junction shapes are to be examined, the No. 1 metallographic paper is replaced by No. 0 emery and the grinding procedure continued through successively finer paper down to 4/0. The final polish is done on a glass plate using Linde* A aluminum oxide powder mixed with water. The jig is used directly on glass with no paper strips this time. The polishing strokes must be in the same direction as was used for grinding. A highly polished surface is not

* Trademark of Linde Air Products Co., New York, N. Y.

necessary if the copper plating technique described below is used to increase contrast.

When the polishing is completed, the transistor is removed from the jig and washed thoroughly with water. It is then etched for about one second in a mixture of 5 parts nitric acid, 5 parts hydrofluoric acid and 2 parts water. After rinsing with water, it is connected in the circuit shown in Figure 8 and immersed in a solution of $\text{Cu}(\text{NH}_3)_4(\text{OH})_2$ (with just enough NH_4OH to clear the solution) contained in a copper cup.² Copper is selectively plated on the p-type germanium regions by depressing the push button for a total of about three seconds. This is normally done by using three one-second pulses separated by an

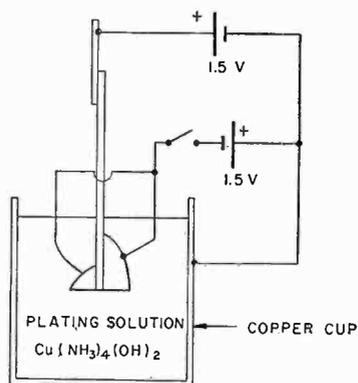


Fig. 8—Plating circuit.

interval of a few seconds to prevent depletion of the electrolyte near the germanium.

The unit is then washed with water and dried. The junction contours can now be seen by viewing the transistor through a microscope having vertical illumination. The junction locations may be seen both as a color difference and as a fine line caused by selective electrolytic etching at the junction interface.

By use of the described procedure, a satisfactory sample can be prepared in about 15 minutes. Multiple jigs that hold three units may also be used.

* The resistivity of the plating solution is 600 to 800 ohm centimeters.

SOME ASPECTS OF THERMAL CONVERSION IN GERMANIUM*

BY

H. KROEMER

RCA Laboratories,
Princeton, N. J.

Summary—It is known that a slight copper contamination on n-type germanium may convert the germanium to p-type upon heat treatment. If, however, the germanium is in contact with a molten metal during the heating process, then the metal soaks up almost all of the copper and no conversion occurs. This is consistent with the high solubility of copper in molten metals. It explains why no conversion has been observed after the heat treatment during the processing of alloy transistors. It also furnishes a method of preventing thermal conversion generally during other kinds of heat-treatments.

WHEN n-type germanium is raised to an elevated temperature (500-800° C) and then cooled, it commonly converts to p-type.¹ This phenomenon is called thermal conversion and is due to two effects. One of these is the production of lattice defects which are quenched in on rapid freezing.^{2, 3} The other is contamination by minute amounts of copper.^{4, 5} The acceptor concentration introduced by lattice defects is ordinarily small, and these defects rapidly disappear if the cooling rate is not too fast. Thermal conversion due to this effect therefore can be neglected in processes involving a slow cooling rate. This paper is concerned only with conversion due to copper.

At temperatures above 500°C, copper can diffuse into germanium at a very high rate due to its outstandingly high diffusion constant. Because of this, the slightest trace of copper on the surface of the ger-

* Presented at the I.R.E.-A.I.E.E. Conference on Semiconductor Research, Philadelphia, Pa., June 22, 1955.

¹ H. C. Theuerer and J. H. Scaff, "Effect of Heat Treatment on the Electrical Properties of Germanium," *Journal of Metals*, Vol. 191, p. 59, January 1, 1951; for a complete literature survey see Reference (5).

² R. A. Logan, "Thermally Induced Acceptors in Single Crystal Germanium," *Phys. Rev.*, Vol. 91, p. 757, August 1, 1953.

³ S. Mayburg and L. Rotondi, "Thermal Acceptors in Vacuum Heat-Treated Germanium," *Phys. Rev.*, Vol. 91, p. 1015, August 15, 1953.

⁴ C. S. Fuller and J. D. Struthers, "Copper as an Acceptor Element in Germanium," *Phys. Rev.*, Vol. 87, p. 526, August 1, 1952.

⁵ C. S. Fuller, J. D. Struthers, J. A. Ditzenberger and K. B. Wolfstirn, "Diffusion and Solubility of Copper in Germanium," *Phys. Rev.*, Vol. 93, p. 1182, March 15, 1954.

manium or in the heated environment will introduce copper acceptor atoms into the germanium upon heating. These acceptors can neutralize the donor impurities and, if present in sufficient numbers, convert the germanium to the opposite conductivity type. Harmful effects on minority carrier lifetime have also been observed when copper is present.

In order to prevent thermal conversion of this kind it is necessary both to clean the germanium very carefully prior to the heat treatment and to eliminate any possibility of copper contamination from the atmosphere or the walls of the furnace during the heat treatment.

Strangely enough, the n-type germanium which is used in the ordinary p-n-p alloy type transistor, although it may receive the necessary heat treatment, has not been known to convert even in the presence of large amounts of copper. Even when copper is intentionally added to the indium dots which are commonly used in these units, or as a plating beneath them,⁶ the transistors show no signs of thermal conversion. This paper offers an explanation of the apparent anomaly.

Although the diffusion constant of copper is very high, its solid solubility in germanium is, chemically speaking, extremely low. The maximum solubility is less than 10^{-4} atomic per cent at about 870°C , dropping to 10^{-5} atomic per cent at about 710°C , and so on.⁴ The solubility of copper in ordinary metals and particularly in molten metals, however, lies in the order of a few per cent or higher, i.e., a difference in solubility of at least 10^4 .

Consider now the case where (1) a metal such as molten indium directly contacts the germanium surface, and (2) the system is heated to temperatures at which the molten contact metal intimately wets the germanium. Under these conditions the copper can readily diffuse through both the contact metal and the germanium, and its distribution over the two parts of the system can reach an equilibrium. Because of their extremely different solubilities, the two parts of the system will then have different copper concentrations. This is obvious if the total amount of copper present is sufficient to saturate both parts of the system with copper. However, under ordinary circumstances, the total amount of copper initially present, whether in the contact metal or the germanium itself, will be small compared to the value necessary for over-all saturation. If one assumes that the solutions of copper in germanium and in the molten metal are thermodynamically ideal solu-

⁶ L. D. Armstrong, "p-n Junctions by Impurity Introduction Through an Intermediate Metal Layer," *Proc. I.R.E.*, Vol. 40, p. 1341, November, 1952.

tions,⁷ then the ratio of the two equilibrium concentrations will be equal to the ratio of the solubilities, and the relative concentrations (ratio of actual concentration to saturation concentration) will be the same in both phases. If there is only a very small amount of copper present (an amount which usually would cause thermal conversion if no molten metal were in contact with the germanium surface), then this small amount will be soaked up almost completely by the molten metal because of its much higher copper solubility. If the amount of copper present or being fed into the system from the walls of the furnace is not too high, there will be only a negligible amount of copper left in the germanium, and thermal conversion will not occur.

If the solution of copper in the molten metal is not an ideal solution the above conclusions do not hold quantitatively. However, they remain qualitatively true because the deviations from the ideal are never so large as to compensate for a solubility ratio of 10^4 or more.

In order to prove the hypothesis, at least qualitatively, the following experiments were carried out:

(1) Several germanium wafers (6-8 ohm-centimeter n-type) were heated for 1 hour at 700°C in an atmosphere of rather impure hydrogen in a furnace in which thermal conversion occurred regularly. Indium dots had first been soldered to some of these wafers at about 250°C . To assure the presence of copper, one of the wafers containing dots had been dipped into $\text{Cu}(\text{NO}_3)_2$ solution before heating. On another, the indium dots were allowed to dissolve some copper by dipping a copper wire into the molten dots during the soldering process.

After heating to 700°C , the wafers were tested with a thermal probe for their conduction type. It was found that all wafers without dots had strongly converted to p type. All wafers with dots remained n type on both sides with the exception of isolated spots more than 1 to 2 millimeters away from the nearest dot. This result supports the original conjecture. That some p-type spots were found far from the dots was to be expected for two reasons: (a) With increasing distance the diffusion gradient becomes flatter to that it requires a much longer diffusion time to remove the copper from more distant spots; (b) The furnace which was used in these experiments was known to be contaminated so that fresh copper was constantly being fed into the Ge-In system. Under these circumstances a stationary diffusion gradient will

⁷ This is true at least for the solution of Cu in Ge. See C. D. Thurmond and J. D. Struthers, "Equilibrium Thermochemistry of Solid and Liquid Alloys of Germanium and of Silicon. II. The Retrograde Solid Solubilities of Sb in Ge, Cu in Ge, and Cu in Si," *Jour. Phys. Chem.*, Vol. 57, p. 831, November 20, 1953.

be established around the In dots and it will be impossible to remove the copper beyond a certain distance.

(2) In a second experiment, some previously converted wafers and some indium-doped p-type wafers (10-15 ohm-centimeter) were fired with dots upon all of them. The previously converted wafers converted back to n-type in the region near the dots, while the doped p-type wafers remained unchanged. This shows that the copper is really taken out of the germanium and that the prevention of conversion is not due to a compensating n-type impurity coming out from the dots.

It appears, then, that indium and probably most molten metals act as a sink for impurities with low solubilities in Ge, especially the rapidly diffusing copper. This explains why no thermal conversion occurs in the process of making alloy junctions, in spite of the fact that the temperature in many cases should be sufficiently high, and even when copper is purposely added. In addition, it appears that thermal conversion during any kind of high-temperature treatment of germanium may be prevented by bringing the germanium into contact with a suitable molten metal during the heating process.

SURFACE TREATMENT OF SILICON FOR LOW RECOMBINATION VELOCITY*†

BY

A. R. MOORE AND HERBERT NELSON

RCA Laboratories,
Princeton, N. J.

Summary—The surface recombination velocity ordinarily attained on silicon surfaces as a result of common etches is of the order of 10^4 centimeters per second. This high value is a serious limitation in the performance of silicon transistor devices due to the loss of minority carriers at the surface. Films of aniline-like aromatic liquids and films of sodium dichromate-like salts have been found which dramatically reduce the recombination velocity by two orders of magnitude. Surface recombination velocities similar to those commonly obtained on germanium (100 centimeters per second) can be obtained for p-type silicon. In case of n-p-n alloy type silicon transistors, surface treatment with sodium dichromate commonly leads to values of the current amplification factor, α_{cb} , which are four times larger than the values observed before the treatment. The surface treatments also eliminate "channeling leakage" at p-n junctions. This is consistent with the view that the films produce their effects by causing the energy bands (in p-type silicon) to curve upwards at the surface.

A tentative hypothesis is proposed to account for the behavior of the surface films.

INTRODUCTION

THE electrical characteristics of semiconductor devices are greatly affected by surface recombination of minority carriers. It is well known that the base-to-collector current amplification factor, α_{cb} , of transistors decreases¹ and the junction saturation current increases² with an increase in the surface recombination velocity, s . With germanium, conventional electrolytic and chemical etches lead to rather low values of s (50-200 centimeters per second). High values of α_{cb} may be obtained with conventional etches, since only a very small fraction of the current carriers injected by the emitter is lost at the surface. In silicon, however, s remains high (5,000-10,000 centimeters per second) after etching with the common etches. Consequently it is more difficult to obtain high values of α_{cb} in the case of silicon devices. It is particularly important, therefore, in the case of silicon to consider other means than etching for the reduction of s .

* Reprinted from *RCA Review*, March, 1956.

† Presented at the I.R.E.-A.I.E.E. Conference on Semiconductor Research, Philadelphia, Pa., June 22, 1955.

¹ A. R. Moore and J. I. Pankove, "Effect of Junction Shape and Surface Recombination on Transistor Current Gain," *Proc. I.R.E.*, Vol. 42, pp. 907-913, June, 1954.

² W. M. Webster, "The Saturation Current in Alloy Junctions," *Proc. I.R.E.*, Vol. 43, pp. 277-280, March, 1955.

Considerations by A. Moore and W. Webster,³ for instance, lead to the conclusion that an increase in the doping impurity concentration at the surface of a semiconductor will tend to inhibit surface recombination. Under these conditions, the minority carriers are prevented from reaching the surface by an internal electric field associated with the variation in doping impurity concentration. In p-type silicon, therefore, a surface treatment which causes the energy bands to curve upwards at the surface should similarly cause a reduction in s . It was thought that this effect may be produced by a chemical treatment which leads to a surface film containing ionic components of high electron affinity. Attempts to produce films of this type on silicon surfaces have achieved a dramatic reduction of s . Surface recombination velocities similar to those commonly obtained on germanium are achieved on p-type silicon by these treatments. A description of the techniques used and the results obtained constitute the subject matter of the present report.

MATERIALS AND PROCESSING TECHNIQUES

Two groups of chemical compounds have been found to produce surface films on silicon which greatly reduce the surface recombination velocity. One is a group of strongly ionic salts which are oxidizing agents. Sodium dichromate is a typical example. The other is a group of aniline-like aromatic liquids. The conditions under which the two groups of compounds produce the desired effect differ considerably and consequently each case will be discussed separately.

The practices involved in the use of sodium dichromate-like salts for the reduction of s on p-type silicon are extremely simple. A solution of the salt in distilled water is prepared and a droplet applied to the area to be treated. The water evaporates and leaves a strongly adsorbed salt film on the surface which leads to an immediate and permanent reduction of s . The concentration of the salt in solution is not critical; values on the order of 1 per cent by weight are normally used. The silicon specimen must be freshly etched before the application of the salt. A hot sodium hydroxide solution is normally used as the etch and good results are obtained with a solution of 7 grams of the hydroxide in 100 cubic centimeters of water heated to 80°C.

When aromatic liquids are used, the practices involved are more complicated. The vapor pressure of the liquid is high so that lasting

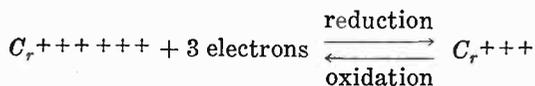
³A. R. Moore and W. M. Webster, "The Effective Surface Recombination of a Germanium Surface with a Floating Barrier," *Proc. I.R.E.*, Vol. 43, pp. 427-435, April, 1955.

reductions of s are obtained only when the silicon specimens are hermetically sealed in containers filled with these liquids. The aromatic liquids inhibit surface recombination only in the presence of an electric field. The difference in potential between the emitter dot and the adjacent surface of the silicon often is sufficient to establish the requisite electrical gradient in a transistor device but in some cases auxiliary electrodes are essential for optimum results. Before immersion in the aromatic liquid, the silicon specimens are etched in the hot sodium hydroxide solution described above.

TENTATIVE HYPOTHESIS OF FILM BEHAVIOR

The exact behavior of the inorganic salts which reduce surface recombination on p-type silicon is not completely understood. Nevertheless, it is of interest to consider a tentative hypothesis based on the observation that these salts are relatively strong oxidizing agents and belong to reversible systems of oxidation-reduction reactions. It is proposed that the oxidation-reduction potential developed in such systems modifies the surface potential in such a manner that minority carriers are inhibited from reaching the surface. It is this effect which reduces the effective recombination rate. Other systems for creating such a barrier within the semiconductor to achieve similar results have been discussed by Moore and Webster.³

The terms oxidation and reduction are used in their general sense, i.e., an oxidation reaction is one in which an element loses electrons and thus increases in valence. Oxygen may or may not be involved. Conversely, in a reducing reaction an element gains electrons to decrease its valence. Thus the dichromate-chromic system (e.g., sodium dichromate) has been found effective in the reduction of surface recombination velocity, and



where the reduction and oxidation directions refer to the chromium.

In a reversible system, both the oxidized and reduced components may be present simultaneously and an equilibrium will be reached between them and the electrons. An inert metallic electrode immersed in a solution of oxidized and reduced components will give up or abstract electrons according to the inclination towards reduction or oxidation, respectively. In our example, the strongly oxidizing dichromate inclines the reaction to reduction (of C_r^{++++}) and in so doing abstracts electrons from the electrode. The potential of the electrode is thereby

raised with respect to the solution. This potential, measured under standard conditions, determines the equilibrium state and, in electrochemistry, is the "standard oxidation-reduction potential." For our example, the standard potential E_0 is 1.3 volts.

It is characteristic of the salts found effective for the reduction of surface recombination on p-type silicon that they have relatively large positive standard potentials (> 0.3 volt). Thus the following systems have been found effective for possible practical use:

System	E_0
$\text{MnO}_4^-:\text{MnO}_2$	1.6
$\text{ClO}^-:\text{Cl}^-$.9
$\text{Cr}_2\text{O}_7^{2-}:\text{Cr}^{+++}$	1.3

In addition many other oxidation-reduction systems, some of a particularly simple type, have been found equally effective in reducing s . Many of these are not suitable for practical use because of chemical instability, excessive moisture effects and the like. However, it is characteristic of all that their standard oxidation-reduction potentials exceed 0.3 volt.

In the case of a metallic electrode, the oxidation-reduction potential is supported in a narrow region of the solution adjacent to the electrode. It cannot be supported in the electrode itself since the electrode is a good conductor. If a semiconductor is substituted for the metallic electrode, the situation is entirely different. With a semiconducting electrode the fixed charge concentration (impurity density) in the semiconductor is generally much lower than the ionic concentration in ordinary electrolytes. Under these conditions it is much easier to support the potential in the semiconductor than in the solution. Thus, we arrive at the conclusion that the surface of a chemically inert semiconductor immersed in an oxidation-reduction solution will charge up to the appropriate standard electrode potential while the interior of the semiconductor will remain unaffected. In the case of a positive electrode (or surface) potential, this creates just under the surface the type of barrier described by Moore and Webster to inhibit surface recombination on p-type material. That is, the surface is effectively more p-type (through the loss of electrons) than the interior. The resultant internal electrostatic field prevents minority carriers from reaching the surface and consequently the effective surface recombination rate is lowered.

It has also been suggested that s may be a function of surface potential because of its effect on the population of surface traps. What-

ever the detailed reasoning it does not seem improbable that adjustment of the surface potential can have a drastic influence on surface recombination.

It is recognized that in the experimental evidence described earlier we do not have the "standard" solutions so that the actual potential will be somewhat different from the standard potential. Moreover, in the absence of definite information to the contrary the possibility of a chemical reaction with the silicon has been neglected. However, in the present hypothesis, it is suggested that these effects would modify the magnitude of the potential but not alter the general argument.

The above argument has been derived solely from observations on inorganic oxidation-reduction systems. If this hypothesis is a general one it should also explain the action of the aromatic liquids; these are equally effective in reducing s but require a small electrolyzing voltage. The chemistry of these materials is complicated and very little is known about their reaction with silicon. However, nitrobenzene forms many intermediate products during electrolytic reduction to aniline and one of these products is quinone. It is suggested that this intermediate product might be the active ingredient. Quinone and hydroquinone form a well-known oxidation-reduction system with a standard electrode potential of 0.718 volt. A mixture of quinone and hydroquinone (quinhydrone) in water solution was tried on p-type silicon (without an electrolyzing potential). This solution was found to be fully as effective as either aniline or nitrobenzene for the reduction of s . The function of the polarizing voltage required for the latter compounds, according to this view, is simply to provide for anodic oxidation or cathodic reduction to the active intermediates at the semiconductor surface.

EXPERIMENTAL PROCEDURES AND RESULTS

The effect of the surface treatments on the recombination velocity was determined either directly by measurements of the s of a particular silicon surface before and after application of the film or indirectly by measurements of the α_{cb} of a silicon transistor before and after application of the film to the emitter area.

The light-pulse photoconductivity technique was used for the direct measurements. Thin wafers of silicon (about 5 mils thick) were prepared for the test by an initial etch in CP4 and a final etch in a hot sodium hydroxide solution. To minimize the contribution of volume recombination to the loss of current carriers, silicon of good bulk lifetime (> 20 microseconds) was used for the direct measurements.

Direct measurements were made on the sodium dichromate, aniline, and nitrobenzene systems with p-type silicon.

In case of the sodium dichromate-p-type silicon system, a permanent reduction of s from a value of about 5,000 centimeters per second to a value less than 100 centimeters per second was found to occur as a direct result of application of the film.

In case of the aniline and the nitrobenzene-p-type silicon system, the silicon specimen was inserted between two condenser plates with films of the aromatic liquid occupying the space between the major surfaces of the silicon specimen and the condenser plates. When aniline was used, the surface recombination velocity was reduced from an approximate value of 5,000 to a value of 100 centimeters per second. This reduction of s , however, was contingent upon the application of a *negative* voltage of about one volt to the condenser plates with respect to the silicon specimen. In the absence of this voltage no reduction of s was observed. When nitrobenzene was used, the value of s decreased

Table I—n-p-n Silicon Transistor Performance Before and After Sodium Dichromate Treatment

Transistor	4 kc Current Amplification Factor (α_{cb})		4 kc Power Gain db		Effective Lifetime (τ_e) μ sec.	
	Before	After	Before	After	Before	After
A No. 1	8.2	29.5	33.0	40.5	.75	1.0
No. 2	7.4	30.0	33.0	38.6	.7	1.1
No. 3	8.4	29.1	33.1	40.2	.7	.9
No. 4	8.5	27.2	30.3	35.3	.7	1.1
B No. 1	4.6	13.2	29.5	37.6	.35	.55
No. 2	5.7	14.0	30.5	36.0	.5	.75
No. 3	4.0	11.5	30.5	36.7	.5	.7
No. 4	5.3	13.3	31.0	36.1	.45	.6

All values of gain measured at $V_c = 6$ volts, $I_E = 1$ milliampere. Power gain measured with resistive input and conjugate matched output.

from 5,000 to 220 centimeters per second as a result of the application of a *positive* voltage of one volt to the condenser plates.

Many of the tests carried out to determine the effect of the surface treatments have involved measurements of α_{cb} before and after applying the film to the emitter area of an alloy type n-p-n silicon transistor. Since the use of sodium dichromate films has shown the most promise for practical applications, a large number of tests have involved this salt. Table I is a summary of the results obtained.

The large increases of α_{cb} observed for the transistors of group A are typical of results obtained with units made from silicon of relatively high bulk lifetime. The transistors of group B were made with base widths equal to those of group A but with silicon of low bulk lifetime. Since bulk recombination contributes materially to the loss

of minority carriers in these units, smaller increases in α_{cb} were obtained. The effective lifetime, τ_e , which is determined by bulk as well as surface recombination was measured by a pulse injection technique.⁴

Similar indirect measurements were used to determine the effects of various other surface films. The results obtained indicate that for silicon of equal bulk lifetime and for transistors of equal base widths, roughly equal increases in α_{cb} are obtained when other strongly ionic salts are substituted for sodium dichromate. This is true, for instance, in the case of magnesium chromate, calcium dichromate, calcium hypochlorite, lithium dichromate and potassium permanganate.

Indirect measurement showed that results similar to those obtained with aniline could also be achieved with pyredine, nitrocyclohexane, 0-nitrotoluene, 2, 4 dinitrofluorobenzene and α -100 naphthalene. In these cases, maximum increases in the current amplification factors were obtained with those aromatic liquids which required an applied negative potential for their "activation." This is consistent with the fact that n-p-n transistors require a negative emitter-to-base bias during operation. Thus with aniline, which direct measurements showed to require a negative potential, the increase in α_{cb} was equal to that with sodium dichromate. On the other hand, the increase was smaller with nitrobenzene, which direct measurements showed to require a positive potential.

A few tests were carried out with n-type germanium and silicon. Sodium dichromate showed no effect while immersion in aniline did increase the current-gain factors of p-n-p transistors.

The effects of the surface films on p-type silicon appear permanent. This observation is based largely on experimental data for transistors using sodium dichromate films. A relatively large number of hermetically sealed n-p-n silicon transistors made with the sodium dichromate film have shown no decrease in current amplification factor on shelf life. A few of these transistors have also been exposed to a temperature ambient of 150°C for as long as 48 hours. A decrease in the current amplification factor occurs as a result of the heating. The rate of decrease is appreciable, however, only during the first hours of heating and in all cases the transistors after baking have shown values of α_{cb} substantially higher than those observed before the application of the film.

OTHER EFFECTS OF SURFACE FILMS

In the course of these investigations, other effects have been observed which are consistent with the view that the films cause the

⁴ L. R. Lederhandler and L. J. Giacoletto, "Measurement of Minority Carrier Lifetime and Surface Effects in Junction Devices," *Proc. I.R.E.*, Vol. 34, pp. 477-483, April, 1955.

energy bands (in p-type silicon) to curve upwards at the surface. The presence of the films at the p-n junctions was found to cause an improvement of the reverse junction characteristics. This effect is illustrated by the curves A and B of Figure 1. Curve A was obtained before and curve B after application of a sodium dichromate film to the silicon surface adjacent to a p-n junction. "Roundhouse" reverse character-

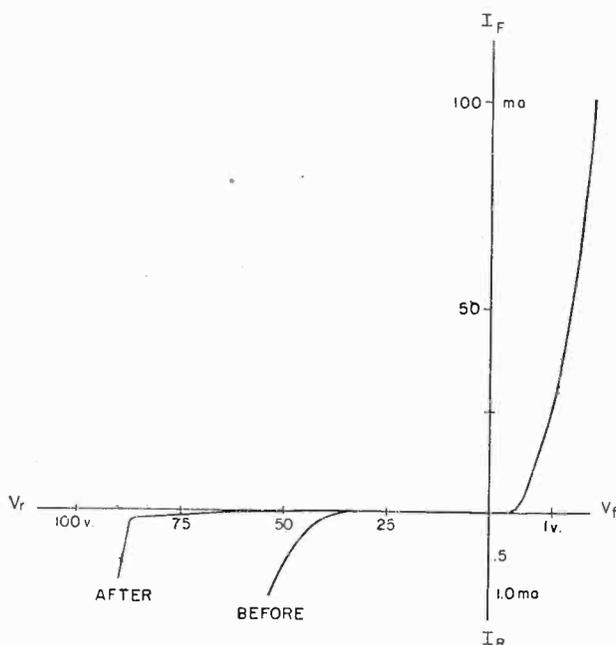


Fig. 1—Diode characteristics of p-n junction on p-type silicon before and after application of sodium dichromate to junction area.

istics of the type shown by curve A are often ascribed to "channeling leakage" at the p-n junction due to the fact that the energy bands normally tend to curve downwards at the surface (in p-type semiconductors). The elimination of the "roundhouse" characteristic in curve B may thus be considered evidence that the application of the surface film has reversed the curvature of the energy bands at the surface.

P-N-P TRANSISTORS USING HIGH-EMITTER-EFFICIENCY ALLOY MATERIALS*†

BY

L. D. ARMSTRONG,‡ C. L. CARLSON,‡ AND M. BENTIVEGNA#

RCA Tube Division,
Harrison, N. J.

Summary—In the past, the performance of p-n-p alloy transistors at high currents has been limited by the decrease of emitter efficiency with current level. The addition of small percentages of gallium or aluminum to indium, for use as the emitter alloy, produces greatly improved high-current characteristics. As compared with pure indium, the use of gallium alloys improves emitter efficiency by about 3.5 times, and the use of aluminum-bearing alloys by about 10 times. Techniques for preparation of the alloys and results of tests on transistors using the various emitters are described. Volume lifetime is measured as a function of injection level to permit comparison with the theoretical equations for current amplification factor. These measurements are discussed briefly, and a revised equation for current amplification factor at high currents is given.

INTRODUCTION

ONE of the most common methods used in making rectifying p-n junctions in semiconductors is the alloy process.^{1, 2} In this technique a metal is alloyed into the semiconductor. When cooling takes place, some of the metal atoms remain in the recrystallized material. Depending on the choice of metal used for alloying, the recrystallized zone may be either "n" or "p" type.

A metal must have certain characteristics to be suitable for alloying into a semiconductor to form rectifying junctions and transistors. It should have low vapor pressure at the alloying temperature so that it will not evaporate and thus spread over the entire surface of the device. It should be soft to minimize mechanical strain during freezing and cooling. It should readily wet the surface of the semiconductor in a controllable fashion. Electrically, the junctions so formed should have low saturation currents, high reverse impedances, and low forward

* Reprinted from *RCA Review*, March, 1956.

† Presented at the *I.R.E.-A.I.E.E. Conference on Semiconductor Device Research*, Philadelphia, Pa., June 20, 1955.

‡ RCA Semiconductor Division, Harrison, N. J.

Formerly RCA Semiconductor Division, Harrison, N. J.; now with CBS Hytron, Danvers, Mass.

¹ R. R. Law, C. W. Mueller, J. I. Pankove, and L. D. Armstrong, "A Developmental Germanium P-N-P Junction Transistor," *Proc. I.R.E.*, Vol. 40, p. 1352, November, 1952.

² R. N. Hall, "Power Rectifiers and Transistors," *Proc. I.R.E.*, Vol. 40, p. 1512, November, 1952.

resistance. The breakdown voltage should be high. If the alloy is intended for use as the emitter of a transistor, the injection efficiency of the junction, γ , should be high.³

Because indium satisfies these requirements well, it has been widely used in the familiar p-n-p transistor. At high current levels, however, the injection efficiency of the emitter junction decreases, causing a reduction in current amplification factor.⁴ In addition, the current amplification factor varies from unit to unit when indium emitters are used, indicating possible variations in emitter efficiency.

Shockley³ has pointed out that emitter efficiency depends on the ratio of majority-carrier concentration on the two sides of the emitter junction. The emitter efficiency and, hence, current amplification factor are higher in transistors having high-conductivity emitter regions. In a p-n-p alloy transistor, the critical region is the p-type layer formed during the recrystallization process. A possible method of improving emitter efficiency involves the use of alloying materials having higher segregation coefficients, i.e., metals which are more soluble in solid germanium. When recrystallization occurs, a higher concentration of these metals remains in the recrystallized germanium, and the desired higher conductivity results.

EXPERIMENTAL WORK WITH GALLIUM AND ALUMINUM ALLOYS

Both gallium and aluminum have the desired characteristics for use as possible p-type "doping" agents. Their segregation coefficients at the melting point of germanium are 100 times greater than that of indium. It is reasonable to expect that they will also be more soluble than indium in solid germanium at the lower temperatures used in alloying. However, the physical properties of both metals have limitations. Gallium, for example, is molten at room temperature and cannot be used for mechanical support of connections. It is also difficult to alloy appreciable quantities of gallium with most other soft metals. Many three-component alloys containing gallium and other soft metals become hard and very brittle, frequently breaking into powder during subsequent processing. Aluminum, on the other hand, quickly forms an oxide surface which makes alloying difficult. When alloyed with germanium, it also forms a hard and brittle eutectic which makes penetration control difficult and causes severe mechanical strains.

Fortunately, it has been found that the advantages of both gallium

³ W. Shockley, *Electrons and Holes in Semiconductors*, D. Van Nostrand & Co., New York, N. Y., 1950; pp. 90-91.

⁴ W. M. Webster, "On the Variation of Junction-Transistor Current Amplification Factor with Emitter Current," *Proc. I.R.E.*, Vol. 42, p. 914, June, 1954.

and aluminum as "doping" agents may be retained when only small percentages of these metals are combined with one or more other components.* With these low concentrations, some of the mechanical difficulties can be minimized.

Reproducible gallium-bearing alloys have been made using indium as a major element, with small percentages of gallium and gold or silver added. The third constituent, gold or silver, acts as a carrier agent for the gallium, and insures that the gallium content is reproducible and consistent. In addition, the use of gold or silver favorably modifies the surface tension of the alloy, improves its wetting properties, and results in more consistent and uniform alloy penetration.

These gallium-bearing alloys exhibit alloying and recrystallization properties similar to those of pure indium. The depth of penetration or alloying depends on the relative amounts of the three materials. In general, the percentages of gallium and carrier metal are not critical. Satisfactory results have been obtained with alloys having gallium concentrations between 0.1 and 0.5 per cent, and gold or silver concentrations between 2 and 10 per cent.

It is more difficult to obtain consistent wetting and alloying of alloys containing 0.5 to 2.0 per cent aluminum in indium. Three methods have produced fairly satisfactory results, although each has disadvantages. In the first method, a dot of indium is first soldered or alloyed to germanium at a relatively low temperature. Another indium dot containing a small percentage of aluminum is then placed on top of the first dot and the combination is fired at a higher temperature. During firing, the aluminum from the second dot mixes with the first, or pre-wetted, dot. In the second method, an aluminum-bearing dot is soldered to germanium with the aid of a chemical flux. Metallic (zinc-chloride) fluxes tend to leach out the aluminum and produce low emitter efficiency. Satisfactory results have been obtained with some organic fluxes. The third method involves the preparation of special alloying dots in which the aluminum-bearing portion is surrounded by or sandwiched between layers of a "good-wetting" indium alloy or pure indium. The external layers wet the germanium at a fairly low temperature, but the aluminum becomes uniformly distributed only at higher alloying temperatures.

Junctions made with gallium or aluminum alloys exhibit essentially the same rectification properties as those made with pure indium. Because the emitter efficiency is higher, the current amplification factor

* A power transistor using gallium in the emitter was described by the Philips Company at Hamburg, Germany, in September, 1954. See also L. J. Tummers, "The Influence of Minority Carrier Injection on Power Transistors," N.T.F., Beiheft No. 1, pp. 31-32, 1955.

is greater at normal currents and decreases at a slower rate with increasing current. If desired, germanium having a lower resistivity may be used with these alloys without significant reduction of current amplification factor. Consequently, more freedom is allowed in the choice of one of the important transistor-design parameters.

ELECTRICAL CHARACTERISTICS

Figure 1 shows typical curves of small-signal common-emitter current amplification factor, α_{cb} , as a function of emitter-current density for transistors using emitters made of (1) indium, (2) indium-silver-gallium alloy and (3) indium-aluminum alloy. These units are identical except for emitter material. The emitter diameter is 0.015 inch and the junction spacing, W , is 0.0013 inch. The data shown in Figure 1 was measured on what is termed a "raw" surface, i.e., a condition of low surface-recombination velocity, s , which is obtained by electrolytic etching. Similar curves are shown in Figure 2 for the same three transistors with the surface modified to increase s to a relatively high value. This modification reduces the current amplification factor greatly at low currents and to a smaller degree at high currents.

It can be seen that the value of α_{cb} is considerably increased by the addition of gallium or aluminum and that this higher value is maintained to much higher emitter currents than is possible with pure indium emitters. Some transistors using aluminum alloy emitters have exhibited α_{cb} values greater than 100 at emitter-current densities of 1,000 amperes per square centimeter. It has also been found that the variation of small-signal current amplification factor (measured at low currents) between transistors is less when gallium-bearing alloys are used than when pure indium emitters are used. The reason for this improved uniformity may be simply that variations in emitter efficiency contribute to the spread of low-current α_{cb} when the emitter efficiency is low. Although transistors using aluminum-bearing alloys are even more variable at present than those using pure indium, the variation is probably due to the processing problems mentioned previously.

THEORETICAL DISCUSSION

As mentioned above, the current amplification factor of a transistor is limited by three factors: (a) surface recombination, (b) volume recombination, and (c) emitter efficiency. In most p-n-p alloy transistors, surface recombination is the major limitation at low current levels, as discussed by Moore and Pankove.⁵ However, all three factors may vary

⁵ A. Moore and J. I. Pankove, "Effect of Junction Shape and Surface Recombination on Transistor Current Gain," *Proc. I.R.E.*, Vol. 42, p. 907, June, 1954.

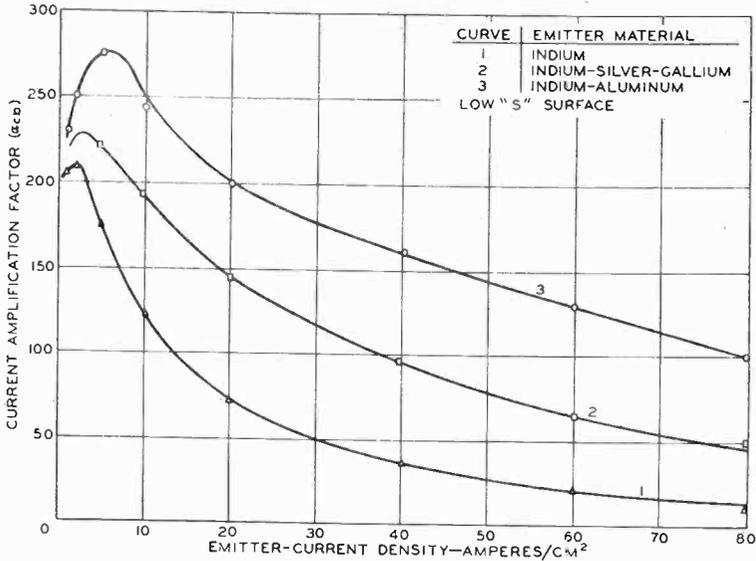


Fig. 1—Variation of small-signal current amplification factor, α_{cb} , with emitter-current density for (1) pure indium, (2) gallium-indium, and (3) aluminum-indium emitters. Surface-recombination velocity, s , is low.

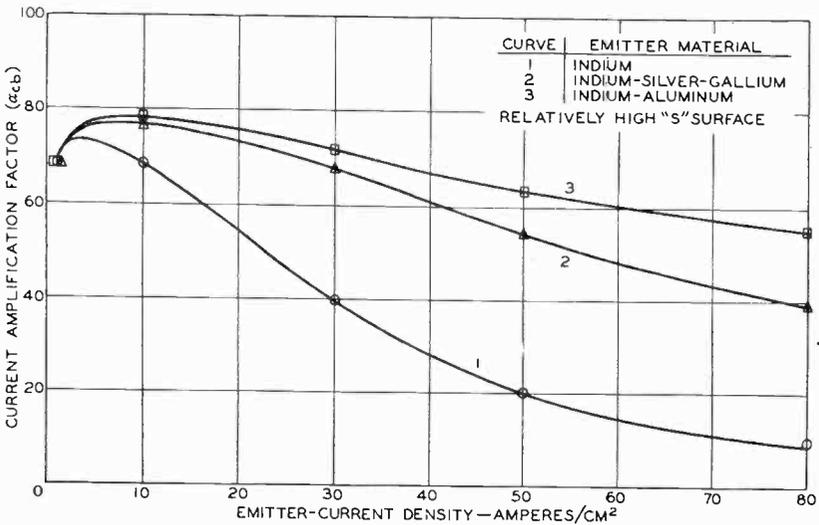


Fig. 2—Variation of small-signal current amplification factor, α_{cb} , with emitter current density for (1) pure indium, (2) gallium-indium, and (3) aluminum-indium emitters. Surface-recombination velocity, s , is high.

with emitter current, as shown by Webster,⁴ Rittner,⁶ and Misawa.⁷ At higher currents, the relative importance of the three factors changes. The percentage of carriers lost from the base by surface recombination decreases by a factor of two due to the development of a small electric field in the base region which aids the flow of these carriers to the collector. Emitter efficiency decreases with increasing emitter current and eventually becomes more important than surface recombination in limiting the current amplification factor, α_{cb} .

It was assumed by Webster that volume lifetime would also decrease with increasing emitter current. This assumption was a consequence of treating volume recombination as a bimolecular process (i.e., the rate of recombination is proportional to the product of hole and electron densities). Recent experiments indicate that this theory is incorrect. Effective lifetime measurements were made on alloy junction diodes in the manner described by Lederhandler and Giacoletto.⁸ The effects of surface recombination were minimized by making the diodes on thick slabs of n-type germanium having a resistivity of 1.5 ohm-centimeters. The diodes were etched in such a way that surface-recombination velocity was very low. Figure 3 shows the volume lifetime of two such diodes as a function of the ratio p/n_0 , where p is the injected density of holes and n_0 is the equilibrium density of electrons in the base material. It can be seen that the lifetime is fairly constant at low injection levels, increases somewhat as the injection level is increased, and then remains essentially constant to quite high injection levels. This data has been confirmed by W. P. Senett of the RCA Semiconductor Division who has measured volume and surface recombination as functions of injection level by a different method in which pulsed light is used as a minority-carrier source. The behavior shown in Figure 3 is qualitatively consistent with the Shockley-Read theory of recombination.⁹ As a point of reference, when the emitter-current density is about 270 amperes per square centimeter, the injection level, p/n_0 , is equal to 50 for a transistor in which the base region has a resistivity of 1.5 ohm-centimeters and a width of 0.0013 inch.

These measurements indicate that the equation given by Webster for small-signal current amplification factor as a function of emitter

⁶ E. S. Rittner, "Extension of the Theory of the Junction Transistor," *Phys. Rev.*, Vol. 94, p. 1161, June, 1954.

⁷ T. Misawa, "Emitter Efficiency of Junction Transistors," *Jour. Phys. Soc. Japan*, Vol. 10, p. 362, May, 1955.

⁸ S. Lederhandler and L. J. Giacoletto, "Measurement of Minority-Carrier Lifetime and Surface Effects in Junction Devices," *Proc. I.R.E.*, Vol. 43, p. 477, April, 1955.

⁹ W. Shockley and W. T. Read, Jr., "Statistics of the Recombinations of Holes and Electrons," *Phys. Rev.*, Vol. 87, p. 835, September, 1952.

current should be revised. Experiments have shown that surface-recombination velocity, s , as well as volume lifetime, is fairly constant at higher injection levels (at least for the surface treatment described above). Only emitter efficiency, therefore, changes appreciably with current, causing the observed reduction in current amplification factor. At high currents ($p/n_0 \gg 1$), the equation for current amplification factor in a p-n-p transistor is given by*

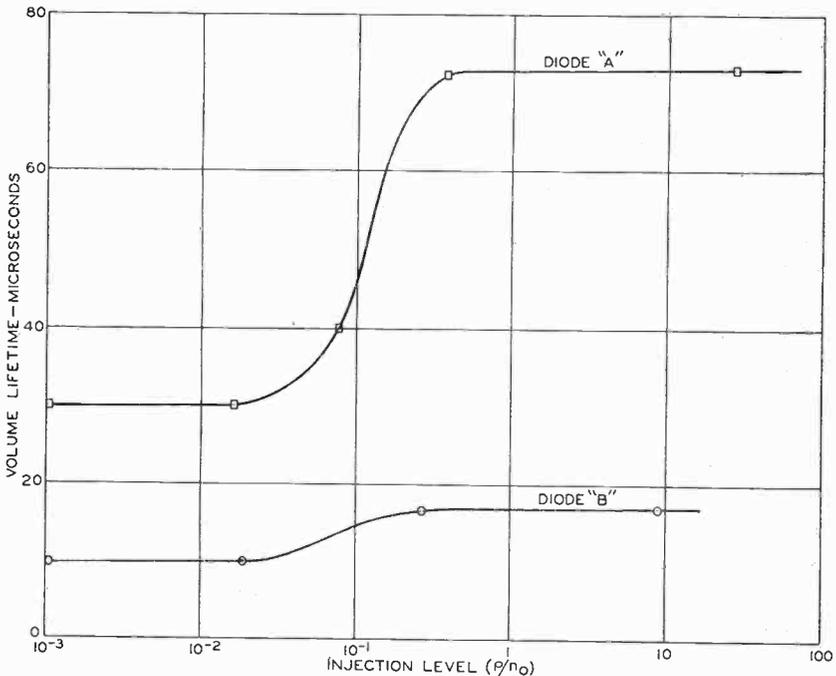


Fig. 3—Variation of volume lifetime with injection level for two alloy-junction diodes.

$$\frac{1}{\alpha_{cb}} = \frac{s A_s W}{2 D_p A} + \frac{W^2}{4 D_p \tau} + \frac{1}{2} \frac{W^2 \mu_c}{A \sigma_e L_e D_p} I_E, \quad (1)$$

where α_{cb} is the small-signal a-c collector-to-base current amplification factor, W is the effective junction spacing, s is the surface recombination velocity, A_s is the area over which surface recombination takes

* Equation (1) supplied by W. M. Webster. The factor 1/2 which appears in the third term does not appear in Reference (4), but is the result of a more rigorous derivation^{6,7} which states that $\sigma_e W / \sigma_e L_e$ should be multiplied by $(1 + p/n_0)$ rather than $(1 + Z)$. At high currents, $(1 + p/n_0)$ approaches $Z/2$.

place, D_p is the diffusion constant for holes in the base region, A is the area of the emitter, σ_e is the conductivity of the emitter region adjacent to the junction, L_e is the diffusion length for electrons in the emitter region, τ is the volume lifetime at high injection levels, μ_e is the electron mobility, and I_B is the d-c emitter current.

The first term in Equation (1) represents the effects of carrier loss due to recombination at the surface. As Moore and Pankove⁵ have shown, this loss occurs mainly in a circular region around the emitter having a width approximately equal to the junction spacing, W . The second term gives the loss of carriers in the base region due to volume recombination, and the third term gives the effect of injection efficiency.

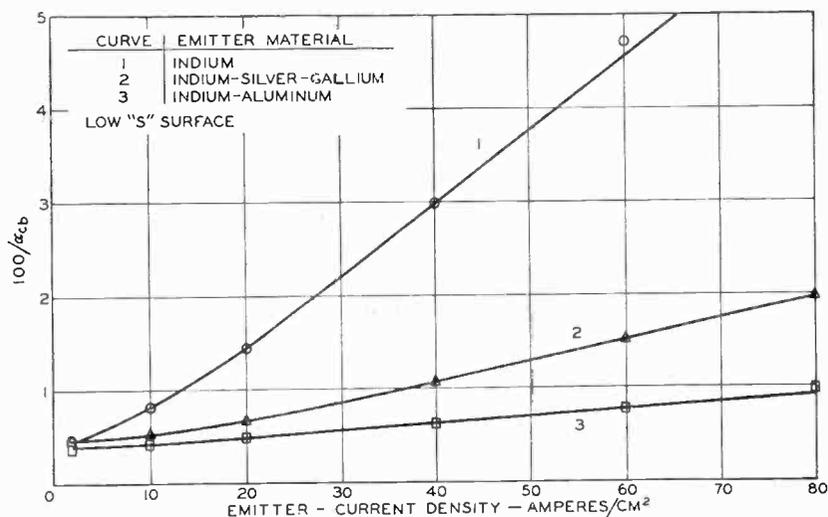


Fig. 4—Curves of $100/\alpha_{cb}$ versus emitter-current density for (1) pure indium, (2) gallium-indium, and (3) aluminum-indium emitters. The value of $\sigma_e L_e$ can be obtained from these curves.

Equation (1) indicates that the product $\sigma_e L_e$ can be computed from the slope of a curve showing $1/\alpha_{cb}$ as a function of I_B provided the emitter area and the junction spacing are known. The data shown in Figure 1 is replotted in Figure 4 to show $100/\alpha_{cb}$ as a function of emitter-current density. The curves in Figure 4 illustrate the difference between the units. The values of $\sigma_e L_e$ obtained from these curves were approximately 0.6 mho for pure indium, 2 mhos when gallium is added, and 6 mhos for aluminum alloys. The values for indium and gallium alloys are relatively consistent, having a spread of about two to one, but wider variations are observed in aluminum alloys. Although values of $\sigma_e L_e$ as high as 30 mhos have been observed, it is felt that 6 mhos is a fairly typical figure.

DISCUSSION

The use of these alloys permits a given p-n-p geometry to be used at much higher currents than previously considered practical. The alloys are very useful, therefore, in power transistors for audio output stages, switching service, and the like. Because high current amplification factor values are maintained to high current levels, high-level operation with good sensitivity and low distortion is possible. With these alloys, it is also possible to make good emitters on low-resistivity base layers, and thus to improve high-frequency devices. A less obvious advantage concerns p-n-p and n-p-n transistors which are symmetrical in electrical characteristics. If two devices are identical in all respects (dimensions, conductivities, etc.) except that one is n-p-n and the other p-n-p, current amplification factor will decrease less rapidly with increasing emitter current in the n-p-n unit by the square of the ratio of electron mobility to hole mobility (4 in germanium).⁴ It is desirable, therefore, to use a higher value of $\sigma_e L_e$ in the p-n-p transistor than in the n-p-n unit so that true complementary symmetry will be maintained at high currents. The flexibility introduced by the new p-type alloys permits closer matching of characteristics.

ACKNOWLEDGMENT

In conclusion, the authors would like to express their appreciation to W. M. Webster, both for valuable discussion and suggestions concerning the problems, and for editorial assistance in the composition of the paper.

RECENT ADVANCES IN POWER JUNCTION TRANSISTORS*

BY

B. N. SLADE

RCA Laboratories,
Princeton, N. J.

Summary—This paper is in two parts. Part I describes improved p-n-p alloy junction power transistors designed primarily for amplifier operation at dissipations of 5 watts per unit. Part II describes improved n-p-n power transistors having characteristics which are similar to those of the p-n-p type. Acceptor and donor elements such as gallium and arsenic, which have high segregation coefficients in germanium, are used in the emitter alloy to improve emitter injection efficiency. With improved emitter efficiency, collector-to-base current ratios (d-c alpha) between 30 and 50 at 1 ampere are consistently obtained on both n-p-n and p-n-p types. With close spacing between junctions, values on the order of 200 at 1 ampere have been achieved.

Modifications in the design of the p-n-p type have resulted in transistors with breakdown voltages greater than 100 volts and other characteristics desirable for use in high-voltage switching circuits such as television deflection.

PART I—P-N-P POWER TRANSISTORS

INTRODUCTION

JENNY AND ARMSTRONG¹ described the early development of p-n-p and n-p-n alloy junction power transistors designed for operation at one watt collector dissipation and at currents on the order of 100 milliamperes. This paper extends some of the power transistor design considerations discussed in that paper and discusses recent developments in p-n-p alloy junction power transistors. These experimental transistors have been developed primarily for use in class A and B power amplifiers at dissipations of approximately 5 watts and at currents of one or more amperes. Two of these transistors operated in class-B push-pull amplifier circuits are capable of 20 watts output.² Some variations in the design of p-n-p power transistors for use in high-voltage switching circuits such as are used for television horizontal deflection are described.

* Presented at the meeting of the I.R.E. Professional Group on Electron Devices, Washington, D. C., October 24-25, 1955.

¹ D. A. Jenny and L. D. Armstrong, "Behavior of Germanium-Junction Transistors at Elevated Temperatures and Power Transistor Design," *Proc. I.R.E.*, Vol. 41, p. 577, March, 1954.

² H. C. Lin, "20 Watt Transistor Audio Amplifier," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 536-546.

Four important factors determine the operating capabilities of the power transistor. The decrease in current amplification factor, α , at high emitter currents limits the useful operating current, while the junction breakdown voltage limits the useful operating range of voltages. The series base-lead resistance,³ $r_{bb'}$, through which the input signal is usually fed, determines how much driving power is wasted. Finally, the increase in junction temperature with increasing collector dissipation determines both the dissipation capabilities of the transistor and the ambient temperature range over which it may operate for a given dissipation.

The first three of these factors are, in many cases, interdependent and must be so considered in designing the transistor for a particular application. The fourth factor is affected by the mechanical construction of the device. The control and improvement of these factors are discussed, and the characteristics of power transistors designed for certain types of circuit applications are described.

CONSTRUCTION AND PROCESSING OF THE P-N-P POWER TRANSISTOR

The mechanical construction of the transistor is shown in Figure 1. The emitter and collector alloy dot diameters are .060 and .100 inch respectively except in the case of the symmetrical transistors in which both dots are .100 inch in diameter. The collector dot is pure indium and the emitter is a gallium-silver-indium alloy.⁴ Transistors with comparable performance have also been made with aluminum-indium emitters.⁴ The n-type germanium pellets range from .009 to .011 inch in thickness and 1.5 to 6 ohm-centimeters in resistivity depending upon the electrical characteristics desired. A nickel-iron alloy base tab with a hole 0.130 inch in diameter is soldered to the emitter side. The junctions are alloyed at 585°C. The collector is soldered to the copper stud with Cerroseal,* an alloy which softens at approximately 100°C and melts somewhat above this temperature. The low melting point enables a solder connection to be made to the copper stud without melting the collector dot, yet does not offer difficulty in operation.

The metallic parts of the assembly are masked with a vinyl lacquer

³ L. J. Giacoletto, "Equipments for Measurements of Junction Transistor Small-Signal Parameters for a Wide Range of Frequencies," *RCA Review*, Vol. 14, p. 269, June, 1953.

⁴ L. D. Armstrong, C. Carlson, and M. Bentivegna, "P-N-P Transistors using High Emitter-Efficiency Alloy Materials," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 144-152.

* Trademark of the Cerro de Pasco Corp., New York, N. Y.

during etching to prevent contact of the copper with the etching solution. The transistor is electrolytically etched in a solution of NaOH, encapsulated with a thin coating of a silicone compound, baked for several hours, and sealed.

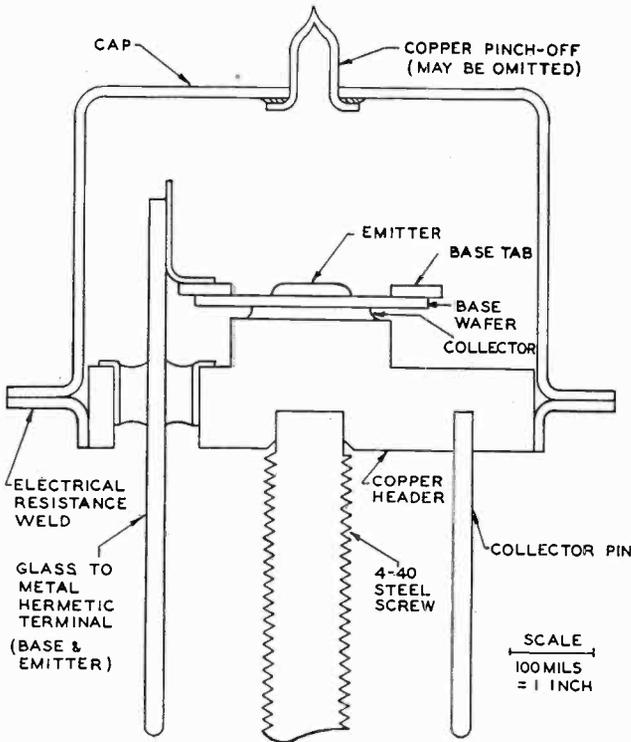


Fig. 1—Power transistor (cross section).

THERMAL DROP AND DISSIPATION

If the copper header closely contacts a good heat sink, the high thermal conductivity of the stud to which the collector dot is soldered limits the temperature rise at the junction. For the mechanical construction used in this transistor, a thermal drop between the collector junction and copper stud of approximately 3° to 5°C per watt is obtained.

To determine the maximum dissipation at which the transistor can operate, the maximum junction temperature the circuit can tolerate as well as the ambient temperature variations expected must be known. For example, if the circuit can tolerate the changes in transistor

characteristics occurring with junction temperatures as high as 85°C, then the transistor can be operated at a dissipation of approximately 15 watts at 25°C if a 4°C per watt thermal drop is assumed. If the equipment has to withstand ambient temperatures as high as 65°C, the transistor could operate at dissipations of up to 5 watts. Methods of determining the power ratings of transistors have been discussed elsewhere.²

VARIATION OF CURRENT AMPLIFICATION FACTOR WITH EMITTER CURRENT

Webster⁵ has shown that the fall-off in current amplification factor, α_{cb} , of a junction transistor at high currents is primarily due to a decrease in emitter efficiency, i.e., a decrease in ratio of hole current to electron current across the emitter junction. Throughout this paper the collector-to-base current ratio, called d-c alpha, rather than the small-signal α_{cb} , will be used as a criterion of high-current performance, but the behavior of the two is qualitatively the same; at high currents the magnitude of d-c alpha at a given current is approximately twice the value of α_{cb} .

Conventional alloy p-n-p transistors having pure indium for both the emitter and collector show a fall-off of d-c alpha which is usually great enough to limit useful operation of these transistors to emitter current densities less than 50 amperes per square centimeter. This d-c alpha fall-off results from a low emitter efficiency caused by a relatively low emitter conductivity.

Recent work⁴ has shown that higher emitter conductivities and reduced d-c alpha fall-off can be obtained by doping the emitter alloy metal with very small concentrations of acceptor impurities such as gallium or aluminum. The improvement in the high-current performance of power transistors using gallium doped emitters is shown in Figure 2 taken on p-n-p units of the construction herein described. The improvement in d-c alpha fall-off of the gallium-doped over the pure-indium emitter case can be seen more readily by the curves of reciprocal d-c alpha versus I_e in Figure 3. By use of the Webster analysis,⁵ it can be shown that the slope of this curve should be inversely proportional to the product of the emitter conductivity, σ_e , and diffusion length, L_e , of the majority carriers in the emitter region. The small slope of the curve for the transistor having the gallium-doped emitter reflects a higher emitter conductivity and efficiency over the pure indium case.

⁵ W. M. Webster, "On the Variation of Junction Transistor Current Amplification Factor with Emitter Current," *Proc. I.R.E.*, Vol. 41, p. 914, June, 1954.

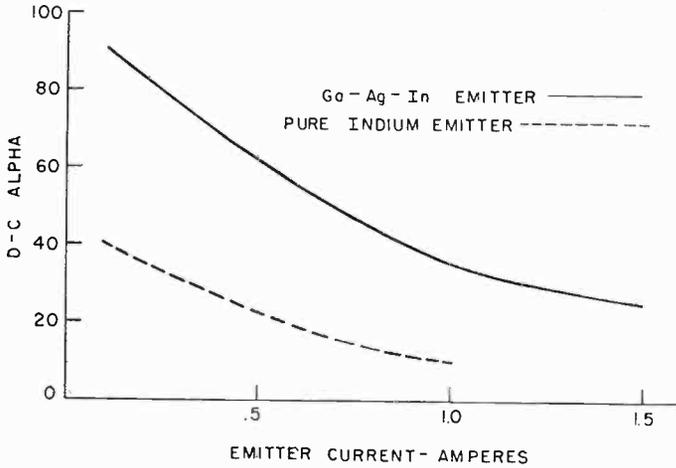


Fig. 2—D-C alpha versus emitter current for p-n-p power transistor.

The germanium resistivity of the base region also affects the alpha fall-off. This occurs because low resistivities result in lower emitter efficiencies at low emitter currents, while at high currents the emitter efficiency is independent of the base resistivity. The effect of germanium resistivity on alpha fall-off can be seen in Table I which shows average values of d-c alpha at 100 milliamperes and 1 ampere for different germanium resistivities. The ratio of d-c alpha values reflects the fall-off in alpha.

The spacing between junctions, W , also has a considerable effect on alpha. The average value of this spacing for the transistors tabulated in Table I is approximately 2.2 mils. By adjusting the spacing W to

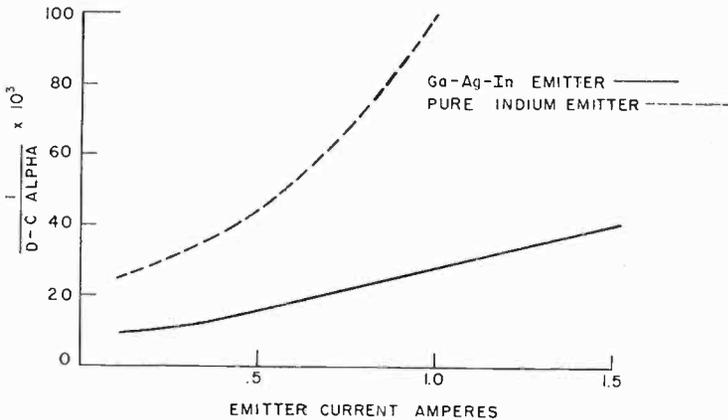


Fig. 3—Reciprocal of d-c alpha versus emitter current for p-n-p power transistor.

values on the order of 1.2 to 1.5 mils, transistors have been made with 1-ampere d-c alphas as high as 200. However, as is shown later, W cannot be decreased without also considering other design factors such as base-lead resistance and emitter-to-collector punch-through.

Table I—High Current D-C Alpha for P-N-P Power Transistors of Various Germanium Base Resistivities

Resistivity (ohm-cm)	100-milliampere d-c alpha	1-ampere d-c alpha	Ratio of 100-milli- ampere d-c alpha to 1-ampere d-c alpha
1-2	47	31	1.5
2-3	63	30	2.1
4-5.5	100	37	2.7

EFFECT OF PROCESSING ON ALPHA

The surface processing before and after encapsulation of the transistor affects d-c alpha and its fall-off. Freshly etched germanium transistor surfaces usually have a very low surface recombination velocity, s . However, this surface condition is generally quite unstable, particularly at high temperatures. During transistor shelf or operational life at elevated temperatures, the value of s usually increases, thus causing a decrease in the value of d-c alpha. Baking and aging schedules have proved to be quite useful in stabilizing s . Although the resultant d-c alpha values are lower than in the freshly etched case, they remain substantially constant during life.

In addition to stabilizing the transistor characteristics, the aging treatments also serve to reduce the fall-off of d-c alpha at high emitter currents. Surface recombination plays an important part in the determination of alpha at small emitter currents, but its importance decreases as the emitter current is increased. Therefore the increase in s during aging tends to flatten the alpha versus I_e curve by reducing the alpha at low currents to a greater extent than at high currents. Table II shows average values of d-c alphas at 100 milliamperes and 1 ampere before and after the baking as well as the ratio of these alphas for each case. The decrease in ratio indicates the decrease in the alpha fall-off as a result of the aging process.

Except in Table II all the d-c alphas which are quoted in this paper were measured after aging.

EFFECT OF EMITTER AREA ON D-C ALPHA

For a high value of emitter current, emitter efficiency can be increased not only by increasing the emitter conductivity, but also by

Table II—High Current D-C Alpha of P-N-P Power Transistors
Before and After Aging

	100-milliamperere d-c alpha	1-ampere d-c alpha	Ratio of 100-milli- ampere d-c alpha to 1-ampere d-c alpha
Before aging	116	39	3.0
After aging	63	28	2.2

increasing the emitter area. Moore and Pankove⁶ have shown that because of the effects of surface recombination, improved alphas can be achieved at low currents if the emitter is somewhat smaller than the collector. Thus most alloy junction transistors, particularly those designed to operate at low power levels, have been made with collectors having larger diameters than the emitters. However, at high injection levels the effect of surface recombination is less important than that of reduced emitter efficiency. Thus by increasing the emitter area to equal the collector area, an increase in high-current alpha can be obtained with a decrease of low-current alpha. This leads to a flatter curve of d-c alpha versus current, as can be seen in Table III. Transistors made with emitter and collector of 0.100-inch diameter are compared to transistors of 0.060-inch diameter emitter and 0.100-inch diameter collector. All transistors were made from the same germanium crystal so that the base material resistivity was approximately constant. It can be seen that the symmetrical geometry has superior high-current d-c alpha and smaller alpha fall-off.

Table III—High Current D-C Alpha of Symmetrical and Nonsymmetrical P-N-P Power Transistors

	Junction Spacing W (mils)	100-milli- ampere d-c alpha	1-ampere d-c alpha	Ratio of 100-milli- ampere d-c alpha to 1-ampere d-c alpha
Symmetrical case	1.7	166	125	1.3
	2.3	66	50	1.3
	2.4	60	40	1.5
	3.0	40	25	1.7
Nonsymmetrical case	1.7	200	60	3.3
	2.3	55	22	2.5
	2.4	35	17	2.0
	3.0	36	10	3.6

⁶ A. R. Moore and J. I. Pankove, "The Effect of Junction Shape and Surface Recombination on Transistor Current Gain," *Proc. I.R.E.*, Vol. 42, p. 907, June, 1954.

In addition to improving the high-current d-c alpha and alpha fall-off, the transistors having equal-area junctions show useful symmetrical properties. Switching applications frequently require the transistor to operate equally well in either direction, that is, with either junction operating as an emitter. Table IV gives some typical data indicating the degree of symmetry which can be obtained with these devices.

Table IV—Electrical Characteristics of Symmetrical P-N-P Power Transistors

Unit	Emitter to collector d-c alpha (1 ampere)	Collector to emitter d-c alpha (1 ampere)	Breakdown voltage		Emitter current with zero collector current ($E_c = -25$ volts)	Collector current with zero emitter current ($E_e = -25$ volts)
			Emitter junction	Collector junction		
a	45	50	70	70	14 μ a	15 μ a
b	50	66	60	65	45 μ a	50 μ a
c	20	25	90	100	12 μ a	13 μ a
d	40	40	50	55	18 μ a	18 μ a

COLLECTOR BREAKDOWN VOLTAGE

Surface leakage generally causes voltage breakdowns at values somewhat less than the theoretical "avalanche" breakdown,⁷ particularly in large-area junction devices. This surface breakdown occurs because of the presence of leakage across the junction either at the edge or underneath the alloy dot. Leakage may occur underneath the dots because of poor wetting of the indium alloy to germanium during the alloying process. These problems become increasingly difficult as the junction size is increased.

Careful surface processing before and after alloying has reduced the leakage to values which are quite reasonable for large-signal applications. Table V shows average values of reverse collector current (COI_c) with the emitter open circuited at 1 volt and 25 volts for several values of germanium resistivity. Saturation current values as calculated from the theory⁸ vary from 2 microamperes for 1 ohm-centimeter germanium to 10 microamperes for 5 ohm-centimeter. The readings tabulated here indicate that (1) some leakage is present even at very

⁷ S. L. Miller, "Avalanche Breakdown in Germanium," *Phys. Rev.*, Vol. 99, p. 1234, August 15, 1955.

⁸ W. M. Webster, "Saturation Current in Alloy Junctions," *Proc. I.R.E.*, Vol. 43, p. 277, March, 1955.

Table V—Reverse Collector Currents of P-N-P Power Transistors of Various Germanium Base Resistivities

Resistivity (ohm-cm)	Collector current with zero emitter current ($E_e = -1$ volt)	Collector current with zero emitter current ($E_e = -25$ volts)
2	11 μ a	31 μ a
2.5	12 μ a	35 μ a
4.8	18 μ a	35 μ a

low voltages since measured currents are somewhat greater than the theoretical values of saturation current, and (2) the collector leakage is relatively independent of the germanium resistivity.

Despite the presence of surface leakage, some relationship between germanium resistivity and breakdown voltages can be shown experimentally (Figure 4). Each point used in establishing this curve represents an average of 50 power transistors. Breakdown voltages greater than 200 volts have been obtained on power transistors made with 5 ohm-centimeter germanium.

The breakdown voltages plotted in Figure 4 were measured at 25°C. The effect of higher junction temperatures on breakdown voltage is shown in Figure 5. Although most of the curves are quite flat to temperatures of 75 to 80°C, some of them break rather sharply at temperatures as low as 50°C. The change in breakdown voltage as a function of temperature must be considered in determining the dissi-

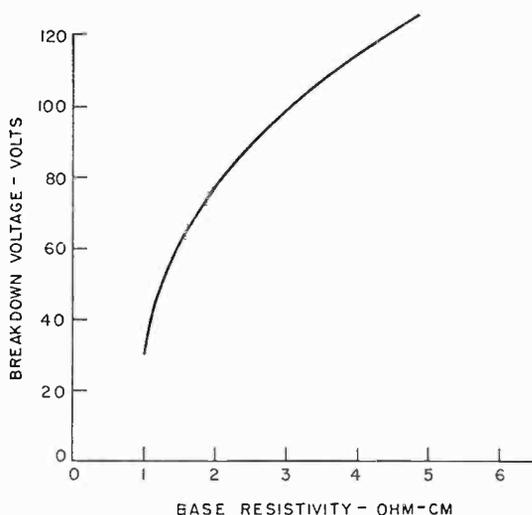


Fig. 4—Collector breakdown voltage versus base resistivity for p-n-p power transistor.

pation capabilities of the transistors and the ambient temperature range over which they can operate.

Emitter-to-collector "punch-through" also may limit the maximum useful voltage of the transistor. The width of the p-n junction depletion layer⁹ varies with the voltage across the layer and the resistivity of the germanium as follows:

$$\text{Width} = K \sqrt{V\rho_b}$$

If the germanium resistivity is sufficiently high and the junction spacing small enough, the collector depletion layer can touch the emitter

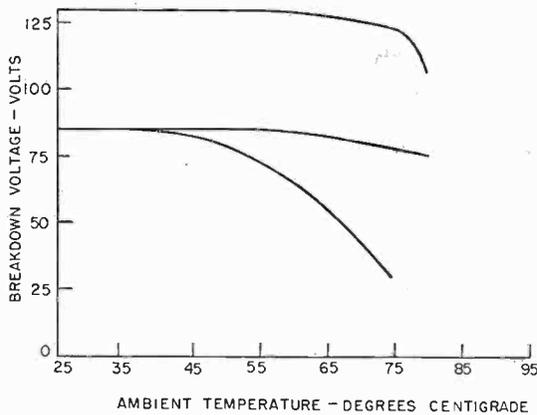


Fig. 5—Breakdown voltage versus ambient temperature for p-n-p power transistor.

junction causing punch-through. For example, if the junction spacing at zero voltage were 0.5 mil at the closest point, and if the base resistivity were 5 ohm-centimeter in resistivity, punch-through would occur at a value less than 50 volts. This is apparent from Table VI which shows some values of depletion layer widths as a function of germanium resistivity and voltage.

Curved or nonuniform junction fronts can aggravate this problem by causing punch-through even though the average W is relatively large. Emitter-to-collector punch-through is therefore a problem at close spacings and high resistivities, and may effectively limit the minimum spacing and the maximum alpha.

⁹ W. Shockley, "P-N Junctions in Semiconductors," *Bell Sys. Tech. Jour.*, Vol. 28, p. 449, July, 1949.

Table VI—Width of Junction Depletion Layer for Various Voltages and Germanium Base Resistivities

Resistivity (ohm-cm)	Voltage	Depletion layer width (mils)
1	10	.13
	30	.22
	50	.28
3.3	10	.23
	30	.40
	50	.52
5.0	10	.28
	30	.48
	50	.62

BASE-LEAD RESISTANCE

The series lead resistance, r_{bb} , is an important factor to be considered in power transistor design. Since the base is usually driven with a low-impedance source, the lead resistance should be as small as possible to minimize the waste of driving power. The series resistance of the germanium between the base ring and the edge of the collector junction, l , is

$$R_1 = \frac{\rho}{2\pi W} \log \frac{r_2}{r_1}$$

where r_2 is the radius of the base tab hole and r_1 is the radius of the collector junction. For the geometry shown in Figure 6 this component of r_{bb} is approximately 1.6 ohms per ohm-centimeter resistivity of the germanium. Most of r_{bb} lies between the junctions, l , and is dependent to a large extent upon junction spacing. Table VII shows average values of r_{bb} as a function of resistivity and junction spacing.

These values were measured at 1 milliampere emitter current. At high emitter currents the slope of the curves in Figure 7 of base voltage versus base current gives an approximate value of the base-lead resist-

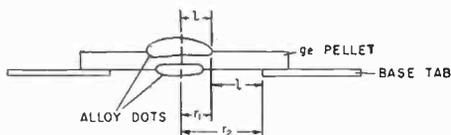


Fig. 6—Junction and pellet assembly of power transistor.

Table VII—Average Base-Lead Resistance of P-N-P Power Transistors for Various Values of Germanium Base Resistivities and Spacing Between Junctions

Germanium resistivity (ohm-cm)	W (mils)	Base-lead resistance (ohms)
4 - 5.5	1.5	95
	2.0	80
	2.5	65
2 - 3	1.5	85
	2.0	55
	2.5	38

ance. The values of $r_{bb'}$ for these transistors decrease from values of 45, 60, and 70 ohms at 1 milliampere to 9, 9, and 11 ohms respectively at 2 amperes. This decrease may be due in large part to the increase in base conductivity between the junctions at high injection levels, but it may also be affected by the electric fields in the base region formed by the electron and hole density gradients at high currents.

DESIGNS FOR SPECIFIC APPLICATIONS

The design considerations discussed in this paper are in many cases interdependent. A change in one factor may improve one characteristic while degrading another. However, the improvement in emitter efficiency through the use of the new alloys discussed earlier has resulted in an increased flexibility in the design of power transistors. Useful values of alpha can be obtained at high currents with wider junction

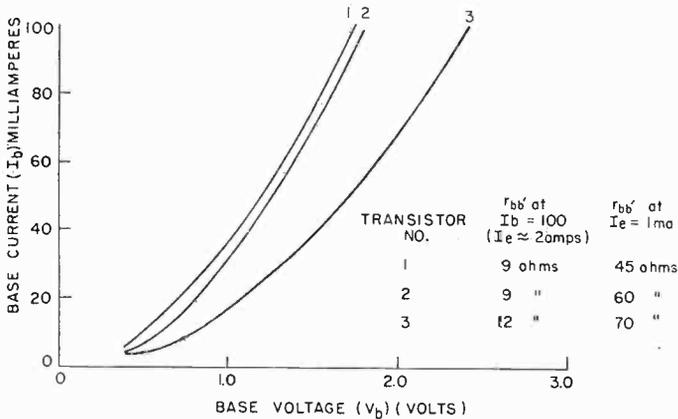


Fig. 7—Base voltage versus base current for p-n-p power transistor.

spacings, thus decreasing base lead resistance and punch-through problems. Furthermore, the current and voltage ranges have been extended.

The considerations described have been directed toward two general fields of applications, (1) class-B audio output circuits for both high- and low-voltage operation, and (2) high-voltage switching circuits such as are used for television horizontal deflection. Table VIII shows some typical characteristics of power transistors designed for these classifications. This is not intended to be a specification sheet. It only summarizes some of the design factors considered in this paper.

Table VIII—Some Electrical Characteristics of P-N-P Power Transistors Designed for Different Applications

	Class-B high-voltage	Class-B low-voltage	High-voltage switching
D-C alpha at 100 milliamperes	80	60	115
D-C alpha at 1 ampere	40	40	45
COI_c at -25 volts (microamperes)	25	25	25
Base-lead resistance (ohms)	45	20	80
Collector breakdown (volts)	65	35	120
Junction spacing (mils)	2.2	2.2	2.2

PART II — N-P-N POWER TRANSISTORS

INTRODUCTION

N-p-n transistors may be substituted in a circuit for p-n-p transistors of equivalent characteristics with only a reversal of the voltage polarities. However, their usefulness is very much greater than this simple substitution when used in conjunction with p-n-p transistors. Since the current and voltage polarities of the n-p-n are opposite to those of the p-n-p, circuits utilizing this complementary feature^{10,11} are possible. This is not possible in vacuum-tube circuitry since a similar complementary device is not available. Furthermore, it has been shown that circuitry employing this complementary principle, such as single-ended push-pull and direct-coupled cascaded amplifiers, makes possible a considerable reduction in the number of circuit components. Thus, much more is to be gained from having both n-p-n and p-n-p types available than the simple addition of one more transistor type.

In addition, the n-p-n power transistor is of interest on its own

¹⁰ A. Aronson, "Transistor Audio Amplifiers," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 515-535.

¹¹ G. Sziklai, "Symmetrical Properties of Transistors and their Applications," *Proc. I.R.E.*, Vol. 41, p. 717, June, 1953.

merits, for the greater mobility of electrons in germanium leads to a smaller decrease in α at high currents than occurs in a similar p-n-p transistor.

In this part the development of the n-p-n power transistor is discussed. This type has been developed primarily for amplifier operation at approximately 5 watts dissipation and may be used in complementary symmetry circuits with p-n-p power transistors having similar characteristics.

PROCESSING

The basic construction of the p-n-p transistor described in Part I is used with minor modifications for the n-p-n type. Alloy dots of 1 per cent arsenic by weight and 99 per cent lead are used for both emitter and collector. The junctions are alloyed for 30 minutes at 750°C and the transistor is etched in a high-current-density electrolytic etch using NaOH as the electrolyte. A lead-tin-indium alloy having a melting point of 295°C is used for soldering the collector to the copper stud. The junction assembly is encapsulated with a silicone grease before sealing.

DONOR ALLOYS FOR N-P-N POWER TRANSISTORS

The donor impurity alloy for the n-p-n transistor should fill three major requirements: (1) the emitter conductivity after alloying should be high in order to obtain a high emitter efficiency; (2) the germanium should be sufficiently soluble in the alloy at the alloying temperatures to assure adequate junction penetration; and (3) the alloy metal should be soft in order to minimize thermal strains between the alloy and the germanium. During the early development of the general purpose n-p-n junction transistor, Jenny¹² considered these requirements and chose a eutectic with the composition 10 per cent by weight of antimony and 90 per cent lead. This alloy is much more ductile than pure antimony. In the fabrication of power transistors, alloy ductility becomes even more important due to the complex wetting and regrowth problems encountered during the alloying of large-area junctions.

An alloy of arsenic and lead appears to be more desirable than lead-antimony in meeting the above requirements. The segregation coefficient of arsenic at the melting point of germanium is approximately 13 times that of antimony.¹³ Acceptor impurities having high segre-

¹² D. A. Jenny, "A Germanium N-P-N Alloy Junction Transistor," *Proc. I.R.E.*, Vol. 41, p. 1728, December, 1953.

¹³ J. A. Burton, "Impurity Centers in Ge and Si," *Proceedings of the International Conference on Semiconductors*, 1954, p. 845.

gation coefficients have resulted in an improvement in emitter conductivity of p-n-p transistors, and some improvement might be expected from the substitution of arsenic for antimony in n-p-n transistors. In contrast to indium, pure lead will dissolve very little germanium at the temperatures of alloying. The addition to the lead of a third element such as arsenic or antimony can have, therefore, a considerable effect on the solubility of germanium. A study of the arsenic-germanium and antimony-germanium phase diagrams (Figures 8 and 9) show that at 750°C for example, arsenic will dissolve approximately 60 per cent of germanium by weight, while antimony will dissolve approximately 32 per cent of germanium. Consequently, although phase diagrams are not available for the As-Pb-Ge and Sb-Pb-Ge ternary systems, it is not unreasonable to expect that a given percentage of arsenic in lead will dissolve more germanium than will an equal percentage of antimony in lead. Studies of penetration of As-Pb and Sb-Pb alloys in germanium at 750°C indicate that this greater As-Pb penetration actually occurs. With the greater solubility of the arsenic alloy, a smaller amount of arsenic than antimony is necessary to get the same junction penetration.

A consistent improvement in the appearance of the recrystallized germanium has been observed with decreasing amounts of arsenic in lead. Figure 10 shows photographs of the recrystallized area of n-p junctions made with 3 per cent and 1 per cent arsenic in lead and with 10 per cent antimony in lead. The alloys with the lowest arsenic content stand out as the most satisfactory, both with respect to the appearance of the recrystallized germanium and to measurements of reverse leakage currents of junctions made with these alloys. N-p-n transistors described in this paper were made with the 1 per cent arsenic and 99 per cent lead alloy which is a suitable compromise between good junction electrical characteristics and sufficient junction penetration.

ELECTRICAL CHARACTERISTICS OF THE N-P-N POWER TRANSISTORS

The improvement in the d-c alpha fall-off of the As-Pb over the Sb-Pb emitter can be seen from Figures 11 and 12 taken on units having the same base resistivities and spacing between junctions.

A comparison of the alpha fall-off of the As-Pb n-p-n transistor with the Ga-In p-n-p type discussed in Part I is made in Table IX, which shows average d-c alpha values for transistors of the same base resistivities (2.5 ohm-centimeters) and spacing between junctions (.0025 inch). The $\sigma_c L_e$ product for the As-Pb emitter calculated from the slope of the curve in Figure 12 is 2.2 mhos which is approximately equal to the 2.0-mho value reported for the Ga-In case.⁴ The lower d-c

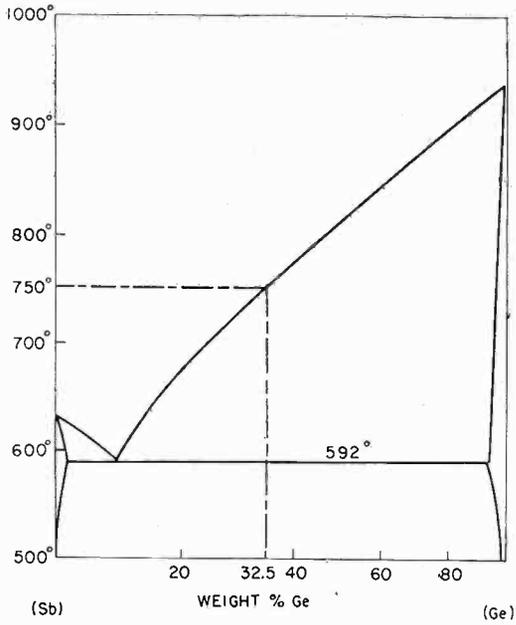


Fig. 8—The system antimony-germanium.

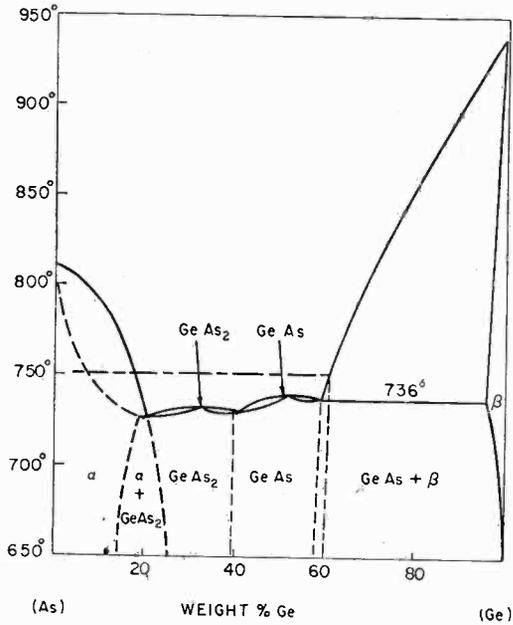
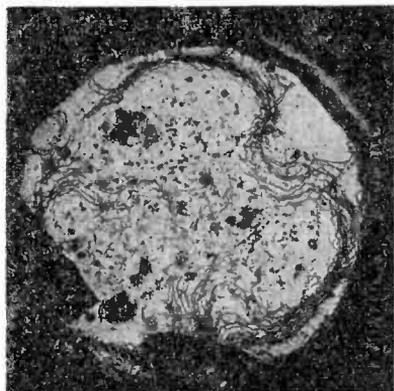
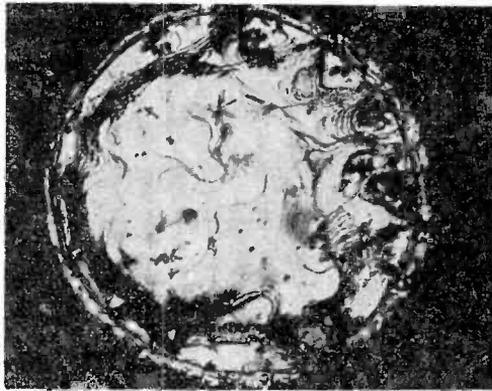


Fig. 9—The system arsenic-germanium.

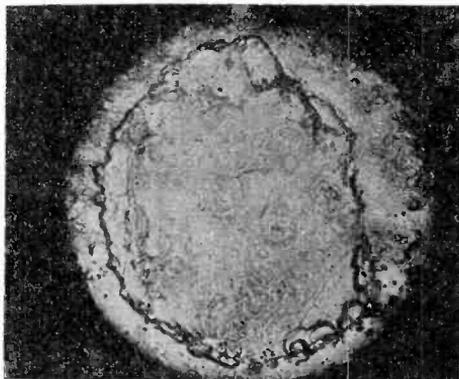
alpha fall-off of the n-p-n type shown in Table IX therefore confirms the expectations of theory,⁵ since the fall-off (measured on a reciprocal-alpha curve) should be inversely proportional to the square of the minority carrier mobility.



(1 per cent arsenic, 99 per cent lead)



(3 per cent arsenic, 97 per cent lead)



(10 per cent antimony, 90 per cent lead)

Fig. 10—Recrystallized n-type germanium of alloy n-p junctions.

D-c alphas of n-p-n transistors between 100 and 140 at 1 ampere have been obtained with spacings of the order of 2 mils and values as high as 200 have been obtained at smaller junction spacings.

Other electrical characteristics of the n-p-n transistor are shown in Table X. These values represent an average of approximately 100 transistors made with 2.5 ohm-centimeter germanium. Electrical characteristics of the p-n-p power transistors made with the same germanium resistivity are also shown for comparison.

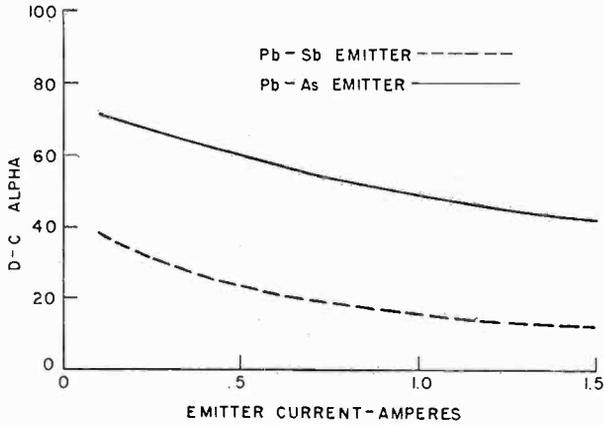


Fig. 11—D-C alpha versus emitter current for n-p-n power transistor.

Table IX—High Current D-C Alpha of N-P-N and P-N-P Power Transistors

	100-milliamperere d-c alpha	1-ampere d-c alpha	Ratio of 100-milli- ampere d-c alpha to 1-ampere d-c alpha
n-p-n	67	41	1.6
p-n-p	45	21	2.1

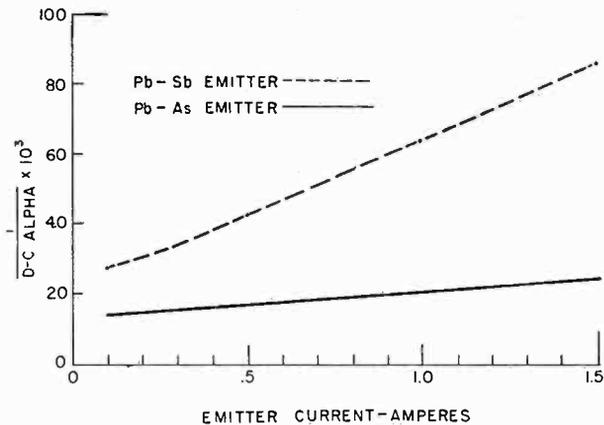


Fig. 12—Reciprocal of d-c alpha versus emitter current for n-p-n power transistor.

Table X—Other Electrical Characteristics of N-P-N and P-N-P Power Transistors

	Collector current with zero emitter current			
	$E_c = -1$ volt (microamperes)	$E_c = -25$ volts (microamperes)	Breakdown (volts)	$r_{bb'}$ (ohms)
n-p-n	19	45	55	50
p-n-p	12	35	60	40

A SILICON N-P-N JUNCTION TRANSISTOR BY THE ALLOY PROCESS*

BY

HERBERT NELSON

RCA Laboratories,
Princeton, N. J.

Summary—Silicon has semiconductor properties very similar to those of germanium and in theory it can be used very much like germanium for the preparation of various types of transistor devices. The wide use of germanium for these devices is partly a matter of convenience, since germanium is easier to work. A comparison of silicon characteristics with those of germanium shows that efforts to master silicon techniques may be amply rewarded through the advantages offered by silicon for high-temperature-device operation.

Silicon n-p-n alloy junction transistors have been prepared with satisfactory electrical performance, even with silicon crystal material which is not of the best quality. The attainable base-to-collector current amplification factor, α_{cb} , has been found to depend greatly upon the minority carrier lifetime of the silicon used. Average α_{cb} values of about 6 have been obtained with silicon having 2-microsecond lifetime; when improved silicon with 20-microsecond lifetime and treatment for lowering surface recombination velocity are used, α_{cb} values of over 100 have been obtained. In general, the performance at room temperature is similar to that of germanium alloy units. At elevated temperatures the low reverse current, which is under 10 microamperes at 150° C, make the silicon units highly advantageous.

INTRODUCTION

IT is appropriate to examine the published physical data on germanium and silicon, as of interest for transistor use. This data is assembled in Table I.

The higher energy gap of silicon leads to the presence in silicon of a far lower density of thermal carriers than in germanium, as shown in Table I. Since the uncontrollable fraction of the output current of a transistor is proportional to the thermal carrier density, it follows that this current is very much smaller in silicon than in germanium. This difference becomes particularly important at high operating temperatures. Calculations show, for instance, that at 150° C, saturation current density values of 1.6 amperes per square centimeter for germanium and 0.0007 ampere per square centimeter for silicon, each with room temperature resistivity of 2 ohm-centimeters in the tran-

* This paper is a revision of one presented at the Semiconductor Research Conference of the Institute of Radio Engineers, June, 1954, Minneapolis, Minn.

sistor range. One may predict, therefore, that at 150° C and at a total collector area of 1 square millimeter that 16 milliamperes of the collector current would be independent of emitter voltage when the base material is germanium, while only a few microamperes of the collector current need be uncontrollable when the base material is silicon. Although in practice the currents may be somewhat higher, the ratio is still significant.

In view of the above, silicon may be preferable to germanium in transistors designed for applications where high operating temperature

Table I—Semiconductor Characteristics of Germanium and Silicon

Characteristic	Germanium	Silicon	Reference
Energy Gap	0.75-0.0001 T volts	1.12-0.0003 T volts	J. Bardeen and W. Shockley, <i>Phys. Rev.</i> , 80, 72, 1950
Density of thermal carriers at 300°K	$2.5 \times 10^{13}/\text{cm}^3$	$6.8 \times 10^{10}/\text{cm}^3$	E. M. Conwell, <i>Proc. I.R.E.</i> , 40, 1329, 1952
Intrinsic resistivity at 300°K	47 ohm-cm	63,600 ohm-cm	E. M. Conwell, <i>Proc. I.R.E.</i> , 40, 1329, 1952
Donor ionization energy	0.01 volt	0.05 volt	G. L. Pearson and J. Bardeen, <i>Phys. Rev.</i> , 75, 865-883, (1949)
Acceptor ionization energy	0.01 volt	0.046 volt	G. L. Pearson and J. Bardeen, <i>Phys. Rev.</i> , 75, 865-883 (1949) E. Burstein et al <i>J. Phys. Chem.</i> , 57, 849, 1953
Electron drift mobility at 300°K for 10 ohm-cm material	3900 ± 100 cm ² /volt sec	1200 ± 100 cm ² /volt sec	M. B. Prince, <i>Phys. Rev.</i> , 92, 681, 1953 and <i>Phys. Rev.</i> , 93, 1204, 1953
Hole drift mobility at 300°K for 10 ohm-cm material	1900 ± 50 cm ² /volt sec	500 ± 50 cm ² /volt sec	M. B. Prince, <i>Phys. Rev.</i> , 92, 681, 1953 and <i>Phys. Rev.</i> , 93, 1204, 1953

and low uncontrolled collector current are important considerations. In some of these applications high-frequency response may not be essential so that the lower mobilities of holes and electrons in silicon are not a serious disadvantage.

As a first step to investigate the advantages offered by silicon for high-temperature operation, a low-power, audio-frequency silicon tran-

sistor has been developed and is the subject of the present paper. Since the work described in this paper was completed, commercial silicon transistors made by the *grown-junction* techniques have been announced. The present paper describes *alloy-junction* transistors which have advantages over grown units in reduction of electrode load resistances and in operation at high injection levels.

METHOD OF FABRICATION

General methods involved in the preparation of alloy transistors have been described¹ and here only the variations in technique necessitated by the use of silicon in place of germanium will be described. P-type silicon of 2.0-4.0 ohm-centimeters resistivity, grown as a single crystal, generally in the [111] direction, is used. Transistor theory indicates that material of high lifetime, preferably greater than 20 microseconds, is necessary for good transistor results. However, much of the work in the early stages was done with silicon of lower quality.

Alloying Materials

The dot materials used to form the junctions were selected after extensive tests in which the metallurgical as well as the electrical nature of alloy junctions in silicon were investigated. A variety of n-type impurity elements and dilutants were tested. As in the case of the n-p-n germanium transistors,² the use of dilutants was found to minimize the introduction of thermal mismatch strains in the neighborhood of the junction. Also, the use of suitable dilutants in the dot material was found valuable in promoting wetting and liquefaction of the silicon during the alloying process. When pure antimony or lead-arsenic alloy is used in forming a junction, wetting is difficult to attain and intersolubility of the elements is very low even at 900° C. When gold is added, wetting and solubility are good even at temperatures as low as 500° C.

As an outcome of these tests, an alloy consisting of 54 per cent by weight of lead, 45.4 per cent of gold, and 0.6 per cent of arsenic was chosen for dot material. Lead was included because it favors spreading of the dot during the alloying process. Its presence also lowers the melting point of the dot material and consequently leads to a more strain-free junction. The gold and arsenic of the dot material are both

¹ R. R. Law, C. W. Mueller, J. I. Pankove and L. R. Armstrong, "A Developmental p-n-p Junction Transistor," *Proc. I.R.E.*, Vol. 40, pp. 1352-1357, November, 1952.

² For detailed discussion of this point see D. A. Jenny, "A Germanium n-p-n Alloy Junction Transistor," *Proc. I.R.E.*, Vol. 41, pp. 1728-1734, December, 1953.

donors. The gold, however, produces donor levels only 0.33 volt above the occupied band in silicon.³ It therefore does not contribute to the formation of an n-type region in the p-type silicon. The arsenic, on the other hand, produces donor levels 0.05 volt below the bottom of the conduction band and converts the recrystallized silicon to high-conductivity n-type material. Though more gold than arsenic is present in the dot, the segregation coefficient of gold in silicon is very small and consequently only a very small percentage of the gold is caught in the recrystallizing silicon.

The lead-gold-arsenic alloy is brittle and consequently difficult to form into disk-shaped dots. Composite dots are consequently used. They are made by superimposing disks of gold upon disks of lead-arsenic alloy. Dots of this type can readily be punched out from a strip of gold and a strip of lead-arsenic alloy bonded together. A gold thickness of 5 mils, a lead-arsenic alloy thickness of 10 mils and a disk diameter of 40 mils are used.

A third dot is used for securing ohmic contact to the base wafer. This dot is punched from a strip consisting of a 2 mil thick core of S No. 46 metal on whose major sides .4-mil-thick layers of lead and tin and .2-mil-thick layers of indium have been deposited by electroplating. The S No. 46 metal is an iron-nickel alloy containing 46 per cent nickel. The thermal coefficient of expansion of this alloy nearly matches that of silicon.

Processing

The silicon base wafer used for the transistor is ground and then etched to approximately 3 mils. A suitable etching solution uses nitric, hydrofluoric, and glacial acetic acid, with bromine.

Before the dots are applied to the silicon surface for alloying, they are rinsed in hydrofluoric acid. This rinse leads to the formation of a fluorine-containing salt film on the dot surface which serves as a flux in promoting effective wetting during the alloying process. The three dots are alloyed onto the silicon in one firing step. A graphite boat and graphite washers are used as often employed in making p-n-p germanium transistors. The base wafer and the dots are assembled as shown in Figure 1. After loading, the assembly jig is fired in a dry mixture of hydrogen and nitrogen for a few minutes at 700° C. After cooling at a rate of 250° C per minute, the transistor is etched in a hot sodium hydroxide solution.

The final etch was originally chosen on the basis of a series of tests in which the reverse characteristics of the junction and the recom-

³E. A. Taft and F. H. Horn, "Gold as a Donor in Silicon," *Phys. Rev.*, Vol. 93, pp. 64 (1954).

bination velocity of the minority carriers in the surface adjacent to the junction were measured after various surface treatments. The evaluation of the surface recombination rate was comparative rather than absolute. An effective lifetime determined by bulk as well as surface recombination was measured by a pulse injection technique.⁴ Changes observed in this effective lifetime due to surface treatments of the diode were then attributed to variations in the rate of surface recombination. Results obtained indicated that the surface recombination velocity characteristic of the diodes tested was generally high and also that it was greatly affected by surface treatments. The effective lifetime value for a particular diode was found to increase as a result of the sodium hydroxide etch referred to above. Increases in effective lifetime was also found to result from the conventional etching

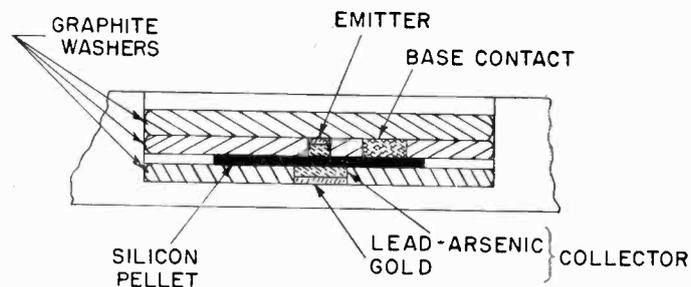


Fig. 1—Silicon transistor parts assembled for alloying.

techniques used with germanium but these increases were lower in magnitude. Rough, absolute measurements of the surface recombination velocity made on special samples of silicon showed qualitative agreement with the diode measurements. As described in another paper,⁵ etching techniques alone do not lower surface recombination in silicon to the values easily obtainable in germanium.

After treatment to reduce surface recombination, the silicon transistors were mounted on metal bases, coated with a silicone resin, and hermetically sealed into metal cans.

ELECTRICAL CHARACTERISTICS

In course of the experimental work with silicon transistors, many units have been tested for electrical characteristics and performance. The best ones have exhibited electrical performance roughly comparable with that of germanium p-n-p alloy units; for these, collector-to-base

⁴S. R. Lederhandler and L. J. Giacoletto, "Measurement of Minority Carrier Lifetime and Surface Effects in Junction Devices," *Proc. I.R.E.*, Vol. 43, pp. 477-483, April, 1955.

current amplification factors (α_{cb}) of 110 and power gains of 40 decibels have been obtained. However, values as low as 6 and 30 decibels, respectively, result when silicon of short lifetime is used. Data presented below represents transistors made without special treatment for reducing surface recombination velocity. The improvement in performance attained through this treatment is described in another paper⁵ to which reference should be made.

Measurements have been made at room temperature and also in a range from 30° to 150° C. The upper limit is not imposed by the silicon but by solder used in mounting and encapsulation. The measurements have been of three types. The first type has been concerned with the determination of hybrid π -equivalent common-emitter circuit parameters;⁶ the second, with the determination of static characteristics; and the third type, with the evaluation of small-signal operating characteristics such as power gain, current gain, and noise factor.

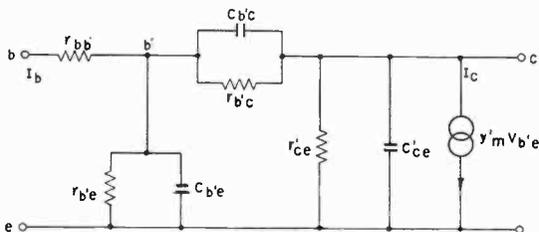


Fig. 2—Single-generator base-input hybrid π -equivalent circuit.

ROOM-TEMPERATURE MEASUREMENTS

The range of values of hybrid π -equivalent common-emitter circuit parameters (Figure 2) measured on a group of transistors is shown in Table II.

For comparative purposes, a range of values for p-n-p germanium transistor is included.¹

The range of values of base-to-collector current amplification factor, α_{cb} , power gain, and noise factor measured for the same group of transistors is shown in Table III.

Room temperature static output characteristics of one of the transistors are shown by the curves labeled A on Figure 3.

The data presented in the above tables and curves highlights the

⁵ A. R. Moore and H. Nelson, "Surface Treatment of Silicon for Low Recombination Velocity," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 136-143.

⁶ C. W. Mueller and J. I. Pankove, "A p-n-p Triode Alloy-Junction Transistor for Radio-Frequency Amplification," *Proc. I.R.E.*, Vol. 42, pp. 387-391, February, 1953.

great similarity of germanium and silicon transistors with regard to room-temperature performance. The tendency towards lower current amplification factor, α_{cb} , of the silicon units is discussed elsewhere.⁵ Because the base-layer thickness was a little smaller for the silicon units, no material inferiority in frequency response is observed with silicon in this comparison.

MEASUREMENTS AT ELEVATED TEMPERATURES

The curves labeled B in Figure 3 show static output characteristics at 150° C of the same silicon transistors for which the curves labeled A show room-temperature results. These characteristics, contrary to findings with germanium, show an increase in α_{cb} with temperature. They also show evidence of the very low reverse current expected with silicon in this comparison.

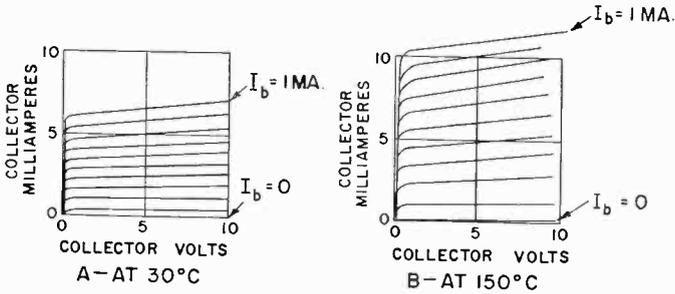


Fig. 3—Output characteristics of a silicon transistor (base current I_b in ten equal steps from 0 to 1 milliampere).

Table II—Hybrid π -Equivalent Circuit Common-Emitter Parameters of Transistors Measured at $E_o = +6$ Volts, $I_B = 1$ milliampere.

Parameter	Range of Values	
	Silicon Transistor	Early p-n-p Germanium Transistor
r_{bb} , ohms	100-300	200-500
$r_{b'e}$ ohms	100-300	600-2000
$C_{b'e}$ $\mu\mu\text{f}$	3200-9000	10,000-25,000
$r_{b'o}$ megohms	0.3-1.2	2
$C_{b'o}$ $\mu\mu\text{f}$	74-129	35
W mils*	0.7-1.4	1-3

* W is the base-layer thickness and is calculated from $C_{b'e} = \frac{q}{kT} \frac{W^2 I_B}{2D}$.

Table III—Gain and Noise-Factor Values of Silicon Transistor without Special Surface Treatment

	Range of Values	
	Silicon Transistor	Early p-n-p Germanium Transistor
4 kc Current amplification factor (α_{cb})	4.0-13.0	20-100
4 kc Power gain db	23.4-34.0	35-48
300 kc Power gain	9.2-14.2	0-10
1 kc Noise factor	14-29	15-25

All values of gain measured at $V_c = 6v$, $I_E = I$ ma. Power gains measured with resistive input and conjugate matched output. Noise factor measured at $E_c = -Iv$, $I_E = I$ ma.

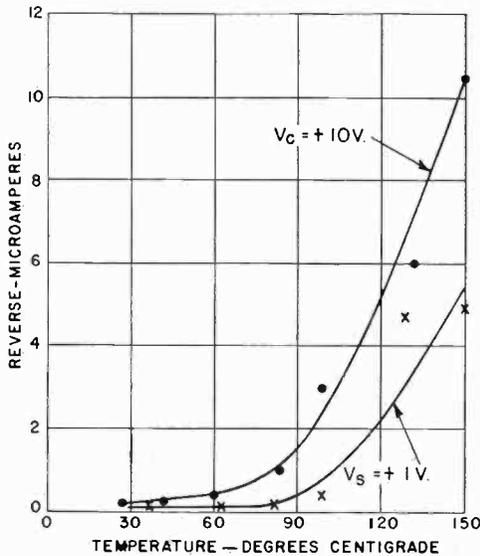


Fig. 4—Reverse current as a function of temperature.

silicon. A plot of this current to the 1-millimeter diameter collector dot is shown in Figure 4. The reverse current apparently contains a surface leakage component as well as the saturation current, yet at 150° C it is far less than the reverse current obtained with germanium at the same temperature.

The base-emitter bias required for a 1-milliamper emitter current in a silicon transistor is shown as a function of temperature by curve

A of Figure 5. Curve B shows comparative data for a p-n-p germanium transistor. That a larger bias is required for the silicon transistor follows directly from a consideration of the diode equation

$$I = I_s \left[\exp \frac{eV}{kT} - 1 \right].$$

The value of the saturation current, I_s , is about six orders of magnitude less for silicon than for germanium and consequently a larger bias voltage is required with the silicon transistor for the same forward current flow. The voltage V across the junction required for any particular forward current flow decreases with temperature even

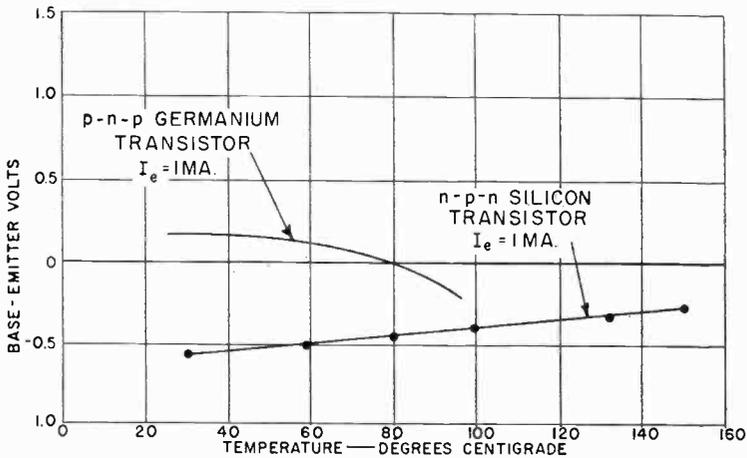


Fig. 5—Base-emitter bias for constant emitter current as a function of temperature.

though the above expression indicates the opposite trend. This is true since I_s increases more rapidly with temperature than the term $\exp eV/kT$ decreases. The greater decrease in base-emitter bias with temperature noted for germanium is, however, largely due to an increase in saturation current flow through r_{bb} . The voltage drop across r_{bb} is in a direction to increase the forward voltage at the p-n junction and calls for a component of bucking voltage in the base-emitter bias.

The variation of power gain of a silicon transistor with temperature is shown in Figure 6. Curve A shows the power gain with input and output impedance set for maximum power gain at 30° C. For comparison, the maximum available power gain for a germanium transistor is shown by curve B in the same figure. The measurements were made at $V_C = 6$ volts and $I_E = 1$ milliamperes.

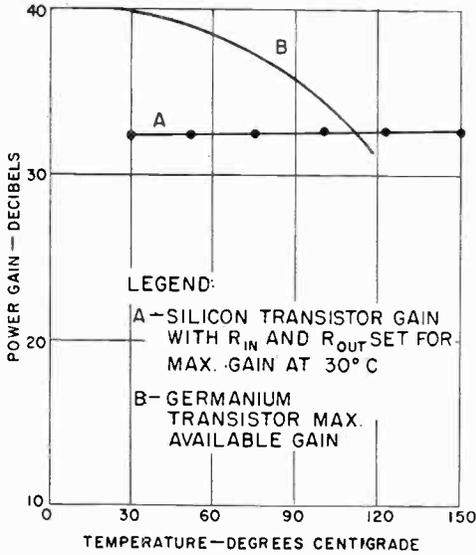


Fig. 6—Power gain versus temperature.

CONCLUSIONS

Silicon techniques have been mastered to the extent that transistors capable of as much as 30-40 decibels gain are achieved. The performance of these transistors at elevated temperatures has been found to accord with expectations.

JUNCTION TRANSISTOR DESIGNED FOR ITERATIVE OPERATION AT LOW FREQUENCIES

BY

L. J. GIACOLETTO

RCA Laboratories,
Princeton, N. J.

Summary—A junction transistor design is considered whose input and output resistances are equal when the transistor is image-matched at low frequencies for maximum power amplification. These transistors can then be iterated in cascade without loss in power amplification. The design considerations determine the transistor base width to achieve the equal input and output resistances. A base width of several microns is indicated as a typical result. Transistors with this base width have been built and verify the design relations.

INTRODUCTION

WHEN the low-frequency operating characteristics of a number of transistors are measured, it is generally found that if the transistors are image-matched for maximum power amplification, the units with high input resistances will have low output resistances and vice versa. With the aid of suitable design equations, this relationship can be verified and attributed to variations in the basic widths of the various transistors. Further, the design equations indicate the value of the base width for achieving equal input and output resistances; in this event, transistors can be coupled together without matching networks so as to effect circuit simplification without loss in power amplification. The design considerations in this article are applicable provided operating frequencies are low enough that transistor reactive effects can be neglected.

ANALYSIS

Consider a common-emitter circuit arrangement image-matched for maximum power amplification. For low-frequency operation the input, g_{in} , and output, g_{out} , are pure conductances related to the input and output self conductances, g_{bb} and g_{cc} respectively, by a common factor containing the forward and reverse current amplification factors, α_{cb} and α_{bc} respectively. Thus¹

¹L. J. Giacoletto, "Terminology and Equations for Linear Active Four-Terminal Networks Including Transistors," *RCA Review*, Vol. 14, pp. 28-46, March, 1953.

$$g_{in} = g_{bb} \sqrt{1 - \alpha_{cb} \alpha_{bc}}, \quad (1)$$

$$g_{out} = g_{cc} \sqrt{1 - \alpha_{cb} \alpha_{bc}}. \quad (2)$$

Therefore, in order that the input and output conductances be equal, it is necessary that $g_{bb} = g_{cc}$.

The hybrid- π equivalent circuit of the junction transistor is shown in Figure 1.² At low-frequencies, the capacitors shown in this equivalent circuit can be neglected, so that

$$g_{bb} = \frac{g_{b'e} + g_{b'c}}{1 + r_{bb'} (g_{b'e} + g_{b'c})}, \quad (3)$$

$$g_{cc} = g_{ce} + g_{b'c} + \frac{r_{bb'} g_{b'c} (g_m - g_{b'c})}{1 + r_{bb'} (g_{b'e} + g_{b'c})}. \quad (4)$$

In these equations the collector leakage conductance, g_b (see Figure 1) is assumed lumped in with $g_{b'c}$. Accordingly, the input and output conductances will be equal if

$$g_{b'e} = g_{ce} [1 + r_{bb'} (g_{b'e} + g_{b'c})] + r_{bb'} g_{b'c} (g_m + g_{b'e}). \quad (5)$$

If desired, the expressions for the transistor parameters as given in Reference (2) can be employed to obtain an equation in terms of material constants and dimensions of the transistor. The resulting equation is not simple enough to be useful. For most design purposes, suitable approximations can be made as follows. $g_{b'c}$ will generally be small enough in comparison with $g_{b'e}$ to be neglected. The same is true of $g_{b'e}$ in comparison with g_m . Transistors that have the desired equality of input and output conductances will usually have a small enough value of $g_{b'e}$ that the product, $r_{bb'} g_{b'e}$, will be small in comparison with unity. Also, for good quality transistors, the triple product, $r_{bb'} g_{b'c} g_m$, will be small in comparison with g_{ce} . In any event, the inequalities enumerated can be examined after the approximate design is carried out provided that $r_{bb'}$ and the leakage conductance component of $g_{b'c}$ can be estimated. As a result of the approximations mentioned above, the condition for equality of input and output conductance, Equation (5), reduces simply to

² L. J. Giacoletto, "Study of p-n-p Alloy Junction Transistor from D-C Through Medium Frequencies," *RCA Review*, Vol. 15, pp. 506-562, December, 1954. This paper should be consulted for the definition and derivation of the various transistor parameters that are used herein.

$$g_{b'e} = g_{ce}. \quad (6)$$

The expressions for these two parameters in terms of the transistor constants and dimensions are²

$$g_{b'e} = -\Lambda(I_B - I_{BS}) \approx \frac{-\Lambda I_E}{\alpha_{CB}} \approx \Lambda I_E \frac{A_e}{A_b} \frac{\sigma_b W_b}{[\sigma_e L_e]_e}, \quad (7)$$

$$g_{ce} \approx I_E \frac{1}{W_b} \frac{\partial W_b}{\partial V_{CB'}}. \quad (8)$$

The quantities in these equations have the following meaning: $\Lambda = q/kT = 39$ at room temperature; I_B and I_E are d-c base and collector currents respectively; I_{BS} = base saturation current, generally small in com-

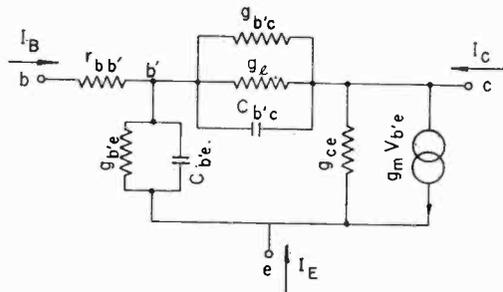


Fig. 1—Hybrid- π common-emitter equivalent circuit.

parison with I_B ; α_{CB} = negative ratio of d-c collector current to d-c base current; A_e/A_b = a ratio generally approximately unity (see Reference (2), p. 513 for details); σ_b , $[\sigma_e]_e$ = base and effective emitter semiconductor conductivities respectively; $[L_e]_e$ = effective diffusion length of minority carriers in the emitter semiconductor; and W_b = base width. According to the above equations, $g_{b'e}$ will be linearly proportional to W_b and g_{ce} will be inversely proportional to W_b . For some value of W_b the two will be equal. The appropriate value of W_b is given by the equation below.

$$W_b^2 = \frac{1}{\Lambda} \frac{A_b}{A_e} \frac{[\sigma_e L_e]_e}{\sigma_b} \frac{\partial W_b}{\partial V_{CB'}}. \quad (9)$$

This equation indicates that the approximate value of W_b will be inde-

pendent of d-c operating current, but will depend upon the d-c operating voltage through $\frac{\partial W_b}{\partial V_{CB'}}$,

$$\begin{aligned} \frac{\partial W_b}{\partial V_{CB'}} &= \frac{1}{2} \left[\frac{2 K_e \epsilon_0 \mu}{\sigma_b V_{CB'}} \right]^{\frac{1}{2}} = \frac{5.05 \times 10^{-6}}{\sqrt{\sigma_b |V_{CB'}|}} \text{ for p-n-p transistor} \\ &= \frac{7.35 \times 10^{-6}}{\sqrt{\sigma_b |V_{CB'}|}} \text{ for n-p-n transistor, (10)} \end{aligned}$$

and because W_b decreases as $|V_{CB'}|$ is increased. The evaluated expressions given in Equation (10) are in meters per volt when σ_b is in mhos per meter and $V_{CB'}$ is in volts.

NUMERICAL CALCULATIONS

Suitable numbers for substitution into Equation (9) are readily available with the possible exception of $[A_b/A_e] [\sigma_e L_e]_e$. Generally, A_b/A_e will be but slightly larger than unity. Thus, for one type of p-n-p transistors,³ $A_b/A_e = 1.4$ has been determined.² In the absence of definite information, it is permissible to assume that this ratio is unity.

$[\sigma_e]_e$ and $[L_e]_e$ as employed above are not the actual quantities associated with the emitter semiconductor, but rather can be considered effective values to take into account surface recombination occurring near the emitter. The relationship between the effective value, $[\sigma_e L_e]_e$ as employed herein and actual $\sigma_e L_e$ for the emitter semiconductor is given by W. M. Webster.⁴ With this method of calculation, A_b/A_e need not be determined separately.

Thus, if I_E is small and if volume recombination is negligible (this should generally be true since W_b is small), then

$$\frac{A_b}{A_e} [\sigma_e L_e]_e = \frac{\sigma_e L_e}{1 + \frac{2 s W_b \sigma_e L_e}{D R_E \sigma_b}}, \quad (11)$$

³ R. R. Law, C. W. Mueller, J. I. Pankove, and L. Armstrong, "A Developmental Germanium p-n-p Junction Transistor," *Proc. I.R.E.*, Vol. 40, pp. 1352-1357, November, 1952.

⁴ W. M. Webster, "On the Variation of Junction-Transistor Current-Amplification Factor with Emitter Current," *Proc. I.R.E.*, Vol. 42, pp. 914-920, June, 1954.

where s = surface recombination velocity, D = diffusion constant of minority carriers, and R_E = radius of emitter.

The results of design calculations for a p-n-p transistor are shown in Figure 2. The curves are plotted in accordance with Equations (7) and (8). The transistor constants and operating conditions employed are given in Figure 2 and are obtained from measured results.² Calculations of g_{bb} and g_{cc} using Equations (3) and (4) indicate that these do not differ appreciably from $g_{b'e}$ and g_{ce} respectively. The intersection point, $W_b = 4.9$ microns (0.19 mils), as shown in Figure 2 can also be obtained with the aid of Equation (9). A transistor with this base

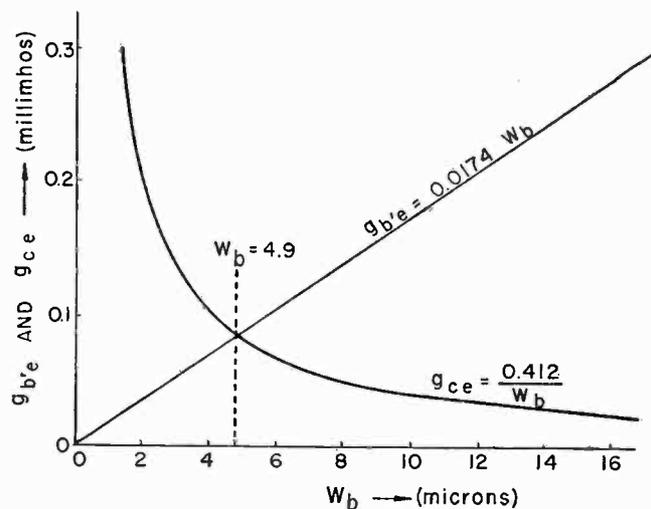


Fig. 2—Variation of $g_{b'e}$ and g_{ce} with base width. Auxiliary data: $T = 25^\circ\text{C}$; $\Lambda = 39$; $I_E = 1$ ma; $V_{CE} = -6$ v; $\sigma_b = 25$ mhos/meter; $r_{bb'} = 250$ ohms; $g_m = 0.0389$ mhos; $g_i = 0.2 \mu$ mhos; $L_b = 470$ microns; $[\sigma_e L_e]_e = 0.04$ mhos; and $A_b/A_e = 1.4$.

width would have $g_{b'e} = g_{ce} = 0.085 \times 10^{-3}$ mhos (11,700 ohms). The d-c collector-to-base current amplification factor would be -460 . The maximum power amplification would be very nearly one quarter of the square of this current amplification factor, or 47 decibels. The base width, $W_b = 4.9$ microns, is the operating width at the V_{CE} collector bias. The base width without applied terminal voltages, W_{b0} , will be larger than the indicated value of W_b by the thickness of the collector junction, $W_c = 2|V_{CE}| \partial W_b / \partial V_{CE}$. For a collector bias, $V_{CE} = -6$ volts, $W_c = 5.0$ microns, so that $W_{b0} = 9.9$ microns.

The transistor design just considered required a rather small base width for equal input and output conductances. Several combined factors either increase the base width or simplify the construction of transistors of the type described herein. First, smaller conductivity base material can be employed to increase W_b . In accordance with Equations (9) and (10), W_b will vary inversely with $\sigma_b^{\frac{1}{2}}$. Second, new alloy materials⁵ provide means for increasing the $[\sigma_e L_e]_e$ factor in Equation (9). Third, the $[\sigma_e L_e]_e$ factor will increase with improvements in etching procedures (lower surface recombination velocities as noted in Equation (11)) and with more favorable emitter geometries.⁶ Fourth, new metallurgical techniques⁷ make transistors with small base widths more feasible than has been the case in the past.

For directly coupled circuits where transformers and chokes are not feasible, the presence of bias source resistances must be included in the preceding design considerations. The conductances associated with the base and collector bias sources can be added to g_{bb} and g_{cc} respectively of Equations (3) and (4), and the design proceed as before. Generally, the conductance of the base bias source will be small in comparison with g_{bb} and can be neglected. The same situation will not usually be true of the collector bias source. In the approximate analysis, the conductance of the collector bias source will be a constant additive term to g_{cc} (Figure 2) so that the base width for equal input and output conductances will occur at a larger base width than formerly. Due to the loss of power to the bias source conductances, the maximum power amplification of the composite circuit will be less than that of the transistor itself.

EXPERIMENTAL RESULTS

Junction transistors were made which verified (very approximately) the design relations. For comparison purposes, measured and computed data on such a p-n-p transistor, with approximately equal input and output conductances, is given below. This transistor was made with the aid of alloy materials described in Reference (5).

Semiconductor: Germanium, n-type, $\sigma_b = 25$ mhos/meter, estimated.

⁵ L. D. Armstrong, C. L. Carlson, and M. Bentivegna, "P-N-P Transistors using High-Emitter-Efficiency Alloy Materials," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 144-152.

⁶ K. F. Stripp and A. R. Moore, "The Effects of Junction Shape and Surface Recombination on Transistor Gain," *Proc. I.R.E.*, Vol. 43, pp. 856-866, July, 1955.

⁷ C. W. Mueller and N. H. Ditrick, "Uniform Planar Alloy Junctions for Germanium Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 121-131.

Operating Point: $V_{CE} = -6$ volts; $I_E = 1$ milliampere.

Frequency: 1 kilocycle.

$W_b = 8.6$ microns (0.34 mils) determined from measurement of diffusion capacitance; see Reference (2).

$g_{ce} = 48$ micromhos computed from Equation (8), and 70 micromhos measured.

$\frac{A_b}{A_c} [\sigma_e L_e]_e = 0.12$ mhos. This was computed using Equation (11) with $\sigma_e L_e = 6$ mhos (see Reference (5)) together with $s = 10$ meters/sec., (estimated), $D = 44 \times 10^{-4}$ meter²/sec., and $R_B = 1910$ microns (7.5 mils).

$g_{b'e} = 71$ micromhos computed from Equation (7), and 56 micromhos measured.

$g_{in} = 66$ micromhos measured.

$g_{out} = 83$ micromhos measured.

Collector-to-base current amplification factor $\alpha_{cb} = -625$ measured.

Maximum power amplification = 48 decibels measured.

Other transistors including n-p-n units have had similar operating characteristics.

CONCLUSIONS

It is possible to design a junction transistor with approximately equal input and output resistances when image-matched for maximum power amplification at low-frequencies. A transistor of this design is characterized by small base width, large current amplification factor, large input resistance, small output resistance, and large power amplification. These transistors can be iterated in cascade without loss in power amplification and with resulting savings in circuit components and additional transistors.

A P-N-P TRIODE ALLOY JUNCTION TRANSISTOR FOR RADIO-FREQUENCY AMPLIFICATION*

BY

C. W. MUELLER AND J. I. PANKOVE

RCA Laboratories Princeton, N. J.

Summary—The performance of alloy junction transistors falls off as the frequency increases. Heretofore, the chief cause of this has been the resistance-capacitance low-pass filter effect in the base-emitter input. The effect is produced by the germanium resistance (between the external base connection and the active junction region) and the emitter-to-base diffusion capacitance. The latter may be extraordinarily high because of the relatively slow diffusion of charge carriers into the base region which must be charged up and discharged with minority carriers. In the new transistor design described, resistance and capacitance have been reduced by using a thick wafer of low-resistance germanium and placing the active junctions on a very thin section produced by drilling a well into the wafer. The junctions are about .0005 inch apart. Other capacitances are reduced by using small diameter junctions. To prevent limitation due to transit-time dispersion, nearly planar junctions are obtained by using germanium-indium alloy discs from which the junctions are formed.

The new unit will give 39 decibels gain (neutralized) at 455 kilocycles, 12 decibels (unneutralized) at 10 megacycles, and has an oscillation limit of 75 megacycles. At 1 megacycle, a noise factor as low as 4 decibels has been obtained.

INTRODUCTION

IN normal broadcast receiver design, an amplifier giving good gain in the frequency region of 0.5-3 megacycles is desirable. In order to fulfill this need, keep power consumption to a minimum, and retain the simplicity and versatility of the triode alloy-junction transistor, a development was undertaken to extend the frequency range of this type of transistor.

To follow the course of the present development, it is desirable to discuss first the important factors which limit the frequency response of the triode transistor such as the TA-153.¹ The important factors stem from a combination of two effects; one inherent to transistor operation, and the other extrinsic to transistor action but arising from

* Reprinted from *RCA Review*, December, 1953; this material was presented at the *I.R.E.-A.I.E.E. Transistor Research Conference*, State College, Pa., July 6, 1953.

¹ R. R. Law, C. W. Mueller, J. I. Pankove, and L. D. Armstrong, "A Developmental Germanium P-N-P Junction Transistor," *Proc. I.R.E.*, Vol. 40, pp. 1352-1357, November, 1952.

the constructional details. In the operation of this type of transistor it is well known that minority charge carriers, after injection by the emitter, flow to the collector by a process of diffusion through the base region and receive little or no aid from electric fields. Such diffusion currents flow under the influence of a density gradient, i.e., different carrier charge densities in different regions of the base layer. Thus, when a signal is applied to the transistor, the number of carriers in the base region must vary in accordance with the signal so that the diffusion current flow to the collector will reproduce the signal in the output circuit. This change in the charge of the base region induced by the applied signal acts as an emitter-to-base capacitance. Furthermore, because this charge flow in and out of the base region is a diffusion flow, an extraordinarily high capacitance results (of the order of 0.01 microfarad in the TA-153 transistor¹).

This capacitance is much greater than that associated with the transition region of the emitter junctions. The "diffusion" capacitance is given approximately by

$$C_{b'e} = \frac{q}{kT} \frac{W^2 I_E}{2D}$$

where

$C_{b'e}$ = input capacitance of the common-emitter representation of Figure 1,

W = effective thickness of the base region,

I_E = d-c emitter current,

D = diffusion constant for minority carriers in the base region,

q = electronic charge,

k = Boltzmann's constant,

T = absolute temperature.

When I_E is expressed in milliamperes, and W in mils, this formula gives the capacitance in farads as

$$C_{b'e} = \begin{array}{l} 2.8 \times 10^{-9} W^2 I_E \text{ (for p-n-p units)} \\ 1.3 \times 10^{-9} W^2 I_E \text{ (for n-p-n units).} \end{array}$$

Since this capacitance is proportional to the square of the thickness of the base region, it is important to minimize this dimension in tran-

sistors designed for high-frequency applications. Note that this capacitance is proportional to the d-c current and does not involve the junction area. It is evident that a measurement of $C_{b'e}$ will provide a measure of the effective thickness of the base region.

Now, in the construction of the alloyed junction transistor, the external base connection to the germanium wafer is made at some distance from the active junction region. The resistance of the germanium between the base connection and the active junction constitutes a series base lead resistance ($r_{bb'}$ of Figure 1) through which the input signal must pass, with consequent attenuation before it can be applied to the intrinsic transistor. Furthermore, the charging current for the emitter-to-base capacitance discussed above must also flow through this resistance. This combination of series resistance and emitter-to-base capacitance constitutes a low-pass filter which is one of the most important limiting factors of transistors of the type under discussion.

The presence of this base lead resistance in addition to diminishing the input signal available to the actual transistor junction, also influences the output impedance; it leads to a reduction in the over-all output resistance and an increase in the over-all output capacitance of the transistor.

In addition to the input circuit effects discussed above, there are further consequences resulting from the fact that charge carriers flow from emitter to collector by a diffusion process. Diffusion flow is essentially a random process with the individual carrier making many collisions and changing direction of motion many times during its travel across the base region. Directivity to this random motion is supplied by the concentration gradient discussed above. Thus, the transit time for the individual carriers must vary widely, giving rise to a dispersive effect in the transfer function (i.e., transconductance). This transit time dispersive effect in the transfer of carriers to the output circuit also limits the frequency response of transistors but, for transistors such as the TA-153, its effect is not as detrimental as the input circuit effects discussed above.

The transit-time dispersion effects in the transfer function have been discussed as resulting only from diffusion phenomena. There may be a similar effect due to different path lengths caused by nonplanar geometries and end effects. The difference in path length may be minimized by small emitter-to-collector spacings, planar junctions, and accurate centering.

This picture of the consequences of diffusion flow in the base region has tended to separate the effects on the input circuit from those con-

cerned with the transfer generator, leading quite naturally to the π equivalent-circuit representation (Figure 4). It may be well to point out that in other representations such as an equivalent T circuit, the transfer generator (alpha) involves both the transfer transit-time dispersion effects and base-charge variation effects. This is as it should be, for alpha is the current gain amplification factor, and must account for changes of current between the input and output. Indeed, the collector-to-emitter current gain cutoff frequency is related to $C_{b'e}$ by

$$f_{ca} = \frac{q}{2\pi kT} \frac{I_B}{C_{b'e}}$$

and a measurement of $C_{b'e}$ at low frequencies may be substituted for a measurement of f_{ca} at considerably higher frequencies.

Another frequency-limiting factor which may become significant as transistors are designed to minimize the above effects is the presence of other capacitances, such as collector-to-base and collector-to-emitter. The collector-to-base capacitance may be reduced by reducing the area of the collector junction. There is also a contribution to this capacitance due to transit-time variation effects caused by a variation of the effective base thickness with the instantaneous collector voltage.

These considerations will be applied to the development of a p-n-p medium-frequency transistor. Similar considerations would also apply to n-p-n transistors.

FACTORS GOVERNING DESIGN

As will now be understood from the preceding discussion, in the design of high-frequency transistors it is important that, in addition to satisfying the requirements for good transistor action at audio frequencies, particular attention must be given to the following parameters:

1. Thickness of base layer.
2. Series base lead resistance.
3. Capacitance.

As pointed out in the general discussion, the value of $C_{b'e}$ (or f_{ca}) is degraded in proportion to the square of the distance between the emitter and collector junctions. This distance should be as small as can conveniently be obtained, and a reasonable value of spacing to aim for is .0005 inch.

In the alloy type transistor, an emitter-to-collector spacing of .0005

inch can be obtained in two ways: (a) by using a thick base wafer and a large impurity alloy penetration, or (b) by using a thin germanium base wafer with small impurity alloy penetration. With method (a), however, the junctions are hemispherical rather than parallel planes, and consequently the possible hole path lengths may vary considerably. In method (b) the junctions are more nearly flat and parallel and this method was therefore chosen. With thin base wafers it is necessary that the impurity alloy depth of penetration be small. Accurate control is then necessary to prevent emitter-to-collector shorts.

The factors governing the penetration of indium by solution into the germanium base material have been discussed.¹ A shallow penetration as desired here can be secured by: (1) a short firing time, i.e., shorter than that required for an equilibrium solution of the alloying action, (2) using a small volume of indium, (3) using a disc of indium-germanium alloy. The method described in this paper is a combination of all three, but depends principally on the latter two.

In order to secure plane parallel junctions, heating at a high temperature is desirable to secure rapid wetting over the entire region of the junction. Various combinations of alloy disc composition and temperature are possible and several have been tried. However, as the atomic percentage of germanium in the indium-germanium alloy is raised above 10 per cent, it is more difficult to secure a homogeneous alloy, and the mechanical properties of the alloy become poorer. The alloy which has been found most advantageous at the present time contains 5 atomic per cent germanium in indium, and this is the alloy used in all experiments described here.

As previously pointed out, it is important that $r_{bb'}$, the resistance of the germanium connecting the active junctions and the base tab, be as small as possible. This resistance is shown in the input equivalent circuit of Figure 1a. The values of this equivalent circuit can be measured by means of specially designed bridges.²

The series resistance, $r_{bb'}$, consists of two regions, as shown in Figure 1b: region "A" directly between the junctions, and region "B" between the junctions and the base connection. Obviously, the resistance of both these regions is reduced by using a base material with the lowest possible resistivity. In the "B" region, no limitation on resistivity exists. However, on the "A" region, two good junctions must be formed.

² L. J. Giacoletto, "Equipments for Measuring Junction Transistor Admittance Parameters for a Wide Range of Frequencies," *RCA Review*, Vol. XIV, pp. 269-296, June, 1953.

Shockley³ has shown that in order to have a good hole emitter, the conductivity of the emitter section must be much higher than that of the base region. Experimentally, it is found that as the resistivity of the base region is decreased, good junctions are more difficult to make. Here, then, an engineering compromise must be made between low base resistance and good junctions, as evidenced by high alpha and high back-resistance values. The resistance of the "A" region, which is directly between the junctions, is also reduced by the carriers injected by the emitter, i.e., conductivity modulation. Considering the present state of the art and the necessary engineering compromises, a resistivity of 0.6 to 0.8 ohm-centimeter was chosen after some exploratory tests.

Let us examine region "B" to see how this resistance can be reduced. In order to obtain an idea of the magnitude of resistance involved in the region "B" for a thin wafer base, consider the simplified geometry shown in Figure 2. The resistance between concentric cylinders is given by

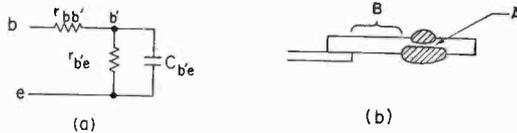


Fig. 1—(a) Equivalent input circuit of the transistor and (b) Transistor cross section showing the physical location of the parts of $r_{bb'}$ on a TA-153 audio-transistor.

$$R = \frac{\rho}{2\pi L} \ln \frac{r_2}{r_1} \text{ ohms}$$

ρ = resistivity of germanium in ohm-centimeters,

r_1, r_2 = radius of the inner and outer cylindrical surfaces,

L = wafer thickness in centimeters.

If r_1 is kept constant at .005 inch (the size of the emitter junction) and r_2 varied from .010 inch, the curves of Figure 2 are obtained.

If such a flat-wafer construction were used for the high-frequency transistor, and $\rho = 3$ ohm-centimeters, we see from Figure 2 that a series resistance of 200 ohms would be introduced to a ring contact .040 inch from the center of the emitter. The curves also show that the contribution of the first few thousandths of an inch around the

³ W. Shockley, "The Theory of P-N Junctions in Semi-Conductors and P-N Junction Transistors," *Bell Sys. Tech. Jour.*, Vol. 28, pp. 435-489, July, 1949.

junctions is important. From these curves can be seen the disadvantages of high-resistivity germanium and of the flat-wafer construction.

Several methods of reducing the resistance of the region "B" are possible. As a guide, methods other than those finally used are discussed first. One way to reduce the resistance of portion "B" is to solder a ring, a few thousands of an inch larger than the junctions, to the germanium that surrounds the junctions. Great care must then be taken to keep the indium and solder from shorting, and to obtain

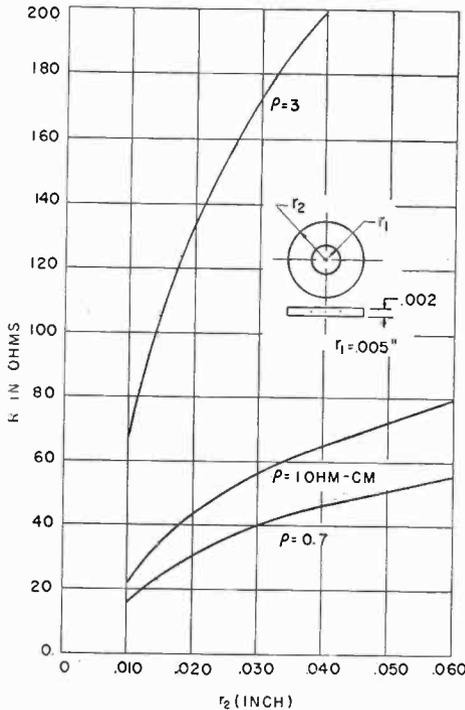


Fig. 2—Resistance between concentric cylinders of germanium.

proper etching after assembly. Another possible procedure is to form junctions, protect the junctions with a suitable lacquer, and then plate a metal on the exposed base wafer. It is necessary to make a plated contact which is ohmic, i.e., does not inject an appreciable number of holes. Such ohmic contact can be made by sand blasting the base and then plating. A word of caution on these methods is in order. As has been shown in another paper,⁴ surface recombination is a critical factor in obtaining low-frequency transistor action. An ohmic contact or

⁴ A. R. Moore and J. I. Pankove, "The Effect of Junction Shape and Surface Recombination on Transistor Current Gain," *Proc. I.R.E.*, Vol. 42, pp. 907-913, June, 1954.

plated region of considerable area too near the emitter junction could lead to such a high recombination rate as to make the transistor unsatisfactory. A compromise is necessary which leaves sufficient low-recombination surface around the junctions, and yet not so much as to leave the base resistance high.

After exploration of the above methods, a design was chosen in which only germanium surrounds the junctions. The structure is shown in Figure 3, which is drawn approximately to scale. Here we see that, as soon as we leave the immediate vicinity of the junctions, the thickness of the wafer increases and thus decreases the series base lead resistance; since only germanium surrounds the junctions, low surface recombination can be attained at the same time.

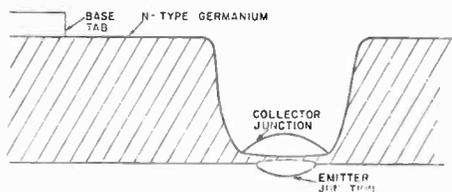


Fig. 3—Cross-sectional view of junction geometry.

The collector-to-base capacitance should be no larger than necessary. The junctions of the alloy type transistor are usually of the abrupt or Schottky type. Formulas for this case are developed by Shockley³ and, when expressed in convenient units, lead to the following equation for the capacitance of an abrupt junction.

$$C = .071 d^2 \frac{1}{\sqrt{\rho_b V_c}} \text{ micromicrofarads}$$

where

d = diameter of the junction in mils,

V_c = collector voltage,

ρ_b = resistivity of the base material in ohm-centimeters.

Thus the diameter of the collector has the greatest influence and should be made small. This should preferably be done without decreasing α_{ce} , collector-to-emitter current amplification factor. In order to keep α_{ce} high, the area of the emitter must be kept to two-thirds or less than that of the collector.⁴

As the area of the emitter is reduced, the ratio of emitter circum-

ference to area increases and surface recombination becomes more important. Consequently etching must be carefully done. Also, as the area of the emitter and collector become small, alignment of emitter and collector is more difficult. As an engineering compromise, the emitter was made .010 inch in diameter and the collector .015 inch in diameter.

ELECTRICAL MEASUREMENTS

The electrical engineer likes to reduce an electron device to an equivalent circuit composed of elements with which he is familiar so that he can apply standard circuit theory. Many types of equivalent circuit are possible. The simplest circuit, with the fewest parameters that are all independent of frequency, is best but not always obtainable. Most of the published literature on transistors is in terms of the T-

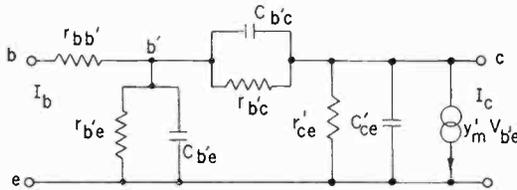


Fig. 4—Single-generator base-input π -equivalent circuit.

equivalent circuit or loop-derived parameters. Many who are familiar with vacuum tube terminology prefer the π -equivalent or nodal-derived parameters. One set of parameters can of course be derived from the other set if the set is complete.⁵

In Figure 4 a single-generator π -equivalent circuit is shown for the common emitter connection. The circuit parameters have been tested and found to be independent of frequency as long as the frequency is not too high. The circuit is believed sufficiently close as long as

$$\frac{\omega W^2}{4 D_p} < 0.2$$

where

ω = angular frequency,

W = spacing between junctions in centimeters,

⁵ L. J. Giacoletto, "Terminology and Equations for Linear Active Four-Terminal Networks Including Transistors," *RCA Review*, Vol. 14, pp. 28-46, March, 1953.

D_p = diffusion constant for holes.

For the transistors described here the circuit is believed to be a reasonable approximation up to a frequency of the order of 3 megacycles. The elements shown in this circuit were measured in bridges previously described.² The values measured on 6 transistors are given in Table I. The value of W , the spacing between junctions, is calculated from $C_{b'e}$ by applying diffusion theory as previously described. The values of the circuit constants in the output were not determined because their effect is negligible in comparison with the other parameters and their measurement is difficult.

Table I— π -equivalent Circuit Parameters
($E_c = -6$ volts, $I_c = -1$ milliampere)

Transistor	A	B	C	D	E	F#	Approx. TA-153 range
$r_{bb'}$ ohms	50	115	60	70	77	100	200-500
$r_{b'e}$ ohms	1250	1500	1400	600	430	500	600-2000
$C_{b'e}$ $\mu\mu\text{fd}$	300	700	900	600	800	2100	10,000-25,000
$r_{b'o}$ megohms	1.2	.10	.14	.86	.13	1.9	2
$C_{b'o}$ $\mu\mu\text{fd}$	10.4	7.7	16.7	9.5	17.0	11.0	35
W^* mils	.33	.50	.56	.46	.52	.86	1-3
f_{ca}^\dagger mc	20.5	8.8	6.9	10.6	7.7	3.0	.4

* Calculated from $C_{b'e}$ as previously described.

† Calculated from $f_{ca} = \frac{q}{2\pi kT} \frac{I_E}{C_{b'e}}$.

This transistor is included to show the deviation to be expected from a higher than normal value of W .

For comparative purposes, a range of values for TA-153 audio p-n-p transistors described in Reference (1) is included. The value of W for the TA-153 transistor was not accurately controlled since the transistor was intended for audio use.

In Table II the low-frequency parameters of the common-base T-equivalent circuit of Figure 5 are shown, again in comparison with those of the TA-153. The parameters r_c and α_{cc} are slightly poorer in the high-frequency transistor because they were compromised to get better high-frequency performance. These low-frequency parameters, of course, give no clues to high-frequency performance, but are included in the data for those who wish to make comparisons.

An important characteristic of a high-frequency transistor is the

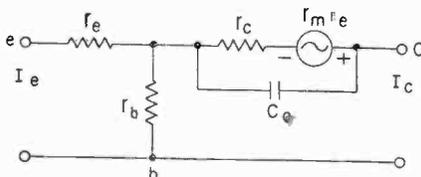


Fig. 5—Low-frequency emitter-input T-equivalent circuit.

single-frequency power gain as a function of frequency. Figure 6 shows a simplified schematic circuit of the test set used to measure power gain. Input matching is made only with the resistive component while the output is conjugate matched by adjusting the capacitance and output impedance. The feedback is not neutralized. In Figure 7 the gain of a transistor with common-emitter and common-base connections is plotted. In Table III the gain of the six transistors is shown for the various conditions described in the footnotes of the table.

Table II—T-equivalent Circuit Parameters
($E_c = -6$ volts, $I_E = 1$ milliampere)

Transistor	A	B	C	D	E	F	Approx. TA-153 range
r_e ohms	4.0	18.0	20.6	14.6	19.4	16.0	5-20
r_b ohms	1000	300	140	270	130	250	400-2000
r_c megohms	.40	.06	.16	.40	.16	.45	.5-3.0
α_{ee}	.977	.982	.976	.955	.952	.963	.95-.99
α_{cb}	41	51	40	21	19	25	20-100
I_{co} μ a	20	10	10	14	18	8	2-20

In Table III the noise factor is also given. There is a considerable variation of one-kilocycle noise factor among various units at the present time, but the one-megacycle noise factor does not vary as much. In Figure 8 the noise factor is plotted as a function of frequency up to one megacycle.

Experiments with a simple oscillator circuit show that the high-frequency transistors described will oscillate at frequencies up to 75 megacycles.

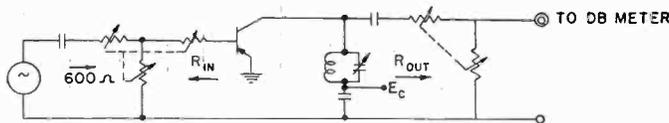


Fig. 6—Simplified diagram of circuit used for measurement of gain versus frequency.

Table III—Gain and Noise Factor Values in Decibels

Transistor	A	B	C	D	E	F	Approx. TA-153 range
4 kc gain*	35.	37.	39.	34.	34.	38.	35 to 48
455 kc gain**	35.	32.	39.	35.	34.	34.	—
1 mc gain†	24.	23.	25.	24.	23.	22.	< 0 to 7.
10 mc gain†	13.	10.	11.	9.	8.	5.	—
1 kc Noise factor‡	14.	23.	33.	24.	16.	20.	15 to 25
1 mc Noise factor#	4.	5.	8.	7.	5.	6.	—

All values of gain measured at $E_c = -6$ volts, $I_B = 1$ milliampere.

* $R_{in} = 500$ ohms, $R_{out} = 30,000$ ohms.

** Input and output conjugate matched and feedback neutralized.

† Input resistive matched and output conjugate matched in circuit of Figure 9.

‡ $E_c = -1$ volt, $I_B = 1$ milliampere, $R_{in} = 560$ ohms.

$E_c = -1$ volt, $I_B = 1$ milliampere, $R_{in} = 100$ ohms.

CONCLUSIONS

A p-n-p triode transistor capable of as much as 39 decibels gain at 455 kilocycles when used in a neutralized, conjugate-matched circuit has been constructed. Without neutralizing, 12 decibels gain can be obtained at 10 megacycles. The noise factor at one megacycle is 4-8 decibels.

A spacing between junctions of .0005 inch and a low base-lead resistance is obtained by using a low-resistivity base section with the collector junction located at the bottom of a well in the germanium. An indium-germanium alloy is used to obtain more planar junctions.

With the continued improvement in the techniques of making and

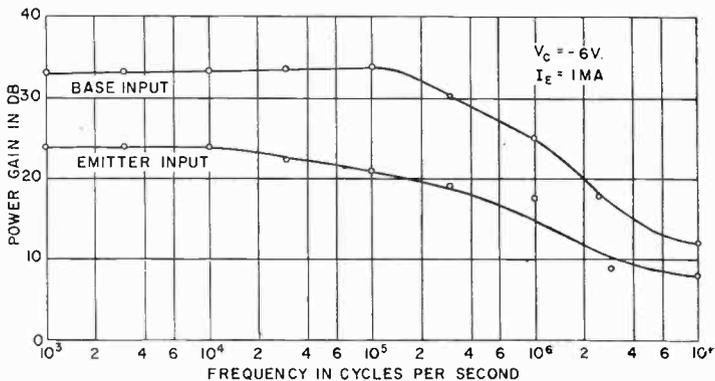


Fig. 7—Single-frequency power gain measured in the circuit of Figure 6.

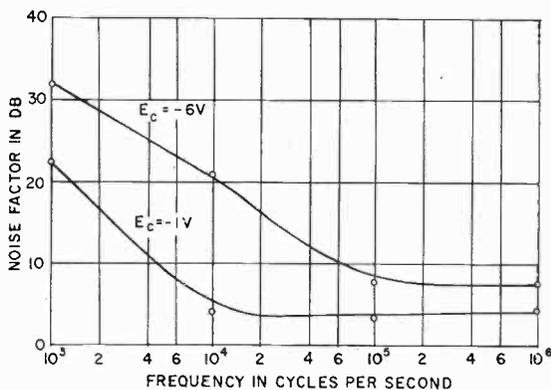


Fig. 8—Noise factor versus frequency; common-emitter connection.

controlling junctions, further improvements by the use of lower resistivity germanium and a reduction in all dimensions are possible.

ACKNOWLEDGMENT

The gain and noise measurements were made on equipment designed and built by P. Werenfels. The neutralized gain measurements at 455 kilocycles were made by D. D. Holmes, and oscillation tests by G. B. Herzog. Analysis of diffusion reactances by H. Johnson was valuable. Mrs. E. Moonan made many of the measurements.

THE DRIFT TRANSISTOR*

BY

H. KROEMER

RCA Laboratories,
Princeton, N. J.

Summary—The transit time of the current carriers diffusing through the base region is an important factor in limiting the frequency response of conventional $p-n-p$ or $n-p-n$ junction transistors. The principles of a transistor are discussed, in which the impurity in the base region decreases exponentially from very high values at the emitter to low values at the collector. This distribution introduces a constant electric drift field parallel to the diffusion direction into the base region. The drift field reduces the transit time and increases the frequency limit of the transistor. In the case of germanium, an improvement of the alpha-cutoff frequency by a factor of eight should be possible for the same value of the base width. At the same time, the high impurity density near the emitter leads to a low base-lead resistance, and the low impurity density near the collector to a low collector capacitance. This results in a further improvement of the transistor performance, since this resistance and capacitance also limit the high-frequency operation of a transistor. In particular, the lower base resistance permits the use of a thinner base, thereby increasing the alpha-cutoff frequency even more.

A comparison between the drift transistor and the $p-n-i-p$ ($n-p-i-n$) transistor is made. The analysis shows that for the same base width the drift transistor should have about 3 decibels more high-frequency amplification than the $p-n-i-p$.

I. INTRODUCTION

IN a series of previous papers¹⁻⁴ the author has given the detailed theory of a junction transistor which has an exponential impurity distribution within the region between the emitter and the collector. This structure exhibits several advantages over previous structures, especially in frequency behavior. The purpose of the present paper is to summarize and extend the conclusions of the earlier work.

A transistor which is to operate at frequencies as high as possible has to fulfill at least three requirements: (a) Short transit time of

* Portions of this paper were presented at the *I.R.E.-A.I.E.E. Conference on Semiconductor Device Research*, in Philadelphia, June 20, 1955.

¹ H. Kroemer, "The Drift Transistor," *Naturwissenschaften*, Vol. 40, p. 578, November, 1953.

² H. Kroemer, "On the Theory of the Diffusion and the Drift Transistors," *Archiv der Elektrischen Übertragung*, Vol. 8, p. 223, 1954.

³ Reference (2), page 363.

⁴ Reference (2), page 499.

the injected carriers from the emitter to the collector, (b) Low collector capacitance, (c) Low base resistance. In an ordinary junction transistor the transit time is determined by the process of the diffusion of the injected carriers through the base region. Diffusion is a comparatively slow process because of its random nature. Compared with a junction transistor of equal collector-emitter spacing, the frequency limit of a point contact transistor is much higher. Since it is actually higher than the α -cutoff frequency of a junction transistor of the same electrode spacing, this cannot be due only to the smaller capacitance of point contacts. In the point-contact unit the transit time is much shorter because the injected carriers, instead of diffusing randomly, are moving in the strong electric field which the collector current builds up around the collector. If it were possible to introduce an electric field in the diffusion direction in a junction transistor, one would expect a considerable increase in frequency limit, especially if the capacitance and the base resistance could be lowered simultaneously. A transistor which contains such a field will be called a drift transistor. A drift transistor with a constant electric field can be obtained by building an exponential impurity distribution into the base region.

Section II gives a general qualitative outline of the mode of operation of a drift transistor, Section III gives a summary of quantitative theoretical data and Section IV compares the drift transistor with the p-n-i-p transistor. All formulas and numbers apply for p-n-p germanium transistors unless otherwise stated. The conversion to the n-p-n type is trivial. In the analysis, the impurity distribution will be assumed to be a true exponential, but the principal results also hold for transistors with slight deviations from this mathematically ideal distribution, especially if the deviations from the exponential law take place on the collector side of the base region rather than on the emitter side.

II. THE EXPONENTIAL IMPURITY DISTRIBUTION

Assume the net impurity concentration in the base region of a p-n-p transistor to be an exponential function. If N_0 is the concentration at $x = 0$, the emitter side of the base region, and F the desired drift field, then this distribution must be²

$$N(x) = N_0 e^{-\frac{qFx}{kT}} \quad (1)$$

To show this, let us assume that all donors are ionized. Then in the region where

$$N(x) \gg n_i, \quad (n_i = \text{intrinsic carrier density}) \quad (2)$$

the local electron density in the base will be practically equal to the net impurity density. Since the electron density is an exponential function of the energy separation between the conduction band and the Fermi level, this separation then varies linearly with position. Because the Fermi level has to be constant, the bands are tilted at such an angle as to produce a homogeneous electric field of strength F in the base (Figure 1). This field is a true electrostatic field which sets itself up within the semiconductor in order to balance the diffusion of majority carriers which would otherwise occur due to the concentration gradient. The drift field is directed from the emitter to the collector, that is, in the direction of the natural diffusion flow. This results in a substantial reduction in carrier transit time. In addition, a sufficiently strong drift field makes the currents less dependent on the base width, w , than in a diffusion transistor. This reduces both the collector conductance and the internal feedback which are caused by the dependence of w on the applied collector voltage.

The base width w — as the term is used in this paper and as shown in Figure 1 — is given by the distance between the emitter and the space-charge region on the collector side and not by the distance between the p-type emitter and collector regions. We will call the latter quantity w_{00} , the “electrode spacing,” while w may be called the “effective base-width.” It is then convenient to discuss the drift transistor in terms of the difference in the potential energy of an electron over the distance w ,

$$\Delta V = qFw = kT \ln \frac{N_0}{N(w)}, \quad (3)$$

rather than in terms of the drift field, F . We call ΔV the built-in drift potential. For a given impurity distribution, that is, for a given drift field F , both w and ΔV still depend upon the electrode spacing, w_{00} . There are two limiting cases to be distinguished, depending on the magnitude of w_{00} :

(a) For small w_{00} 's, that is, when

$$\frac{kT}{qF} \ln \frac{N_0}{n_i} - w_{00} \gg \frac{kT}{qF},$$

Equation (2) is fulfilled throughout the whole middle region. Because of the comparatively high impurity density at $x = w_{00}$, the collector

depletion layer will then be thin compared to w_{00} (for not too high voltages). Therefore $w \approx w_{00}$, $\Delta V \approx qFw_{00}$.

(b) For large w_{00} 's, when

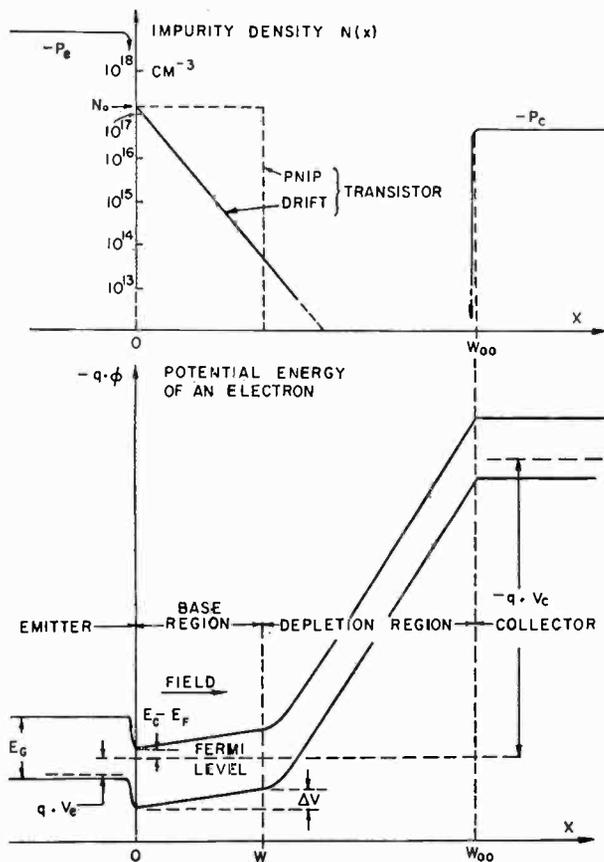


Fig. 1—The impurity and potential distribution in a drift transistor. For comparison, the impurity distribution for a p-n-i-p transistor is shown in the upper drawing. (See Section V.)

$$w_{00} \approx \frac{kT}{qF} \ln \frac{N_0}{n_i} \gg \frac{kT}{qF},$$

the collector side of the middle region has a very low impurity concentration. The collector depletion layer then becomes thick. It penetrates through the whole "quasi-intrinsic" region until it hits the more highly doped parts of the base. The effective basewidth, then, depends only

very weakly upon w_{00} , while the depletion layer thickness varies strongly with w_{00} .

The actual computation of the position of w and its voltage dependence is very difficult. The usual methods of computing a depletion layer thickness are not applicable in this case where the impurity density at $x = w$ is not very large compared to n_i and to the density of the holes travelling through the transistor. A new computation which includes these effects has not yet been made. For the purpose of this paper it will therefore be assumed that in the case of a large w_{00} the effective base width, w , is given by the distance where the Fermi level has come within kT of the middle of the forbidden band.

This case of a large w_{00} incorporates the largest possible drift potential, ΔV . For a given value of w it therefore constitutes the case with the shortest transit time. For this reason, and because this paper is primarily interested in the highest degree of improvement obtainable in a transistor by introducing a drift field, we will restrict ourselves for the rest of the paper to this case of a "full" drift-transistor where w_{00} is much greater than w .

Since the capacitance of the collector junction is inversely proportional to the depletion layer thickness, $w_{00} - w$, it follows that a full drift transistor automatically has a low collector capacitance, fulfilling the second of the above requirements.

The actual value of ΔV depends upon N_0 . An upper limit for N_0 and ΔV can be estimated by the following considerations: In order to obtain a good emitter efficiency, the impurity density in the base at $x = 0$, N_0 , has to be small compared to P_e , the net acceptor concentration in the emitter region. One may, for example, require for the current amplification factor α_{ce} a value of at least .98, or for the base-to-collector alpha, $\alpha_{cb} = \alpha_{ce}/(1 - \alpha_{ce})$, a value of at least 50. The separation of the conduction band from the Fermi level at $x = 0$ then has to be of the order of $kT \times \ln 50 \approx 4kT$ larger than the separation of the valence band from the Fermi level in the emitter region. The latter separation can be assumed to be not less than kT , resulting in a value for $E_c - E_F \geq 5kT$ and in an impurity density $N(0) \leq N_c \cdot e^{-5} \approx 1.6 \times 10^{17} \text{ cm}^{-3}$. Here N_c is the effective density of states in the conduction band, assumed to be $2.4 \times 10^{19} \text{ cm}^{-3}$. Together with the above assumption that, at $x = w$, the separation between the Fermi level and the middle of the band gap is kT , this results in the condition for the maximum built-in drift potential;

$$\Delta V = (1/2) E_G - 6kT. \quad (4)$$

In germanium we have $E_G \approx 0.7$ electron volt $\approx 28 kT$ and therefore

$\Delta V \leq 8 kT$, corresponding to a doping ratio $N_0/N(w) \leq e^8 \approx 3,000$.

The above value for N_0 is a rather high value, much higher than in ordinary "diffusion transistors" with uniform base doping. In the case of uniform doping, such a high concentration would result in an extremely low collector breakdown voltage (< 1 volt) and an extremely high collector capacitance.

The high doping near $x = 0$ leads to a very low base resistance thereby fulfilling the third requirement mentioned at the beginning. Together with the decrease in collector conductance mentioned earlier, this makes it possible to use a much thinner base region than in a diffusion transistor.

IV. THEORY OF THE DRIFT TRANSISTOR

1. Junction Conductances and Capacitances

For the density p of holes in the base at $x = 0$, the same assumption is made as in a diffusion transistor, namely

$$p(0) = p_0 e^{\frac{qV_e}{kT}}, \quad (5)$$

where p_0 is the equilibrium density at $x = 0$ ($p_0 = n_i^2/N_0$) and V_e is the emitter voltage. In a diffusion transistor the hole density $p(x)$ decreases linearly from this value to zero at $x = w$, resulting in a diffusion current density $j = qD_p p(0)/w$ (Figure 2a). For simplicity, recombination is neglected. However, in a drift transistor with sufficiently strong drift field, the current will be carried mostly by the field and the hole density at $x = 0$ will be swept out and will be constant over most of the base region (Figure 2b). The resulting current density is therefore

$$j_p = q\mu_p p(0) F = \frac{qD_p p(0)}{w} \frac{qFw}{kT}. \quad (6)$$

This is larger than the current density in a diffusion transistor by a factor of $\Delta V/kT$. Just short of the end of the base the injection density drops to zero because of the holes falling into the collector. At $x = w$, $p(x) = 0$ and the total current is diffusion current. Because it has to be equal to the value given by Equation (6), the concentration gradient at $x = w$ has to be $p'(w) = p(0)qF/kT$.

By Equations (5) and (6), the total injection current depends exponentially upon V_e . Therefore the small-signal a-c admittance of the

emitter at a given bias current is the same as in a diffusion transistor:

$$\frac{\partial j_e}{\partial V_e} = \frac{q}{kT} j_e. \quad (7)$$

The complete a-c admittance of the emitter including the diffusion capacitance is found to be*

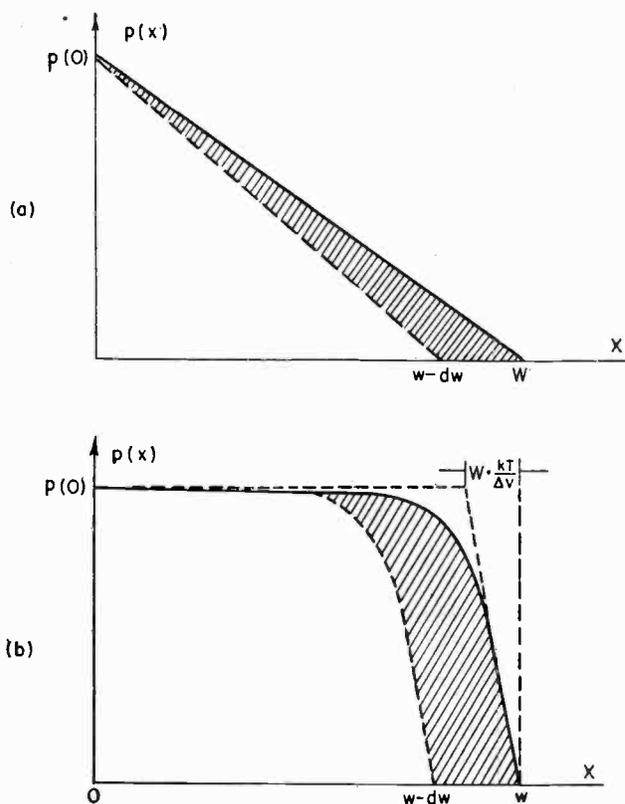


Fig. 2—Injected hole density in the base region: (a) diffusion transistor, (b) drift transistor. The shaded area shows the change in injected carrier density when the base region narrows with increasing collector voltage (Section IV, 1).

$$\left(\frac{\partial j_e}{\partial V_e} \right)_{V_c} = j_e \frac{q}{kT} \frac{1 + \Omega}{2}, \quad (8)$$

* Equations (8), (11), (13), and (21) follow from Equations (45), (46), and (48) of Reference (2) by neglecting the term $w^2/L^2 \ll 1$ wherever it occurs together with the term $(\Delta V/2kT)^2 \gg 1$.

where

$$\Omega = \sqrt{1 + i\omega \frac{w^2}{D_p} \left(\frac{2kT}{\Delta V} \right)^2} \quad (9)$$

For not too high frequencies the square root can be expanded into a series and one obtains the diffusion capacitance per unit area,

$$C_e = j_e \frac{q}{kT} \frac{w^2}{D_p} \left(\frac{kT}{\Delta V} \right)^2, \quad (10)$$

which is smaller by the factor $3(kT/\Delta V)^2$ than the value for a diffusion transistor with the same w . For $\Delta V = 8kT$ this factor is about .05.

Actually, the total emitter capacitance includes, in addition to this diffusion capacitance, the junction transition capacitance. The transition capacitance is not always negligible in the drift transistor because of the much higher doping in the base at $x = 0$. Quantitative values depend on the steepness of the emitter junction; for an abrupt junction with $N_0 = 10^{17} \text{ cm}^{-3}$, the transition capacitance is of the order of 1,000 micromicrofarads per mm^2 , while the diffusion capacitance for $j_e = 10$ milliamperes per mm^2 , $w = 10^{-3}$ centimeters and $\Delta V = 8kT$ is only 150 micromicrofarads per mm^2 . A high transition capacitance is very undesirable, because the capacitive emitter current is majority carrier current and does not contribute to the collector current. This results in a much more rapid decrease in α with frequency than in a comparable transistor with a lower transition capacitance. It is therefore desirable that the emitter junction be graded to some extent and of small area. For a given current this means a high current density, and a low ratio of transition to diffusion capacitance. There is, however, an upper limit for the current density which is derived in Section IV, 3.

The effective base width, w , of a transistor decreases slowly with increasing collector voltage. In a diffusion transistor this causes an increase of the diffusion gradient within the base and thereby of the collector current if the emitter voltage is held constant. Or, if the emitter current and thereby the gradient is held constant, it causes a decrease in emitter voltage and a slight increase in current-amplification factor, α , and therefore in collector current (Figure 2a). In other words, the base-thickness variations cause both an internal feedback and a finite collector conductance.⁵⁻⁷ In a drift transistor the hole dis-

⁵ J. M. Early, "Effects of Space-Charge Layer Widening in Junction Transistors," *Proc. I.R.E.*, Vol. 40, p. 1401, November, 1952.

⁶ J. M. Early, "Design Theory of Junction Transistors," *Bell Sys. Tech. Jour.*, Vol. 32, p. 1271, November, 1953.

⁷ R. L. Pritchard, "Frequency Variations of Junction-Transistor Parameters," *Proc. I.R.E.*, Vol. 42, p. 786, May, 1954.

tribution is substantially uniform near $x = 0$ and a decrease in w results in a shift of the distribution to the left rather than in a change of its shape (Figure 2b). Consequently, the emitter voltage and the collector current change much less than in a diffusion transistor. The detailed theory gives the feedback conductance

$$\left(\frac{\partial j_e}{\partial V_c} \right)_{V_e} = j_e \frac{1}{w} \frac{dw}{dV_c} \frac{\Delta V}{kT} \Omega e^{-\frac{\Delta V}{2kT}} (1 + \Omega) \quad (11)$$

At low frequencies this reduces to

$$\left(\frac{\partial j_e}{\partial V_c} \right)_{V_e} = j_e \frac{1}{w} \frac{dw}{dV_c} \frac{\Delta V}{kT} e^{-\frac{\Delta V}{kT}}, \quad (12)$$

a value which is smaller by the factor

$$\frac{\Delta V}{kT} e^{-\frac{\Delta V}{kT}} \ll 1 \quad (\approx 1/375)$$

than the value for a diffusion transistor with the same w and dw/dV_c . The collector conductance is

$$\left(\frac{\partial j_c}{\partial V_c} \right)_{V_e} = -j_c \frac{1}{w} \frac{dw}{dV_c} \frac{\Delta V}{kT} \left[\Omega e^{-\frac{\Delta V}{kT}} + \frac{\Omega - 1}{2} + \left(\frac{w}{L} \frac{kT}{\Delta V} \right)^2 \right] \quad (13)$$

The low-frequency value of Equation (13),

$$\left(\frac{\partial j_c}{\partial V_c} \right)_{V_e} = -j_c \frac{1}{w} \frac{dw}{dV_c} \left[\frac{\Delta V}{kT} e^{-\frac{\Delta V}{kT}} + \frac{w^2}{L^2} \frac{kT}{\Delta V} \right], \quad (14)$$

is smaller by the factor

$$\frac{\Delta V}{kT} e^{-\frac{\Delta V}{kT}} + \frac{kT}{\Delta V} \frac{w^2}{L^2} \ll 1$$

than the diffusion-transistor value. For a sufficiently narrow base width and high lifetime the second term is negligible. The resulting

extremely low value is due to the fact that in Equation (13) for $\omega = 0$ the term $\Omega - 1$ vanishes. With increasing frequency, however, it rapidly outgrows the exponential function, resulting not only in a capacitive term but in an increase in collector conductance. By expanding the square root and introducing the α -cutoff frequency f_α , which is defined in Equation (20), one finds that the conductance increases according to

$$\begin{aligned} \left(\frac{\partial j_c}{\partial V_c} \right)_{V_e} &= -j_c \frac{1}{w} \frac{dw}{dV_c} \left[\omega \frac{w^2}{2D_p} \left(\frac{2kT}{\Delta V} \right)^{3/2} \right]^2 \\ &= -j_c \frac{1}{w} \frac{dw}{dV_c} \frac{1}{2} \left(\frac{f}{f_\alpha} \right)^2. \end{aligned} \quad (15)$$

This is still much smaller than the value for the diffusion transistor.

The collector diffusion capacitance, which also follows from Equation (14) by expansion of the square root, is

$$C_c = -j_c \frac{1}{w} \frac{dw}{dV_c} \frac{w^2}{2D_p} \left(\frac{2kT}{\Delta V} \right). \quad (16)$$

C_c is smaller by the factor $3kT/\Delta V$ than the diffusion transistor value. However, the collector diffusion capacitance is less than or in the order of the already low transition capacitance of the collector. The latter is therefore the more important capacitance. As shown in Reference (3) the ratio of the two is

$$C_{\text{diff}}/C_{\text{trans}} = \frac{-j_c w}{qD_p N_0} \frac{kT}{\Delta V} e^{\frac{\Delta V}{kT}}. \quad (17)$$

For $j_c = 10$ milliamperes per mm^2 , $N_0 = 1.5 \times 10^{17} \text{ cm}^{-3}$, $w = 10^{-3} \text{ cm}$ and $\Delta V = 8kT$, this ratio is 0.37. For high current density and high w , the diffusion capacitance may overcome the transition capacitance.

2. The α -Cutoff Frequency and the Current Amplification Factor

The mean velocity of the holes travelling through the base of a drift transistor is

$$v_{\text{drift}} = \mu_p F = \frac{2D_p}{w^2} \frac{\Delta V}{2kT} \quad (18)$$

compared to the value for a diffusion transistor of

$$v_{\text{diff}} = \frac{2D_p}{w^2}. \quad (19)$$

(One obtains the latter value by dividing the current density by the average value of the injected charge density, which is $1/2 qp(0)$.) This result corresponds to a transit-time decrease by a factor of $2kT/\Delta V$. The phase shift of an a-c signal decreases by the same amount. One might expect, therefore, an increase of the α -cutoff frequency f_a by a factor of $\Delta V/2kT$. Actually, however, f_a is defined by the decrease in amplitude rather than by an increase in phase shift. A detailed analysis^{2, 3} which is too long to be given here, shows that the α -cutoff frequency is improved by the factor $(\Delta V/2kT)^{3/2}$, resulting in

$$f_a = \frac{1}{\pi} \frac{D_p}{w^2} \left(\frac{\Delta V}{2kT} \right)^{3/2}. \quad (20)$$

For $\Delta V = 8kT$ this improvement factor is eight, and with $w = 10^{-3}$ centimeter, an α -cutoff frequency of about 100 megacycles results.

Equation (20) follows from the complete expression for the transport factor

$$\beta = \frac{2\Omega}{1 + \Omega} e^{-\frac{w^2}{L^2} \frac{kT}{\Delta V}} e^{-\frac{\Delta V}{2kT} (\Omega - 1)} \quad (21)$$

To obtain Equation (21) the following assumptions were made:

- (a) The lifetime, τ , is constant throughout the base region.
- (b) The surface recombination is neglected.
- (c) No recombination nor avalanche multiplication occurs within the collector depletion layer.
- (d) The transit time through the depletion layer is negligible.

The first three conditions affect the numerical value of α but do not affect its frequency behaviour very much. The depletion layer transit time, τ_c , however, which is negligible in ordinary transistors, may set the ultimate frequency limit in a drift transistor because the depletion layer may be much wider than in an ordinary transistor.

Initially, τ_c decreases with increasing collector voltage because the electric field strength in the depletion layer increases. But there is a lower limit for τ_c because the carrier mobility decreases with increasing field until eventually, above a certain critical field strength, F_c , the carrier velocity reaches an upper limit, v_∞ , and remains constant for

higher fields.^{8,9} The lowest obtainable value for τ_c is $\tau_c = w_c/v_{cs}$, where w_c is the depletion layer thickness. This value may become comparable to the base transit time if the latter is sufficiently small. In order to obtain the best possible frequency response from a drift transistor with a very low w it is therefore desirable to apply a collector voltage such that throughout most of the depletion layer the actual field strength is about F_c . A rough estimate of this desirable voltage is

$$-V_c \approx w_c F_c. \quad (22)$$

For holes in germanium, F_c is about 10^4 cm⁻³ and $v_{cs} = 5 \times 10^6$ cm/sec leading to the requirement that the minimum operation voltage of the transistor should be one volt for every micron of depletion layer thickness. The transit time is then 2×10^{-11} second per micron. Similar to the transport factor for the base region, a transport factor, β_c , can be defined for the collector depletion layer. Provided Equation (22) is fulfilled, β_c is given by⁸

$$\beta_c = \frac{1 - e^{-i\omega\tau_c}}{i\omega\tau_c}. \quad (23)$$

For a low frequency, β_c becomes $e^{-\frac{i\omega\tau_c}{2}}$, which means that the depletion layer merely causes an additional phase lag with $\tau_c/2$ as the time constant, without amplitude damping. With increasing frequency, however, β_c will decrease according to

$$|\beta_c|^2 = \left| \frac{\sin \frac{\omega\tau_c}{2}}{\frac{\omega\tau_c}{2}} \right|^2. \quad (24)$$

For $f = \omega/2\pi = 1/2\tau_c$, $|\beta_c|^2$ is down to about 0.4 and we will denote $1/2\tau_c$ as the transition layer limiting frequency, f_c .

For a depletion layer width of 1 mil, $f_c = 1,000$ megacycles. This effect is therefore negligible for a transistor with $w = 10^{-3}$ cm where the α -cutoff frequency for a drift transistor was shown to be 100 megacycles. With decreasing w , however, f_c sets an ultimate limit to the

⁸ J. M. Early, "The PNIP and NPIN Junction Transistor Triode," *Bell Sys. Tech. Jour.*, Vol. 33, p. 517, May, 1954. Early there considers the influence of τ_c for the pnp-transistor, but the same considerations hold here.

⁹ E. F. Ryder, "Mobility of Holes and Electrons in High Electric Fields," *Phys. Rev.*, Vol. 90, p. 766, June, 1953.

obtainable over-all frequency behaviour. With $w_c = 10^{-3}$ cm this occurs for $w \leq 0.2 \times 10^{-4}$ centimeter. Since w_c cannot be decreased indefinitely because of increasing C_c , one should expect that the maximum possible frequency obtainable with a germanium drift transistor lies in the order of a few thousand megacycles.

On the other hand, if $f_c \gg f_a$, it is not necessary to apply the full voltage V_c in order to utilize the full frequency range possible with the given value of w .

3. High-Current Behaviour.

Like a diffusion transistor, the drift transistor changes its properties for high current densities, but in a different way. In a diffusion transistor, the emitter efficiency, γ , decreases with increasing current, resulting in α fall-off.¹⁰ This effect need not be taken into consideration here because, long before the γ reduction becomes important, the effective mobility of the holes in the base layer decreases. This is explained as follows:

When the injected hole density increases, the electron density increases by the same amount in order to neutralize the injected holes. The drift field exerts forces in opposite directions upon electrons and holes thereby trying to separate the two kinds of carriers. This, however, sets up a balancing "polarization" field retarding both the electron and the hole motion. The effect is the same as if the mobility had decreased. Quantitatively, one finds⁴ that the transport equation for the holes becomes

$$j_p = q \frac{N\mu_p}{N+p} Fp - q \frac{N+2p}{N+p} D_p \text{grad } p, \quad (25)$$

where N is given by Equation (1). That means the holes behave as if their mobility were given by

$$\frac{N\mu_p}{N+p}, \quad (26)$$

while the effective diffusion constant would be

$$\frac{N+2p}{N+p} D_p. \quad (27)$$

¹⁰ W. M. Webster, "On the Variation of Junction-Transistor Current-Amplification Factor with Emitter Current," *Proc. I.R.E.*, Vol. 42, p. 914, June, 1954.

For high-level injection, when $p \gg N$, the effective mobility decreases to zero while the effective constant doubles.

The latter result is already known from Webster's work.¹⁰ Because of the mobility decrease a drift transistor will eventually work as a diffusion transistor when the current increases.

The mobility decrease first sets in at the collector side of the base region because $N(x)$ is lowest there. On the other hand, the current near $x = w$ is diffusion current rather than drift current so that the μ effect will be balanced initially by the apparent increase of D_p . Since the apparent mobility goes down to zero while the diffusion constant only doubles, the mobility decrease will eventually become more important. This restricts the useful current range to values where the ratio $p(x)/N(x)$ remains less than unity everywhere in the base. The maximum value of this ratio occurs shortly before the end of the base, namely at

$$\tilde{x} \approx w - \frac{kT}{qF} \ln 2, \quad (28)$$

where

$$\frac{p(\tilde{x})}{N(\tilde{x})} = \frac{1}{4} \frac{p_0}{N_0} e^{\frac{qV_c + \Delta V}{kT}}. \quad (29)$$

Replacing p_0 by the current density (Equation (6)) and requiring* that $p \leq N$, the current density limit is found to be

$$j_p \leq \frac{4qD_p}{w} N_0 \frac{\Delta V}{kT} e^{-\frac{\Delta V}{kT}}. \quad (30)$$

Inserting our standard numbers as an example gives $j_p \leq 110$ milliamperes per mm^2 .

4. The Base-Lead Resistance.

The base-lead resistance r_{bb} consists of two parts: The internal base resistance of the semiconductor under the emitter area, and a second part external to the junction areas. It is conceivable that various techniques may be employed which may make the external

* In Reference (4) the author required $p \ll N$ instead of $p < N$. The stronger inequality is unnecessary because of the increase of the diffusion constant.

resistance much less than the internal resistance. This is not the case in many present transistors. However, in order to confine our arguments to the basic structures we will consider the internal resistance to represent a limiting condition and utilize this as a basis of comparison of transistor structures.

If one denotes the integrated transverse conductivity (ohm^{-1}) of the base layer by S , then for circular electrodes^{4,5}

$$r_{bb'} = \frac{1}{8\pi S},$$

where

$$S = q\mu_n N_0 w \frac{kT}{\Delta V}. \quad (31)$$

For our standard dimensions, $r_{bb'} = 3.7 \Omega^\dagger$. This is a very low value. If the d-c bias current is small compared to 6.8 milliamperes, the small-signal potential drop over the emitter junction,

$$\delta V_e = \frac{kT}{q} \frac{\delta i_e}{i_e},$$

is small compared to the potential drop over $r_{bb'}$. In such cases, the base resistance is negligible.

V. COMPARISON WITH THE P-N-I-P TRANSISTOR

1. General

It has been shown in the previous sections that the drift transistor is a "true" high-frequency transistor in the sense that it fulfills the three requirements for such a transistor mentioned at the beginning. The question then arises as to how it compares with other conceivable high-frequency structures.

It can be stated quite generally that any transistor which is to fulfill the three initial requirements has to utilize a nonuniform impurity distribution within the base region, with a high impurity density at the emitter side and a low impurity density at the collector side. The various conceivable structures differ in the way in which this transition takes place. It can be assumed that the decrease in impurity distribution within the base region, with a high impurity

[†] This calculation assumes a constant mobility. The effect of impurity density on mobility and the consequent modification of $r_{bb'}$ is considered in Section V, 3.

does not provide any advantage. All the remaining possibilities can be considered as intermediate cases between two limiting special distributions. These structures are the drift transistor structure and the p-n-i-p structure.⁸ It is therefore sufficient to compare these two structures only.

The drift transistor is a limiting case in the following sense: For a given effective base width, w , a distribution which leads to a *constant* drift field results approximately in the shortest possible carrier *transit time*. This statement would hold rigorously if no diffusion in addition to the drift were to take place.

The p-n-i-p transistor has a constant impurity density in the base region proper, and an intrinsic region inserted between the base region and the collector region (see the dotted line in Figure 1). The p-n-i-p transistor is another limiting case, because for a given effective base-width, w , a high and constant impurity density within the base leads to the lowest possible *base resistance*.

While the thick depletion layer in a drift transistor is an automatic and necessary consequence of the principle of utilizing a maximum possible drift field, the introduction of an intrinsic layer into the ordinary p-n-p transistor was arrived at through different reasoning. In spite of these differences, both structures are similar. Roughly speaking, a drift transistor with the same effective base width, w , has a higher α -cutoff frequency but also a higher internal base resistance. This makes it difficult to find a suitable basis for comparison. One could make the comparison on the basis of equal base width, equal α -cutoff frequency, equal base resistance or even on a quite different basis. While there may be instances when other bases of comparison may be desirable, a suitable compromise for a general comparison appears to be on the basis of equal structural parameters, w , w_{00} and N_0 .

2. The Carrier Transit Time

According to Equations (18) and (20) the transit velocity in a drift transistor should be higher by the factor of $\Delta V/2kT$, and the f_a value by the factor $(\Delta V/2kT)^{3/2}$, than those values in a p-n-i-p transistor with the same w . In the derivation of this result the assumption of equal mobilities was made. Actually the mobility depends to some extent on the impurity density N and decreases with increasing N . This effect is no longer negligible for the very high values of N which should be used in high-frequency versions of both types of transistors. While in the p-n-i-p device this high value is maintained throughout the whole base, $N(x)$ drops very rapidly with distance from the emitter in a drift transistor. The average mobility is there-

fore higher in the drift transistor, and consequently, the v and f_a ratios are even greater than $\Delta V/2kT$ and $(\Delta V/2kT)^{3/2}$. Because the regions of different mobilities are connected in series, the correction factor entering into the formula is in the first order given by the ratio of the reciprocal mobilities. One can assume the reciprocal mobility to be a linear function of the impurity content, i.e.,

$$\frac{1}{\mu_p} = \frac{1}{\mu_{p0}} \left[1 + \frac{N(x)}{C_p} \right]. \quad (32)$$

Taking $C_p = 6 \times 10^{16} \text{ cm}^{-3}$, Equation (32) fits the experimental data of Prince.¹¹ For $N_0 = 1.5 \times 10^{17}$ the average mobility in the p-n-i-p device then drops by a factor of 3.5.

In the case of an exponential distribution, the average value of $1/\mu$ is given by

$$\left(\frac{1}{\mu_p} \right) = \frac{1}{\mu_{p0}} \left[1 + \frac{N_0}{C_p} \frac{kT}{\Delta V} \right]. \quad (33)$$

Here, the average mobility decreases only by a factor of 1.3. This means that the improvement in transit time for the drift transistor over the p-n-i-p is actually better by a factor of $3.5/1.3 = 2.7$ than predicted above on the basis of equal mobilities.

3. Base Resistance

According to Equation (30) the internal base resistance of the drift transistor is greater by a factor of $\Delta V/kT$ than that of the p-n-i-p. Again, however, the average mobility in the drift transistor is higher so that the base resistance in the drift transistor is not as much greater as indicated by Equation (30). The situation is different from the above calculation for the transit time in two respects:

1. The majority carrier mobilities rather than the minority mobilities play a role. As has been pointed out by Prince, the majority carrier mobility decreases with increasing impurity content only half as much as the minority carrier mobility. In other words, the electron (hole) mobility in an n-type (p-type) semiconductor is higher than in a p-type (n-type) of the same impurity content. Instead of Equation (32), we have, therefore,

$$\frac{1}{\mu_n} = \frac{1}{\mu_{n0}} \left[1 + \frac{N(x)}{2C_n} \right], \quad (34)$$

¹¹ M. B. Prince, "Drift Mobilities in Semiconductors. I. Germanium," *Phys. Rev.*, Vol. 92, p. 681, November, 1953.

where $C_n = 10^{17}$ cm⁻³ fits Prince's data. For $N_0 = 1.5 \times 10^{17}$, the average electron mobility in the p-n-i-p transistor then drops by a factor of 1.75.

2. Because the regions of different mobility are now in parallel, the weighed average value of the mobility itself, not of its reciprocal, is to be considered. One finds for $\Delta V \gg kT$ in the drift case,

$$\mu_n = \frac{\int N(x) \cdot \mu [N(x)] dx}{\int N(x) dx} = \mu_{n0} \frac{2C_n}{N_0} \ln \left[1 + \frac{N_0}{2C_n} \right]. \quad (35)$$

According to Equations (34) and (35) the mean electron mobility in a drift transistor decreases only by 1.34. The base resistance ratio given above is therefore reduced by a factor of $1.75/1.34 = 1.3$.

4. Conclusions

In order to obtain a really satisfactory comparison between the two types of transistors it would be necessary to distinguish between many different circuit conditions like neutralized and unneutralized circuits, ohmic and inductive load matching and so on. Since a complete comparison is much too complicated for practical purposes, the best practical basis for comparison presently available seems to be the well-known figure of merit^{9,12,13}

$$\text{Figure of merit} = 1/r_{bb} \cdot C_c \tau, \quad (36)$$

where r_{bb} , C_c and τ denote base lead resistance, collector capacitance, and transit time, respectively. As C_c is the same in both structures for equal w_c , we have to compare the product of base resistance and transit time. According to the two preceding sections the figure-of-merit ratio in the two cases is equal to

$$\frac{1}{2} \frac{1 + N_0/C_p}{1 + (kT/\Delta V) (N_0/C_p)} [1 + N_0/2C_n] \frac{2C_n}{N_0} \ln [1 + N_0/2C_n]. \quad (37)$$

Inserting standard numbers this becomes 1.8. The figure of merit for a drift transistor is better by almost a factor of two than for a p-n-i-p with corresponding geometry.

¹² L. J. Giacoletto, "The Study and Design of Alloyed-Junction Transistors," 1954 I.R.E. Convention Record, Part 3, pp. 99-103.

¹³ L. J. Giacoletto, "Study of p-n-p Alloy Junction Transistor from D-C through Medium Frequencies," RCA Review, Vol. XV, p. 506, December, 1954.

As pointed out by Giacoletto,¹³ the basis for the figure of merit used above is no longer valid when the base-lead resistance is small compared to the forward resistance of the emitter junction (see Section IV, 4). Under these conditions, it is believed that the high-frequency performance will be largely determined by the α -cutoff frequency and the capacitance alone. On this basis the drift transistor will be even better compared to the p-n-i-p than indicated by the figure of merit used above because the α -cutoff frequency in a drift transistor is much higher. The detailed calculations under these conditions remain to be carried through.

A VARIABLE-CAPACITANCE GERMANIUM JUNCTION DIODE FOR UHF*†

By

L. J. GIACOLETTO AND JOHN O'CONNELL

RCA Laboratories,
Princeton, N. J.

Summary—A semiconductor junction when biased in the reverse (non-conducting) direction is a capacitance which can be varied by the bias voltage. Such a voltage-variable capacitance has many possible uses. In particular automatic frequency control (a-f-c) at ultra-high frequencies (UHF) is attractive provided the electrical losses are sufficiently small. In this paper the design, construction, and measurement of a junction diode useful throughout the UHF range are considered. The diode consists of a 0.020-inch dot of indium alloyed on to a 0.002-inch-thick wafer of 0.1 ohm-centimeter n-type germanium and mounted with low-inductance connections. It has greater control sensitivity and better electrical characteristics than an electron reactance tube and, in addition, the operating power required by the diode is trivial in comparison with that of an electron tube. Typically, the performance at 6-volt bias is as follows: a capacitance of 38 micro-microfarads; a capacitance change of 3 micromicrofarads per volt; lead inductance of 2.6 millimicrohenries; effective series resistance of 0.5 ohm. Such a variable capacitor has a very high Q at the lower radio frequencies, decreasing to $Q = 17$ at 500 megacycles. Only about 1 microwatt of d-c control power is needed. A Q of 36 at 500 megacycles was obtained in one of the better units having a series resistance of 0.23 ohm.

GENERAL DISCUSSION

A JUNCTION of two dissimilar semiconductors constitutes, generally, a junction diode with a variety of electrical properties. If the diode is biased in the reverse (nonconducting) direction, the mobile charge carriers are moved away from the junction, leaving uncompensated fixed charges in a region near the junction. The width, and hence the electrical charge, of this region (space-charge layer) depends on the applied voltage giving rise to a junction transition capacitance whose small-signal value is shown as C_s in Figure 1. The variation of a bias voltage is accompanied by a change in current and gives rise to a small-signal conductance, g , across C_s as seen in Figure 1. At frequencies below a few hundred kilocycles, the parallel combination of C_s and g are sufficient to define the small-signal characteristics of the diode. At higher frequencies the inductance of the leads, L_o , and the series resistance of the semiconductors, r_s , become significant. The stray lead capacitance may also be significant, but in the units to be described is small enough to be neglected. Thus, the

* Reprinted from *RCA Review*, March 1956.

† This material was presented at the Conference on Electron Devices, Washington, D. C., October 24-25, 1955.

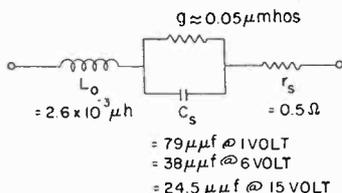


Fig. 1—Diode equivalent circuit.

complete small-signal equivalent circuit of the reverse biased diode applicable from very-low to ultra-high frequencies is as shown in Figure 1.

The factors which must be determined in the design of a diode for ultra-high frequencies are the semiconductor material, its conductivity, the junction area, the height of the semiconductor cylinder between the junction plane and the base plane of the wafer, W (see Figure 2), and the lead length. The semiconductor material is chosen to give the highest Q at high frequencies, and of presently available semiconductors, n-type germanium is well suited. The germanium conductivity should be high to decrease electrical losses but an upper limit is determined by the maximum reverse bias. The desired capacitance for a particular bias voltage determines the area of the junction. The series resistance is determined by the junction area and the dimension W of Figure 2. W should be as small as possible to decrease electrical losses, but a practical minimum value is imposed by the percentage of "short through" rejects during construction of the diodes. The diode leads should have a large diameter and short length to keep the lead inductance to a minimum.

This paper first considers the equations which govern the selection of design factors and which illuminate diode performance. Following this, the construction of a practical diode is described, and its measured

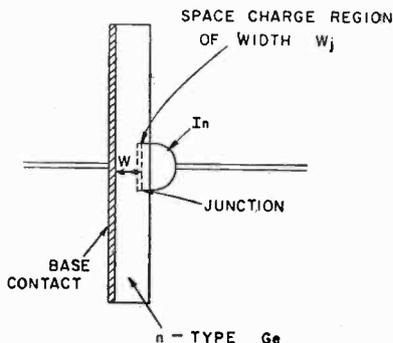


Fig. 2—Diode geometry.

performance is given. An Appendix contains the comparison of measurements with diode theory.

DESIGN RELATIONS

The semiconductor junctions to be considered herein are made by alloying indium (p-impurity) on n-germanium.¹ This process gives an abrupt transition between the p-type indium-enriched germanium and the n-type germanium so that the resulting junction operates in accordance with a theory developed by Schottky.² Similar junctions can be made by alloying n-impurities on p-germanium, or by using a different semiconductor.

Transition Capacitance

The junction diode transition capacitance can be formulated as for a conventional parallel-plane capacitor. Thus

$$C_s = \frac{K \epsilon_0}{W_j} A \quad \text{farads,} \quad (1)$$

where

K = relative permittivity = 16 for germanium,

ϵ_0 = permittivity of free space $\left(\frac{1}{36\pi} \times 10^{-9} \text{ farad/meter} \right)$

A = junction area in square meters, and

W_j = effective width of junction in meters.

The effective junction width is voltage dependent and for the case where the conductivity of the p-type indium enriched germanium is much greater than that of the n-type germanium, is

$$W_j = \sqrt{-(V_0 + V) \frac{2K\epsilon_0\mu_n}{\sigma_n}} \quad \text{meters,} \quad (2)$$

¹ R. R. Law, C. W. Mueller, J. I. Pankove, and L. Armstrong, "A Developmental Germanium P-N-P Junction Transistor," *Proc. I.R.E.*, Vol. 40, pp. 1352-1357, November, 1952.

² For a summary of Schottky's work and complete references, see J. Joffe, "Schottky's Theories of Dry Solid Rectifiers," *Electrical Communication*, Vol. 22, pp. 217-225, 1945. For details of analysis, see W. Shockley, "The Theory of P-N Junctions in Semiconductors and P-N Junction Transistors," *Bell Sys. Tech. Jour.*, Vol. 28, pp. 435-489, July, 1949.

where

μ_n = mobility of electrons in n-germanium in $\frac{\text{meter}^2}{\text{volt sec.}}$

= 0.39 for intrinsic germanium

= 0.30 for $\sigma_n = 1,000$ mhos per meter germanium
used for the present diodes,³

V_0 = contact voltage in volts,

V = applied bias voltage in volts measured with n-germanium
as reference (negative for reverse bias),

σ_n = conductivity of n-germanium in mhos per meter

= 1,000 for the germanium used for the present diode.

The change in diode capacitance with temperature will be negligible, provided the diode voltage is constant, since the temperature-dependent quantity in Equation (2), mobility, cancels out for doped germanium where $\sigma_n \propto \mu_n$.

The fractional change in C_s as a function of V can be formulated from Equations (1) and (2) as

$$\frac{\Delta C_s}{C_s} = - \frac{\Delta V}{2 (V_0 + V)}. \quad (3)$$

The fractional change in C_s is one half the fractional change of applied voltage and is independent of diode dimensions and material properties, except that at small bias voltages the material properties may be significant in determining the contact voltage, V_0 .

Contact Voltage

The voltage, V_0 , of Equation (2) is an internal contact potential developed between the n- and p-type semiconductors. This voltage will be a few tenths of a volt negative and therefore negligible except at very small applied voltages. The contact voltage can be considered as composed of two parts: a contact voltage of the p-type germanium on intrinsic germanium; and a contact voltage of intrinsic germanium on n-type germanium. Since the p-type germanium conductivity approaches that of a metal, the first part of the contact voltage is very

³ M. B. Prince, "Drift Mobilities in Semiconductors I. Ge," *Phys. Rev.*, Vol. 93, pp. 681-687, November 1, 1953.

nearly half of the germanium voltage gap, namely, -0.36 volt. The second half of the contact voltage can be formulated analytically with the aid of usual junction equations in terms of the properties of the intrinsic and n-type germanium. The net result is a value between zero and -0.36 volt. For the diode herein described, the calculation gives a total contact voltage of -0.54 volt, which is in agreement with measurements (see Appendix).

Series Resistance

The diode series resistance, r_s , is due to the wafer of germanium. When the junction diameter is much larger than W , the distance to the base contact (Figure 2), the series resistance can be rather accurately formulated on the basis of a semiconductor cylinder of area A and height W provided that the diode is biased in the reverse direction. Thus

$$r_s = \frac{W - W_j}{\sigma_n A} \quad \text{ohms.} \quad (4)$$

W_j (Equation 2) will generally be negligible in comparison with W . The temperature dependency of r_s will be that of σ_n and will be generally small.⁴ In the temperature range of ± 100 degrees centigrade, r_s will increase about 0.6 per cent per degree centigrade increase in temperature provided $\sigma_n > 100$ mhos per meter.

The series resistance for the forward biased diode is current dependent and decreases rapidly with increasing current.

Lead Inductance

The diode lead inductance is an important factor which cannot be neglected at ultra-high frequencies. The lead inductance can be computed reasonably accurately from the formula⁵ for a straight length of wire:

$$L_0 = 5.08 l \ln \left(\frac{4l}{d} \right) \quad \text{millimicrohenries,} \quad (5)$$

where l is the length of the wire and d is the wire diameter in inches. Diodes constructed as described in this paper require interpretation in determining appropriate values of l and d , since the wire cross

⁴ P. G. Herkart and J. Kurshan, "Theoretical Resistivity and Hall Coefficient of Impure Germanium Near Room Temperature," *RCA Review*, Vol. 14, pp. 427-440, September, 1953.

⁵ Frederick E. Terman, *Radio Engineers' Handbook*, McGraw-Hill Book Co., New York, N. Y., p. 48, 1943.

section is not uniform. The bulk of the inductance will generally be due to the small wire contacting the diode dot.

Maximum Junction Voltage

The maximum reverse bias (breakdown voltage) which may be applied to a junction is determined by an avalanche breakdown phenomenon which causes a precipitous increase in the junction current. The breakdown voltage, V_B , is a function of the impurity density, N_D (impurities per cubic meter), of the semiconductor material. According to measurements⁶ on n-type germanium

$$V_B = -22 \times 10^{16} (N_D)^{-0.725} \approx -2650 (\sigma_n)^{-0.725} \text{ volts,} \quad (6)$$

where σ_n is in mhos per meter.

Q Figure of Merit

It is common practice to use Q (the ratio of series reactance to series resistance) as a figure of merit for capacitors. The diode Q , taking into account both series resistance and shunt conductance, is

$$Q_d = \frac{\omega C_s}{g + r_s (g^2 + \omega^2 C_s^2)}. \quad (7)$$

If the diode parameters are independent of frequency (this is a good assumption particularly in the case of r_s and C_s), Q_d has a maximum when

$$\omega = \frac{1}{C_s} \sqrt{\frac{g}{r_s} (1 + r_s g)}. \quad (8)$$

The maximum Q_d is

$$Q_d (\text{max.}) = \frac{1}{2} \left[r_s g (1 + r_s g) \right]^{-1}. \quad (9)$$

At high frequencies, where the shunt conductance can be neglected, Q_d can be determined from Equations (1), (2), and (4). The result is

$$Q_d = \frac{1}{\omega C_s r_s} = \frac{1}{\omega (W - W_j)} \sqrt{- (V_0 + V) \frac{2\mu_n \sigma_n}{K\epsilon_0}}. \quad (10)$$

⁶ S. L. Miller, "Avalanche Breakdown in Germanium," *Phys. Rev.*, Vol. 99, pp. 1234-1241, August 15, 1955.

From this equation it can be concluded that:

(a) Q_d is independent of junction area and varies inversely with frequency.

(b) For a large Q_d , σ_n should be made as large as possible — the upper limit being determined by the junction breakdown voltage, Equation (6), or by the fact that μ_n decreases as σ_n is made larger. For n-type germanium, the maximum Q_d occurs when σ_n is approximately 10,000 mhos per meter, and for this conductivity, $V_B \approx -3$ volts; W should be as small as construction techniques permit; and a large operating bias should be employed.

(c) Q_d can be increased by selecting a semiconductor material with the largest value of μ/K . Of the two commonly used semiconductor materials (germanium and silicon), n-type germanium has the largest value of μ/K . For this reason the presentation herein has centered around n-type germanium and this material is used in the diode described below.

A loss-free capacitor in series with the diode can be used as an impedance transforming means to increase the effective Q . This series combination decreases the net losses at the expense of the amount of variable capacitance. Placing a capacitor in series with the diode has other advantages such as blocking the d-c bias voltage, limiting a-c voltage across the diode, and eliminating lead inductive reactance at one frequency.

At low frequencies and when the diode is biased in the reverse direction, Q_d is determined by Equations (1) and (2) and the diode conductance, g , which is determined in practice by a leakage conductance. Thus:

$$Q_d = \frac{\omega C_s}{g} = \frac{\omega A}{g} \sqrt{\frac{K\epsilon_0 \sigma_n}{-2\mu_n (V_0 + V)}} \quad (11)$$

Saturation Current

The saturation current is determined by the geometry and surface and volume recombination of hole-electron pairs. A more exact calculation can be carried out, but with germanium of good quality the contribution from volume recombination is small and may be neglected. Further, the contribution from the base contact surface will outweigh the contributions of the free surfaces because of its higher surface recombination velocity, s , and, also, in a structure designed for UHF, because of its proximity to the junction. The saturation current may be found by formulating an equation for the terminal current, $I = -I_s$, which flows when the diode is biased in the reverse direction.

Under these conditions the hole density at the junction will be zero and will increase approximately linearly to a value p_n at the base contact. Analytically this requires $sW/D_p \gg 1$, where D_p is the diffusion constant of holes in n-type germanium ($D_p = 44 \times 10^{-4}$ square meter per second for intrinsic germanium). It is believed that the s of the metallic base contact is large enough to satisfy the inequality if W is greater than 25 microns (0.001 inch). The actual value of s under these conditions does not enter into the calculation. The saturation current is thus determined by the hole density gradient $p_n/(W - W_j)$ and is

$$I_s = q A D_p \frac{p_n}{W - W_j} = \frac{q^2 n_i^2 \mu_n A D_p}{\sigma_n (W - W_j)} \text{ amperes,} \quad (12)$$

where q is the carrier charge ($q = 1.60 \times 10^{-19}$ coulombs) and n_i is the carrier density in intrinsic material ($n_i = 2.4 \times 10^{19}$ carriers per cubic meter for germanium at room temperature).

The magnitude of I_s together with the bias voltage, V , will determine the amount of bias power required to operate the diode in the reverse direction. I_s will increase rapidly with temperature⁷ (approximately 10 per cent per degree centigrade near room temperature) due primarily to the increase in n_i^2 .

Frequency Effects

The frequency dependencies of r_s and C_s are intimately associated with fundamental properties of the semiconductor⁸ provided the diode is biased in the reverse direction. In this event, r_s and C_s will be frequency independent as long as the semiconductor displacement current proportional to $2\pi f K \epsilon_0$ is much less than the conduction current proportional to σ . For impure germanium as used herein of conductivity, $\sigma_n = 1,000$ mhos per meter, the displacement and conduction currents become equal at $f = 1.1 \times 10^{12}$ cycles. Hence, the formulas for r_s and C_s are valid throughout the present range of interest. At higher frequencies, electrical losses due to skin resistance of the leads may be appreciable, but throughout UHF correction therefor is not required.

In the forward direction the diode performance changes rapidly with frequency and different formulas must be used.⁹

⁷ L. J. Giacoletto, "Power Transistors for Audio Output Circuits," *Electronics*, Vol. 27, pp. 144-148, January, 1954.

⁸ William Shockley, *Electrons and Holes in Semiconductors*, D. Van-
Nostrand Co., Inc., New York, N. Y., pp. 199-200.

⁹ W. Shockley, *op. cit.*, pp. 309-318.

Noise

The noise associated with the diode can be determined by introducing into the equivalent circuit of Figure 1 a mean-square voltage generator, $\overline{V}^2 = 4kTr_s\Delta f$, in series with r_s and a mean-square current generator, $\overline{I}^2 = 2q(I + 2I_s)\Delta f$ in shunt with g . In these formulas, k is Boltzmann's constant (1.38×10^{-23} joules per degree Kelvin) and T is temperature in degrees Kelvin. These noise generators are independent of frequency when the diode is biased in the reverse direction ($I = -I_s$). A $1/f$ type of noise generator may also be significant if diode leakage is appreciable.¹⁰

CONSTRUCTION

The diode for which measurements will be given has a parallel plane geometry with axial leads, and is designed to give good performance through ultra-high frequencies. In accordance with the design relations, n-type germanium is used since this semiconductor gives better performance than p-type germanium or p- or n-type silicon. A minimum germanium resistivity of 0.1 ohm-centimeter was selected to permit operation over a useful range of bias voltages. This resistivity material will permit a bias voltage as large as 18 volts. A junction diameter of 20 mils (area of 20×10^{-6} square centimeters) was selected to achieve the desired range of diode capacitance. Lead inductance is kept small by using short, large diameter lead wires.

The completed diode is shown in the photograph of Figure 3. For comparison purposes, a standard 1N82(K3E) UHF mixer diode is also shown.

Parts Preparation

The disassembled view of the diode showing the various component parts is shown in Figure 4. The base stud is made of Kovar[†] or Therlo^{*} to match the germanium thermal expansion and is designed so that the diode can be screwed directly to a chassis. A 0.002-inch depression on the top of this stud aids in positioning the base wafer. The depression is covered with about 0.001 inch of high purity tin-lead-antimony solder. The wafers are chemically etched to 0.002 inch thickness. The dots are punched cylinders of indium 0.015 inch by 0.015 inch diameter. The nickel wire has one end balled and coated with indium and the wire is lightly tinned with a low-melting solder. The top stud has a

¹⁰ W. H. Fonger, "A Determination of $1/f$ Noise Sources in Semiconductor Diodes and Triodes," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 239-295.

[†] Trademark of the Westinghouse Electric Corp., Pittsburgh, Pa.

^{*} Trademark of the Driver-Harris Co., Harrison, N. J.

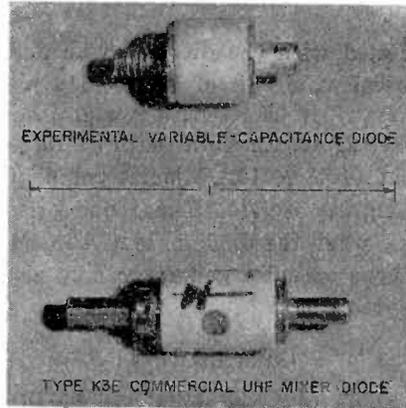


Fig. 3—Photograph of variable capacitance diode and type K3E commercial UHF mixer diode.

0.041-inch-diameter hole through the center, and the inside of this hole is tinned with the same low-melting solder. The top stud is screwed to the ceramic body and bonded in place with Araldite† to form a sub-assembly. The indium dot, germanium wafer, and base stud are also processed as a subassembly. This subassembly is made by mounting the base stud, germanium wafer, and indium dot together with the aid of carbon jigs.¹ The assembled unit is fired for five minutes at 550°C. in an atmosphere of dry hydrogen both to solder the germanium wafer to the stud and to alloy the dot into the germanium.

Assembly

A jig as shown in Figure 5 is used to facilitate assembly and etching. The brass stud ceramic subassembly is inserted and held in place with the set screw. The base stud subassembly is mounted and held in place with a set screw made of insulating material. The wire is inserted through the jig hole, adjusted so that the indium coated balled

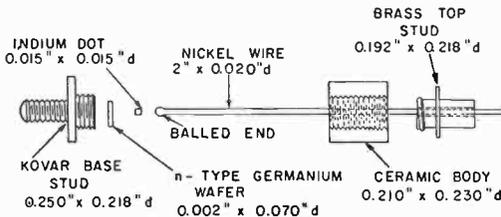


Fig. 4—Disassembled drawing of diode.

† Trademark of Ciba Inc., New York, N. Y.

end is contacting the indium dot, and then held in place with the set screw. A hot jet of hydrogen gas is used to solder the wire and dot together.

The diode is next electrolytically etched by dipping the base stud end only of the jig into the electrolyte. After etching, the diode is washed, dried, and then coated with a protective coating. The threads of the base stud are next coated with Araldite and the brass stud ceramic subassembly lowered and threaded onto the base stud. The wire and brass stud are soldered together with the same low-melting

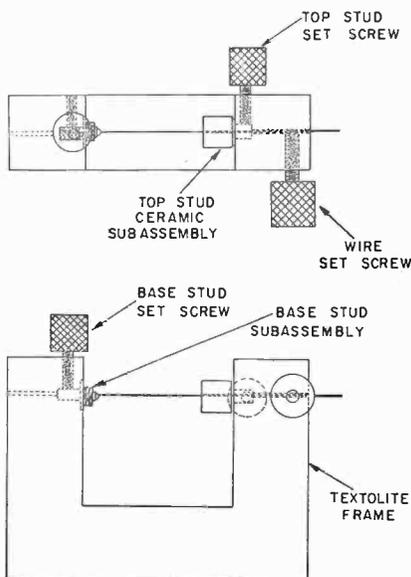


Fig. 5—Diode fabrication jig.

solder used to tin these parts. At this point, the diode is a completed unit and may be removed from the jig.

MEASURED RESULTS ON A TYPICAL UNIT

As a variable reactance device, the characteristics of greatest interest are the capacitance versus reverse bias voltage (measured at low frequencies), the total reactance with variation of bias voltage and frequency, and the loss. Figure 6 shows measured data at 1 and 2 megacycles where lead inductance and loss resistance are negligible. The solid curve was computed from the design formulas and agrees with the data. At the nominal bias of -6 volts, the capacitance is 38

micromicrofarads, and the slope is about 3 micromicrofarads per volt. Over the useful range, up to -16 volts, the capacitance varies inversely with the square root of the bias voltage from about 160 to 25 micromicrofarads.

The terminal reactance as a function of frequency, with the bias as a parameter, is plotted in Figure 7. The measured points agree with the solid curves which were calculated from the theoretical junction capacitance (see Equations 1 and 2) and a lead inductance

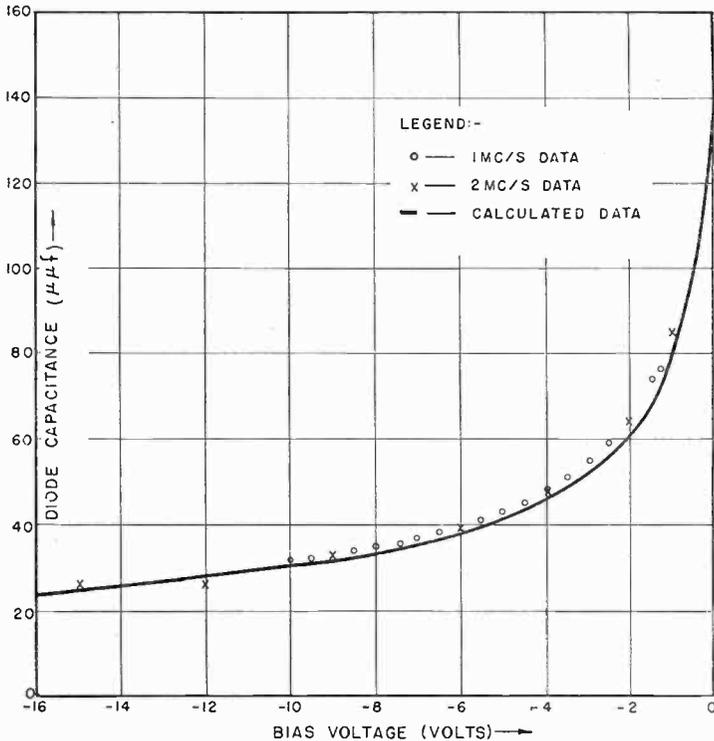


Fig. 6—Diode capacitance versus bias voltage.

of 2.6 millimicrohenries. It is seen that the diode is useful as a controllable reactance well into the UHF range.

The loss resistance was measured in equipment specially designed for the purpose and was found to be approximately constant with frequency and bias variations. Typical measured results correspond to a series resistance of about 0.5 ohm, and the Q at 500 megacycles, with -6 volts bias, is about 17. Since all the data confirms the type of equivalent circuit shown in Figure 1, it is possible to compute the

diode behavior over a wide range of frequencies and operating conditions.

Since the units described in the paper are experimental units, some variation was encountered between units, particularly in the series resistance. In one of the better units, the series resistance was 0.23 ohm resulting in a Q of 36 at 500 megacycles.

In the Appendix, comparison of the measured results with the theoretical design relations is discussed in more detail.

CONCLUSIONS

A junction diode has been described which has attractive operating

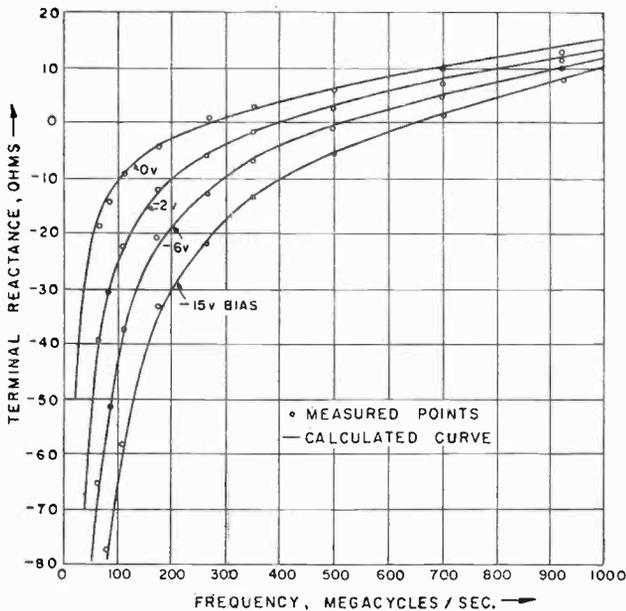


Fig. 7—Terminal reactance versus frequency.

characteristics over a wide range of frequencies including ultra-high frequencies. Diodes of the type described can be used for automatic frequency control,¹¹ frequency selection, mixing, voltage-controlled tuning, frequency modulation, and as dielectric amplifiers. It is found that the diodes follow theory very closely so that the design relations can be used for designing diodes for various applications.

ACKNOWLEDGMENT

The writers would like to acknowledge the following contributions

¹¹ W. Y. Pan and O. Romanis, "Automatic-Frequency Control of Television Receivers using Junction Diodes," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 598-608.

by their colleagues: I. Sochard* assisted in the assembly of the diodes and carried out most of the measurements. R. Braden helped considerably on measurement problems and built the special equipment required for measuring the series resistance. B. Goldstein provided the 1-megacycle data in Figure 6.

APPENDIX—COMPARISON OF MEASUREMENTS WITH DIODE THEORY

It is of interest to compare the junction theory with measured data. The agreement in general is exceptionally good.

Consider first the forward volt-ampere characteristics of the diode. The current flow in an ideal diode is given by

$$I = I_s (e^{\Lambda V} - 1), \quad (13)$$

where Λ is q/kT (equal to 38.8 volts⁻¹ at 25° C). In many conventional diodes, the series ohmic resistance is not negligible and the theoretical exponential diode equation is not found. Figure 8 shows the measured forward characteristics of one of the present diodes. This figure indicates that the characteristics follow rather exactly the ideal diode Equation (13) with the slope indicating an exponential factor $\Lambda = 37.3$, rather than the theoretical value of 38.8 at $T = 25^\circ$ C. In accordance with Equation (13) the zero voltage intercept of the straight line on Figure 8 is the saturation current, $I_s = 0.15$ microampere.

Another method of determining I_s is to evaluate¹² the diode conductance, $g_0 = \Lambda I_s$, at zero bias voltage and current. From the measured $g_0 = 7.2$ micromhos, $I_s = 0.186$ microampere is computed using $\Lambda = 38.8$ for room temperature, $T = 25^\circ$ C. A third method of determining I_s is to measure the diode current for a reverse bias. By this means $I_s = 0.17$ microampere was measured at $V = -1$ volt. The last two methods of measuring I_s include current flow due to any leakage across the junction; this current is not included in the first method of evaluating I_s . Therefore the difference between these values of I_s can be used as a rough estimate of the leakage conductance, which is about 1 micromho by this method. The value is in reasonable agreement with direct measurement of diode conductance with reverse bias which

* Now with the National Bureau of Standards.

¹² L. J. Giacoletto, "Equipments for Measuring Junction Transistor Admittance Parameters for a Wide Range of Frequencies," *RCA Review*, Vol. 14, pp. 269-296, June, 1953.

ranges from 1 to 0.03 micromho for bias voltages ranging from -1 to -6 volts.

With aid of Equation (13), the diode conductance, g , is

$$g = \frac{\partial I}{\partial V} = \Lambda (I + I_s) + \frac{I}{I_s} \frac{\partial I_s}{\partial V}. \tag{14}$$

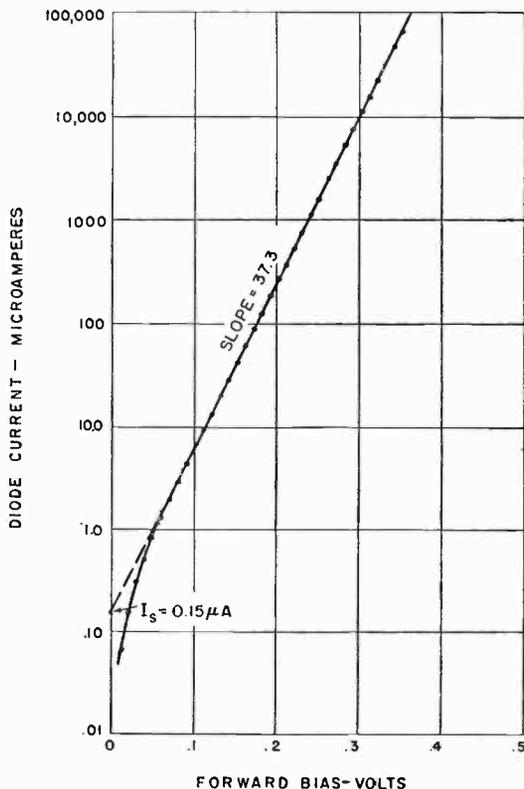


Fig. 8—Diode current versus forward bias voltage.

In the forward direction, the diode conductance is dominated by the first term in Equation (14); in the reverse direction, provided $-V$ is greater than a few tenths of a volt, $I = -I_s$, and g is given by the second term in Equation (6). The formulation of the second term can be carried out with the aid of Equations (2) and (12). When this formulation is carried out and evaluated it is found that the resulting conductance is about an order of magnitude smaller than measured conductances. It is therefore concluded that the measured conductance is due to a leakage conductance, g_l . The diode conductance is of rela-

tively minor importance which is fortunate since g_i is not a designable constant.

When the diode bias is smaller than a few tenths of a volt in the reverse direction or when the diode is biased in the forward direction, a diode diffusion capacitance, C_d , must be added to the transition capacitance (Equations (1) and (2)) to obtain the total junction capacitance. The diode diffusion capacitance when W is small compared with the diffusion length of minority carriers is

$$C_d = \Lambda (I + I_s) \frac{W^2}{2D_p} \text{ farads.} \quad (15)$$

This diffusion capacitance relation is useful to determine the value W , i.e., the thickness of the semiconductor between junction and soldered connection. For example, on the present diode, $C_d = 2120$ micromicrofarads at $I = 285$ microamperes was measured, and $W = 4.1 \times 10^{-5}$ meter is obtained. This value of W agrees reasonably well with 4.6×10^{-5} meter calculated from a resistance measurement (see below).

A plot of diode transition capacitance with bias voltage is shown in Figure 9. The data of this figure is the same as for Figure 6 but has been replotted to indicate that the transition capacitance obeys exactly the formulation of Equations (1) and (2). The slope of the line should,

according to theory, be $\frac{2\mu_n}{A^2 K \epsilon_0 \sigma_n} = -1.06 \times 10^{-4} (\mu\mu f)^{-2} V^{-1}$. This

can be compared with a measured slope of $-1.0 \times 10^{-4} (\mu\mu f)^{-2} V^{-1}$.

The intercept on the voltage axis in Figure 9 gives the contact voltage, $V_0 = -0.5$ volt. This value can be compared with the computed value, -0.54 volt, obtained using the equation

$$V_0 = -0.36 - \frac{1}{\Lambda} \ln \frac{N_d}{n_i} \text{ volt.} \quad (16)$$

N_d is the impurity density in the doped n semiconductor and can be determined from the approximate expression, $\sigma_n \approx q \mu_n N_d$.

The frequencies used for the data of Figure 9 were low enough that lead inductive reactance is negligible. At higher frequencies the lead inductive reactance cannot be neglected. A plot of the type shown in Figure 10 can be used in determining the lead inductance. Here, the intercept at zero net bias voltage corresponding essentially

with zero diode impedance gives the lead inductive reactance. Due to small measurement errors the intercepts do not all give the same lead inductance. An average of the several inductance values indicates $L_0 = 2.6$ millimicrohenries. According to theory this inductance for a wire diameter of 0.020 inch would correspond with a wire length of 0.15 inch. This wire length corresponds closely with the best estimate for the geometry employed. Also, the straight-line plots of Figure 10 indicate that stray capacitance is negligible for the range of operation considered here.

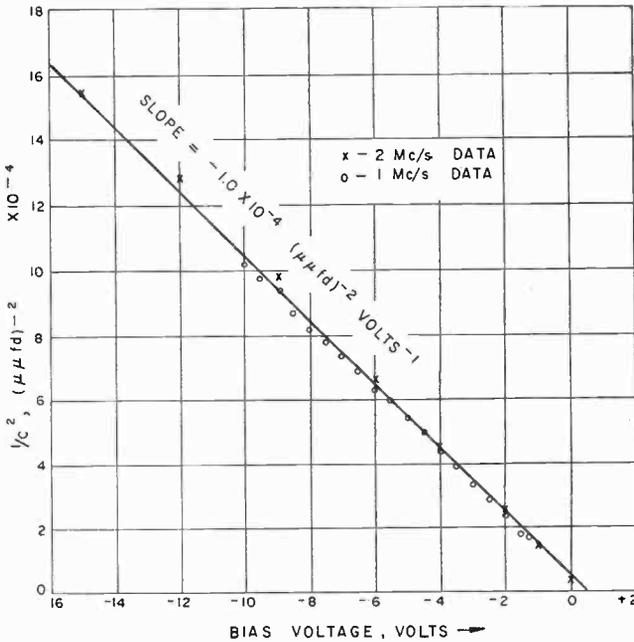


Fig. 9— $1/C^2$ versus bias voltage.

Using the measured value of the series resistance for this particular unit, i.e., $r_s = 0.23$ ohm, Equation (4) indicates that $W = 4.6 \times 10^{-5}$ meter. From the original wafer thickness of 5.1×10^{-5} meter, a penetration during alloying of 0.5×10^{-5} meter (0.2 mil) is indicated. This amount of penetration is somewhat less than expected on the bases of alloying temperature and dot geometry.¹³ Other diodes had series resistance values that were larger than could be explained by Equation (4). Thus it appears that there is an extraneous variable contribution to the series resistance. This is believed to be a contact resistance.

¹³ L. Pensak, "Calculations of Alloying Depth of Indium in Germanium," *Transistor I*, RCA Laboratories, Princeton, N. J., 1956, pp. 112-120.

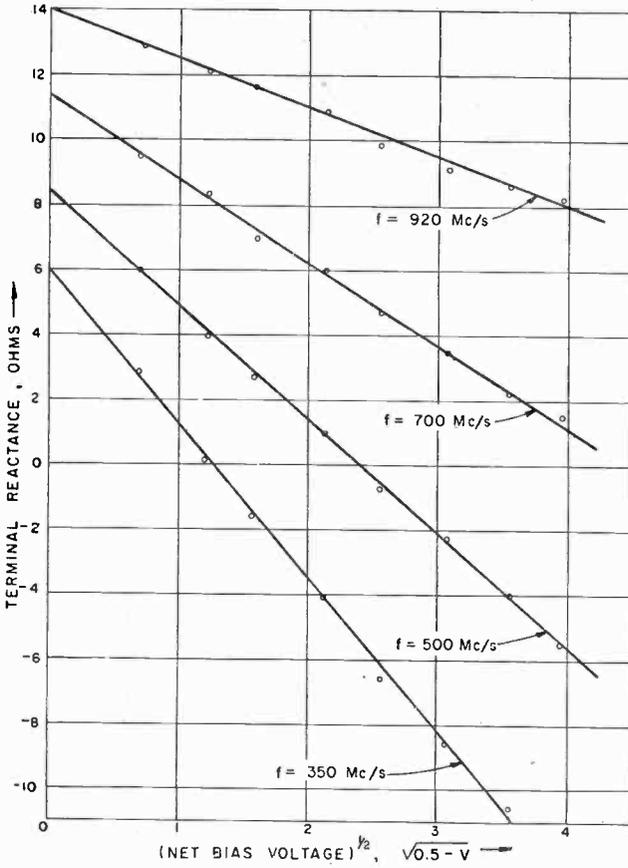


Fig. 10—Terminal reactance versus $(\text{net bias voltage})^{1/2}$.

Calculations of diode Q (Equation 7) using $r_s = 0.23$ ohm, $C_s = 38$ micromicrofarads and $g = 0.03 \times 10^{-6}$ mho corresponding to -6 volts bias indicate values of $Q_d = 175, 36,$ and 18 at $100, 500,$ and $1,000$ megacycles. Equations (8) and (9) indicate a maximum Q of $6,015$ at $f = 1.5$ megacycles.

A DETERMINATION OF $1/f$ NOISE SOURCES IN SEMICONDUCTOR DIODES AND TRIODES*

BY

WILLIAM H. FONGER

RCA Laboratories,
Princeton, N. J.

Summary—Generators representing $1/f$ noise sources in forward- and reverse-biased diodes and in triodes are experimentally determined and related. The generators result from two independent sources. One source, surface noise, has been interpreted by D. O. North as thermal fluctuations in surface potentials. The second, leakage noise, has long been associated with leakage currents across the periphery of the collector. These two sources, with their multi-generator representations and very different dependences on the biases and temperature, account nicely for the complicated $1/f$ noise observed in junction devices.

I. EXTENDED SUMMARY

THIS report describes the low-frequency noise observed in junction diodes and triodes (transistors). Simplified circuit schematics for these devices, valid at low frequencies, are shown in Figures 1(a), 2(a), and 3(a). See Table I for the notation. It will be useful here to appreciate parallels between the forward-biased diode (hereafter called forward diode for simplicity) and the emitter portion of the triode and between the reverse-biased diode (hereafter called reverse diode) and the collector portion of the triode. Thus emitter, base, and collector notations have been used for the diode devices too. The triode schematic is convenient for the common-emitter connection but can be used for any connection.

There are three types of noise in junction devices:

1. *White Noise*. This noise, very well understood theoretically, has its origins in the random atomic-scale processes which underlie macroscopic quantities such as diffusion constant, bulk lifetime, and surface recombination velocity. The name white noise is inappropriate at high frequencies, but will be retained for the low-frequency work here. The generators which must be appended to the schematics to represent this

* The work described here was summarized by the author at the 1955 IRE-AIEE Conference on Semiconductor Device Research, University of Pennsylvania, and by Dr. D. O. North at the Massachusetts Institute of Technology summer program *Noise in Electron Devices*, Cambridge, Mass., July, 1955.

Table I—Notation

da	element of area	q	electronic charge
A	base surface area	r	radial coordinate; pair recombination rate per unit surface area
b	external base point	ζ	three-space coordinate
b'	internal base point	$R_{bb'}$	d-c base-lead resistance
B	collector junction area	$r_{bb'}$	a-c base-lead resistance
c	collector	R_c	large dot radius
D	pair diffusion constant	R_e	small dot radius
e	emitter	R_{eq}	equivalent noise resistance
f	frequency	$R_{eq}^{b'e}$	portion of R_{eq} due to surface-noise $di_{b'e}$ generator
ΔF	contribution to noise factor	R_{eq}^s	portion of R_{eq} due to surface noise
g	pair generation rate	R	see Table III
dg	small-signal pair generation rate	s	surface recombination velocity
g_m	triode transconductance	S	superscript denoting surface noise
G_{eq}	equivalent noise conductance	dv	noise voltage generator
I_b	base current	$V_{b'b}$	base-lead voltage
dI_b	small-signal base current	$dV_{b'b}$	small-signal base-lead voltage
I_c	collector current	$V_{cb'}$	V_c , collector voltage
dI_c	small-signal collector current	$dV_{cb'}$	small-signal collector voltage
I_{cs}	collector saturation current	V_0	thermal-equilibrium potential of n-type region relative to p-type region
I_e	emitter current	$V_{eb'}$	emitter voltage
dI_e	small-signal emitter current	$dV_{eb'}$	small-signal emitter voltage
I_{es}	emitter saturation current	w_e	probability of pair reaching emitter
I_{ec}	emitter-collector saturation current	w_c	probability of pair reaching collector
I_L	leakage current	W	base width
dI_L	small-signal leakage current	X	see Table III
dI_{LN}	noise component of leakage current	Y	see Table III
di	noise current generator	z	axial coordinate
L	$(WD/2s)^{1/2}$, diffusion length beyond R_c	α_{cb}	collector-base current amplification factor
n_n	thermal-equilibrium electron density in base	ρ	base resistivity
p	pair density	ρ_n	thermal-equilibrium base resistivity
p_n	thermal-equilibrium hole density in base	$\psi(f)$	quantity characterizing noisiness of surface
p_p	hole density in emitter		
Δp	$p - p_n$, excess pair density		
p_o	excess pair density near emitter		
dp_o	small-signal excess pair density near emitter		

Typical values for the units studied here are

$$D = 45 \text{ cm}^2/\text{sec}$$

$$L = 0.016 \text{ cm}$$

$$R_c = 0.057 \text{ cm}$$

$$R_e = 0.023 \text{ cm}$$

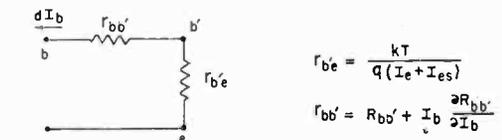
$$W = 0.004 \text{ cm}$$

$$s = 350 \text{ cm/sec}$$

$$\rho_n = 3 \text{ ohm-cm}$$

$$\psi(f) = 10^{-7}/f \text{ cm}^4/\text{sec}$$

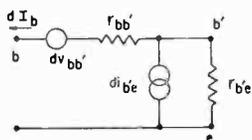
noise are shown in Figures 1(b), 2(b), and 3(b). All generators are uncorrelated. These representations can be derived from North's mesh model¹ or from the argument followed by Giacoletto.² The triode representation can be transformed identically into that described by



(a) SIGNAL SCHEMATIC

$$r_{b'e} = \frac{kT}{q(I_e + I_{es})}$$

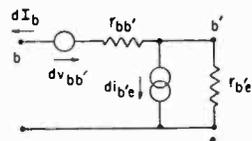
$$r_{bb'} = R_{bb'} + I_b \frac{\partial R_{bb'}}{\partial I_b}$$



(b) WHITE NOISE

$$\overline{di_{b'e}^2} = 2q(I_e + 2I_{es})df$$

$$\overline{dv_{bb'}^2} = 4kTR_{bb'}df$$



(c) 1/f SURFACE NOISE

$$di_{b'e} = q w_e dg$$

$$dv_{bb'} = I_b \left(\frac{\partial R_{bb'}}{\partial I_b} q w_e + m \right) dg$$

Fig. 1—Circuit schematics for forward diode.

van der Ziel,³ provided one uses the correspondences listed in Table II. The older representation of Montgomery and Clark,⁴ while incomplete, is of historical significance.

2. *1/f Surface Noise.* This noise is known empirically from its strongly current-dependent contribution to triode noise factor. The

¹ D. O. North, "A Physical Theory of Noise in Transistors," presented at the 1955 IRE-AIEE Conference on Semiconductor Device Research, University of Pennsylvania; also, "Transistor Noise," presented at the Massachusetts Institute of Technology summer program *Noise In Electron Devices*, Cambridge, Mass., July, 1955.

² L. J. Giacoletto, "A Noise Factor Formulation for Junction Transistors," presented at the 1954 IRE-AIEE Conference on Semiconductor Device Research, University of Minnesota; also, "The Noise Factor of Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 296-308.

³ A. van der Ziel, "Note on Shot and Partition Noise in Junction Transistors," *Jour. Appl. Phys.*, Vol. 25, p. 815, June, 1954, Case (b).

⁴ H. C. Montgomery and M. A. Clark, "Shot Noise in Junction Transistors," *Jour. Appl. Phys.*, Vol. 24, p. 1337, October, 1953.

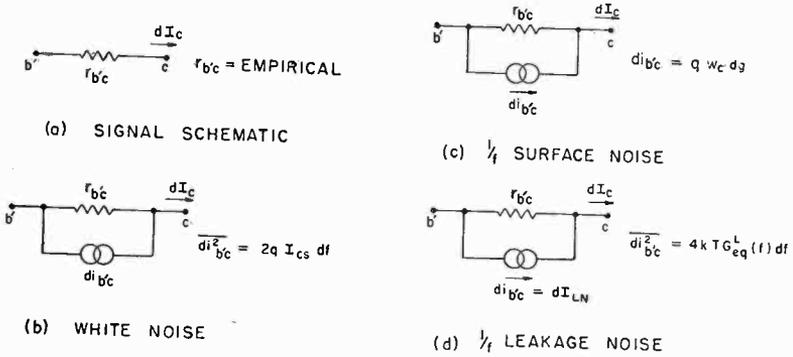


Fig. 2—Circuit schematics for reverse diode.

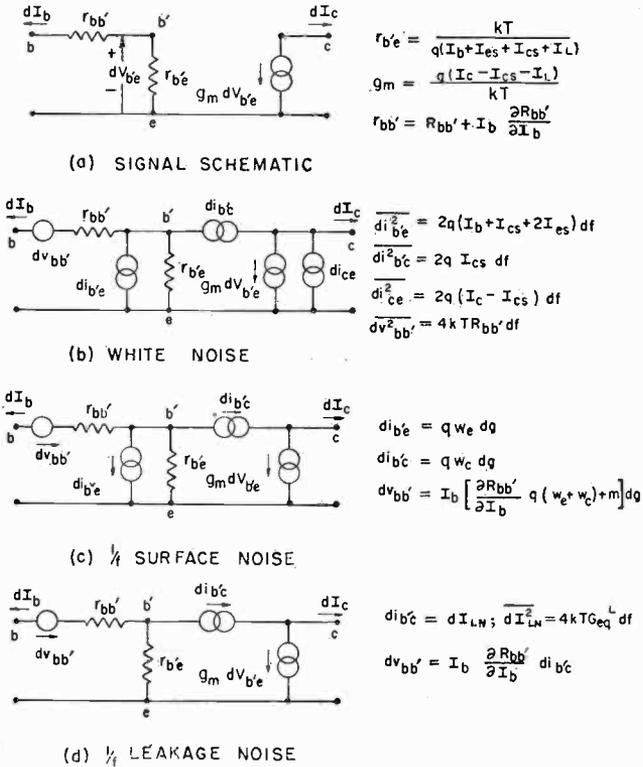


Fig. 3—Circuit schematics for triode.

name stems from a physical model for the noise proposed by North.¹ According to this model, thermal fluctuations in surface potentials perturb the surface recombination velocity, s . This variable s may be

alternatively viewed as a constant s plus an *apparent* noisy surface generation of pairs proportional to the local excess d-c pair density. Because noisy generations at distant surface elements are uncorrelated, one may introduce a phenomenological spectral density $\psi(f)$ defined by

$$\overline{dg^2} = (p - p_n)^2 \psi(f) df, \quad (1)$$

where $\overline{dg^2}/df$ is the spectral density of the pair generation rate per unit *area* and $(p - p_n)$ is the excess d-c pair density. Equation (1) is the noise counterpart of the well-known equation

$$r = (p - p_n) s$$

for average recombination. ψ , the counterpart of s , characterizes the noisiness of a surface.

Disregarding the atomic-scale processes behind s , one can still use it profitably as an empirical constant. This is the viewpoint adopted here with respect to ψ . In so far as Equation (1) may be regarded as a phenomenological model for surface noise, a principal portion of this paper is devoted to showing that this model explains all observed features of surface noise, except those concerning the value of ψ itself. Any detailed model predicting Equation (1) with the correct value for ψ would satisfactorily explain all surface-noise measurements reported here. ψ will be found empirically to vary something like $1/f$, to differ in magnitude among different samples, and to be relatively insensitive to p (decreases somewhat with increasing p).

Table II—Correspondences to Transform van der Ziel's White-Noise Representation into that of Figure 3(b).

van der Ziel	Figure 3(b)
$(1 - \alpha) I_s$	$I_b + I_{cs} + 2I_{es}$
αI_s	$I_c - I_{cs}$
I_{co}	I_{cs}
R_b	$r_{bb'}$
R_b	$R_{bb'}$
$\frac{1}{(1 - \alpha)} r_s$	$r_{b'e}$
$\alpha \frac{1}{r_s}$	g_m
R_g	R_a

As a preliminary to surface noise, consider a low-frequency *point* source of pairs of small-signal strength dg at some position r in the device base. It is shown in Appendix I that the generators which must be appended to the schematics to represent such a source are those shown in Figures 1(c), 2(c), and 3(c). All generators in each of these schematics are perfectly correlated. The $di_{b'e}$ generators result from dg pairs dissociating at the emitter. w_e is the probability that dg pairs reach the emitter. The $di_{b'c}$ generators result from dg pairs dissociating at the collector. The m term of $dv_{bb'}$ results from direct base conductivity modulation by dg pairs. The $\partial R_{bb'}/\partial I_b$ term results indirectly from base conductivity modulation; it is explained in Appendix I.

These perfectly correlated generators represent a point source of pairs. To obtain the generators representing surface noise, one must combine the many uncorrelated generators arising from the many uncorrelated sources described by Equation (1). For example, the summed $di_{b'e}$ generator has the spectrum

$$\sum d\bar{i}_{b'e}^2 = \sum q^2 w_e^2 dg^2 = q^2 \psi(f) df \int_A da w_e^2 (p - p_n)^2,$$

where A is the recombination surface where p departs from p_n . Because m , w_e , and w_c depend on r , the summed generators are imperfectly correlated. Because of the $(p - p_n)^2$ factors, the strengths of all generators increase rapidly with emitter current. ψ characterizes all generators.

3. *1/f Leakage Noise.* When reverse biases are applied to junctions, they sometimes pass appreciable leakage currents in excess of their saturation currents. A somewhat erratic 1/f noise (leakage noise) appears concurrently and increases rapidly with d-c leakage. No detailed model for leakage is available at present.

It is shown in Appendix I that the generators which must be appended to the schematics to represent such a leakage-noise source are those shown in Figures 2(d) and 3(d). The two generators in the triode schematic are perfectly correlated. The $di_{b'c}$ generators portray the noisy flow across the collector junction. $G_{eq}^L(f)$ is the empirical equivalent noise conductance characterizing this flow. G_{eq}^L varies something like 1/f and increases rapidly with d-c leakage. The $dv_{bb'}$ generator, analogous to the similar surface-noise generator, results indirectly from base conductivity modulation.

A satisfactory experimental proof of the complete noise representation above would consist, in the writer's opinion, of the following steps:

1. Verify that the representation accounts for the observed noise at all operating points.
2. Find a sequence of operating points such that each generator separately has an opportunity to dominate the observed noise.
3. At each position of dominance somehow verify the predicted location of the generator in the schematic.

This ambitious program is only partly accomplished, as follows:

1. *White Noise.* This noise is viewed principally as an annoying background. On those occasions when it dominates the observed noise, its total magnitude is checked. No attempt is made to determine the locations of individual generators.

2. *1/f Surface Noise.* This noise should be large when the excess pair density is large, and hence should be much larger in the forward than in the reverse diode. In fact, surface noise has yet to be detected with certainty in the reverse diode. The surface-noise generators

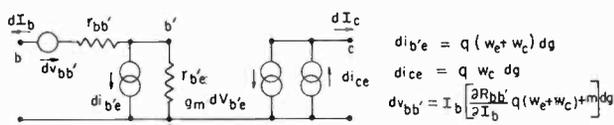


Fig. 4—Alternative representation for triode surface noise.

shown in Figure 1(c) account for the 1/f noise observed in forward diodes at all operating currents. While the $di_{b'e}$ and $dv_{bb'}$ generators should dominate the noise at low and high currents respectively, there is no way to choose between these generators experimentally. Nevertheless, the forward-diode measurements provide excellent evidence that surface noise is excited by pair injection and do prove the existence of a generator in *at least* one of the two positions cited.

The double-ended triode affords better opportunity for locating generators. The $di_{b'e}$ generator may be rewritten as two generators—one in the $b'e$ position, which may then be absorbed into the $di_{b'e}$ generator already there, and one in the ce position—as shown in Figure 4. The resulting di_{ce} generator should contribute negligibly to the observed noise except at extremely low emitter currents and has yet to be detected with certainty. At low and moderate currents, the composite $di_{b'e}$ generator should dominate the noise. This is verified by the observed ratio of input to output surface noise. At high currents, the input noise in the common-emitter connection should be importantly modified by the $dv_{bb'}$ generator. This is verified indirectly. The output noise continues to behave as predicted by $di_{b'e}$. The input noise does not. This difference must be attributed to an additional

$dv_{bb'}$ generator, the one generator which would concurrently contribute negligibly to output noise. Since the input noise behaves in detail as predicted, both the detailed behavior of $dv_{bb'}$ and the oneness of ψ for both $di_{b'e}$ and $dv_{bb'}$ are verified.

These results can be extended to the forward diode as follows: The alloy-junction triode is often constructed with a small emitter and a large collector. One can invert such a unit and use the large junction as the emitter and the small junction as the collector. In this condition, the d-c pair density near recombination surfaces hardly changes if the collector voltage is removed. Thus one can locate, as described above, the two surface-noise generators in the inverted triode. If the collector circuit is then opened, the surface noise across the resulting forward diode should be practically identical with that previously observed at the input of the common-emitter triode. This is observed, and one surely infers that the two generators located in the triode are also present in the forward diode. Moreover, this establishes the oneness of ψ for both diode and triode. The model of course presumes no relation between ψ and the collector voltage.

The generators representing surface noise were derived from a point-source representation. These generators are therefore qualitatively common to *all* models which can ultimately be decomposed into noisy point sources of pairs. Such models for which the spectral intensities of the sources are proportional to $(p - p_n)^2$ also predict the same dependence of noise on current. Generators representing such alternative models differ, however, in quantitative details. The observed oneness of ψ for the normal and inverted triodes (where the d-c pair density distributions are very different) significantly supports the surface-noise model. Two other formally possible source distributions are shown to be deficient in this respect. North originally proposed the surface-noise model only on the basis of its theoretical plausibility plus the empirical fact that triode $1/f$ noise often increases rapidly with current. The experimental establishment of the generator arrangement and of the oneness of ψ for different d-c pair density distributions came later.

In the detailed studies, the m term of the $dv_{bb'}$ generators has been neglected. The computation of this term is tedious, and the final expressions involve constants which are only approximately known. The $\partial R_{bb'}/\partial I_b$ term can be derived empirically from $r_{bb'}$ measurements. If one neglects the m term, such derived values go far toward explaining individual differences in the $dv_{bb'}$ generators for different units, thereby adding credibility to the theory. This result obtains because the error in neglecting m is modest.

3. *1/f Leakage Noise.* The large 1/f noise observed in leaky reverse diodes may surely be represented by a generator across the collector junction, that is, by a $di_{b,c}$ generator.

In the triode connection, this $di_{b,c}$ generator should dominate the noise at low and moderate emitter currents. This is verified by the observed ratio of input to output leakage noise. The observed magnitude of the triode noise is consistent with that expected from reverse-diode determinations of $G_{eq}^L(f)$. The model presumes no relation between G_{eq}^L and the emitter current. At high currents, the input noise in the common-emitter connection should be importantly modified by the dv_{bb} generator. Our measurements proved to be indecisive for establishing this. Detection of a leakage-noise dv_{bb} generator meets the complication that surface noise is large at high currents.

The importances of these sources for present-day junction triodes are as follows: The white noise establishes a minimum noise factor of a few decibels. At 1 kilocycle and at commonly used collector voltages ($V_c \approx 5$ volts), leakage noise should be small compared to the white noise for nearly all units. This situation no longer obtains at $V_c \approx 20$ volts. At 1 kilocycle and at commonly used emitter currents ($I_e \approx 1$ milliamperes), surface noise spreads the observed noise factors over a range of approximately 15 decibels just above the white-noise level. For low-noise applications at subaudio frequencies, units with small G_{eq}^L and ψ should be chosen and should be operated at moderate collector voltages and low emitter currents.

The measurements reported here utilized potted p-n-p alloy-junction germanium transistors with the geometry shown in Figure 23(a) and with the typical characteristics listed in Table I. Noise studies were made with the units connected successively as forward diodes, reverse diodes, normal triodes, and inverted triodes.

II. FORWARD-BIASED DIODES

When operated as a forward diode, the emitter and collector electrodes of the sample were tied together. The tied electrodes will be referred to as the emitter, the n-type region as the base, and the whole unit as a *double* diode. The circuit used for measuring noise in the diode is shown in Figure 5. The observed noise is characterized here by the equivalent noise resistance R_{eq} .

The expected contribution of the white noise to R_{eq} , from Figure 1(b), is

$$R_{eq}^{WH} = \frac{1}{2} \frac{(I_c + 2I_{es})}{(I_e + I_{es})} r_{b'e} + R_{bb'}. \quad (2)$$

The generators required to represent a point source of pairs in the base, the basis for the surface-noise representation, have been shown in Figure 1(c). Contributions from the many uncorrelated sources described by Equation (1) have been summed in Appendix I. The spectrum of the summed $di_{b'e}$ generator is⁵

$$\sum \overline{di_{b'e}^2} = \frac{I_b^2}{\pi s^2} \frac{(R_c^2 - R_e^2 + R_c L)}{(R_c^2 - R_e^2 + 4R_c L)^2} \psi(f) df. \quad (3)$$

Since $I_b \partial R_{bb'}/\partial I_b$ is independent of the position of generation, neglect of the m term makes the summed $dv_{bb'}$ and $di_{b'e}$ generators perfectly correlated. In this approximation, the surface-noise contribution to R_{eq} is

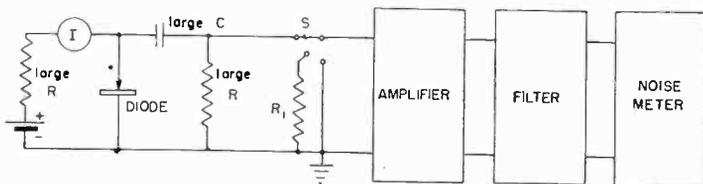


Fig. 5—Circuit used for measuring noise in the forward diode. By judicious use of switch S , the diode open-circuited noise voltage can be compared with that of the resistor R_1 . The amplifier consisted of battery-powered, RC-coupled 6AG5 pentodes. Three different RC filters were used, covering broad frequency ranges around 100, 1,000, and 6,000 cycles, respectively. These filters do not suffice to determine frequency spectra in detail, but readily distinguish between white and $1/f$ noise. The noise meter was a Hewlett-Packard Model 400-C voltmeter.

$$R_{eq}^S = \frac{1}{4kTdf} \left(r_{b'e} + I_b \frac{\partial R_{bb'}}{\partial I_b} \right)^2 \left(\sum \overline{di_{b'e}^2} \right), \quad (3a)$$

$$= \frac{1}{4\pi s^2} \frac{kT}{q^2} \left[\frac{1}{(1 + I_{cs}/I_b)} + \frac{qI_b}{kT} I_b \frac{\partial R_{bb'}}{\partial I_b} \right]^2 \frac{(R_c^2 - R_e^2 + R_c L)}{(R_c^2 - R_e^2 + 4R_c L)^2} \psi(f).$$

$r_{bb'}$ and $R_{bb'}$, the a-c and d-c base-lead resistances respectively, are related by

$$r_{bb'} = R_{bb'} + I_b \frac{\partial R_{bb'}}{\partial I_b}. \quad (4)$$

Figure 6 shows measured values of $r_{bb'}$ versus I_b for six double diodes.

⁵ For the forward diode, I_b and I_c are identical. In particular situations, diode quantities will be identified by names appropriate to the corresponding triode quantities.

Derived values of $R_{bb'}$ and $(qI_b/kT) (I_b \partial R_{bb'}/\partial I_b)$, obtained by numerical integration of Equation (4), are also shown. $I_b \partial R_{bb'}/\partial I_b$ varies only slightly with current in the region of interest and ranges from -20 to -80 ohms for the different units. This result is made plausible in Appendix II. In Equation (3a), the variation of R_{eq}^s with current is contained in the quantity in the square brackets. An expression for this quantity, representative of the units studied here, is

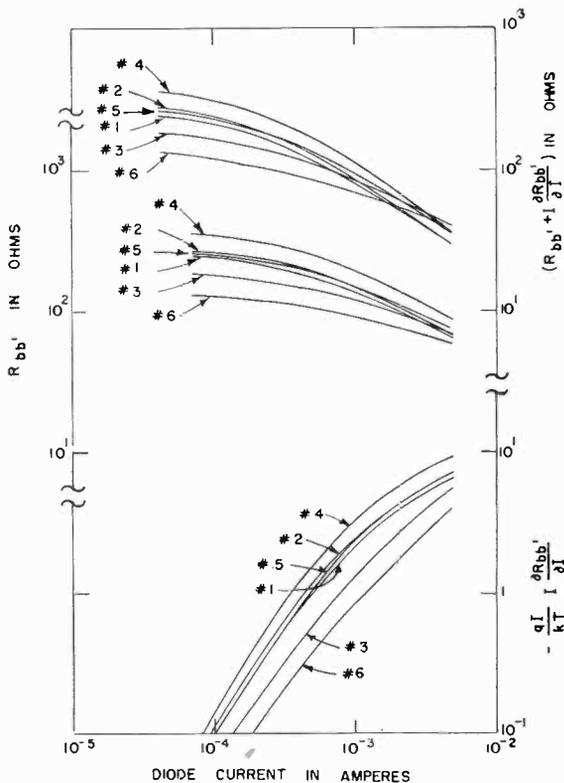


Fig. 6—Measured values of $r_{bb'}$ and derived values of $R_{bb'}$ and $(qI_b/kT) (I_b) (\partial R_{bb'}/\partial I_b)$ for six forward double diodes.

$$\left[\frac{1}{1 + 2 \times 10^{-6}/I_b} - \frac{I_b}{5 \times 10^{-3}} \right]^2 \quad (3b)$$

Expression (3b), together with its component terms, is plotted in Figure 7. The first term, from $di_{b'o}$, dominates the noise at low currents; the second, from $dv_{bb'}$, dominates at high currents. The principal features of the total noise are a rapid rise at low currents, a broad

maximum, a steep minimum where the $di_{b'e}$ and $dv_{bb'}$ generators interfere destructively, and a final rapid rise at high currents.⁶

Figure 8 shows observed R_{eq} 's versus I_b for two double diodes. Measurements at 100, 1,000, and 6,000 cycles are shown by circles, triangles, and squares respectively. The expected contributions from white noise, from Equation (2), are shown by solid curves. The excess noise, the observed noise minus expected white noise, is shown by dotted curves at each frequency. The excess noise varies something like $1/f$ and exhibits the dependence on current expected for surface

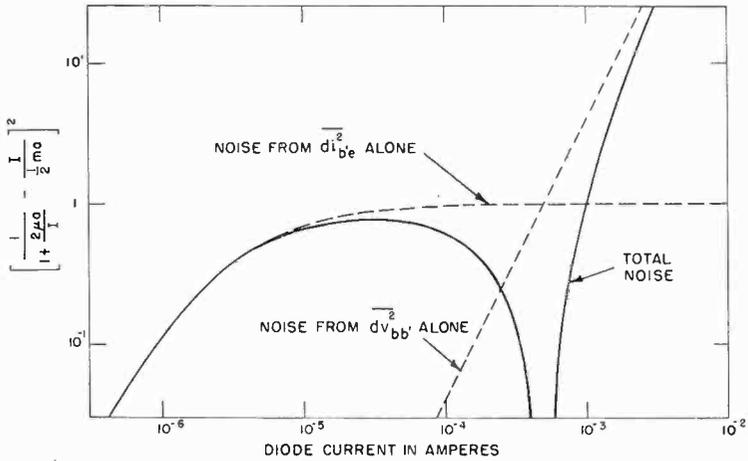


Fig. 7—Expected current dependence of surface-noise contribution to forward-double-diode equivalent noise resistance.

noise. When the $1/f$ noise is small, the white noise checks the theory very well.

From the measured saturation currents and the data of Figure 6, individual predictions of Equation (3a) were computed for the individual units. These are shown by dot-dash curves in Figure 8 with $\psi(f)$ adjusted, in each instance, to match the observed excess noise at 100 cycles. The observed and predicted locations of the interference minima near 10^{-3} ampere agree well. Indeed, considering the neglect of the m term and various other idealizations underlying Equation (3a) (see Appendix I), the observed and predicted dependences of $R_{eq}^{1/f}$ on current agree as well as could reasonably be expected. Similar

⁶ If the m term were included in the theory, the correlation between the $di_{b'e}$ and $dv_{bb'}$ generators would drop from -1 to ≈ -0.9 . The minimum in the total noise would then extend only to a level about one order of magnitude below the saturation level of noise due to $di_{b'e}$ alone. The total noise at currents well above the minimum would increase by a factor of approximately 2.

results were obtained with other units. Investigation of the 1/f noise is understandably more difficult when ψ is small.

$\sum \overline{di_{b,c}^2}$ and R_{eq}^S , see Equations (3) and (3a), are proportional to the ratio $\psi(f)/s^2$. This is a direct consequence of the model. If s is raised without changes in I_b and ψ , $(p - p_n)$ and the noisy genera-

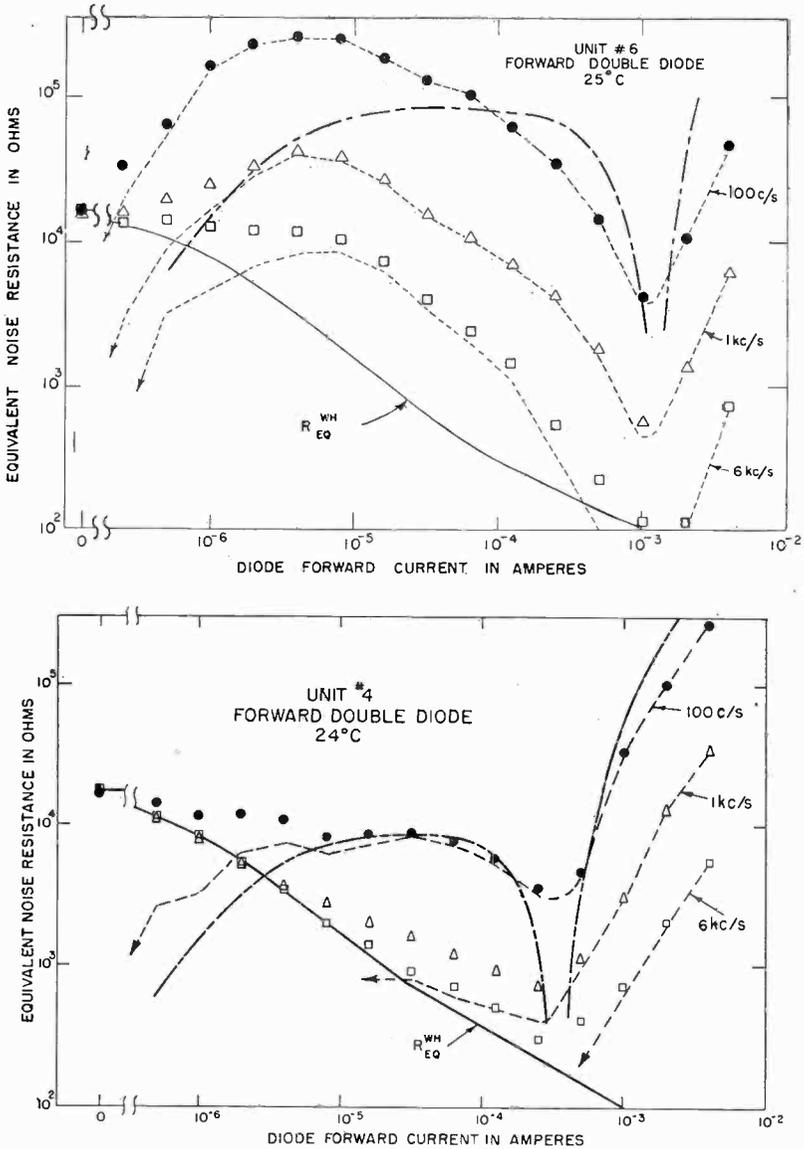


Fig. 8—Observed equivalent noise resistances of two forward double diodes.

tion described by Equation (1) become smaller. s and the geometrical factors entering Equations (3) and (3a) did not vary much among the units studied here.

The complicated dependence of forward-diode $1/f$ noise on current has been plausibly explained by the two well-correlated generators of North's surface-noise model. In the triode connection, where nature kindly permits a noise measurement at the midpoint b' between these two generators, one is able to verify the existence and predicted properties (see Figure 7) of both generators separately.

III. REVERSE-BIASED DIODES

For the reverse-diode connection, the units were connected sometimes as single, sometimes as double diodes. The observed noise is characterized here by the equivalent noise conductance G_{eq} .

The expected contribution of the white noise to G_{eq} , from Figure 2(b), is

$$G_{eq}^{WH} = \frac{1}{2} \frac{q}{kT} I_{cs}. \quad (5)$$

The contribution of the $1/f$ leakage noise is taken, completely phenomenologically, as $G_{eq}^L(f)$.

Figure 9 shows observed 1-kilocycle values of G_{eq} versus collector voltage for six units. The data is shown by solid curves where the white-noise level [Equation (5)] was realized, by dotted curves where it was exceeded. The excess noise varied something like $1/f$ and was somewhat erratic (pops, bursts, etc.), though these considerations are not demonstrated in the figure.

The 1-kilocycle noise data of Figure 9 is replotted in Figure 10 in the form of excess noise, i.e., leakage noise, versus excess current, i.e., d-c leakage in excess of the saturation current. The empirical relationship

$$G_{eq}^L(f) = 4 \times 10^{13} I_L^2/f, \quad (6)$$

where I_L is expressed in amperes and f in cycles, characterizes all data within plus or minus one order of magnitude, irrespective of the operating voltage, dot diameter, or particular unit. Leakage current and noise are not simple functions of reverse voltage, but depend in a complicated way upon the reverse-bias history preceding measurement. They also increase with temperature. Equation (6) applies

approximately, however, irrespective of the temperature or of the unit history preceding measurement.

In most triode applications, leakage noise is comparable to the white noise when $G_{eq}^L \approx 10^{-3}$ mho. In order to obtain triodes with negligible 1-kilocycle leakage noise, d-c leakage must be held, according to Equation (6), to a few tenths of a microampere.

The mechanism of leakage in commercial p-n-p alloy-junction transistors is unknown. Leakage surely occurs at the perimeter of the junction, as it is so sensitive to surface treatments. It is not thought

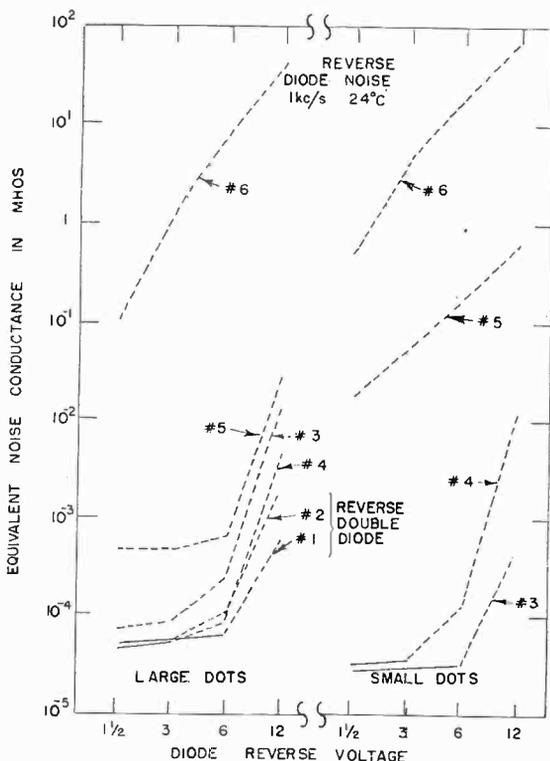


Fig. 9—Observed equivalent noise conductances of six reverse diodes.

to be associated with large-area channels such as have been studied by Brown,⁷ McWhorter and Kingston,⁸ and others in connection with grown junctions. While observed hysteresis effects are reminiscent of what Brown has called "conditioning" a channel, there is no extra

⁷ W. L. Brown, "n-Type Surface Conductivity on p-Type Germanium," *Phys. Rev.*, Vol. 91, p. 518, August, 1953.

⁸ A. L. McWhorter and R. L. Kingston, "Channels and Excess Reverse Current in Grown Germanium p-n Junction Diodes," *Proc. I.R.E.*, Vol. 42, p. 1376, September, 1954.

junction capacitance associated with large leakage in the present case. The junction small-signal capacitance remains a single-valued $V^{-1/2}$ function of reverse bias, as expected from Schottky's theory,⁹ even though leakage hysteresis may be exploited to alter leakage currents at given voltages by tens of microamperes or hundreds of percents.

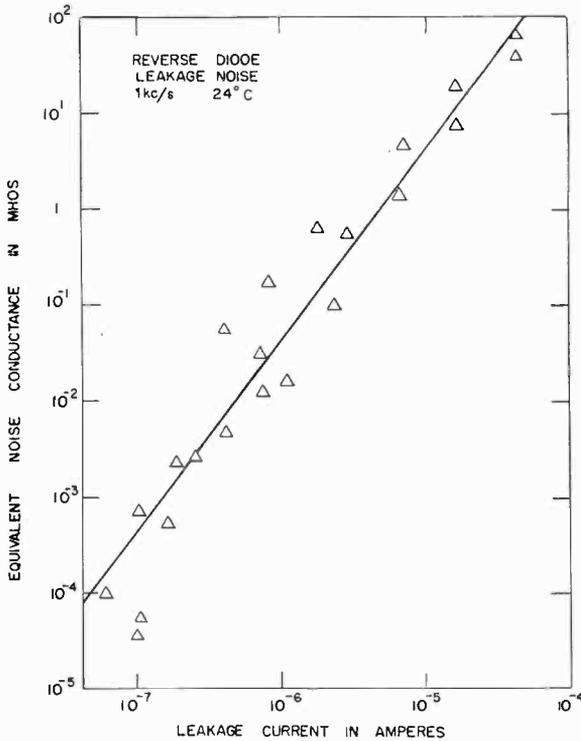


Fig. 10—Excess equivalent noise conductances of six reverse diodes.

IV. TRIODES

A. General

The common-emitter circuit used for measuring noise at the triode output and input is shown in Figure 11. The load impedance R_L and source impedance R_G were 5,000 and 475 ohms respectively for all measurements. The observed output noise is characterized here by the conventional noise factor F_{OUT} . The observed input noise is characterized either by the equivalent noise resistance R_{eq} of the input considered as a two-terminal device or by an input noise factor F_{IN} . In

⁹ W. Schottky, "Vereinfachte und erweiterte Theorie der Randschichtgleichrichter," *Zeits. f. Phys.*, Vol. 118, p. 539, 1941-1942.

analogy with F_{OUT} , F_{IN} is defined as the ratio of the total noise power per unit bandwidth available at the *input* terminals to that portion of this power due to thermal noise in the *source* resistance R_G . F_{IN} is a convenient quantity experimentally; R_{eq} is convenient for bringing out the correspondence between the triode input and the forward diode. For the simplified triode of Figure 3(a) in the common-emitter connection, R_{eq} and F_{IN} are related by

$$F_{IN} - 1 = \Delta F_{IN} = R_G R_{eq} / (r_{bb'} + r_{b'e})^2.$$

In the common-emitter connection, the expected contributions of the white noise to F_{OUT} and F_{IN} , from Figure 3(b), are

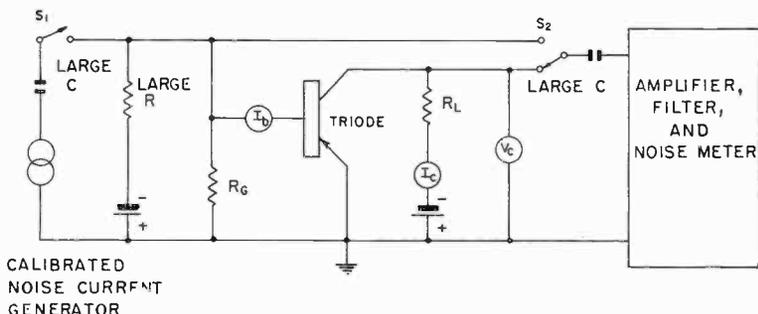


Fig. 11—Common-emitter circuit used for measuring noise at the triode input and output. The calibrated noise current generator is in parallel with the noise current generator, not shown here, which represents thermal noise in the source resistance R_G . The input and output noise factors are determined by use of the two switches.

$$F_{OUT}^{WH} = 1 + \frac{R_{bb'}}{R_G} + \frac{q}{2kT} R_G \left[1 + \frac{r_{bb'}}{R_G} \right]^2 \left[(I_b + I_{cs} + 2I_{cs}) + \frac{(I_c - I_{cs})}{(g_m R)^2} + \left(1 + \frac{1}{g_m R} \right)^2 I_{cs} \right], \tag{7}$$

$$F_{IN}^{WH} = 1 + \frac{R_G}{(1 + r_{bb'}/r_{b'e})^2} \left[\frac{R_{bb'}}{(r_{b'e})^2} + \frac{q}{2kT} (I_b + 2I_{cs} + 2I_{cs}) \right].$$

F_{IN}^{WH} was never large enough compared to the background amplifier noise to verify accurately, but it was formally subtracted from the total input noise when 1/f noise was under study.

The generators required to represent a point source of pairs in the base, the basis for the surface-noise representation, have been shown in Figure 4. Contributions from the many uncorrelated sources de-

scribed by Equation (1) have been summed in Appendix I. The spectrum of the summed composite $di_{b'e}$ generator is¹⁰

$$\left. \begin{aligned} \sum \overline{di_{b'e}^2} &= \frac{(I_b + I_{cs} + I_L)^2}{\pi s^2} \frac{(R_c^2 - R_e^2 + R_c L)}{(R_e^2 - R_e^2 + 4R_e L)^2} \psi(f) df, \\ &= \frac{(I_c - I_{cs} - I_{cc} - I_L)^2}{\pi} \left(\frac{W}{D}\right)^2 \frac{(R_c^2 - R_e^2 + R_c L)}{R_e^4} \psi(f) df, \end{aligned} \right\} \begin{array}{l} \text{Inverted} \\ \text{Triode} \end{array} \quad (8)$$

$$\left. \begin{aligned} \sum \overline{di_{b'e}^2} &= \frac{(I_b + I_{cs} + I_L)^2}{8 s^2} \frac{1}{R_e W} \psi(f) df, \\ &= 2 \frac{(I_c - I_{cs} - I_{cc} - I_L)^2}{\pi^2 D^2} \left(\frac{W}{R_e}\right)^3 \psi(f) df. \end{aligned} \right\} \begin{array}{l} \text{Normal} \\ \text{Triode} \end{array} \quad (8a)$$

The alternative forms arise as follows: $\sum \overline{di_{b'e}^2}$ is proportional to $\psi(f) df$, to the square of the excess d-c pair density p_0 just inside the emitter, and to the geometrical extent of this pair density distribution — an area something like πR_e^2 for the inverted triode, $2\pi R_e W$ for the normal triode. In shifting to directly observed quantities, one may express p_0 in terms of either base current ($I_b \propto sp_0$) or collector current ($I_c \propto D p_0/W$). The former is useful for bringing out the correspondence between the triode and forward diode. See Equation (3). The latter is useful too since the triode operating point is usually specified by I_c rather than I_b .

It is demonstrated in Appendix I that the summed di_{cc} generator may be neglected here. The m term of $dv_{bb'}$ will again be neglected, with consequences similar to those discussed in Section II for the forward diode. In this approximation, in the common-emitter connection the surface-noise contributions to F_{OUT} and F_{IN} are

$$\Delta F_{\text{OUT}}^S = \frac{R_G}{4kTdf} \left(1 + \frac{r_{bb'}}{R_G}\right)^2 \left[1 - \frac{I_b}{(R_G + r_{bb'})} \frac{\partial R_{bb'}}{\partial I_b}\right]^2 \left(\sum \overline{di_{b'e}^2}\right), \quad (8b)$$

$$\Delta F_{\text{IN}}^S = \frac{R_G}{4kTdf} \frac{1}{(1 + r_{bb'}/r_{b'e})^2} \left[1 + \frac{I_b}{r_{b'e}} \frac{\partial R_{bb'}}{\partial I_b}\right]^2 \left(\sum \overline{di_{b'e}^2}\right).$$

¹⁰ For the inverted triode, emitter and collector notations refer to those electrodes actually used as emitters and collectors—the functional electrodes. There are two exceptions: R_c and R_e continue to mean the radii of the large and small dots respectively.

Both formulas apply formally, as do the schematics of Figures 3 and 4, to both the inverted and normal triodes.

The relative contributions of $dv_{bb'}$ and $di_{b'e}$ to F_{OUT} and to F_{IN} are contrasted by the $\partial R_{bb'}/\partial I_b$ and unity terms respectively in the square brackets of Equations (8b). The quantity $I_b \partial R_{bb'}/\partial I_b$ for the inverted triode is practically identical with the corresponding quantity for the forward double diode and hence has a value of about -50 ohms in the region of interest for the units studied here. The same quantity for the normal triode, while formally different analytically (see Appendix II), should have, and indeed proved experimentally to have, a comparable value. $(R_G + r_{bb'})$ was always much larger than 50 ohms here, so the contribution of $dv_{bb'}$ to F_{OUT} was negligible. On the other hand, $r_{b'e}$ was sometimes as small as 10 ohms here, in which case $dv_{bb'}$ should dominate ΔF_{IN}^S . For $r_{b'e} \approx 50$ ohms, $di_{b'e}$ and $dv_{bb'}$ interfere strongly at the input, and ΔF_{IN}^S should experience a steep minimum analogous to that discussed in the forward diode section.¹¹

The generators required to represent leakage noise have been shown in Figure 3(d). In the common-emitter connection, the leakage noise contributions to F_{OUT} and F_{IN} are

$$\Delta F_{OUT}^L = \left(1 + \frac{r_{bb'}}{R_G} \right)^2 \left[1 + \frac{1}{g_m R} - \frac{I_b}{(R_G + r_{bb'})} \frac{\partial R_{bb'}}{\partial I_b} \right]^2 R_G G_{ea}^L(f),$$

$$\Delta F_{IN}^L = \frac{1}{(1 + r_{bb'}/r_{b'e})^2} \left[1 + \frac{I_b}{r_{b'e}} \frac{\partial R_{bb'}}{\partial I_b} \right]^2 R_G G_{ea}^L(f).$$

Other than $\sum \overline{di_{b'e}^2}/4kTdf$ being replaced by G_{ea}^L , the formulas differ from those characterizing surface noise [Equations (8b)] only by the additional $1/g_m R$ term in the bracket of ΔF_{OUT}^L . An extra term arises because the $di_{b'e}$ leakage-noise generator makes *two* contributions to F_{OUT} : one, its direct contribution at c ; the other, its contribution at b' which is amplified and fed to the output through the g_m generator. The composite $di_{b'e}$ surface-noise generator makes only the latter type of contribution. When the triode gain is high, $g_m R \gg 1$. The contribution of the $dv_{bb'}$ leakage-noise generator to F_{OUT} was negligible here. The contribution of $dv_{bb'}$ to F_{IN} should be important only at high currents. ΔF_{IN}^L too should experience a steep minimum when $r_{b'e} \approx 50$ ohms.

The contributions to common-emitter F_{OUT} and F_{IN} and the ratios

¹¹ In so far as $dv_{bb'}$ does contribute to F_{OUT} , it always *increases* the output noise. $dv_{bb'}$ and $di_{b'e}$ are located so as to interfere destructively at b , but constructively at b' . See Figure 4 and recall that $\partial R_{bb'}/\partial I_b$ is negative.

$\Delta F_{IN}/\Delta F_{OUT}$ appropriate to surface and leakage noise respectively are listed in Table III, where the contributions of the $dv_{bb'}$ generators to F_{OUT} have been neglected. Contributions to common-emitter F_{OUT} and F_{IN} and ratios $\Delta F_{IN}/\Delta F_{OUT}$ appropriate to *single* generators with arbitrary spectra in the $b'e$, ce , $b'e$, and bb' positions respectively are also listed. Subsequent studies to determine the locations of triode noise sources will use the $\Delta F_{IN}/\Delta F_{OUT}$ ratios of Table III extensively. The noise factors involve $r_{b'e}$, $r_{bb'}$, $I_b \partial R_{bb'}/\partial I_b$, and $g_m R$. $r_{b'e}$ was computed from the formula listed in Figure 3(a); $r_{bb'}$ was obtained by a bridge measurement;¹² $I_b \partial R_{bb'}/\partial I_b$ was obtained by numerical integration of Equation (4); and $g_m R$ was computed from the observed triode voltage gain.

B. Examples of Surface Noise

The measurements reported in this section were obtained with units with negligible leakage noise.

Observed common-emitter F_{OUT} 's versus current are shown for two normal triodes in Figure 12 and for two inverted triodes in Figure 13. Measurements at 100, 1,000, and 6,000 cycles are shown by circles, triangles, and squares respectively. The expected contributions from white noise, from Equation (7), are shown by solid curves. The excess noise, the observed noise minus expected white noise, is shown by dotted curves at each frequency. The excess noise varies something like $1/f$ and exhibits the dependence on current expected for surface noise. When the $1/f$ noise is small, the white noise checks the theory fairly well.

Observed common-emitter $\Delta F_{IN}^{1/f}$'s, observed F_{IN} 's minus expected white noise contributions, are shown for two normal triodes in Figure 14 and for two inverted triodes in Figure 15. Observations at 100, 1,000, and 6,000 cycles are shown by circles, triangles, and squares respectively.

The solid curves in Figure 14 were obtained by multiplying observed $\Delta F_{OUT}^{1/f}$'s for the appropriate units by the $\Delta F_{IN}/\Delta F_{OUT}$ ratios appropriate to surface noise (see Table III). Considering the neglect of the m term and various other idealizations underlying the theory (see Appendix I), these solid curves agree with the observed $\Delta F_{IN}^{1/f}$'s as well as could reasonably be expected.

The dotted curves in Figure 14 were obtained by multiplying the observed 100-cycle $\Delta F_{OUT}^{1/f}$'s by the $\Delta F_{IN}/\Delta F_{OUT}$ ratios appropriate to a generator in the $b'e$ position (see Table III). At low currents, where

¹² L. J. Giacoletto, "Equipments for Measuring Junction Transistor Admittance Parameters for a Wide Range of Frequencies," *RCA Review*, Vol. XIV, p. 269, June, 1953.

Table III—Noise Factor Contributions from Various Noise Sources.

NOISE SOURCE	ΔF_{OUT}	ΔF_{IN}	$\Delta F_{IN}/\Delta F_{OUT}$
SURFACE NOISE	$\frac{R_G (\sum \overline{d_{bb',e}^2})}{4kTXdf}$	$\frac{R_G Y (\sum \overline{d_{bb',e}^2})}{4kTdf} \left(1 + \frac{I_b}{r_{b'e}} \frac{\partial R_{bb'}}{\partial I_b}\right)^2$	$XY \left(1 + \frac{I_b}{r_{b'e}} \frac{\partial R_{bb'}}{\partial I_b}\right)^2$
LEAKAGE NOISE	$\frac{R_G G_{eq}^L}{X} \left(1 + \frac{1}{g_m R}\right)^2$	$R_G G_{eq}^L Y \left(1 + \frac{I_b}{r_{b'e}} \frac{\partial R_{bb'}}{\partial I_b}\right)^2$	$XY \left(1 + \frac{I_b}{r_{b'e}} \frac{\partial R_{bb'}}{\partial I_b}\right)^2 \left(1 + \frac{1}{g_m R}\right)^{-2}$
$d_{ib'e}$	$\frac{R_G \overline{d_{ib'e}^2}}{4kTXdf}$	$\frac{R_G Y \overline{d_{ib'e}^2}}{4kTdf}$	XY
$d_{ib'c}$	$\frac{R_G \overline{d_{ib'c}^2}}{4kTXdf} \frac{1}{(g_m R)^2}$	0	0
$d_{ib'c}$	$\frac{R_G \overline{d_{ib'c}^2}}{4kTXdf} \left(1 + \frac{1}{g_m R}\right)^2$	$\frac{R_G Y \overline{d_{ib'c}^2}}{4kTdf}$	$XY \left(1 + \frac{1}{g_m R}\right)^{-2}$
$d_{ib'b'}$	$\frac{\overline{d_{ib'b'}^2}}{4kTR_G df}$	$\frac{Y \overline{d_{ib'b'}^2}}{4kTR_G df} \left(\frac{R_G}{r_{b'e}}\right)^2$	$Y \left(\frac{R_G}{r_{b'e}}\right)^2$

In these expressions it is convenient to use quantities R, X, and Y defined by

$$\frac{1}{R} = \frac{1}{(R_G + r_{bb'})} + \frac{1}{r_{b'e}}; \quad X = \left(1 + \frac{r_{bb'}}{R_G}\right)^{-2}; \quad Y = \left(1 + \frac{r_{bb'}}{r_{b'e}}\right)^{-2}$$

TRANSISTORS I

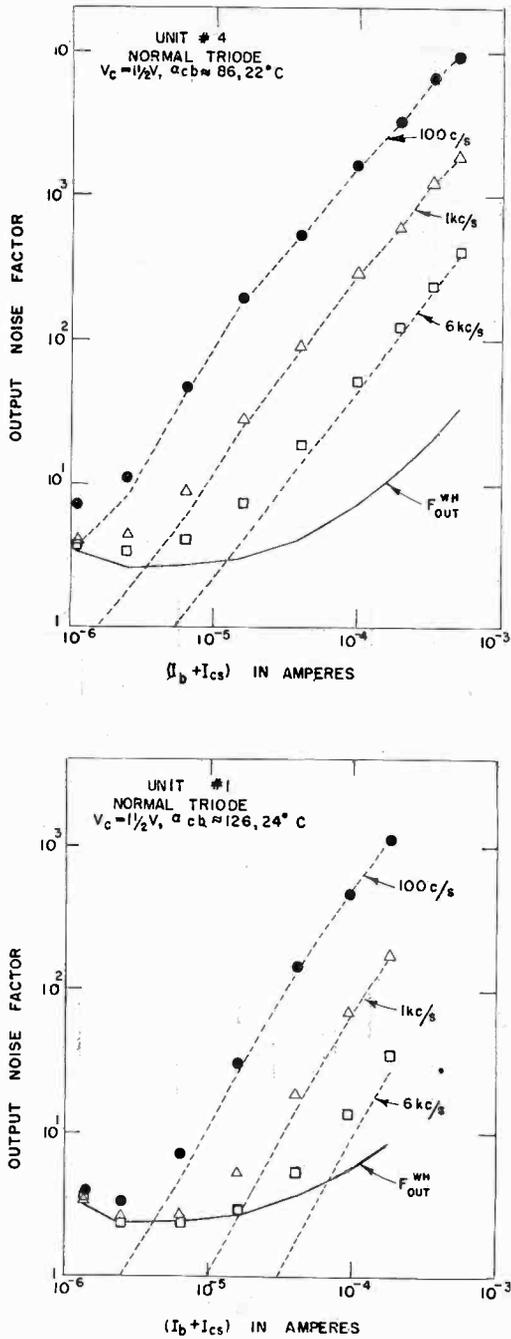


Fig. 12—Observed output noise factors of two normal triodes.

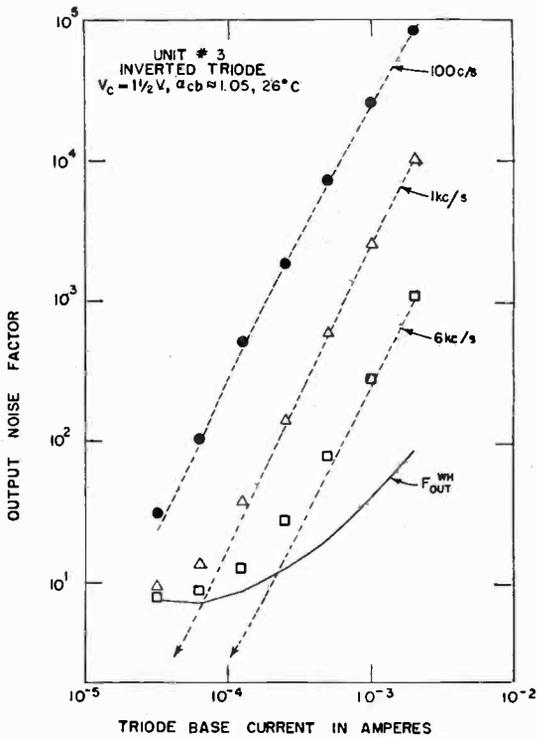
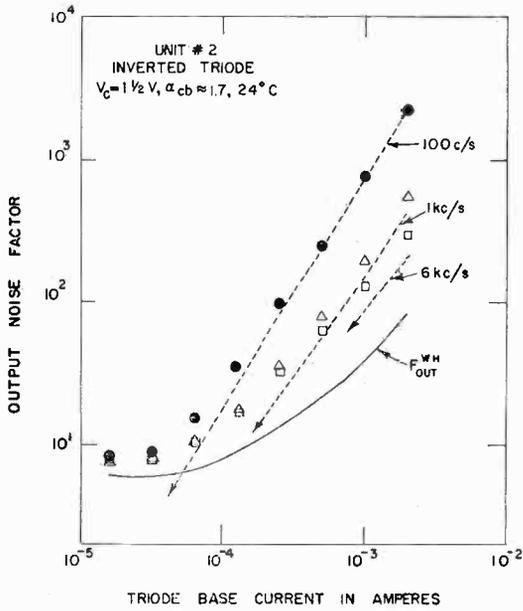


Fig. 13—Observed output noise factors of two inverted triodes.

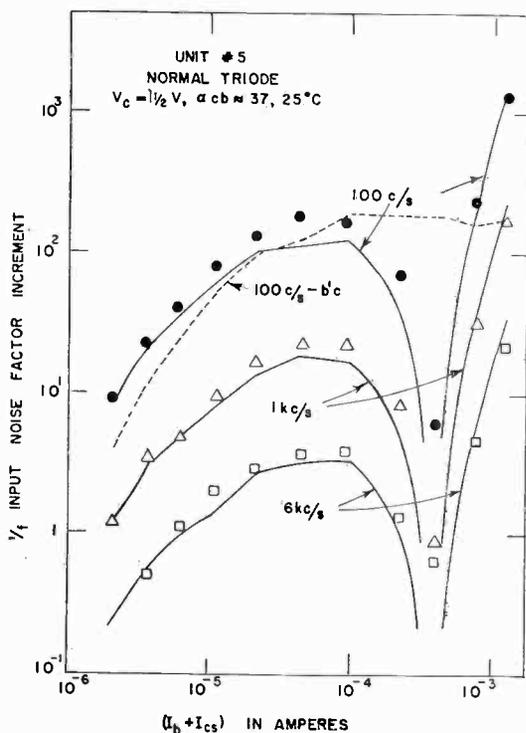
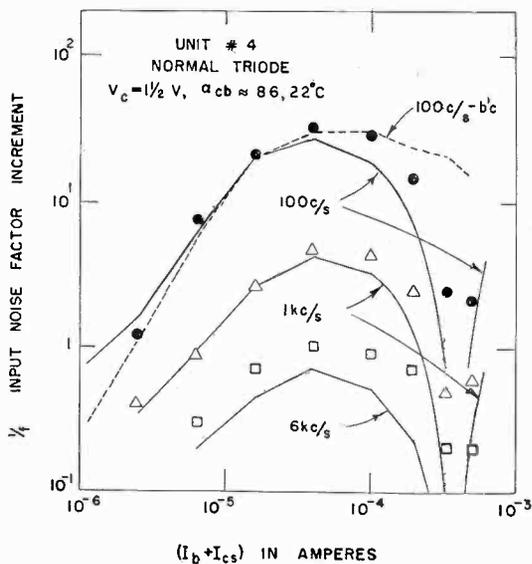


Fig. 14—Observed input noise factors of two normal triodes.

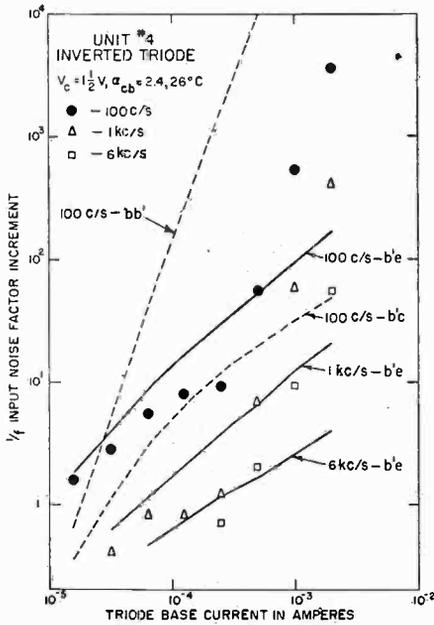
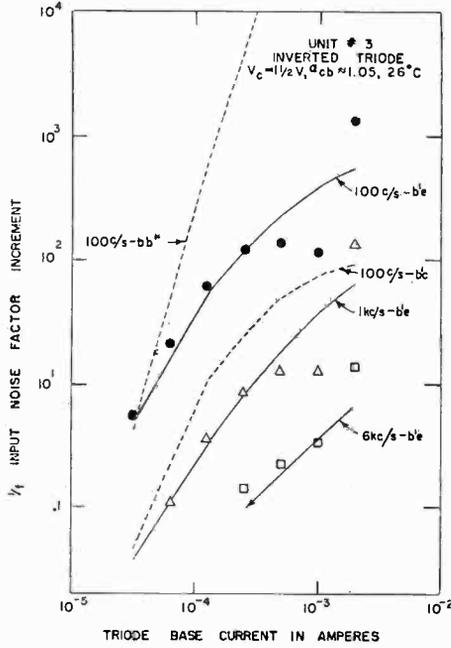


Fig. 15—Observed input noise factors of two inverted triodes.

surface noise is represented essentially by a $di_{b'e}$ generator, the differences between the dotted curves and 100-cycle solid curves indicate one's ability to distinguish, via the $\Delta F_{IN}/\Delta F_{OUT}$ ratio, between $di_{b'c}$ and $di_{b'e}$ noise representations. These differences are small, though the observed $\Delta F_{IN}^{1/f}$'s are more in agreement with the $di_{b'e}$ representation. At high currents, where discrimination between $di_{b'c}$ and $di_{b'e}$ representations is really poor, the departures of the 100-cycle solid curves from the dotted curves indicate how, according to the surface-noise model, the $dv_{bb'}$ generator should modify the input noise from that due to the $di_{b'e}$ generator alone. This modification accounts substantially for the observed dips in $\Delta F_{IN}^{1/f}$. In particular, with the aid of empirically derived $I_b \partial R_{bb'}/\partial I_b$, the model predicts accurately the locations of the interference minima. Similar results were obtained for other units. The analysis is understandably more difficult when the $1/f$ noise is small. When α_{cb} is large, it is sometimes inconvenient to attain the high base currents required to observe the postminimum rise in ΔF_{IN}^s .

The solid and dotted curves in Figure 15 were obtained by multiplying observed $\Delta F_{OUT}^{1/f}$'s for the appropriate inverted triodes by the $\Delta F_{IN}/\Delta F_{OUT}$ ratios appropriate to single generators in the $b'e$, $b'c$, and bb' positions, as indicated. [For the simplified triode of Figure 3(a), a di_{ce} generator produces no noise at the input in the common-emitter connection.] The differences between these curves indicate one's ability to distinguish, via the $\Delta F_{IN}/\Delta F_{OUT}$ ratio, between alternative single-generator noise representations. At low currents, the observed $\Delta F_{IN}^{1/f}$'s coincide decisively with the solid ($b'e$) curves, verifying the prediction of the surface-noise model.

The decision here for a $di_{b'e}$ representation at low currents properly encourages a similar representation for surface noise in the normal triode, where the discrimination between $di_{b'e}$ and $di_{b'c}$ representations is poorer. Comparing Figures 12 and 13 and Figures 14 and 15, we conclude that the same noise phenomenon is being studied in both the normal and inverted triodes.¹³

¹³ The interference minimum in ΔF_{IN}^s is more pronounced for the normal triode than for the inverted triode. This is explained as follows: The summed $dv_{bb'}$ and composite $di_{b'e}$ generators are imperfectly correlated only because of the m term. For the forward double diode or inverted triode, noise pairs generated well beyond R_c live approximately 10 microseconds ($\tau \approx W/s$) for the units studied here, contribute noticeably to m during this period, and yet contribute only weakly to the $di_{b'e}$ generator ($w_s + w_c \ll 1$). For the normal triode, noise pairs generated well beyond R_c live approximately $1/2$ microsecond ($\tau \approx W^2/D$) for the units studied here, therefore contribute less to m during this period, and yet contribute fully to the $di_{b'e}$ generator ($w_s + w_c = 1$). Thus the normal-triode m term is both smaller in size and better correlated with the summed $di_{b'e}$ generator than is the inverted-triode m term.

In discussing inverted-triode surface noise at high currents, it is appropriate to consider forward-double-diode noise too. For a given unit at a given base current, the d-c pair density distribution near recombination surfaces in the inverted triode is practically identical with that in the forward double diode (see Appendix I). Thus the common-emitter inverted-triode input and forward double diode should appear practically identical with respect to both signal (input impedance) and noise (white and surface, but not leakage noise).

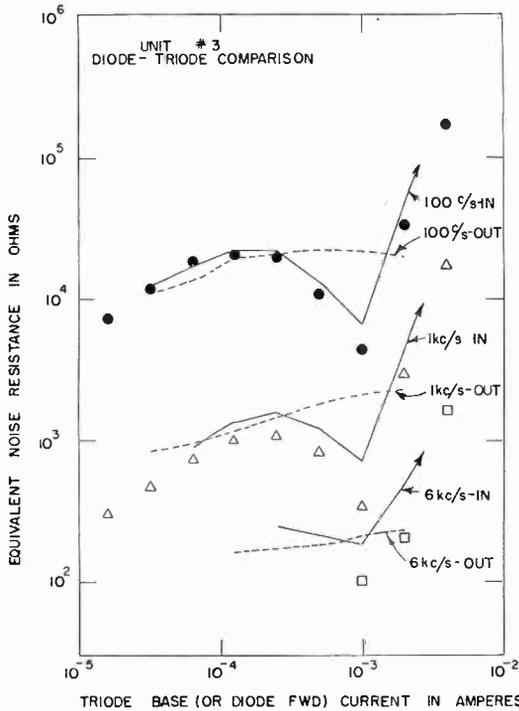


Fig. 16a—Comparison of excess noise observed in forward-double-diode and inverted-triode connections, unit #3.

Figure 16 shows observed $R_{eq}^{1/f}$'s, observed R_{eq} 's minus expected white-noise contributions, versus I_b for two forward double diodes. Measurements at 100, 1,000, and 6,000 cycles are shown by circles, triangles, and squares respectively. Observed common-emitter inverted-triode input $R_{eq}^{1/f}$'s are shown by solid curves. These diode and triode equivalent noise resistances agree within experimental error, and one may draw the appropriate conclusions cited on page 246.

The summed $di_{b,e}$ surface-noise generator, when converted by Thevenin's theorem to the corresponding noise voltage generator in

series with $r_{b'e}$, can be characterized by an equivalent noise resistance $R_{eq}^{b'e}$ defined by

$$r_{b'e}^2 \left(\sum \overline{di_{b'e}^2} \right) = 4 kT R_{eq}^{b'e} df.$$

One can obtain $R_{eq}^{b'e}$ explicitly, according to the model, by substituting for $r_{b'e}$ from Figure 3(a) and for $\sum \overline{di_{b'e}^2}$ from Equations (8) or (8a). Like $\sum \overline{di_{b'e}^2}$, $R_{eq}^{b'e}$ is proportional to ψ . Except at the very lowest currents, $R_{eq}^{b'e}$ should be independent of current. On the other

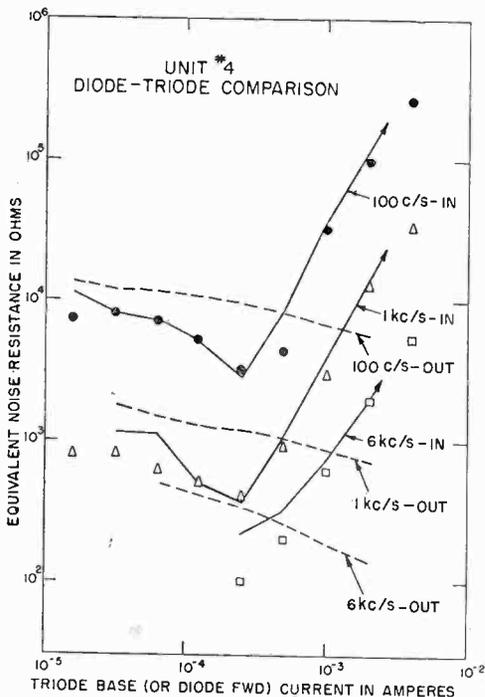


Fig. 16b—Comparison of excess noise observed in forward-double-diode and inverted-triode connections, unit #4.

hand, since surface noise at the triode output is due entirely to $di_{b'e}$ (see Table III), experimental $R_{eq}^{b'e}$'s can be computed from observed $\Delta F_{OUT}^{s'}$'s.

The dotted curves in Figure 16 show values of $R_{eq}^{b'e}$ computed from observed inverted-triode $\Delta F_{OUT}^{1/f}$'s. These $R_{eq}^{b'e}$'s should be practically identical with that portion of common-emitter inverted-triode input R_{eq} (of forward-double-diode R_{eq}) due to the $di_{b'e}$ surface-noise generator alone. See Figures 4 and 1(c). These dotted curves are related to the solid curves and data points in Figure 16 approximately as

predicted in Figure 7. This proves that forward-diode and common-emitter inverted-triode input noise due to $di_{b'e}$ alone is approximately independent of current and that the added complications in the total 1/f noise are due to a dv_{bb} generator with the properties predicted. Similar results were obtained for other units. $R_{eq}^{b'e}$ often decreases somewhat with current as shown in Figure 16b.

Figure 17 shows values of $R_{eq}^{b'e}$ computed from observed ΔF_{OUT}^S 's for five normal triodes. Here, where the measurements extend to somewhat higher injections, the decrease of $R_{eq}^{b'e}$ with current is more pronounced. Evidently ψ decreases somewhat with increasing pair injection.

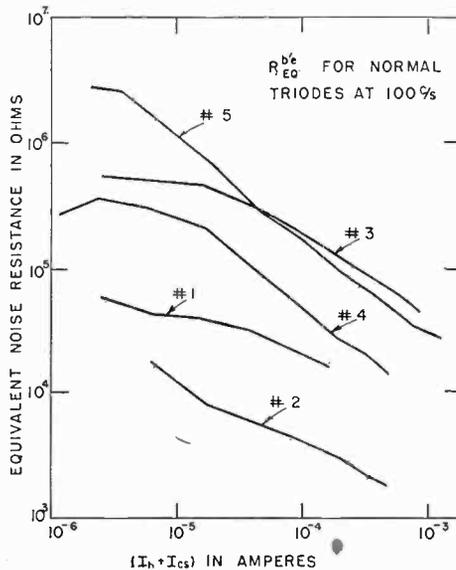


Fig. 17— $R_{eq}^{b'e}$ computed from observed ΔF_{OUT}^S for five normal triodes.

The computation of $R_{eq}^{b'e}$, according to the model, was outlined above. Provided $p_o \gg p_n$ (which was satisfied for the measurements here), the expected ratio of normal- to inverted-triode $R_{eq}^{b'e}$ is

$$\frac{R_{eq}^{b'e} \text{ (normal triode)}}{R_{eq}^{b'e} \text{ (inverted triode)}} = M = \frac{\pi}{8} \frac{1}{R_c W} \frac{(R_c^2 - R_e^2 + 4R_c L)^2}{(R_c^2 - R_e^2 + R_c L)} \quad (9)$$

Since ψ has been cancelled out, experimental verification of M would amount to substantiating the oneness of ψ for the inverted and normal triodes. Using the typical values listed in Table I, M should be about

47 for the units studied here. In computing M from observed ΔF_{OUT}^S 's, one can eliminate the small dependence of ψ on injection by dividing values of $R_{eq}^{b'e}$ observed at comparable pair densities, that is, at equal collector currents. Figure 18 shows values of M computed thusly from observed 100-cycle values of ΔF_{OUT}^S for five units. The predicted ratio of approximately 47 is borne out. Individual differences in W , R_G , etc., do not explain observed differences in M for different units. According to the preceding figure, ψ varies two orders of magnitude among different, but supposedly similar, units. Since the normal triode samples ψ over a surface area approximately 1/50 the size of that sampled by the inverted triode, the existing agreements in M come rather as a pleasant surprise.

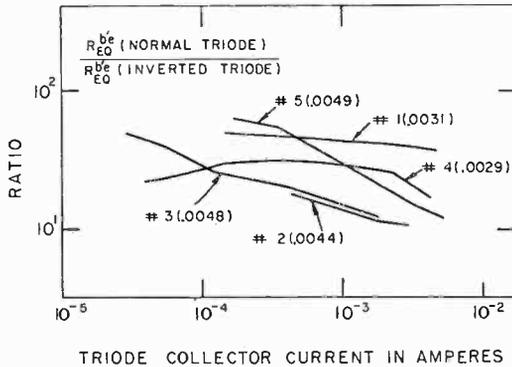


Fig. 18—Ratios of normal- to inverted-triode $R_{eq}^{b'e}$ computed from observed 100-cycle values of ΔF_{OUT}^S at equal collector currents. W values, in centimeters, are listed for the individual units.

Collector voltage does not enter the surface-noise model directly, and surface noise should be relatively independent of this bias. The observed noise actually depends weakly on V_c through the base width W . Normal-triode ΔF_{OUT}^S at constant I_c is proportional to $(1 + r_{bb'}/R_G)^2 W^3$ [see Table III and Equation (8a)]; $r_{bb'}$ is proportional to W^{-1} ; and W decreases with increasing V_c . Thus

$$\frac{\partial \Delta F_{\text{OUT}}^S}{\partial V_c} = \Delta F_{\text{OUT}}^S \left[3 - 2 \frac{r_{bb'}/R_G}{(1 + r_{bb'}/R_G)} \right] \frac{1}{W} \frac{\partial W}{\partial V_c}. \quad (10)$$

From depletion-layer theory for abrupt junctions,⁹

$$\frac{\partial W}{\partial V_c} = -5 \times 10^{-5} \left[\frac{\rho_n}{(V_o + V_c)} \right]^{1/2} \text{ cm/volt}, \quad (10a)$$

where ρ_n is expressed in ohm-centimeters and $(V_o + V_c)$ in volts, for n-type germanium. Figure 19 shows one example of this weak dependence on voltage. Observed 1-kilocycle values of F_{OUT} versus V_c are shown for Unit #4 at $I_c = 3$ milliamperes, where surface noise dominates the output noise (see Figure 12). For this unit, $r_{bb'}/R_G = 1.18$, $W = .0029$ centimeter, $\rho_n \approx 3$ ohm-centimeters, and Equations (10) and (10a) predict slopes of ΔF_{OUT}^s of -5.7 and -2.8 per cent per volt at $V_c = 1$ and 4 volts respectively. These slopes are realized within experimental error. Leakage noise caused the observed noise to increase beyond $V_c \approx 8$ volts.

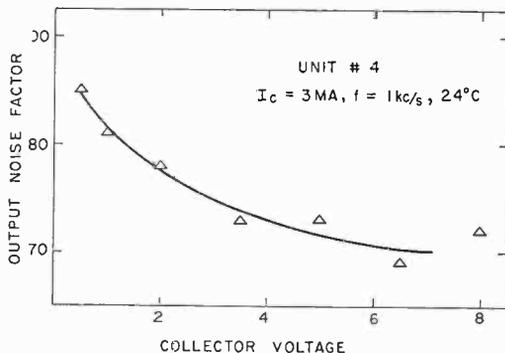


Fig. 19—Observed output noise factor versus collector voltage.

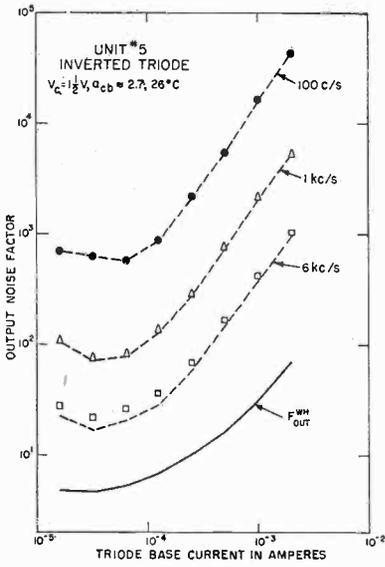
C. Example of Mixed Noise

The analysis of inverted-triode unit #5, which exhibits both large surface and large leakage noise, is presented in Figure 20.

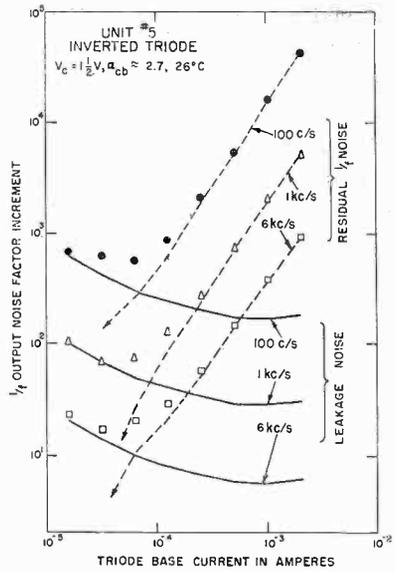
Figure 20(a), analogous to Figure 13, shows observed common-emitter F_{OUT}^s 's versus current. The excess noise, shown by dotted curves, varies something like $1/f$, is relatively independent of I_b at low currents, but increases rapidly at high currents.

$G_{eq}^L(f)$ for this unit was obtained from reverse-diode studies of the functional collector. Expected leakage-noise contributions to common-emitter F_{OUT} and F_{IN} were then computed using Table III. These computed values accounted for the excess triode noise observed at low currents. This is shown for the output noise in Figure 20(b). The $\Delta F_{\text{OUT}}^{1/f}$'s of Figure 20(a) are shown by data points, the ΔF_{OUT}^L 's predicted from reverse-diode measurements by solid curves, and the residual noise, the excess noise minus expected leakage noise, by dotted curves. The residual noise varies something like $1/f$ and exhibits the dependence on current expected for surface noise.

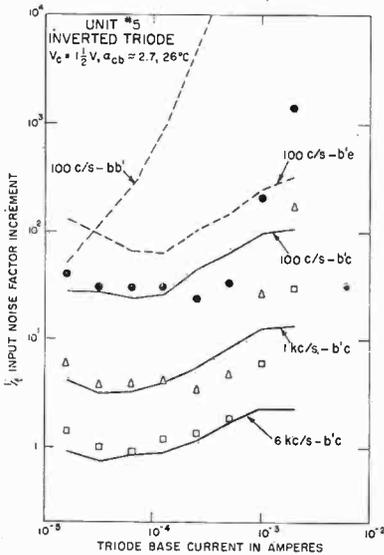
Figure 20(c), analogous to the Figure 15, shows that leakage noise, as expected, is represented accurately by a $di_{b'c}$ generator at low emitter



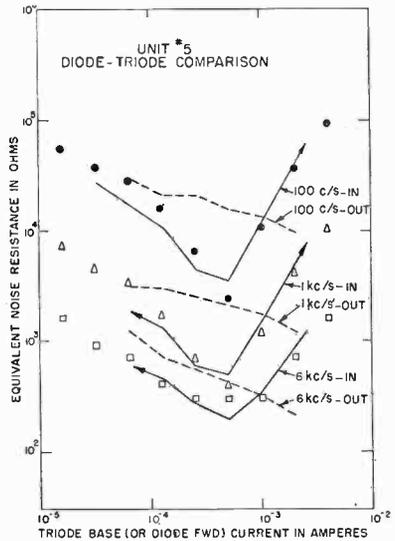
(a)



(b)



(c)



(d)

Fig. 20—Analysis of unit exhibiting both surface and leakage noise.

currents. Common-emitter $\Delta F_{IN}^{1/f}$'s observed at 100, 1,000, and 6,000 cycles are shown by circles, triangles, and squares respectively. The solid and dotted curves were obtained by multiplying the $\Delta F_{OUT}^{1/f}$'s of Figure 20(a) by $\Delta F_{IN}/\Delta F_{OUT}$ ratios appropriate to single generators in the locations indicated. The observed $\Delta F_{IN}^{1/f}$'s, dominated at low currents by leakage noise, coincide at low currents with the solid ($b'c$) curves. Similar results were obtained for other units.

Figure 20(d), analogous to Figure 16, shows that the residual noise, as expected, is surface noise. Observed forward-double-diode $R_{eq}^{1/f}$'s, observed common-emitter inverted-triode input $R_{eq}^{1/f}$'s (residual 1/f noise only), and observed inverted-triode $R_{eq}^{b'e}$'s [computed from the residual $\Delta F_{OUT}^{1/f}$'s of Figure 20(b)] are shown by data points, solid curves, and dotted curves respectively. The statements concerning Figure 16 apply here word for word.

Triode leakage noise increases strongly with collector voltage, is somewhat erratic, and exhibits hysteresis effects. It is preferable to study these effects in the reverse-diode connection, where surface noise is small.

D. Temperature Dependence

The dependence of G_{eq}^L and ψ on temperature near room temperature was investigated.

Direct-current leakage in germanium p-n-p alloy-junction units increases with temperature according to an activation energy of approximately 0.2 electron volt. Since $G_{eq}^L \propto I_L^2$ [see Equation (6)], G_{eq}^L should exhibit an activation energy of approximately 0.4 electron volt. The temperature dependence of ψ is initially open.

Figure 21 shows observed 1-kilocycle values of F_{OUT} versus temperature for four normal triodes. The operating points were chosen so that either leakage noise (unit #6) or surface noise (other three units) dominated the observed noise. The temperature dependence of F_{OUT} , if appreciable, would be contained in the phenomenological quantities G_{eq}^L and ψ . For unit #6, F_{OUT} increases with temperature according to an activation energy of approximately 0.46 electron volt, as expected. For the other three units, F_{OUT} is insensitive to temperature. The activation energy of ψ is (0 ± 0.1) electron volt near room temperature.

Besides providing preliminary temperature data on G_{eq}^L and ψ , the present section explains why triode noise factor sometimes increases with temperature, sometimes does not. Change with temperature is appreciable only when leakage noise is prominent.

E. More About ψ

The surface-noise generators are qualitatively common to all models which can ultimately be decomposed into noisy point sources of pairs and are especially closely related to models wherein the point sources increase in strength rapidly with injection. Two other formally possible source distributions will be considered here.

Instead of surface noise, one could propose that the noise is a bulk phenomenon characterized, in analogy with Equation (1), by an apparent pair generation rate per unit *volume* with spectrum

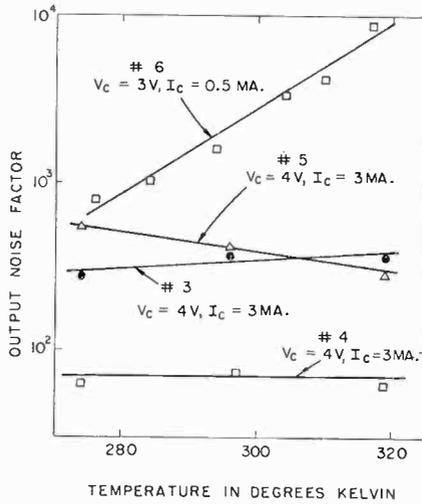


Fig. 21—Observed output noise factors versus temperature for four normal triodes.

$$\overline{dg^2} = (p - p_n)^2 \psi'(f) df, \quad (11)$$

where ψ' , varying something like $1/f$ and decreasing somewhat with injection, characterizes the noisiness of the bulk. Summing over the many uncorrelated sources described by Equation (11) (the computations are similar to those described in Appendix I for surface noise), one obtains the following results:

In the forward-double-diode connection, Equations (3) and (3a) would have ψ replaced by

$$\psi(f) \rightarrow \psi'(f) W \frac{(R_c^2 + R_e L/2)}{(R_c^2 - R_e^2 + R_e L)}. \quad (11a)$$

The ratio of common-emitter inverted-triode input R_{eq}^S to forward-

double-diode R_{eq}^s would change from unity for surface noise to

$$\frac{(R_c^2 - \frac{2}{3} R_e^2 + R_c L/2)}{(R_e^2 + R_c L/2)} \approx 0.91 \quad (11b)$$

for the units studied here. The difference between these two ratios could not be resolved by the measurements here, although units could be constructed for which this difference would be larger. The ratio M of normal- to inverted-triode $R_{eq}^{b'c}$ would be larger than that quoted in Equation (9) for surface noise by the factor

$$\pi R_c / 6W \approx 3.2 \quad (11c)$$

for the units studied here. The smaller value appropriate to surface noise is more in agreement with the measurements. See Figure 18.

Similarly, one could propose that the noise is associated with the emitter junction itself and is characterized by an apparent pair generation rate per unit *area of junction* with spectrum

$$\overline{dg^2} = (p - p_n)^2 \psi''(f) df. \quad (12)$$

Summing over the many uncorrelated sources described by Equation (12), one finds that the relations corresponding to Equations (11a), (11b), and (11c) are, in the present case,

$$\psi(f) \rightarrow \psi''(f) \frac{(R_c^2 + R_e^2)}{(R_c^2 - R_e^2 + R_c L)}, \quad (12a)$$

$$\frac{R_c^2}{(R_c^2 + R_e^2)} \approx 0.86, \quad (12b)$$

$$\frac{\pi R_c}{2W} \frac{(R_c^2 - R_e^2 + R_c L)}{R_e^2} \approx 10.1, \quad (12c)$$

respectively. The ratio in Equation (12b) could not be resolved by the measurements here from the unity ratio appropriate to surface noise. The ratio M , being approximately 10 times larger than that appropriate to surface noise, is inconsistent with the measurements.

In terms of I_c , and for $I_c \gg I_{cs}$, the surface-noise contribution to

normal-triode output noise is [see Table III and Equation (8a)]

$$\Delta F_{\text{OUT}}^s = \frac{1}{2\pi^2} \frac{R_G}{kT} \left(1 + \frac{r_{bb'}}{R_G} \right)^2 \left(\frac{I_c}{D} \right)^2 \left(\frac{W}{R_e} \right)^3 \psi(f). \quad (8c)$$

Since surface noise is the principal $1/f$ noise plaguing most transistor applications, Equation (8c) is of considerable importance. It is interesting to contrast this formula with that for the collector-base current amplification factor (see Appendix I):

$$\alpha_{cb} = \frac{\pi}{4} \frac{R_e}{W} \frac{D}{sW}. \quad (13)$$

The base width W , the emitter radius R_e , and the diffusion constant D enter both formulas such that high alpha and low noise go hand in hand. This is not because surface noise enters the scene *after* input signals are amplified, but rather because factors which can lower the excess pair density near base recombination surfaces lower both the average base current and the noisy pair current described by Equation (1). Low ψ yields low noise just as low s yields high alpha. On etched germanium surfaces ψ varies more than two orders of magnitude—from 6×10^{-12} to 3×10^{-9} cm⁴/sec at 1 kilocycle for the units studied here. This wide variation does not appear to be correlated with s .

One can reduce ΔF_{OUT}^s by constructing units with small W/R_e and by operating them at low I_c . Because ψ decreases somewhat with injection, ΔF_{OUT}^s is typically proportional to $I_c^{(3/2 \pm 1/2)}$. But, in the beginning, one can at least hope for some large-scale reduction in ψ —perhaps something like that gained in s by etching sand-blasted surfaces.

V. RELATED WORKS OF OTHERS

Inspection of the existing noise literature shows that surface and leakage noise are not peculiar to the units studied here, but rather are common to all junction devices.

In the earliest days of the junction triode, collectors were poor and leakage noise was large. Thus noise factors usually increased with collector voltage and were insensitive to emitter current.¹⁴

The detailed measurements reported by Montgomery¹⁵ can be ex-

¹⁴ R. L. Wallace and W. J. Pietenpol, "Some Circuit Properties and Applications of n-p-n Transistors," *Bell Sys. Tech. Jour.*, Vol. 30, p. 530, July, 1951.

¹⁵ H. C. Montgomery, "Transistor Noise in Circuit Applications," *Proc. I.R.E.*, Vol. 40, p. 1461, November, 1952.

plained in detail by the models above.

Montgomery discusses triode noise in terms of very general four-pole representations and carefully stresses that preference of a particular representation is somewhat subjective. In the end, he made measurements in the common-base connection and analyzed them in the hybrid V - I representation. The representations of Figure 3, when transformed to the V - I representation for the common-base connection, yield the following values for V_1 , I_2 :

1. *Leakage Noise.*

$$V_1^L = -R_{bb'} dI_{LN}, \quad I_2^L = dI_{LN}. \quad (14)$$

Thus

$$\overline{V_1^2}^L = R_{bb'}^2 \overline{I_2^2}^L, \quad \overline{I_2^2}^L = 4kT G_{eq}^L(f) df, \quad \overline{V_1 I_2}^L = -R_{bb'} \overline{I_2^2}^L. \quad (14a)$$

2. *Surface Noise.* For a point source of pairs of small-signal strength dg in the triode base,

$$V_1^S = (r_{eb'} - \alpha_{ce} r_{bb'}) qw_e dg + I_b \left[\frac{\partial R_{bb'}}{\partial I_b} q(w_e + w_c) + m \right] dg - r_{bb'} qw_e dg, \\ I_2^S = \alpha_{ce} qw_e dg + qw_e dg, \quad (15)$$

where $r_{eb'}$, as opposed to $r_{b'e}$, is given by Equation (34), and where the collector-emitter current amplification factor α_{ce} is given by

$$\alpha_{ce} = g_m r_{eb'} = \frac{(I_c - I_{cs} - I_L)}{(I_e + I_{cs})}.$$

If one neglects the m term and takes $\alpha_{ce} \approx 1$, Equations (15) simplify to

$$V_1^S = r_{eb'} qw_e dg - R_{bb'} q(w_e + w_c) dg, \quad I_2^S = q(w_e + w_c) dg. \quad (15a)$$

For this common-base connection, there is an interference minimum in V_1^S when $w_e r_{eb'} = (w_e + w_c) R_{bb'}$. This minimum is not duplicated for leakage noise. Thus differentiation between leakage and surface noise by the $\Delta F_{IN}/\Delta F_{OUT}$ method is easier in the common-base connection. The minimum occurs at moderate emitter currents ($I_e \approx 10^{-4}$ ampere) and is due to a feedback effect of $r_{bb'}$, not to interference between the $di_{b'e}$ and $dv_{bb'}$ generators. $dv_{bb'}$, if one neglects the m term, acts here only to bring $R_{bb'}$ rather than $r_{bb'}$ into Equations (14), (14a), and (15a).

Montgomery reported measurements on grown-junction triodes. For this geometry the base surface parallels the z direction, and the following relations obtain:

$$w_e = 1 - z/W, \quad w_c = z/W, \quad \Delta p = (p_o + p_n)(1 - z/W) - p_n,$$

where $z = 0$ is the emitter junction.

For Montgomery's units, ψ decreased somewhat with injection. In analyzing his data, it is therefore convenient to take

$$\psi(f) = A(f)/|p - p_n|, \quad (1a)$$

where $A(f)$ varies something like $1/f$ but is independent of p . Summing over the many uncorrelated sources described by Equation (1), one obtains, provided $p_o \gg p_n$, the following V - I representation for surface noise:

$$\begin{aligned} \overline{V_1^2}^S &= \left(\frac{1}{2} r_{eb'}^2 - \frac{4}{3} r_{eb'} R_{bb'} + R_{bb'}^2 \right) \overline{I_2^2}^S, \\ \overline{I_2^2}^S &= \sum \overline{di_{b'e}^2}, \quad \overline{V_1 I_2}^S = \left(\frac{2}{3} r_{eb'} - R_{bb'} \right) \overline{I_2^2}^S, \end{aligned}$$

where the summed composite $di_{b'e}$ generator is given by

$$\sum \overline{di_{b'e}^2} = \sum q^2 (w_e + w_c)^2 \overline{dg^2} = I_c \frac{qW^2}{DR_e} A(f) df.$$

Because of Equation (1a), $\sum \overline{di_{b'e}^2}$ increases only as the first power of the current. The cross-product $\overline{V_1 I_2}^S$ versus current changes sign at $r_{eb'} = (3/2) R_{bb'}$, and the strength of V_1 passes through a pronounced minimum nearby.

3. White Noise.

$$\begin{aligned} \overline{V_1^2}^{WH} &= (r_{eb'} - \alpha_{ce} r_{bb'})^2 2q(I_b + I_{cs} + 2I_{es}) df \\ &+ [r_{cb'} + (1 - \alpha_{ce}) r_{bb'}]^2 2q(I_c - I_{cs}) df + r_{bb'}^2 2qI_{cs} df + 4kTR_{bb'} df, \\ \overline{I_2^2}^{WH} &= \alpha_{ce}^2 2q(I_b + I_{cs} + 2I_{es}) df + (1 - \alpha_{ce})^2 2q(I_c - I_{cs}) df + 2qI_{cs} df, \\ \overline{V_1 I_2}^{WH} &= \alpha_{ce} (r_{eb'} - \alpha_{ce} r_{bb'}) 2q(I_b + I_{cs} + 2I_{es}) df \\ &- (1 - \alpha_{ce}) [r_{eb'} + (1 - \alpha_{ce}) r_{bb'}] 2q(I_c - I_{cs}) df - r_{bb'} 2qI_{cs} df. \end{aligned}$$

When these three independent noise sources are present simultaneously, the total V - I generators have the magnitudes

$$\overline{V_1^2}^T = \overline{V_1^2}^L + \overline{V_1^2}^S + \overline{V_1^2}^{WH}, \quad \overline{I_2^2}^T = \overline{I_2^2}^L + \overline{I_2^2}^S + \overline{I_2^2}^{WH}, \quad (16)$$

and the correlation

$$r_{12}^T = \frac{\overline{V_1 I_2}^L + \overline{V_1 I_2}^S + \overline{V_1 I_2}^{WH}}{(\overline{V_1^2}^T \overline{I_2^2}^T)^{1/2}}. \quad (16a)$$

Montgomery provides V_1 , I_2 , and r_{12} data for a n-p-n triode in his Figure 5. Figure 22 shows his observed values replotted versus I_c at $V_c = 0.2, 0.5,$ and 10.0 volts respectively. The solid curves show the predictions of Equations (16) and (16a) under the assumptions

$$\begin{aligned} \frac{1}{df} \left(\sum \overline{di_{b'e}^2} \right) &= 206.5 \text{ decibels below } 1 \text{ ampere}^2/\text{cycle at } I_c = 1 \text{ milli-ampere,} \\ R_{bb'} &= 100 \text{ ohms,} \\ 4kTG_{ea}^L &= 0 \text{ at } V_c = 0.2 \text{ volts,} \\ &= 212.5 \text{ decibels below } 1 \text{ ampere}^2/\text{cycle at } V_c = 0.5 \text{ volt,} \\ &= 200.5 \text{ decibels below } 1 \text{ ampere}^2/\text{cycle at } V_c = 10.0 \\ &\text{volts.} \end{aligned} \quad (17)$$

For reasonable values of α_{ce} , $\overline{I_2^2}^{WH}$ and $\overline{V_1 I_2}^{WH}$ contribute negligibly to $\overline{I_2^2}^T$ and $\overline{V_1 I_2}^T$ respectively, and $\overline{V_1^2}^{WH}$ is given accurately by

$$\overline{V_1^2}^{WH} = r_{eb}^2 2q(I_c - I_{cs})df + 4kTR_{bb'}df.$$

Thus the precise value of α_{ce} is not needed. If the white noise had been neglected, the predictions of Equations (16) and (16a), under the assumptions of Equations (17), would be given by the dotted curves in Figure 22.

It is concluded that Montgomery's Figure 5 measurements are rather plausibly explained by the models. The I_2 measurements in his Figure 6 can be similarly explained. Thus the same models explain noise in both p-n-p alloy- and n-p-n grown-junction transistors. Since correlation between input and output noise had not been measured here, Montgomery's data provides a significant additional verification of the theory.

The writer feels that transistor noise representations based on general four-pole concepts are less convenient than the specific representations of Figures 3 and 4 based on the physical mechanisms underlying noise. The noise generators of Figures 3 and 4 may be written without change into any signal schematic containing the internal base point b' .

Montgomery and Clark,⁴ in a paper concerned primarily with white noise, show a good example of triode $\Delta F_{\text{OUT}}^{1/f}$ increasing rapidly with emitter current. The surface-noise model would explain this.

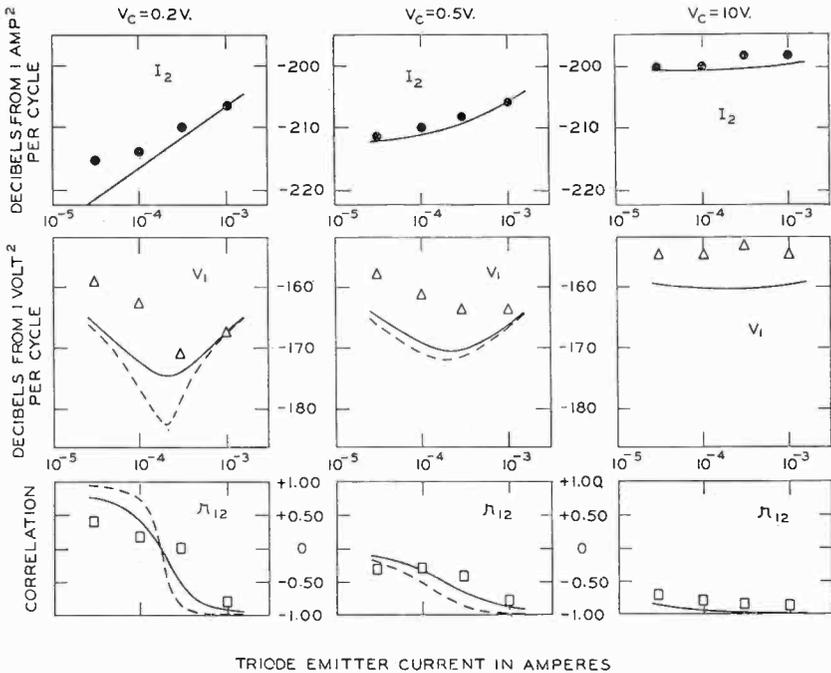


Fig. 22—Analysis of Montgomery's n-p-n grown-junction-triode noise measurements.¹⁵

Bargellini and Herscher¹⁶ describe noise measurements on p-n-p and n-p-n alloy-junction triodes. For their samples $\Delta F_{\text{OUT}}^{1/f}$ increases rapidly with emitter current at low collector voltages and increases rapidly with collector voltage beyond $V_c \approx 10$ volts. The increase with current is explained by surface noise, that with voltage by leakage noise. Indeed, their Figures 10 and 11 are qualitatively identical with Figures 12 and 19 respectively of this paper, and should be interpreted

¹⁶ P. M. Bargellini and M. B. Herscher, "Investigations of Noise in Audio Frequency Amplifiers Using Junction Transistors," *Proc. I.R.E.*, Vol. 43, p. 217, February, 1955.

identically. These authors incorrectly associate noise which increases with I_c with the collector junction.

Anderson and van der Ziel¹⁷ describe noise measurements on forward-biased junction diodes. Their Figure 5 shows 1-kilocycle values of $n^{1/f}$ (noise ratio n minus expected white-noise contributions) versus current. This curve can be extended to 1.0 milliampere from data available in their Figure 3. If one multiplies $n^{1/f}$ by $(r_{b'e} + r_{bb'})$, the current dependence of resulting $R_{eq}^{1/f}$ is well explained by Figure 7 herein, that is, by the surface-noise model. Anderson and van der Ziel's measurements did not extend to sufficiently high currents to reveal the postminimum rise in R_{eq}^s .

Montgomery¹⁸ describes noise measurements on semiconductor diodes. In his Figure 2, the dotted curve labeled B shows an example of junction-diode noise versus forward current. His ordinate must be squared, then multiplied by $(1/4kT)(r_{b'e} + r_{bb'})^2$ to obtain diode equivalent noise resistance. If one performs this transformation, the current dependence of resulting R_{eq} is again explained by the surface-noise model.

Montgomery,¹⁸ Kennedy,¹⁹ and McWhorter²⁰ have studied 1/f noise in reverse diodes. Their observed noise strengths differ from the predictions of Equation (6) above by ± 20 decibels. McWhorter, working with grown-junction diodes, distinguishes between 1/f leakage and 1/f channel noise. His channel noise strengths in particular are low compared to the predictions of Equation (6).

The surface- and leakage-noise models above may be used to discuss 1/f noise in point-contact devices, in photo junctions, and in semiconductor filaments. Indeed, the surface-noise model above, in its phenomenological aspects, and the model used by Montgomery¹⁸ and Shockley²¹ to discuss noise in germanium filaments have some common features.

ACKNOWLEDGMENTS

D. O. North, besides providing the surface-noise model, collaborated

¹⁷ R. L. Anderson and A. van der Ziel, "On the Shot Effect of P-N Junctions," *Transactions of the I.R.E. Professional Group on Electron Devices*, I, p. 20, November, 1952.

¹⁸ H. C. Montgomery, "Electrical Noise In Semiconductors," *Bell Sys. Tech. Jour.*, Vol. 31, p. 950, September, 1952.

¹⁹ D. Kennedy, "Gaseous Ambients and Diode Noise," presented at the 1954 IRE-AIEE Conference on Semiconductor Device Research, University of Minnesota.

²⁰ A. L. McWhorter, "1/f Noise and Related Surface Effects in Germanium," MIT Lincoln Laboratory Report No. 80, May, 1955.

²¹ W. Shockley, *Electrons and Holes in Semiconductors*, D. Van Nostrand Co., Inc., 1950, p. 342.

in the design and interpretation of the experiments. The author also benefited from numerous discussions with L. J. Giacoletto. C. C. Cheng helped in the later stages of the work.

APPENDIX I — CIRCUIT SCHEMATICS

With the aid of certain simplifying approximations, some relations useful in analyzing diode and triode noise will be computed here.

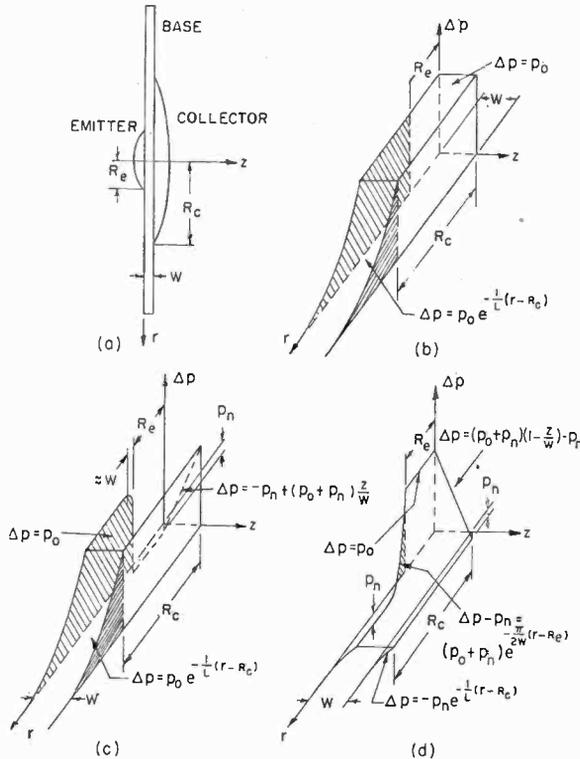


Fig. 23—Idealized p-n-p alloy-junction transistor. (a) Simplified parallel-plane geometry. Excess d-c pair density distribution in forward-double-diode (b), inverted-triode (c), and normal-triode (d) connections.

The p-n-p alloy-junction transistor will be represented by the idealized parallel-plane geometry of Figure 23(a). The conductivities of the p-type regions are assumed very large. Volume recombination of pairs is neglected. Pair motion other than via a quasi-steady diffusion is neglected. Thus the equation governing pair flow simplifies to Laplace's equation (hereafter called the diffusion equation). Collector-voltage modulation of the base width is also neglected. See Table I

for the notation. Constants such as s are assumed strictly constant, i.e., independent of position and the excess pair density Δp .

The forward double diode will be treated first. Provided $sW/D \ll 1$, Δp is nearly independent of the axial coordinate z . Out to the edge of the larger dot, Δp has a nearly constant value p_0 . Beyond R_c , provided $L \ll R_c$, Δp satisfies the equation

$$\left(\frac{d^2}{dr^2} - \frac{1}{L^2} \right) \Delta p = 0,$$

and the boundary condition $\Delta p = 0$ at $r = \infty$. Thus the spatial dependence of Δp is described approximately by the configuration shown in Figure 23(b). The location and magnitude of net surface recombination are indicated by the shaded areas.

Δp is related to the junction voltage and current by two fundamental equations. The first asserts Fermi-Dirac equilibrium of minority carriers across the narrow junction;

$$p_0 + p_n = p_p \exp [-q(V_0 - V_{cb'}) / (kT)]. \quad (18)$$

The second states that diode current equals the net recombination current;

$$I_e = I_b = \int_A da \, q s \Delta p = \pi q s p_0 (R_c^2 - R_e^2 + 4R_c L), \quad (19)$$

where A is the base recombination surface. The junction d-c and low-frequency a-c characteristics follow immediately;

$$I_e = I_b = I_{es} (\exp [qV_{cb'} / (kT)] - 1), \quad (20)$$

$$dI_e = dI_b = I_{es} \frac{dp_0}{p_n} = \frac{1}{r_{b'e}} dV_{cb'}, \quad (21)$$

where

$$I_{es} = \pi q s p_n (R_c^2 - R_e^2 + 4R_c L) \quad (22)$$

and where $r_{b'e}$ is given by the formula listed in Figure 1(a). Equation (21) is stated in alternative forms which will be alternatively useful.

The d-c base-lead resistance is defined as the ratio of voltage to current in the base-lead arm;

$$R_{bb'} = \frac{V_{b'b}}{I_b} \quad (23)$$

$R_{bb'}$ depends upon the excess pair density Δp which in turn is characterized by the single parameter p_0 . Taking differentials,

$$dV_{b'b} = R_{bb'} dI_b + I_b dR_{bb'} \quad (24)$$

Now

$$dR_{bb'} = \frac{dR_{bb'}}{dp_0} dp_0 = \frac{dR_{bb'}}{dp_0} \frac{p_n}{I_{es}} dI_b, \quad (25)$$

where Equation (21) was used to eliminate dp_0 . The coefficient of dI_b in Equation (25) is really $dR_{bb'}/dI_b$, so that Equation (24) may be rewritten

$$dV_{b'b} = r_{bb'} dI_b, \quad (24a)$$

where $r_{bb'}$ is given by Equation (4). The dependence of $R_{bb'}$ on I_b has been stated explicitly through the intermediary p_0 . A case will be considered shortly where Equation (21) does not hold, and then Equation (24a) will have to be modified.

Equations (21) and (24a) are restated in the small-signal equivalent circuit of Figure 1(a).

The diode equations considered thus far summarize the average statistical reactions of electrons and holes to applied biases (signals). Spontaneous fluctuations from the mean or essentially internal stimulations such as provided by light or particle bombardment must be represented in the circuit schematic by additional generators.

As a preliminary to surface noise, the generator representation of a point source of pairs of strength g located at some position \underline{r} in the diode base will be derived.

Let p_1 be the solution to the diffusion equation for pairs excited by the source g at \underline{r} subject to the surface recombination condition $r = sp_1$ and the condition $p_1 = 0$ at the emitter. That is, p_1 represents the density distribution of pairs as they flow from the source to the emitter or to the base surface. The distribution p_1 will have some definite shape characterized by a single parameter, say its amplitude p_m at \underline{r} . The probability $w_e(\underline{r})$ that source pairs reach the emitter is related to p_1 by

$$1 - w_e = \frac{1}{g} \int_A da sp_1 \quad (26)$$

Let the diode be terminated by an arbitrary impedance and battery.

The battery and the flow engendered by the source g combine to bias the junction to some voltage V_{eb} , which, in accordance with the fundamental Equation (18), will specify a constant excess pair density p_0 just inside the emitter. Let p_2 be the solution to the diffusion equation for pairs excited by the bias V_{eb} , subject to the surface recombination condition $r = s(p_2 - p_n)$. That is, Δp_2 is the pair density distribution shown in Figure 23(b).

Solutions to the diffusion equation are summable. The total excess pair density is

$$\Delta p = p_1 + \Delta p_2. \quad (27)$$

The fundamental Equation (19) stating that diode current equals the net recombination current reads, in the present case,

$$I_e = I_b = -qg + \int_A da \, qs \Delta p. \quad (19a)$$

Using known surface integrals of p_1 , Δp_2 , one finds

$$I_e = I_b = -qw_e g + I_{es} \frac{p_0}{p_n}. \quad (19b)$$

The junction d-c and low-frequency a-c characteristics (allowing for a small-signal variation in g) follow immediately;

$$I_e = I_b = -qw_e g + I_{es} (\exp [qV_{eb}/(kT)] - 1), \quad (20a)$$

$$dI_e = dI_b = -qw_e dg + I_{es} \frac{dp_0}{p_n} = -qw_e dg + \frac{q(I_e + I_{es} + gw_e g)}{kT} dV_{eb}, \quad (21a)$$

where I_{es} is still given by Equation (22). Equation (21a) is stated in alternatively useful forms. Equations (20a) and (21a) are well-known. For example, they occur in the theory of diode batteries energized by light or particle bombardment.

The special case where the point source has no d-c component will be of primary interest here. That is, g is zero but dg is not. Thus, Equation (20a) simplifies to Equation (20) and Equation (21a) simplifies to

$$dI_e = dI_b = -qw_e dg + I_{es} \frac{dp_0}{p_n} = -qw_e dg + \frac{1}{r_{b'e}} dV_{eb}, \quad (21b)$$

where $r_{b'e}$ is still given in Figure 1(a).

For the base-lead arm, Equations (23) and (24) remain unchanged. Since $g = 0$, $R_{bb'}$ still depends on the single parameter p_0 and hence is the same unique function of I_b . But small-signal changes in $R_{bb'}$ depend on small-signal changes in both p_1 and Δp_2 . The equation corresponding to Equation (25) reads, in the present case,

$$\begin{aligned} dR_{bb'} &= \left. \frac{\partial R_{bb'}}{\partial p_0} \right|_{p_m=0} dp_0 + \left. \frac{\partial R_{bb'}}{\partial p_m} \right|_{p_0} dp_m, \\ &= \left. \frac{\partial R_{bb'}}{\partial p_0} \right|_{p_m=0} \frac{p_n}{I_{cs}} (dI_b + qw_e dg) + \left. \frac{\partial R_{bb'}}{\partial p_m} \right|_{p_0} \frac{dp_m}{dg} dg, \end{aligned} \quad (25a)$$

where this time Equation (21b) was used to eliminate dp_0 . The coefficient of $(dI_b + qw_e dg)$ in Equation (25a) is readily identified with that of dI_b in Equation (25), namely, $\partial R_{bb'}/\partial I_b |_{p_m=dg=0}$. Thus Equation (25a) reads

$$dR_{bb'} = \frac{\partial R_{bb'}}{\partial I_b} (dI_b + qw_e dg) + m dg, \quad (25b)$$

where

$$m(\underline{r}, p_0) = \left. \frac{\partial R_{bb'}}{\partial p_m} \right|_{p_0} \frac{dp_m}{dg}. \quad (28)$$

Substituting Equation (25b) in (24), the equation corresponding to Equation (24a) reads, in the present case,

$$dV_{b'b} = r_{bb'} dI_b + I_b \left(\frac{\partial R_{bb'}}{\partial I_b} qw_e + m \right) dg, \quad (24b)$$

where $r_{bb'}$ is still given by Equation (4). Note that $\partial R_{bb'}/\partial I_b$ and m are both negative, that is, do not produce compensating effects. The m term arises from direct base conductivity modulation by source pairs. The $\partial R_{bb'}/\partial I_b$ term, due indirectly to base conductivity modulation, has its origin in the disturbed connection, see Equation (21b), between incremental pair density and incremental base current.

Equations (21b) and (24b) are restated in the small-signal equivalent circuit of Figure 1(c).

Provided $sW/D \ll 1$ and $L \ll R_c$, w_e is given by

$$\begin{aligned}
 w_e &= 1, & r < R_c, \\
 &= e^{-(r-R_c)/L}, & r > R_c.
 \end{aligned}
 \tag{29}$$

The point-source representation is readily generalized to describe surface noise. According to North's model, one must combine the effects of many uncorrelated point sources distributed over the base surface. The spectral densities of the sources per unit surface area are stated phenomenologically by Equation (1). The summed $di_{n'e}$ generator has the spectrum

$$\overline{\sum di_{b'e}^2} = \sum q^2 w_e^2 \overline{dg^2} = q^2 \psi(f) df \int_A da w_e^2 (\Delta p)^2.$$

Integrating over the excess d-c pair density distribution shown in Figure 23(b), one obtains Equation (3).

The applicability of the above discussion to the 1/f noise observed in forward diodes is discussed in Section II.

The reverse double diode will now be treated. The previous treatment of the forward double diode can be reapplied here. One needs only to replace the subscript e everywhere by the subscript c . Because of the sign convention used here, I_e must be replaced by $-I_c$; also, R_e and R_c are retained as the radii of the large and small dots respectively. For saturation biasing,

$$\begin{aligned}
 p_0 &\rightarrow -p_n, \\
 I_c &= -I_b \rightarrow I_{cs}, \\
 r_{b'c} &\rightarrow \infty.
 \end{aligned}$$

Since I_{cs} is so small and $r_{b'c}$ is so large, a leakage path shunting the junction, neglected in the forward diode, must sometimes be invoked to explain observed I_c and $r_{b'c}$. Formally, one writes

$$\begin{aligned}
 I_c &= I_{cs} + I_L, \\
 dI_c &= dI_L = -\frac{1}{r_{b'c}} dV_{cb'},
 \end{aligned}
 \tag{30}$$

where I_L is an empirical excess d-c current, $r_{b'c}$ an empirical a-c resistance. This resistance is still large compared to $r_{bb'}$, and the latter plays no role here in the reverse diode. Equation (30) is restated in the small-signal equivalent circuit of Figure 2(a).

The leakage current is very noisy, however. To allow for this, one adds to dI_L a term dI_{LN} accounting phenomenologically for instantaneous fluctuations from the mean;

$$dI_c = dI_L = -\frac{1}{r_{b'c}} dV_{cb'} + dI_{LN}. \quad (30a)$$

Equation (30a) is restated in the small-signal equivalent circuit of Figure 2(d).

Regarding surface noise, if there is no change in ψ at very small pair densities, Equation (3), with $di_{b'e}$ replaced by $di_{b'c}$ and with I_b replaced by I_{cs} , should apply to the $di_{b'c}$ generator shown in Figure 2(c). Since I_{cs} is so small, that is, since the pair density in the reverse diode is disturbed by such a small amount $-p_n$ from its thermal-equilibrium value, this generator will be very small.

The inverted triode will now be treated. The spatial dependence of the excess pair density Δp is described approximately by the configuration shown in Figure 23(c). In front of the functional collector, Δp rises linearly with z from $-p_n$ to p_0 at the emitter. At the side of collector, Δp rises from $-p_n$ to p_0 in a radial distance of approximately W . In this paper Δp will be taken as $p_0 - (p_0 + p_n) e^{-\pi(r - R_e)/2W}$ along the base surface just beyond R_e . This amounts to representing Δp just beyond the functional collector by the first two terms of its series expansion. Beyond $r \approx R_e + W$, Δp is identical with the corresponding pair density distribution for the forward double diode. The location and magnitude of net surface recombination are indicated by shaded areas in Figure 23(c).

As a preliminary to noise, the inverted-triode circuit schematic, including generators representing a point source of pairs of strength g at \underline{r} in the triode base and a leakage current I_L across the collector junction, will be derived. Let $w_e(\underline{r})$, $w_c(\underline{r})$ be the probabilities that source pairs reach the emitter and collector respectively.

Let p_1 be the solution to the diffusion equation for pairs excited by the source g at \underline{r} subject to the surface recombination condition $r = sp_1$ and the conditions $p_1 = 0$ at the emitter and at the collector. That is, p_1 represents the density distribution of pairs as they flow from the source to the junctions or to the base surface. The distribution p_1 will have some definite shape characterized by a single parameter, say its amplitude p_m at \underline{r} . w_e and w_c are related to p_1 by

$$1 - w_e - w_c = \frac{1}{g} \int_A da sp_1, \quad w_c = \frac{1}{g} \int_B da D \frac{\partial p_1}{\partial z}, \quad (26a)$$

where B is the area of the functional collector.

Let the triode be terminated by arbitrary impedances and batteries. The batteries and the flows engendered by the sources g and I_L combine to bias the emitter junction to some voltage V_{eb} , which, in accordance with the fundamental Equation (18), will specify a constant excess pair density p_0 just inside the emitter. Let p_2 be the solution to the diffusion equation for pairs excited by the bias V_{eb} , subject to the surface recombination condition $r = s(p_2 - p_n)$ and the condition $p_2 = 0$ at the collector. That is, Δp_2 is the pair density distribution shown in Figure 23(c).

Solutions to the diffusion equation are summable. The total excess pair density is

$$\Delta p = p_1 + \Delta p_2. \quad (27a)$$

The fundamental equation corresponding to Equation (19) states, in the present case, that triode base current equals the net recombination current minus the leakage current;

$$I_b = -qg + \int_A da \, qs \Delta p - I_L. \quad (19c)$$

Using known surface integrals of p_1 , Δp_2 , one finds

$$I_b = -q(w_e + w_c)g + I_{es} \frac{p_0}{p_n} - I_{cs} - I_L \quad (19d)$$

where

$$I_{es} = \pi q s p_n \left(R_c^2 - R_e^2 + 4R_c L - \frac{4}{\pi} R_e W \right), \quad (22a)$$

$$I_{cs} = 4q s p_n R_e W \ll I_{es}. \quad (22b)$$

There are two more fundamental triode equations. The first states that the collector current equals the flow of minority carriers across the collector junction plus the leakage current;

$$I_c = \int_B da \, qD \frac{\partial(\Delta p)}{\partial z} + I_L. \quad (31)$$

The second states the nonaccumulation of charge;

$$I_e = I_c + I_b.$$

Using known surface integrals of p_1 , Δp_2 , one finds

$$I_c = qw_c g + I_{ec} \frac{(p_0 + p_n)}{p_n} + I_{cs} + I_L, \quad (31a)$$

where

$$I_{ec} \approx \pi R_e^2 q D \frac{p_n}{W}. \quad (22c)$$

The triode d-c characteristics follow immediately:

$$I_b = -q (w_e + w_c) g + I_{es} (\exp [qV_{eb'}/(kT)] - 1) - I_{cs} - I_L, \quad (20b)$$

$$I_c = qw_c g + I_{ec} \exp [qV_{eb'}/(kT)] + I_{cs} + I_L,$$

$$I_e = -qw_e g + I_{ec} \exp [qV_{eb'}/(kT)] + I_{es} (\exp [qV_{eb'}/(kT)] - 1).$$

The emitter and collector do not appear symmetrically in these equations because saturation biasing of the collector has been assumed and an emitter leakage current has been neglected. Allowing for small-signal variations in g and I_L [for I_L a twofold variation of the kind described by Equation (30a) is envisioned], the triode low-frequency a-c characteristics follow immediately;

$$\begin{aligned} dI_b &= -q (w_e + w_c) dg + I_{es} \frac{dp_0}{p_n} + \frac{1}{r_{b'e}} dV_{cb'} - dI_{LN}, \\ &= -q (w_e + w_c) dg + \frac{q [I_b + I_{cs} + I_{cs} + I_L + q (w_e + w_c) g]}{kT} dV_{eb'} \\ &\quad + \frac{1}{r_{b'e}} dV_{cb'} - dI_{LN}, \end{aligned} \quad (21c)$$

$$dI_c = qw_c dg + \frac{q (I_c - I_{cs} - I_L - qw_c g)}{kT} dV_{eb'} - \frac{1}{r_{b'e}} dV_{cb'} + dI_{LN}, \quad (32)$$

$$dI_e = -qw_e dg + \frac{q (I_e + I_{es} + qw_e g)}{kT} dV_{eb'}. \quad (33)$$

Of these three equations, only Equation (21c) has alternatively useful forms here. The triode d-c and a-c characteristics above are of course only straightforward extensions of Shockley's theory of the junction triode.²²

²² W. Shockley, "The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors," *Bell Sys. Tech. Jour.*, Vol. 28, p. 435, July, 1949.

The special case where the point source has no d-c component will be of primary interest here. Thus the triode a-c equations simplify to

$$dI_b = -q(w_e + w_c)dg + \frac{1}{r_{b'e}} dV_{eb'} + \frac{1}{r_{b'c}} dV_{cb'} - dI_{LN}, \quad (21d)$$

$$dI_c = qw_c dg + g_m dV_{eb'} - \frac{1}{r_{b'c}} dV_{cb'} + dI_{LN}, \quad (32a)$$

$$dI_e = -qw_e dg + \frac{1}{r_{eb'}} dV_{eb'}, \quad (33a)$$

where $r_{b'e}$ and g_m are given by formulas listed in Figure 3(a) and where

$$r_{eb'} = kT/q(I_e + I_{es}). \quad (34)$$

For the base-lead arm, Equations (23) and (24) remain unchanged. Since $g = 0$, $R_{bb'}$ still depends on the single parameter p_0 but, because of the d-c leakage current I_L , is a unique function of $(I_b + I_L)$. Small-signal changes in $R_{bb'}$ depend on small-signal changes in both p_1 and Δp_2 . The equation corresponding to Equation (25a) reads, in the present case,

$$dR_{bb'} = \frac{\partial R_{bb'}}{\partial p_0} \bigg|_{p_m=0} \frac{p_m}{I_{es}} \left[dI_b - \frac{1}{r_{b'c}} dV_{cb'} + q(w_e + w_c)dg + dI_{LN} \right] + \frac{\partial R_{bb'}}{\partial p_m} \bigg|_{p_0} \frac{dp_m}{dg} dg, \quad (25c)$$

where this time Equation (21c) was used to eliminate dp_0 . The coefficient of $[dI_b - dV_{cb'}/r_{b'c} + q(w_e + w_c)dg + dI_{LN}]$ is readily identified as $\partial R_{bb'}/\partial I_b |_{p_m = aV_{cb'}/r_{b'c} = dg = dI_{LN} = 0}$. Thus the equation corresponding to Equation (24b) reads, in the present case,

$$dV_{b'b} = r_{bb'} dI_b - \frac{I_b}{r_{b'c}} \frac{\partial R_{bb'}}{\partial I_b} dV_{cb'} + I_b \left[\frac{\partial R_{bb'}}{\partial I_b} q(w_e + w_c) + m \right] dg + I_b \frac{\partial R_{bb'}}{\partial I_b} dI_{LN}, \quad (24c)$$

where $r_{bb'}$ and m are still given by Equations (4) and (28) respectively.

Equations (21d), (32a), and (24c), with $dV_{cb'}/r_{b'c} = 0$, are re-stated in the small-signal equivalent circuits of Figures 3(a), (c), and (d). $dV_{cb'}/r_{b'c}$ was neglected because $r_{b'c}$ was large and the load impedance R_L was small for the measurements here. The $dV_{cb'}$ term in Equation (24c) should not be confused with a similar term having to do with collector-voltage modulation of the base width W ,²³ which has been neglected here. The $dV_{cb'}$, dg , and dI_{LN} terms of Equation (24c) are all due to base conductivity modulation by pairs, W all the while being considered constant.

From Equations (21d) and (32a), with $dV_{cb'}/r_{b'c} = dg = dI_{LN} = 0$, the collector-base current amplification factor α_{cb} is

$$\alpha_{cb} = \frac{dI_c}{dI_b} = g_m r_{b'e} = \frac{(I_c - I_{cs} - I_L)}{(I_b + I_{es} + I_{cs} + I_L)} = \frac{I_{ec}}{I_{es}} \approx \frac{D}{sW} \frac{R_e^2}{(R_c^2 - R_e^2 + 4R_c L)} \quad (35)$$

According to this formula α_{cb} should be, and it indeed proved experimentally to be, approximately 2 for the units studied here.

Provided $sW/D \ll 1$ and $L \ll R_c$, w_e and w_c , on the base surface, are given approximately by

$$\begin{aligned} w_e &= 1 - w_c, & w_c &= e^{-\pi(r-R_e)/2W}, & r &< R_c, \\ &= e^{-(r-R_e)/L}, & &= 0, & r &> R_c. \end{aligned} \quad (29a)$$

The representation above is readily generalized to describe $1/f$ noise.

Leakage noise is described by the dI_{LN} generators in Figure 3(d). The two generators are perfectly correlated. In the absence of a suitable theory, the equivalent noise conductance of the $di_{b'c}$ generator must be determined empirically.

The point-source representation, described by the dg generators in Figure 3(c), is applicable to surface noise. It is convenient to replace the schematic of Figure 3(c) by the equivalent schematic of Figure 4, which will provide the substance for all subsequent discussion. According to North's model, one must combine the effects of many uncorrelated point sources distributed over the base surface. The spectral densities of the sources per unit surface area are stated phenomenologically by Equation (1). The summed composite $di_{b'e}$ generator has the spectrum

²³ Proposed by J. N. Shive. See J. M. Early, "Effects of Space-Charge Layer Widening in Junction Transistors," *Proc. I.R.E.*, Vol. 40, p. 1401, November, 1952; also, "Design Theory of Junction Transistors," *Bell Sys. Tech. Jour.*, Vol. 32, p. 1271, November, 1953.

$$\begin{aligned} \sum \overline{di_{b'e}^2} &= q^2 \psi(f) df \int_A da (w_e + w_c)^2 (\Delta p)^2, \\ &= q^2 \psi(f) df \pi (R_c^2 - R_e^2 + R_c L) p_0^2 (1 - a_1) \end{aligned} \quad (36)$$

where

$$a_1 = \frac{2}{\pi} \frac{R_e W}{(R_c^2 - R_e^2 + R_c L)} \left[3 + 2 \left(\frac{p_n}{p_0} \right) - \left(\frac{p_n}{p_0} \right)^2 \right].$$

For $p_0 \cong 0.26 p_n$, $|a_1| \leq 8R_e W / \pi (R_c^2 - R_e^2 + R_c L) \approx 0.064$ for the units studied here. Thus, except at extremely small injections, a_1 may be neglected. Substituting for p_0 either in terms of base current [Equation (19d) with $g = 0$] or in terms of collector current [Equation (31a) with $g = 0$], one obtains the alternative formulas of Equation (8).

If the m term is neglected, the summed $dv_{bb'}$ generator is $I_b \partial R_{bb'} / \partial I_b$ times the summed composite $di_{b'e}$ generator. Because w_e , w_c depend on r , the summed di_{ce} generator is not perfectly correlated with these two. Neglecting the m term, in the common-emitter connection the contribution of all three dg generators to F_{OUT}^s is

$$\begin{aligned} \Delta F_{\text{OUT}}^s &= \frac{R_G}{4kTdf} \left[1 + \frac{r_{bb'}}{R_G} \right]^2 \\ &\sum \left[1 - \frac{I_b}{(R_G + r_{bb'})} \frac{\partial R_{bb'}}{\partial I_b} + \frac{1}{g_m R} \frac{w_c}{(w_e + w_c)} \right]^2 q^2 (w_e + w_c)^2 \overline{dg^2}, \\ &= \frac{R_G}{4kTdf} \left[1 + \frac{r_{bb'}}{R_G} \right]^2 \left[A_1^2 + 2A_1 \frac{a_2}{g_m R} + \frac{a_3}{(g_m R)^2} \right] \sum \overline{di_{b'e}^2}, \end{aligned} \quad (37)$$

where $\sum \overline{di_{b'e}^2}$ is given by Equation (36) and where

$$A_1 = 1 - \frac{I_b}{(R_G + r_{bb'})} \frac{\partial R_{bb'}}{\partial I_b},$$

$$a_2 = \frac{4}{3\pi} \frac{R_e W}{(R_c^2 - R_e^2 + R_c L)} \frac{1}{(1 - a_1)} \left[1 - \left(\frac{p_n}{p_0} \right) + \left(\frac{p_n}{p_0} \right)^2 \right],$$

$$a_3 = \frac{1}{3\pi} \frac{R_e W}{(R_c^2 - R_e^2 + R_c L)} \frac{1}{(1 - a_1)} \left[1 - 2 \left(\frac{p_n}{p_0} \right) + 3 \left(\frac{p_n}{p_0} \right)^2 \right].$$

The A_1^2 term of Equation (37) is the contribution of the $di_{b'e}$ and $dv_{bb'}$ generators alone and is very close to unity. The $A_1 a_2$ term arises

from the correlation between these two generators and the di_{ce} generator. It is always positive. The a_3 term is the contribution of the di_{ce} generator alone. For $p_0 \cong 0.36 p_n$, both a_2 and a_3 are smaller than the limit $8R_e W/\pi(R_e^2 - R_e^2 + R_e L)$ quoted for $|a_1|$ above. Since $g_m R$ was greater than 0.42 for all the inverted-triode measurements reported here, the $A_1 a_2$ and a_3 terms may be neglected. That is, the di_{ce} generator may be neglected.

The applicability of the above discussion to the $1/f$ noise observed in inverted triodes is discussed in Section IV.

The normal triode will now be treated. The spatial dependence of the excess pair density Δp is described approximately by the configuration shown in Figure 23(d). In front of the small dot, Δp drops linearly with z from p_0 to $-p_n$ at the collector. Beyond R_e , Δp is quite different from the corresponding density distribution in the forward diode or inverted triode. At the side of the emitter, Δp falls from p_0 to $-p_n$ in a radial distance of approximately W . In this paper Δp will be taken as $-p_n + (p_0 + p_n) e^{-\pi(r-R_e)/2W}$ along the base surface just beyond R_e . After the rapid drop, Δp retains the value $-p_n$ out to R_c , after which its absolute value decreases exponentially to zero (characteristic length L). The location and magnitude of net surface recombination are again indicated by the shaded areas. Where Δp is negative there is a net surface generation; this has not been indicated in Figure 23.

As a preliminary to noise, the normal-triode schematic, including generators representing a point source of pairs of strength g at \underline{r} in the triode base and a leakage current I_L across the collector junction, will be derived. The previous treatment of the inverted triode can be reapplied here. Indeed, up to Equation (36), the corresponding inverted-triode text can be read word for word here, with but three alterations:

1. When Figure 23(c) is cited, refer instead to Figure 23(d).
2. The formulas for the emitter and collector saturation currents must be reversed. That is, instead of Equation (22a), read

$$I_{es} = 4qsp_n R_e W, \quad (22d)$$

and instead of Equation (22b), read

$$I_{cs} = \pi qsp_n (R_c^2 - R_e^2 + 4R_e L - \frac{4}{\pi} R_e W) \gg I_{es}. \quad (22e)$$

This changes the formula for the collector-base current amplification

factor from Equation (35) to Equation (13). According to Equation (13) α_{cb} should be approximately 145 for the units studied here. Measured values ranged from 16 to 126, showing that an idealized parallel-plane geometry with no bulk recombination was not quite realized for the units here.

3. The formulas for the probabilities w_e and w_c must be reversed. That is, instead of Equation (29a), read

$$\begin{aligned} w_e &= e^{-\pi(r-R_e)/2W}, & w_c &= 1 - w_e, & r < R_e, \\ &= 0, & &= e^{-(r-R_e)/L}, & r > R_e. \end{aligned} \quad (29b)$$

In the normal-triode case, the summed composite $di_{b'e}$ generator has the spectrum

$$\sum \overline{di_{b'e}^2} = q^2 \psi(f) df 2R_e W p_0^2 (1 + a_4), \quad (36a)$$

where

$$a_4 = \frac{\pi}{2} \frac{(R_c^2 - R_e^2 + R_e L)}{R_e W} \left(\frac{p_n}{p_0} \right)^2 - 3 \left(\frac{p_n}{p_0} \right)^2 - 2 \left(\frac{p_n}{p_0} \right).$$

For the units studied here, $a_4 \approx 59 (p_n/p_0)^2 - 2 (p_n/p_0)$. Since p_0/p_n was greater than 18 for all normal-triode measurements reported here, a_4 may be neglected. Substituting for p_0 , one obtains the formulas of Equation (8a).

Neglecting the m term, in the common-emitter connection the contribution of all three dg generators to F_{OUT} is still given by Equation (37), but where $\sum \overline{di_{b'e}^2}$ is given by Equation (36a) and where a_2 and a_3 have the revised values

$$a_2 = \frac{1}{3} \frac{1}{(1 + a_4)} \left[1 - 4 \left(\frac{p_n}{p_0} \right) - 11 \left(\frac{p_n}{p_0} \right)^2 + \frac{3\pi}{2} \frac{(R_c^2 - R_e^2 + R_e L)}{R_e W} \left(\frac{p_n}{p_0} \right)^2 \right],$$

$$a_3 = \frac{1}{6} \frac{1}{(1 + a_4)} \left[1 - 6 \left(\frac{p_n}{p_0} \right) - 25 \left(\frac{p_n}{p_0} \right)^2 + 3\pi \frac{(R_c^2 - R_e^2 + R_e L)}{R_e W} \left(\frac{p_n}{p_0} \right)^2 \right].$$

The physical interpretation of the A_1^2 , $A_1 a_2$, and a_3 terms is unchanged. For the units studied here,

$$a_2 \approx \frac{1}{3} \frac{1}{(1 + a_4)} \left[1 - 4 \left(\frac{p_n}{p_0} \right) + 175 \left(\frac{p_n}{p_0} \right)^2 \right],$$

$$a_3 \approx \frac{1}{6} \frac{1}{(1 + a_4)} \left[1 - 6 \left(\frac{p_n}{p_0} \right) + 347 \left(\frac{p_n}{p_0} \right)^2 \right].$$

Since p_0/p_n was greater than 18 and $g_m R$ was greater than 0.88 for all normal-triode measurements reported here, the $A_1 a_2$ and a_3 terms may be neglected. That is, the di_{ce} generator may be neglected.

The applicability of the above discussion to the $1/f$ noise observed in normal triodes is discussed in Section IV.

APPENDIX II — BASE CONDUCTIVITY MODULATION

$I_b \partial R_{bb'}/\partial I_b$ has been found empirically to be nearly independent of current in the region of interest — heavy overload where p_0 is larger than the base thermal-equilibrium majority-carrier density. This can be explained by a short argument. Consider the forward-double-diode connection [Figure 23(b)]. Small-signal base conductivity modulation occurs in a ring-shaped region of radial thickness λ just beyond R_c . λ is a conceptual length, a few times L , which will eventually drop out of the formulas. This situation obtains because, first, injection is so heavy in front of the dots that small-signal additions make no appreciable contribution to the conductivity there and, second, injection does not extend beyond λ . Provided $\lambda \ll R_c$, the portion of d-c base-lead resistance $R_{bb'}$ arising from the λ region is

$$R_{bb'}^\lambda = \frac{1}{2\pi R_c W} \int_{R_c}^{R_c + \lambda} \rho \, dr,$$

$$= \frac{\rho_n}{2\pi R_c W} \int_{R_c}^{R_c + \lambda} \frac{dr}{\left[1 + \left(1 + \frac{\mu_p}{\mu_n} \right) \left(\frac{p_0}{n_n} \right) e^{-(r-R_c)/L} \right]},$$

$$\approx \frac{\rho_n}{2\pi R_c W} \left\{ \lambda - L \ln \left[1 + \left(1 + \frac{\mu_p}{\mu_n} \right) \left(\frac{p_0}{n_n} \right) \right] \right\},$$

where it has been assumed that $e^{-\lambda/L} \ll 1$ in the last step. Taking the derivative, one now loses λ ;

$$\frac{\partial R_{bb'}}{\partial I_b} \approx \frac{\partial R_{bb'}^\lambda}{\partial I_b} = \frac{-L\rho_n}{2\pi R_c W} \frac{\left(1 + \frac{\mu_p}{\mu_n}\right) \frac{1}{n_n} \frac{\partial p_0}{\partial I_b}}{\left[1 + \left(1 + \frac{\mu_p}{\mu_n}\right) \left(\frac{p_0}{n_n}\right)\right]}$$

Using Equation (19),

$$I_b \frac{\partial R_{bb'}}{\partial I_b} = \frac{-L\rho_n}{2\pi R_c W} \frac{1}{(1 + I_1/I_b)}, \quad (38)$$

where

$$I_1 = \frac{\pi s (R_c^2 - R_e^2 + 4R_c L)}{(\mu_n + \mu_p) \rho_n} \approx 0.43 \times 10^{-3} \text{ ampere} \quad (39)$$

for the units studied here. For $I_b \gg I_1$, $I_b \partial R_{bb'}/\partial I_b$ has the constant value $-L\rho_n/2\pi R_c W \approx -34$ ohms for the units studied here.

The identical analysis applies to the inverted triode.

For the normal triode, the area $R_c W$ is replaced by an area of approximately $(1/2) R_e W$, the length L is replaced by the length $(2/\pi) W$, and Equation (19) is replaced by Equations (19d), with $g = 0$ and $I_b \gg I_{cs} + I_L$, and (22d). The final result is

$$I_b \frac{\partial R_{bb'}}{\partial I_b} = \frac{-2\rho_n}{\pi^2 R_e} \frac{1}{(1 + I_2/I_b)}, \quad (38a)$$

where

$$I_2 = \frac{4s R_e W}{(\mu_n + \mu_p) \rho_n} \approx 0.8 \times 10^{-5} \text{ ampere} \quad (39a)$$

for the units studied here. The normal triode is heavily overloaded at much lower base currents. For $I_b \gg I_2$, $I_b \partial R_{bb'}/\partial I_b$ has the constant value $-2\rho_n/\pi^2 R_e \approx -26$ ohms for the units studied here.

THE NOISE FACTOR OF JUNCTION TRANSISTORS

BY

L. J. GIACOLETTO

RCA Laboratories,
Princeton, N. J.

Summary—Formulas are developed for the noise factor of junction transistors. These formulas contain shot-noise current generators for the intrinsic transistor, a thermal-noise generator for the base-lead resistance, and an arbitrary noise generator for the $1/f$ sources of noise. The results include dependency upon d-c biases, frequency, and source admittance. The development follows logically from a similar treatment for diode noise and is supported with some measurements of a radio-frequency transistor. When the $1/f$ sources of noise are negligible, the noise factor for normal conditions is largely determined by the shot noise of the d-c base current and approaches unity (zero decibels) as a minimum value when the collector-to-base current amplification factor is large.

INTRODUCTION

THE noise factor of a junction transistor was considered by Montgomery and Clark.¹ Their development was in reasonable agreement with measurements at small currents. Discrepancies that appeared at larger currents were explained by van der Ziel² with the aid of a rather artificial concept of partition noise. The noise factor formulation herein³ differs in many respects from the prior developments. The results given here were originally obtained by means of heuristic reasoning based upon shot noise as used in diode noise treatment. The diode results have been found to be in reasonable agreement with measurements.⁴

Recently North⁵ has developed a mesh model for studying diffusion-recombination noise. He considers noise arising from recombination

¹ H. C. Montgomery and M. A. Clark, "Shot Noise in Junction Transistors," *Jour. Appl. Phys.*, Vol. 24, pp. 1337-1338, October, 1953.

² A. van der Ziel, "Note on Shot and Partition Noise in Junction Transistors," *Jour. Appl. Phys.*, Vol. 25, pp. 815-816, June, 1954.

³ Originally presented at 1954 *Conference on Semiconductor Device Research*, University of Minnesota, June 28-30, 1954.

⁴ R. L. Anderson and A. van der Ziel, "On the Shot Effect of P-N Junctions," *Trans. I.R.E. Professional Group on Electron Devices, PGED-1*, pp. 20-24, November, 1952.

⁵ D. O. North, "A Physical Theory of Noise in Semiconductors and Transistors," 1955 *Conference on Semiconductor Device Research*, University of Pennsylvania, June 20-22, 1955.

of individual carrier pairs and obtains the net terminal noise by a summation process. North's results at low frequencies are the same as the shot-noise formulation given herein. His results indicate that the shot-noise generators as given herein must be modified at high frequencies when the transit time through the transistor base becomes significant. However, this modification is not necessary for normal use as the transistor performance as a function of frequency has been greatly curtailed before the discrepancies became significant. North's theory also includes excess noise originating at the surface of the transistor. This source of noise has been carefully investigated by Fonger⁶ who finds that it has an approximate $1/f$ frequency dependency. This source of noise is not specifically considered here. However, Fonger's study indicates that it can, in part, be lumped with the $1/f$ leakage noise source contained in the present development. This method of consideration greatly simplifies the development since all low-frequency noise is contained in one generator and is particularly appropriate since neither the $1/f$ surface noise nor the $1/f$ leakage noise can be formulated analytically. The error introduced by this method of consideration will normally be insignificant.

In view of the above considerations, the formulas given herein can be expected to give good results over a wide range of operating conditions and should prove useful to both device and circuit designers. The development has usually been left in a general form to include dependency upon d-c bias voltages, frequency, and source admittance. Particular operating conditions can be inserted into the equations as desired to get simplified results.

DIODE NOISE

The method of inserting noise generators into the transistor equivalent circuit can be more easily understood in connection with a diode. A junction diode can be characterized by a conductance, g_0 , or a current coefficient, I_0 . The two are related by a constant, $\Lambda = q/kT$, where, in this paper, q (carrier charge) has a suitable sign determined such that ΔV is positive when the junction is biased in the forward direction. The current coefficient, I_0 , is a measure of the current (electron + hole current) flowing in one direction across the junction barrier. For thermal equilibrium (d-c bias voltage and current are zero), there is an equal current flowing in the opposite direction across the junction barrier. Therefore the terminal current, $I = I_0 - I_0 = 0$.

⁶ W. H. Fonger, "A Determination of $1/f$ Noise Sources in Semiconductor Diodes and Triodes," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 239-295.

The flow of the two currents arises from random motions of the carriers so that the mean-square noise current, \bar{I}_n^2 , of the junction for thermal equilibrium can alternatively be formulated as due to thermal (Johnson) noise of the conductance, g_0 , or due to the sum of the shot noise of the two currents. Thus, $\bar{I}_n^2 = 4kTg_0\Delta f = 4qI_0\Delta f$. When a d-c bias voltage, V , is applied to the junction, the current in one direction is altered by an exponential factor while the current in the opposite direction is unaltered. Accordingly, $I = I_0e^{\Delta V} - I_0$. Assigning shot noise to each of the two currents, $\bar{I}_n^2 = 2qI_0e^{\Delta V}\Delta f + 2qI_0\Delta f = 2q(I + 2I_0)\Delta f$. This result indicates that: (1) when the junction is biased in the reverse direction, \bar{I}_n^2 is the shot noise of the d-c terminal current, I_0 ; (2) When the junction is biased in the forward direction, \bar{I}_n^2 is the shot noise, approximately, of the d-c terminal current, I ; and (3) When the junction is unbiased, \bar{I}_n^2 is the shot noise of twice the current flow in one direction across the junction. These results are entirely analogous to those found in an electron tube diode.⁷

PASSIVE TRANSISTOR

For thermal equilibrium (d-c bias voltages and currents are zero) the transistor is a three-terminal passive device and can be characterized by three independent positive conductances, g_{EB} , $-g_{EC}$, g_{CB} , or three independent current coefficients, I_{EB} , $-I_{EC}$, I_{CB} , related to the conductances by the factor, Λ . The result is a delta arrangement of three conductances as shown in Figure 1a. The negative designation of the $-g_{EC}$ conductance arises from the usual convention of currents and voltages (see Figure 2). The noise properties of the passive transistor can be specified in terms of three shunt uncorrelated mean-square noise current generators as shown in Figure 1a. By Nyquist's noise theorem these three current generators can be written in terms of their associated conductances which conductances can in turn be expressed in terms of current coefficients or saturation currents. Thus

$$\begin{aligned}\bar{I}_{eb}^2 &= 4kTg_{EB}\Delta f = 4q(I_{EE} + I_{EC})\Delta f = -4qI_{ES}\Delta f, \\ \bar{I}_{ec}^2 &= -4kTg_{EC}\Delta f = -4qI_{EC}\Delta f, \\ \bar{I}_{cb}^2 &= 4kTg_{CB}\Delta f = 4q(I_{CC} + I_{EC})\Delta f = -4qI_{CS}\Delta f.\end{aligned}\tag{1}$$

⁷ K. S. Knol and G. Diemer, "Theory and Experiments on Electrical Fluctuations and Damping of Double-Cathode Valves," *Philips Research Reports*, Vol. 5, pp. 131-152, April, 1950.

The significance of the conductors, current coefficients, and saturation currents and their relationship to transistor material and dimensions can be obtained from published articles.^{8, 9} The results of Equations (1) can be interpreted in terms of shot-noise formulation. Thus \bar{I}_{eb}^2 arises from equal and opposite currents, $-I_{ES}$, across the emitter junction; \bar{I}_{cb}^2 arises from equal and opposite currents, $-I_{CS}$, across the collector junction; and \bar{I}_{ec}^2 arises from equal and opposite currents, I_{EC} , through the base and across both junctions.

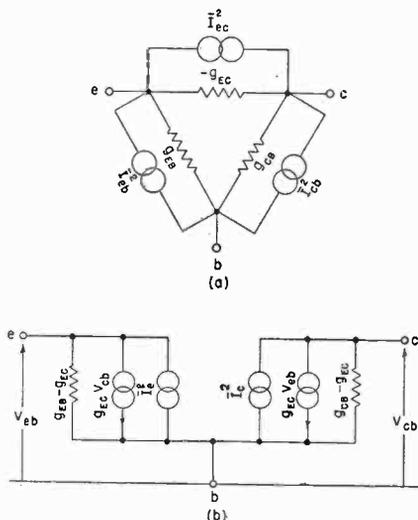


Fig. 1—Transistor equivalent circuit without applied biases.

An alternative representation of the equivalent circuit of Figure 1a is as shown in Figure 1b. The noise generators here can be formulated from those given in Equation (1). The results are

$$I_e^2 = 4q I_{EE} \Delta f,$$

$$I_c^2 = 4q I_{CC} \Delta f. \tag{2}$$

I_{EE} and I_{CC} represent total currents in one direction across the emitter and collector junctions respectively composed of currents to the base, I_{ES} and I_{CS} respectively, and transfer currents I_{EC} . That is, the emitter

⁸ L. J. Giacoletto, "Study of P-N-P Alloy Junction Transistor from D-C Through Medium Frequencies," *RCA Review*, Vol. 15, pp. 506-562, December, 1954.

⁹ W. Shockley, M. Sparks, and G. K. Teal, "P-N Junction Transistors," *Phys. Rev.*, Vol. 83, pp. 151-162, July 1, 1951.

and collector total junction conductances, g_{BE} and g_{CC} respectively, consist of self conductances, g_{ES} and g_{CS} , respectively and a transfer conductance, $-g_{EC}$. The noise generators in Figure 1b have some correlation between them (via I_{EC}) whereas the generators in Figure 1a are all uncorrelated.

ACTIVE TRANSISTOR

When d-c bias voltages, $V_{EB'}$ and $V_{CB'}$, are applied to the emitter and collector junctions respectively, the currents in one direction are altered by an exponential factor while the similar currents in the opposite direction remain unaltered. Shot noise is assigned to these cur-

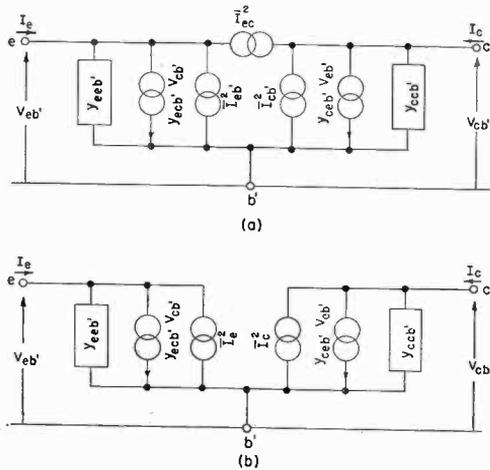


Fig. 2—Common-base equivalent circuit with noise generators.

rents in the same manner as was done above for a single junction. The resulting noise generators as shown in Figure 2a can be formulated with the description of the transistor operation given above or with the aid of the following transistor d-c equations:⁸

$$I_B = I_{EB}(e^{\Delta V_{EB'}} - 1) + I_{BC}(e^{\Delta V_{CB'}} - 1) = -I_{ES}(e^{\Delta V_{EB'}} - 1) - I_{EC}(e^{\Delta V_{EB'}} - e^{\Delta V_{CB'}}), \tag{3}$$

$$I_C = I_{EC}(e^{\Delta V_{EB'}} - 1) + I_{CC}(e^{\Delta V_{CB'}} - 1) = -I_{CS}(e^{\Delta V_{EB'}} - 1) - I_{EC}(e^{\Delta V_{EB'}} - e^{\Delta V_{CB'}}),$$

Thus

$$\begin{aligned}\bar{I}_{cb'}^2 &= -2qI_{ES}(e^{AV_{EB'}} + 1) \Delta f \approx -2q(I_B - I_{BS} + I_{ES}) \Delta f \approx -2qI_B \Delta f, \\ \bar{I}_{cc'}^2 &= -2qI_{EC}(e^{AV_{EB'}} + e^{AV_{CB'}}) \Delta f \approx -2q(I_C - I_{CS}) \Delta f \approx -2qI_C \Delta f, \\ \bar{I}_{cb}^2 &= -2qI_{CS}(e^{AV_{CB'}} + 1) \Delta f \approx -2qI_{CS} \Delta f \approx 0.\end{aligned}\quad (4)$$

The first approximation in these equations is valid when the d-c collector voltage is large enough for saturation, i.e., when $-\Delta V_{CB'} > 1$; the second approximation is valid when, in addition, the d-c terminal currents are large in comparison with the saturation currents. The above noise generators are shown in Figure 2a superimposed upon the common-base equivalent circuit of the transistor. The equivalent circuit here as well as some others to follow does not include certain extrinsic elements to be considered subsequently; this is the reason for the b' designations. With the aid of previously derived equations for the transistor parameters⁸ it can be shown that the conductances and associated noise current generators of Figure 1a are all accommodated in Figure 2a. Thus, the circuit of Figure 2a reverts to that of Figure 1a when bias voltages are made zero.

An alternative representation of the circuit in Figure 2a is shown in Figure 2b. Here, using Equation (3) and the first expressions in Equation (4),

$$\begin{aligned}\bar{I}_e^2 &= 2q(I_B - 2I_{ES}) \Delta f - 4qI_{EC} e^{AV_{CB'}} \Delta f, \\ \bar{I}_c^2 &= 2q(I_C - 2I_{CS}) \Delta f - 4qI_{EC} e^{AV_{EB'}} \Delta f.\end{aligned}\quad (5)$$

These noise generators, shown in Figure 2b, are strongly correlated. For this reason, the noise factor can be more easily formulated from Figure 2a. Here, in accordance with the previous logical development, the noise generators are completely uncorrelated. Accordingly, the spot noise factor for the intrinsic transistor of Figure 2 when operated in a common-base circuit between a current generator of internal admittance $Y_g = G_g + jB_g$, and an arbitrary load admittance is

$$\begin{aligned}F_{b'} &= 1 + \frac{\Delta}{2G_g} \left[-I_{ES}(e^{AV_{EB'}} + 1) - \left| \frac{Y_g + y_{eeb'}}{y_{ceb'}} \right|^2 I_{CS}(e^{AV_{CB'}} + 1) \right. \\ &\quad \left. - \left| \frac{Y_g + y_{eeb'} + y_{ceb'}}{y_{ceb'}} \right|^2 I_{EC}(e^{AV_{EB'}} + e^{AV_{CB'}}) \right].\end{aligned}\quad (6)$$

The common-emitter equivalent circuit with noise generators can be obtained by simple transformations.^{8, 10} The results are shown in Figure 3. Using Equation (3) and the first expressions in Equation (4) the noise generators in Figure 3b are

$$\begin{aligned} \bar{I}_b^2 &= -2q(I_B - I_{BS}) \Delta f, \\ \bar{I}_c^2 &= 2q(I_C - I_{CS}) \Delta f - 4qI_{EC} e^{\Delta V_{EB'}} \Delta f. \end{aligned} \quad (7)$$

These noise generators are correlated to some extent so that the noise factor is best formulated from Figure 3a where the noise generators are completely uncorrelated. Thus, the spot noise factor for the intrinsic transistor of Figure 3 when operated between a current generator of internal admittance, $Y_g = G_g + jB_g$, and an arbitrary load admittance is

$$\begin{aligned} F_{e'} &= 1 + \frac{\Lambda}{2G_g} \left[-I_{BS}(e^{\Delta V_{EB'}} + 1) \right. \\ &\quad - \left| \frac{Y_g + y_{b'b'e'}}{y_{cb'e'}} \right|^2 I_{EC}(e^{\Delta V_{EB'}} + e^{\Delta V_{CB'}}) \\ &\quad \left. - \left| \frac{Y_g + y_{b'b'e'} + y_{cb'e'}}{y_{cb'e'}} \right|^2 I_{CS}(e^{\Delta V_{CB'}} + 1) \right]. \end{aligned} \quad (8)$$

A development similar to that for the common-base and common-emitter circuits can be pursued to obtain the spot noise factor of the intrinsic transistor in a common-collector circuit. The result is

$$\begin{aligned} F_{e'} &= 1 + \frac{\Lambda}{2G_g} \left[-I_{CS}(e^{\Delta V_{CB'}} + 1) \right. \\ &\quad - \left| \frac{Y_g + y_{b'b'c'}}{y_{cb'c'}} \right|^2 I_{BC}(e^{\Delta V_{EB'}} + e^{\Delta V_{CB'}}) \\ &\quad \left. - \left| \frac{Y_g + y_{b'b'c'} + y_{eb'c'}}{y_{eb'c'}} \right|^2 I_{BS}(e^{\Delta V_{EB'}} + 1) \right]. \end{aligned} \quad (9)$$

¹⁰ L. J. Giacoletto, "Terminology and Equations for Linear Active Four-Terminal Networks Including Transistors," *RCA Review*, Vol. 14, pp. 28-46, March, 1953.

COMPLETE TRANSISTOR

The development to this point has been applicable to the intrinsic transistor. Actual transistors have extrinsic elements which must be added into the equivalent circuits together with their associated noise. First there is a base-lead resistor, $r_{bb'}$, connecting the external base, b , to the internal base, b' . This resistor has a thermal mean-square noise voltage, $4kT r_{bb'} \Delta f$, associated with it. The second extrinsic element is a collector-to-internal base leakage conductor, g_l . This leakage conductor is one source of the $1/f$ "flicker" noise spectrum of junction transistors. Its noise has not been analytically formulated so it is represented generally as a mean-square noise current generator,

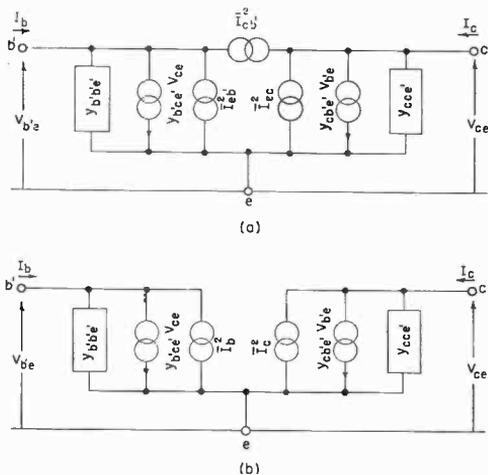


Fig. 3—Common-emitter equivalent circuit with noise generators.

$\bar{I}_{g_{ln}}^2$, or as an equivalent noise conductance, \bar{g}_{ln} . As mentioned in the Introduction, the $1/f$ surface noise can also be lumped into the $\bar{I}_{g_{ln}}^2$ current generator. The $1/f$ leakage noise will be strongly dependent upon the leakage current whereas the $1/f$ surface noise will be strongly dependent upon the emitter current. Thus, in general $\bar{I}_{g_{ln}}^2$ will vary with both collector voltage and current.

The complete common-emitter equivalent circuit including appropriate noise generators, all of which are completely uncorrelated, is shown in Figure 4. When operated between a voltage generator of internal impedance, $Z_g = R_g + jX_g = 1/Y_g = [G_g + jB_g]^{-1}$ and an arbitrary load impedance, the spot noise factor is

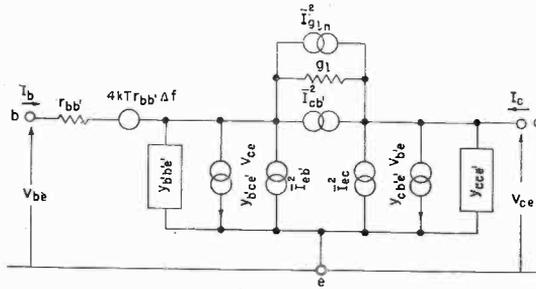


Fig. 4—Complete common-emitter equivalent circuit with noise generators.

$$\begin{aligned}
 F_e = & 1 + \frac{r_{bb'}}{R_g} + \frac{\Lambda |r_{bb'} + Z_g|^2}{2R_g} \left\{ -I_{ES} (e^{\Lambda V_{EB'}} + 1) \right. \\
 & - \left| \frac{(1/[r_{bb'} + Z_g]) + y_{b'b'e'} + g_l}{y_{cb'e'} - g_l} \right|^2 I_{EO} (e^{\Lambda V_{EB'}} + e^{\Lambda V_{CB'}}) \\
 & + \left. \left| \frac{(1/[r_{bb'} + Z_g]) + y_{b'b'e'} + y_{cb'e'}}{y_{cb'e'} - g_l} \right|^2 \right. \\
 & \left. \left[-I_{CS} (e^{\Lambda V_{CB'}} + 1) + \frac{2\bar{g}_{ln}}{\Lambda} \right] \right\}. \tag{10}
 \end{aligned}$$

The noise factors for the common-base and common-collector circuits can be formulated in the same manner as for the common-emitter circuit. The complete equivalent circuits with uncorrelated noise generators are shown in Figures 5 and 6. The spot noise factors for these circuits (see Reference (8) for definitions of the transistor parameters) are

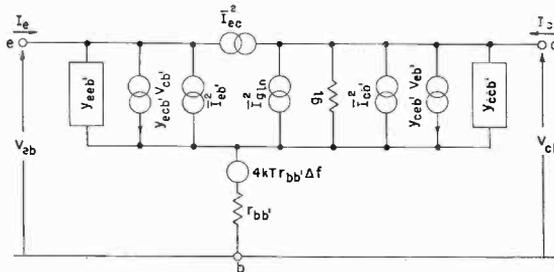


Fig. 5—Complete common-base equivalent circuit with noise generators.

$$\begin{aligned}
 F_b = 1 + |y_{ceb'} - r_{bb'}\Delta|^{-2} & \left\{ |\Delta + Y_g (y_{ceb'} + y_{ccb'} + g_l)|^2 \frac{r_{bb'}}{G_g} \right. \\
 - \frac{\Lambda}{2G_g} |Y_g + y_{eeb'} + y_{ceb'} + r_{bb'}Y_g (y_{eeb'} & \\
 + y_{ccb'} + y_{ceb'} + y_{ccb'} + g_l)|^2 I_{EC} (e^{\Lambda V_{EB'}} + e^{\Lambda V_{CB'}}) & \\
 - \frac{\Lambda}{2G_g} |y_{ceb'} + r_{bb'}Y_g (y_{ceb'} + y_{ccb'} + g_l)|^2 I_{ES} (e^{\Lambda V_{EB'}} + 1) & \\
 + \frac{\Lambda}{2G_g} |Y_g + y_{ceb'} + r_{bb'}Y_g (y_{ceb'} & \\
 + y_{ccb'})|^2 \left[-I_{CS} (e^{\Lambda V_{CB'}} + 1) + \frac{2\bar{g}_{in}}{\Lambda} \right] & \left. \right\} \quad (11)
 \end{aligned}$$

where $\Delta = y_{eeb'} (y_{ccb'} + g_l) - y_{ceb'} y_{eeb'}$; and

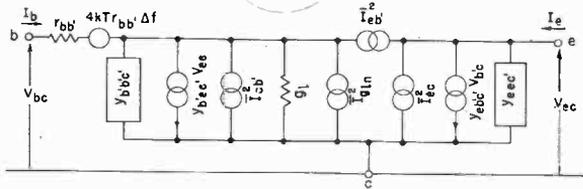


Fig. 6—Complete common-collector equivalent circuit with noise generators.

$$\begin{aligned}
 F_c = 1 + \frac{r_{bb'}}{R_g} + \frac{\Lambda}{2R_g} & \left\{ |r_{bb'} + Z_g|^2 \left[-I_{CS} (e^{\Lambda V_{CB'}} + 1) + \frac{2\bar{g}_{in}}{\Lambda} \right] \right. \\
 - \left| \frac{1 + (r_{bb'} + Z_g) (y_{b'b'c'} + g_l)}{y_{eb'c'}} \right|^2 I_{EC} (e^{\Lambda V_{EB'}} + e^{\Lambda V_{CB'}}) & \\
 - \left| \frac{1 + (r_{bb'} + Z_g) (y_{b'b'c'} + y_{eb'c'})}{y_{eb'c'}} \right|^2 I_{ES} (e^{\Lambda V_{EB'}} + 1) & \left. \right\} \cdot \quad (12)
 \end{aligned}$$

The preceding noise factor equations have general validity to cover a wide range of circuit operation. For certain restricted circuit conditions suitable approximations may be made to simplify the applica-

tion of the equations. This will be illustrated in connection with Equation (10), but similar simplifications may be effected with Equations (11) and (12). Thus, if the second approximations in Equation (4) are used, if operation at low frequencies only is considered, and if both leakage conductance and equivalent noise conductance, are zero, Equation (10) becomes

$$F_e \approx 1 + \frac{r_{bb'}}{R_g} + \frac{(r_{bb'} + R_g)^2}{2R_g} \left[|\Delta I_B| + \left(\frac{1}{r_{bb'} + R_g} + |\Delta I_B| \right) |\Delta I_C|^{-1} \right]$$

If, further, the generator is image matched to the transistor input for maximum low-frequency power amplification so that $1/R_g \approx |\Delta I_B| \ll 1/r_{bb'}$, then $F_e \approx 1 + 1/2 [1 + (4/|\alpha_{cb}|)]$. For a transistor with a large current amplification factor, α_{cb} , $F_e \approx 1.5$ (1.76 decibels). The noise factor will be even smaller if the generator is mismatched to the transistor. Thus, at low frequencies and using the same approximations as above, the minimum noise factor, $F_e \approx 1 + (1/\sqrt{|\alpha_{cb}|})$ and is obtained when $G_g \approx (|\Delta I_C|/\sqrt{|\alpha_{cb}|})$; the minimum noise factor approaches unity when $|\alpha_{cb}|$ is large. Measurements made on recently manufactured junction transistors have indicated spot noise factors of 2 decibels at 1 kilocycle.¹¹ These transistors are very close to their theoretical limit since they contain extrinsic sources of noise.

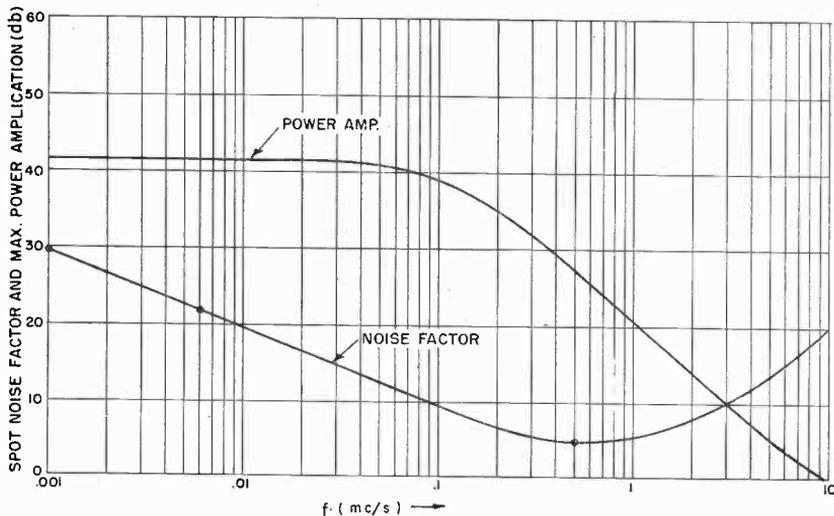


Fig. 7—Spot noise factor and maximum power amplification versus frequency for a radio-frequency transistor.

¹¹ J. W. England, "Noise Considerations for P-N-P Junction Transistors." *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 309-321.

MEASUREMENTS

Figure 7 shows curves of maximum power amplification and associated spot noise factor for a radio-frequency transistor of a type recently described¹² in a common-emitter circuit with partial neutralization so that the net feedback is resistive. Measured spot noise factors are shown by the datum points at 1, 6, and 500 kilocycles. The measured value at 1 kilocycle was used to determine \bar{g}_{in} , and the measured value at 6 kilocycles was used to determine the $f^{-1.04}$ frequency dependence of \bar{g}_{in} . The measured value at 500 kilocycles is therefore the only value

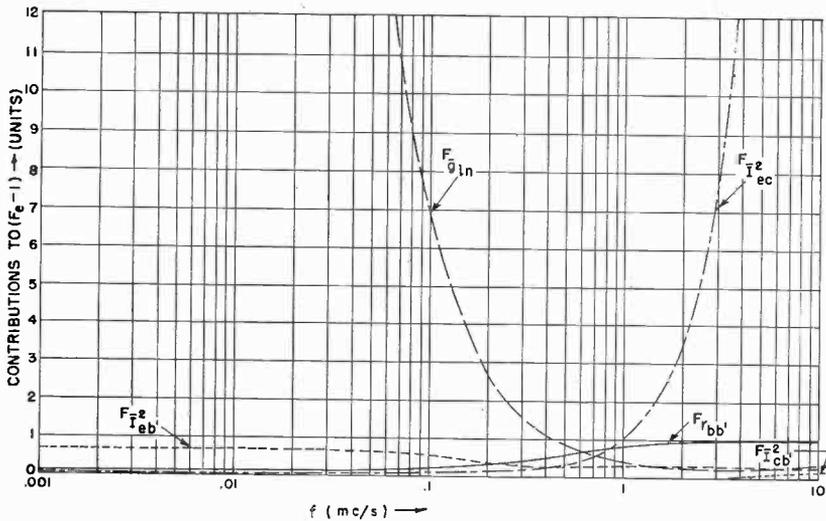


Fig. 8—Individual noise contributions to $(F_e - 1)$.

that can be used to check the formulation of Equation (10). The agreement in this case is quite good. Similar good agreements have been obtained in other comparisons.

In Figure 8 the five individual noise contributions to $F_e - 1$ are shown on a linear scale. It is noted that the noise due to \bar{g}_{in} predominates at low frequencies, whereas the noise due to \bar{I}_{ec}^2 predominates at high frequencies. The noise due to \bar{I}_{cb}^2 is negligible throughout the range of frequencies.

¹² C. W. Mueller and J. I. Pankove, "A P-N-P Triode Alloy Junction Transistor for Radio-Frequency Amplification," *RCA Review*, Vol. 14, pp. 586-598, December, 1953; also *Proc. I.R.E.*, Vol. 42, pp. 386-391, February, 1954.

The operating conditions of the transistor and the data¹³ used to compute the noise factor with Equation (10) and approximations given in Equation (4) are as follows: $V_{CE} = -6$ volts; $I_C = -1$ milliamperes; $I_B = -31.8$ microamperes (for use in Equation (4) this value must be corrected by the addition of -0.84 microampere corresponding to current flow due to g_l); $I_{CS} = -0.75$ microampere; $I_{BS} = 1.27$ microampere; $I_{ES} = -0.52$ microampere; $g_l = 0.14$ micromho; $\overline{g_{ln}} = 0.87$ mho at 1 kilocycle with an $f^{-1.04}$ frequency dependency. The remaining factors of Equation (10) can be computed⁸ using the following transistor parameters: $r_{bb'}$ = 100 ohms; $g_{b'e} = 0.98$ millimho; $C_{b'e} = 1800$ micromicrofarads; $g_{b'c} = 0.13$ micromho (this does not include g_l); $C_{b'c} = 14$ micromicrofarads; $g_{ce} = 13.8$ microhms; and $g_m = 34.7$ millimhos.

ACKNOWLEDGMENT

The measured noise factors shown in Figure 7 were supplied by L. A. Freedman and W. H. Fonger of RCA Laboratories.

¹³ L. J. Giacoletto, "Performance of a Radio-Frequency Alloy Junction Transistor in Different Circuits," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 431-457.

NOISE CONSIDERATIONS FOR P-N-P JUNCTION TRANSISTORS

By

JOHN W. ENGLUND

RCA Semiconductor Division,
Harrison, N. J.

Summary—This paper discusses the various sources of noise in p-n-p alloy-junction transistors, and reports the progress achieved in the reduction of noise in alloy-junction transistors. Although the inherent high noise factor of point-contact transistors (50 to 70 decibels at a frequency of 1 kilocycle) precludes their use in low-level audio equipment, junction transistors have a considerably lower noise factor (10 to 35 decibels for older "noisy" units, and 2 to 15 decibels for low-noise units described herein), and are now practical for these applications. Because junction transistors introduce no appreciable microphonism, eliminate heater-supply hum, and provide a low noise level, it is now possible to improve the useful sensitivity of transistorized audio equipment, such as hearing aids, microphone preamplifiers, wire recorders, and the like by an appreciable factor.

NOISE COMPONENTS

TRANSISTOR noise consists predominantly of three kinds: shot noise, $1/f$ surface noise, and $1/f$ leakage noise.^{1, 2} In addition, extraneous components of noise have been observed which are attributed to poor emitter- and collector-lead connections and internal microscopic breaks and strains resulting from thermal cycling. Although the noise contributions other than that due to shot noise are often confined to a limited audio spectrum, they may be prevalent in any part of the audio spectrum at any particular instant. In low-noise junction transistors, noise from the above sources has been reduced to a low value.

TEST METHODS

For a study of noise existing in junction transistors in the common-emitter connection, two methods of noise measurement have been used, as shown in Figure 1. In the first method, a signal generator is used for wide-band and medium-band noise measurements, together with a thermal or peak-to-peak indicating device. In the second method, a

¹ W. H. Fonger, "A Determination of $1/f$ Noise Sources in Semiconductor Diodes and Triodes," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 239-295.

² L. J. Giacoletto, "The Noise Factor of Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 296-308.

noise diode is used for wide-band, medium-band, and narrow-band noise measurements, together with a thermal indicating device. The peak-to-peak indicating device measures the peak-to-peak noise voltage generated by the transistor and referred to the input terminals, which are terminated in a resistance R_g , with a reference level of one microvolt r-m-s. The noise value obtained (decibels above one microvolt) may also be described as the r-m-s power required from a signal source, terminated in a resistance R_g , to provide a signal-to-peak-noise power ratio of unity. Although much less power is required to provide a signal-to-noise power ratio of unity when a thermal indicating device

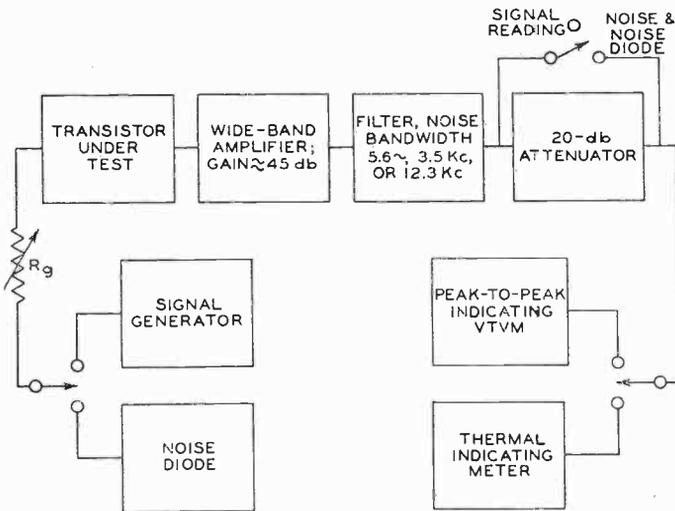


Fig. 1—Block diagram of test methods used for noise measurement.

is used, the peak-to-peak method offers some advantage in “listening” tests.³ Moreover, this method is preferred for use with “noisy” transistors because the ratio of peak-to-r-m-s value of transistor noise is not a constant for these units, but varies considerably from unit to unit. Improved low-noise junction transistors have almost a constant peak-to-r-m-s value of noise so that in the evaluation of these units, a noise-factor measurement based on noise output power has equal significance with a measurement using peak-to-peak noise voltage. Unless otherwise specified, noise values given in this article are measured with a peak-to-peak indicating device and are expressed in decibels above one microvolt.

³C. M. Burrill, “Progress In The Development of Instruments For Measuring Radio Noise,” *Proc. I.R.E.*, Vol. 29, p. 433, August, 1941.

NOISE DATA

1. *Oscilloscope Presentations:* Figure 2 shows three oscillographic traces representing noise. The trace at the upper left shows "white" noise generated by a noise diode. The trace at the lower left represents a low-noise junction transistor having a low peak-to-r-m-s value of noise. The similarity between the type of noise in a low-noise transistor and "white" noise is clearly evident. The trace at the right illustrates noise for a high-noise transistor having a high peak-to-r-m-s value of noise.

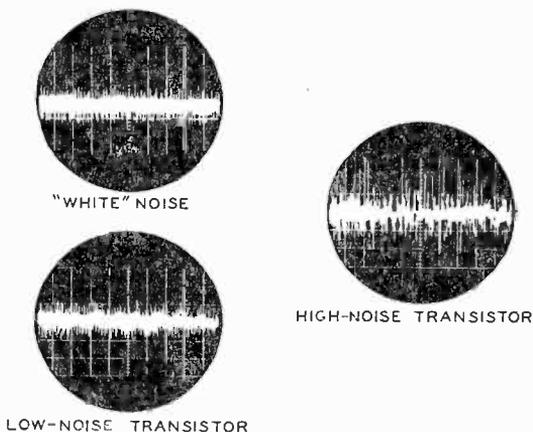


Fig. 2—Oscillographic traces showing "white" noise and noise observed in two transistors.

2. *Noise Distribution:* Bar graphs showing the distribution of noise-level values for older "noisy" junction transistors and improved low-noise junction transistors are shown in Figure 3. It is evident that the sources of extraneous noise mentioned previously have been greatly reduced in the low-noise junction units. In addition, low-noise junction transistors have a normal distribution of noise as compared to the abnormal rectangular distribution for the older "noisy" designs.

3. *Noise Factor versus Frequency:* The noise factor of a transistor as a function of frequency is shown in Figure 4 for two "noisy" junction transistors (A and B) and three low-noise junction transistors (C, D, and E). For all units, the noise factor follows a $1/f^n$ relationship in the lower end of the audio spectrum.^{4, 5} The curves shown in

⁴ R. L. Wallace and W. J. Pietenpol, "Some Circuit Properties and Applications of N-P-N Transistors," *Bell Sys. Tech. Jour.*, Vol. 30, p. 530, July, 1951.

⁵ P. L. Bargellini and M. B. Herscher, "Investigation of Noise In Audio-Frequency Amplifiers Using Junction Transistors," *Proc. I.R.E.*, Vol. 43, p. 217, February, 1955.

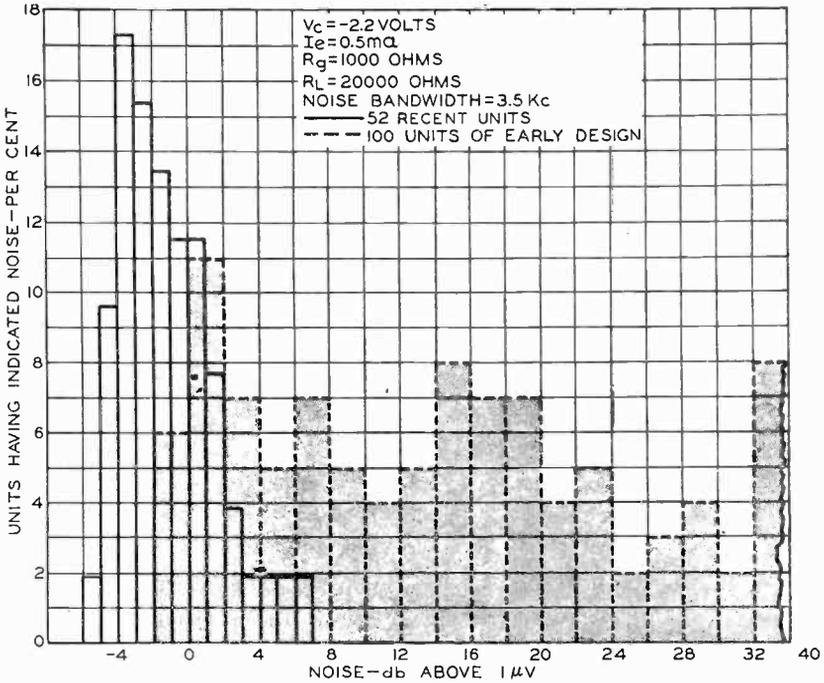


Fig. 3—Description of noise for low-noise junction transistors as compared with that for “noisy” designs (shaded area).

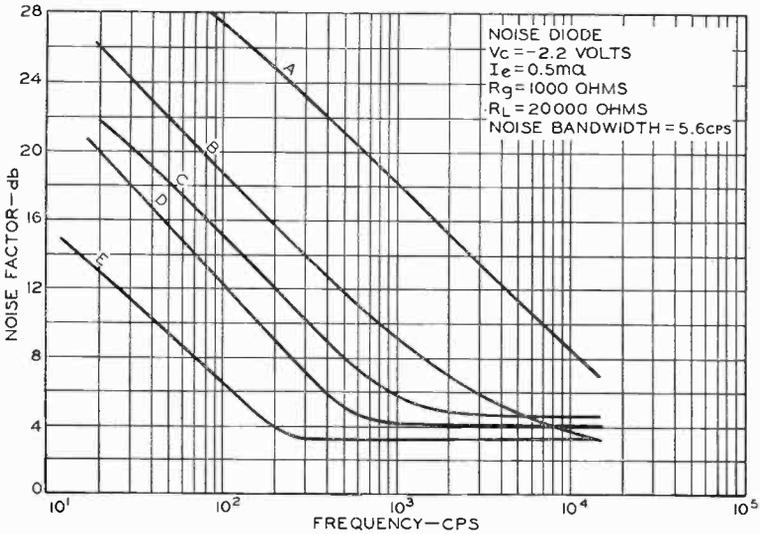


Fig. 4—Noise factor as a function of frequency for two “noisy” designs (A and B) and three low-noise units (C, D, and E).

Figure 4 for transistors C, D, and E indicate that the frequency range for $1/f$ noise has been reduced considerably on recent junction transistors. The possibility of the reduction of surface noise and leakage noise through construction techniques and geometry is discussed by W. H. Fonger.¹

The flat portion of the curves shown in Figure 4 is attributed to shot noise contributed by the input and output junctions and thermal noise in the base-lead resistance. For unit A, a very "noisy" transistor, there is no indication of a flat portion up to a frequency of 15 kilocycles. For unit B, a "noisy" transistor, there is some indication of a flat portion beginning at a frequency above 10 kilocycles. For the low-noise transistors, C, D, and E, the flat portion extends over a considerable part of the audio spectrum, from approximately 200 to 1,000 cycles up to 15 kilocycles.

The absence of a flat portion in the curves for the "noisy" transistors A and B in the audio range up to 15 kilocycles can be explained by the presence of additional noise sources, as mentioned previously, which also follow a $1/f^n$ noise relationship. One of the additional noise sources may be attributed to the leakage current, I_L , of a transistor, which is defined as the difference between the cutoff collector current, COI_c , at the operating voltage and the saturation current, I_s . Figure 5 shows a curve of cutoff collector current, COI_c , as a function of collector voltage, V_c . The leakage-current factor, I_L , is omitted from the equation shown in Figure 5 because the contribution of this leakage current to the total measured current is negligible at the low collector voltage (-25 millivolts) used in the determination of saturation current. Because the cutoff collector current varies exponentially with voltage at low collector voltages, as shown in Figure 5, the saturation current, I_s , may be determined by dividing the current measured at -25 millivolts by 0.632.

In the curves of cutoff collector current, COI_c , versus collector voltage, V_c , shown in Figure 6, the effect of leakage noise sources is indicated by the difference in slope between the low-noise unit, E, and the high-noise unit, A. The correlation between leakage and noise for the transistors shown in Figure 6 and Figure 7 is as follows:

Transistor	Leakage (microamperes)	Noise (db above 1 microvolt)
A	0.95	14.8
B	0.63	5.1
C	0.39	0
D	0.36	-3.68
E	0.34	-4.0

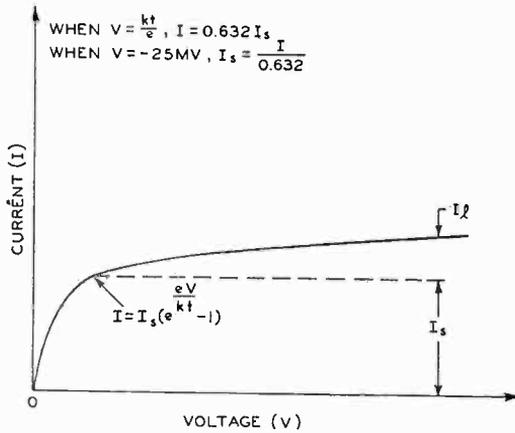


Fig. 5—Typical collector characteristic for junction transistor.

Good correlation between leakage current and noise power was also obtained in tests on a large number of units, as shown in Figure 7.

4. *Crest Factor*: The peak-to-r-m-s value of transistor noise, termed crest factor (*C*), is determined using a medium bandwidth of 3.5 kc, maximum response at 1,000 cycles. One method for determining

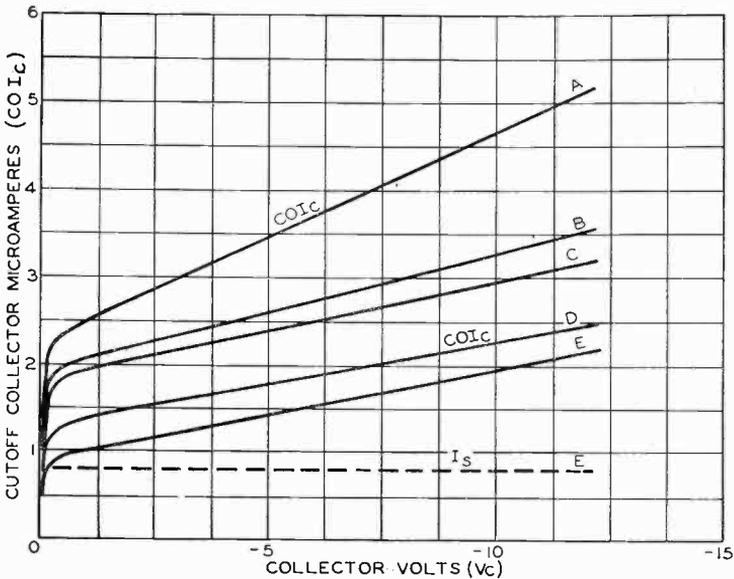


Fig. 6—Cutoff collector current (COI_c) as a function of collector voltage (V_c) for two "noisy" designs (A and B) and three low-noise units (C, D, and E).

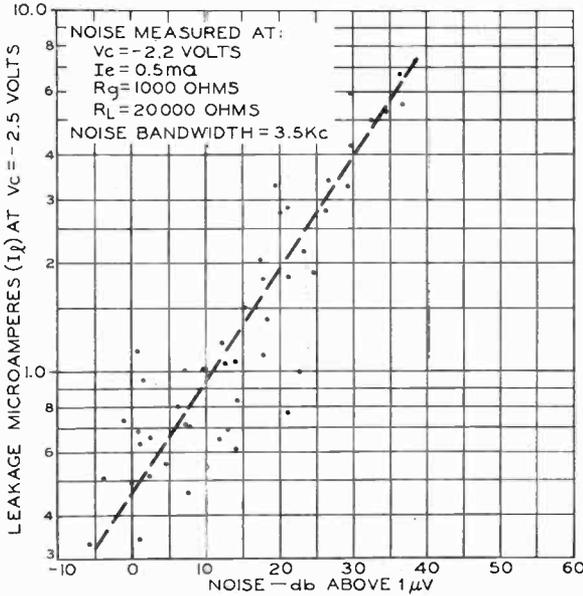


Fig. 7—Leakage current as a function of noise power in decibels above 1 microvolt.

crest factor, shown in Figure 8, utilizes a signal generator together with a peak-to-peak indicating meter,

$$\text{noise factor (NF)} = 20 \log \left[\frac{e_{\text{sig peak}}}{C} \right] \cdot \frac{1}{e_{R_D}}$$

and a noise diode together with a thermal indicating meter,

$$NF = 20 \log \sqrt{20R_g I_d}$$

The crest factor, C , is then determined by equating the two noise factors. A second method, shown in Figure 9, uses a noise diode together with a peak-to-peak meter,

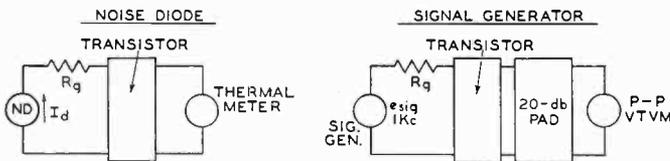


Fig. 8—Test method used for determination of crest factor.

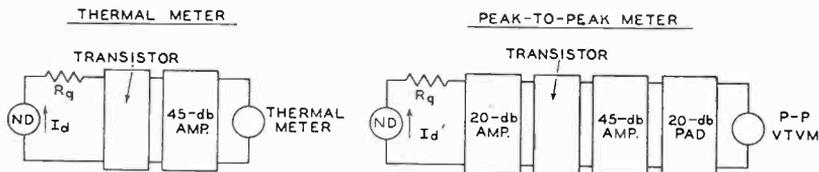


Fig. 9—Test method used for determination of crest factor.

$$NF = 20 \log \left[\frac{e_{\text{peak noise}}}{C} \right] \frac{1}{e_{Rq}}$$

and a noise diode together with a thermal meter

$$NF = 20 \log \sqrt{20R_q I_d}$$

Again, the crest factor is determined by equating the noise factors. Figure 10 shows the constant relationship between crest factor and noise in decibels above 1 microvolt for low-noise junction transistors, and the variant relationship for “noisy” transistors.

In Figure 11, noise factor is plotted as a function of noise. If the crest factor were a constant equal to 3.7, as reported by V. D. Landon,⁶ the straight line shown in Figure 11 for the low-noise transistors would coincide with the dotted straight line. However, the crest factor deter-

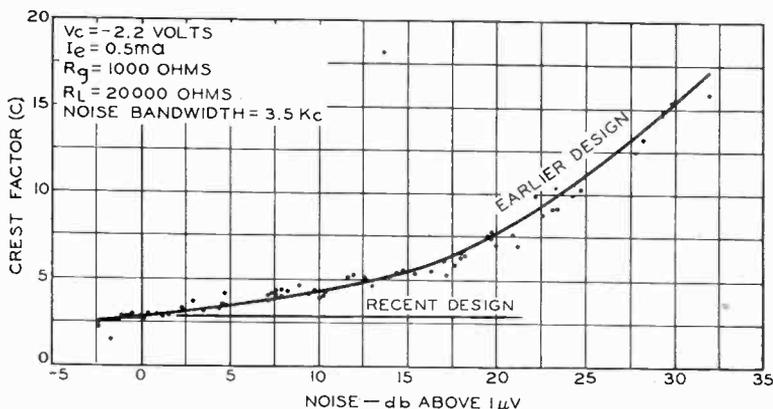


Fig. 10—Crest factor, C , as a function of noise in decibels above 1 microvolt.

⁶ V. D. Landon, “The Distribution of Amplitude With Time In Fluctuation Noise,” *Proc. I.R.E.*, Vol. 29, p. 50, February, 1941.

mined by the methods shown in Figures 8 and 9 is lower than 3.7 because of the integration constant of the peak-to-peak meter. Because the crest factor is not a linear function of noise factor for transistors having high noise, a peak-to-peak noise measurement is preferable for these units, and is more consistent with "listening" tests than a thermal measurement. Crest-factor values for the five test transistors are as follow:

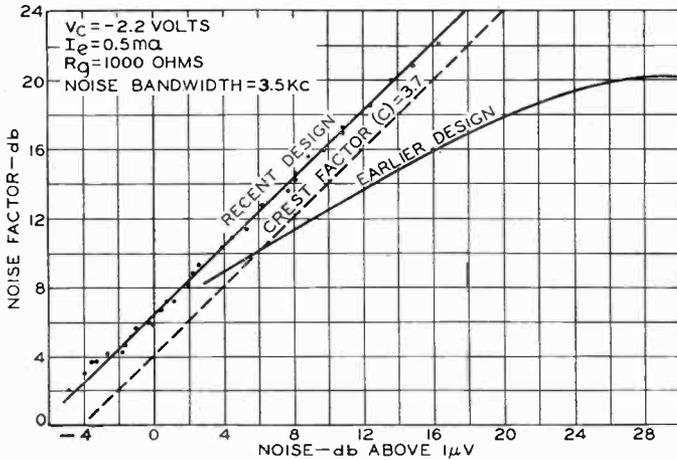


Fig. 11—Noise factor in decibels as a function of noise in decibels above 1 microvolt.

Transistor	Crest Factor
A	4.05
B	3.6
C	2.84
D	2.73
E	2.8

The crest factor for a number of low-noise transistors was found to be almost constant, with an average value of 2.81. For these units, therefore, either a thermal or peak-to-peak indicating device may be used to indicate the noise. Tests were also made on low-noise transistors over a wide noise bandwidth (12.3 kilocycles per second) with a source resistance of 518 ohms. Average noise factors for units E, D, and C were 3.7, 4.5, and 6.5 decibels, respectively. These low values of noise factor are comparable to those of the best low-noise tubes,

and make the use of low-noise transistors very attractive for high-quality, low-noise audio applications.

5. *High-Frequency Noise*: To be complete, a description of the noise in audio junction transistors should include noise above the audio spectrum. The noise factor has been reported to increase with frequency above the audio range. This increase is attributed to a decrease in power gain with no appreciable reduction of the major contribution of noise due to the collector or output circuit.

6. *Noise versus Time*: A phenomenon which has not yet been fully explained, the increase of noise with operating time, has been observed in many "noisy" junction transistors and in a very small percentage of low-noise units. Oscillographic traces indicate that when transistor noise measured peak-to-peak increases with time, the peak-to-r-m-s value of the noise (crest factor) also increases, the average value of the noise remaining approximately constant. It has also been observed that the greatest change in peak-to-peak noise level occurs within the first hour of operation, after which the level remains stable. If, however, the circuit is momentarily interrupted, the noise returns to the value measured initially. Microscopic examinations of cross sections of alloyed junctions in transistors which gave an increase in noise level with time showed improper wetting of the connector wires to the emitter and collector dots. In addition, some of the units examined did not appear to have proper alloying or uniform wetting of the dots to the germanium wafer. It was also observed that units having a high initial leakage current were more likely to display an increase in noise with operating time, and that the leakage current in such transistors also increases. With continued application of circuit voltages, the increasing leakage current could account for the change in measured noise with operating time. The change in leakage, due to possible contamination of the germanium surface and the accompanying shunting resistance, is probably a large contributing factor in the noise-vs.-time phenomenon.

CIRCUIT CONSIDERATIONS

When junction transistors are used in typical low-level audio applications, in which low noise is the prime consideration, certain circuit requirements must be considered. First, the source resistance must be chosen for minimum noise factor. The curves of noise factor as a function of generator impedance shown in Figure 12 indicate that the minimum noise factor is obtained when a source resistance of 400 to 600 ohms is used at the conditions specified. This value of optimum

source resistance for lowest noise is approximately one-half of the value required for input match.

For the "noisy" junction transistors, operation at as low a collector voltage as possible was also required. The collector voltage for these transistors should be no greater than 4 volts, as shown by the curve A in Figure 13. Because the low-noise transistors are much less dependent upon collector voltage (see curve E), these units may be operated at collector voltages up to approximately 20 volts without a consequent increase in noise factor.

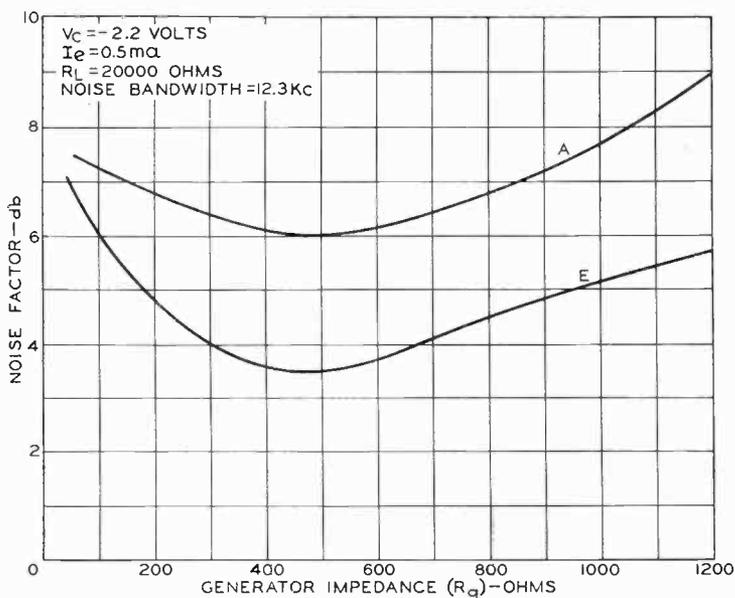


Fig. 12—Noise factor as a function of source resistance for "noisy" transistor (A) and a low-noise unit (E).

For low-noise applications, the emitter current should also be maintained at as low a value as permissible, limited only by the signal-handling requirements specified for the intended application. The dependence of noise factor on emitter current is illustrated by the curves shown in Figure 14. The value of emitter current (0.2 milliampere) at which the noise factor is a minimum agrees with the calculated and measured results reported by W. H. Fonger.¹ The curves also follow the same general shape, except in the case of unit E, where $1/f$ leakage noise predominates.

ENVIRONMENTAL CONSIDERATIONS

The $1/f^n$ surface noise is insensitive to temperature. However, the

components due to thermal noise generated by the base-lead resistance, shot-noise sources, and $1/f$ leakage-noise sources show an increase with temperature.

The leakage current of most plastic-encapsulated transistors also tends to increase with life, causing a consequent increase in noise. This increased leakage indicates that these plastics are not impervious to contaminating agents. The necessity for improved sealing methods was pointed up by experience on plastic-encapsulated units. The hermetic seal used on low-noise transistors provides complete protection against the penetration of contaminants, thus contributing to the

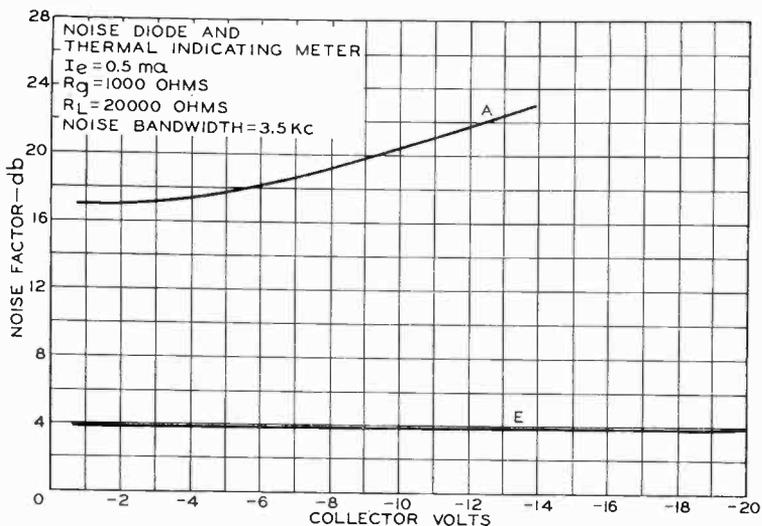


Fig. 13—Noise factor in decibels as a function of collector voltage for a “noisy” transistor (A) and a low-noise unit (E).

reliability and continued low noise of this transistor despite wide variations in environment.

CONCLUSION

The particular method of noise measurement preferred for a junction transistor is determined on the basis of experience obtained from evaluation of a large number of units by the various methods described above. If the noise distribution is widespread, a special selection test should be performed with a peak-to-peak indicating device to obtain units for low-noise applications consistent with “listening” tests. A wide-band noise-factor measurement using a noise diode is the preferred method for noise evaluation of low-noise junction transistors.

One of the inherent advantages of transistors for low-level audio applications is their low input impedance (400 to 600 ohms) in the common-emitter connection. Because of this low impedance, transistors have an advantage over vacuum tubes when low-impedance, low-level devices such as magnetic microphones and pickups are to be used. The use of a transistor eliminates the necessity for the critical and expensive input transformer normally used with tubes to raise the signal-source impedance level. Inherent lack of appreciable microphonism,

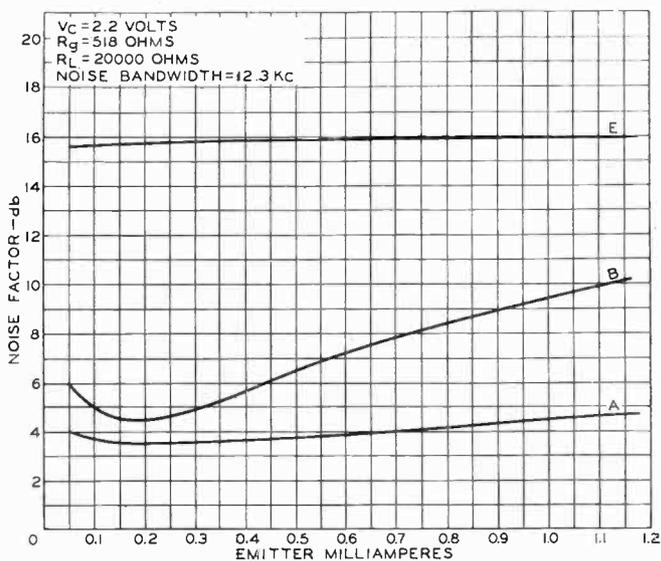


Fig. 14—Noise factor as a function of emitter current for two “noisy” transistors (A and B) and a low-noise unit (E).

absence of heater-supply hum, and comparable or lower noise level, as compared with vacuum tubes, are other advantages which make the junction transistor desirable for low-noise audio applications.

As a result of the increasing knowledge of factors contributing to noise in p-n-p junction transistors, and the possibility of noise reduction by means of careful construction techniques, reliable, low-noise, junction transistors can now be used to extend the useful sensitivity range of transistorized audio equipment.

A TEST SET FOR TRANSISTOR PERFORMANCE MEASUREMENT AT 455 KILOCYCLES

BY

DAVID D. HOLMES, LARRY A. FREEDMAN, AND THOMAS M. SCOTT

RCA Laboratories,
Princeton, N. J.

Summary—This paper describes a laboratory test set for the measurement of transistor performance at 455 kilocycles. A common-emitter neutralized test circuit is employed, conjugate matching being provided at input and output. The test set affords measurement of power gain, input and output resistance, neutralizing resistance and capacitance, and noise factor.

The general principles of measurement are discussed, and the electrical and mechanical details of the test set and certain auxiliary equipment are described.

METHOD OF GAIN MEASUREMENT

THE method of transistor power-gain measurement utilized in the apparatus described in this paper is shown in Figure 1. The transistor under test is represented as a four-pole network

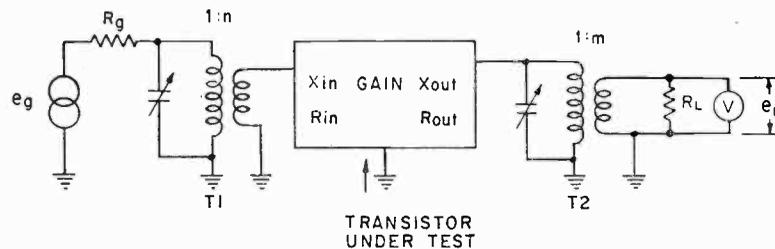


Fig. 1—Simplified circuit illustrating method of power-gain measurement.

having input resistance and reactance, gain, and output resistance and reactance. (Feedback effects are discussed in a subsequent section.) A test signal is supplied to the transistor from a signal generator having an open-circuit voltage, e_g , and an internal resistance, R_g , via a tuned transformer, T1, having a turns ratio 1-to- n . The input circuit is tuned to the frequency of the test signal. If the transformer turns ratio is now adjusted so that $n^2 R_g = R_{in}$, then maximum available input power, $e_g^2/4R_g$, will be delivered to the transistor.

Similarly, for the output circuit, the transformer, T2, is tuned to the test-signal frequency, and its turns ratio 1-to- m , is adjusted so

that $m^2 R_{\text{out}} = R_L$. The adjustments of tuning and matching of the input and output circuits for maximum power transfer may be made with reference to the indication of a single voltmeter connected across the load resistor, maximum voltmeter deflection indicating optimum adjustment of the various controls.

The generator and load resistances may be chosen to facilitate the calibration of the voltmeter so that it provides a direct indication of power gain. For the matched condition, the power input to the transistor is $e_g^2/4R_g$, while the power into the load resistor is e_L^2/R_L . If R_L is made equal to $4R_g$, then the power into the load will be equal to $e_L^2/4R_g$. Thus, for a transistor having unity gain, the generator and load voltages are equal; if a voltmeter, calibrated in decibels, is placed across the signal generator and the signal generator output is adjusted for an indication of 0.0 decibels, the voltmeter, when placed across the load resistor, indicates power gain directly.

The above discussion applies for the case in which transformers having no losses are employed. An equivalent circuit for a generator coupled to a load by means of a tuned transformer is shown in Figure 2. The elements R_g , R_T , and R_L represent the generator resistance, the transformer tuned resistance, and the load resistance respectively. Maximum load power is obtained when R_L is equal to the parallel resistance of R_g and R_T . This power is less than the maximum available power by an amount depending upon the relative magnitudes of R_T and R_g . For the output transformer used in this test set, the power loss is less than 1.5 decibels. The input transformer loss is negligible, being less than 0.2 decibel.

FEEDBACK AND NEUTRALIZATION

The apparatus utilizes a neutralized (or unilateralized) common-emitter test circuit. The operation of the neutralizing circuit may be visualized by referring to Figure 3. Figure 3a shows a single-generator common-emitter π -equivalent circuit of the transistor. Feedback from the output to point b' occurs via the parallel combination of $C_{b'e}$ and $r_{b'e}$. This feedback may be neutralized by deriving a reverse-phase voltage from the output which is then fed back through a parallel RC circuit, R_f and C_f , to the base connection, point b , as shown in Figure 3b. The presence of $r_{bb'}$ renders the neutralization dependent upon frequency, and influences the requisite parameters in the neutralizing circuit. The extent to which $r_{bb'}$ may render the neutralization frequency dependent is a function of the relative magnitudes of $r_{bb'}$ and $z_{b'e}$. The consequences of imperfect neutralization depend upon the transistor voltage gain, and the driving impedance, and must be

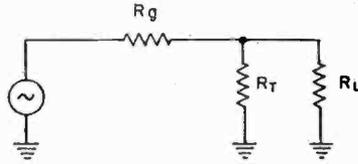
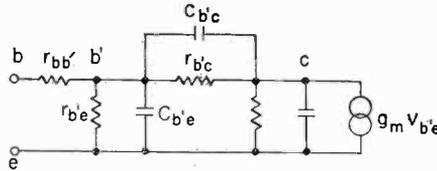


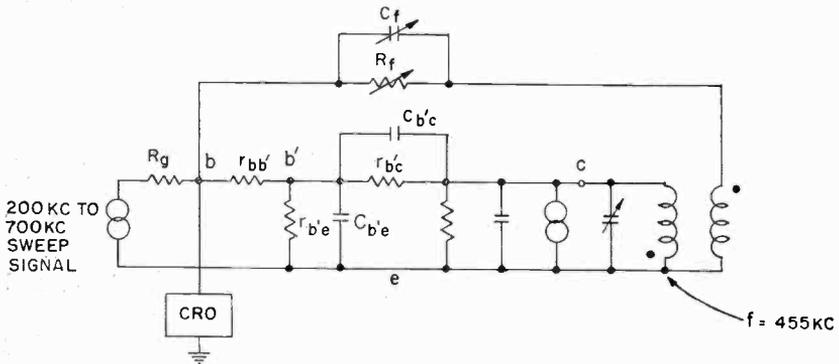
Fig. 2—Equivalent circuit of a generator coupled to a load by means of a tuned transformer.

evaluated for a given type of transistor. For the type of r-f transistor described by Mueller and Pankove¹ these are second order at 455 kilocycles, and can usually be neglected.

Figure 3b illustrates a method for determining whether or not the circuit is properly neutralized. A 200 to 700 kilocycle sweep signal is fed to the base of the transistor through a generator resistance, R_g .



(a) Single generator common-emitter π -equivalent circuit of an alloy-junction transistor.



(b) Transistor equivalent circuit with illustration of a method employed for neutralization adjustment.

Fig. 3—Transistor feedback and neutralization.

¹ C. W. Mueller and J. I. Pankove, "A p-n-p Triode Alloy Junction Transistor for Radio-Frequency Amplification," *RCA Review*, Vol. XIV, pp. 586-598, December, 1953.

An oscilloscope, connected between base and ground, is employed to provide an indication of the manner in which the transistor input impedance varies as a function of frequency. Several representative oscillograms are shown in Figure 4. In Figure 4a, the unneutralized case, the input impedance is found to increase at frequencies somewhat below collector circuit resonance, indicating positive feedback. Approximately at resonance and above, the input impedance decreases, indicating negative feedback.

Figure 4b shows the input impedance as a function of frequency for the neutralized case. This is substantially the same presentation which one would obtain if the output were short-circuited.

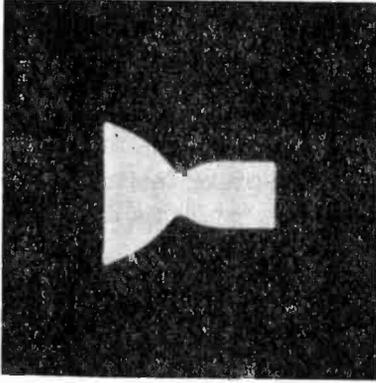
Figure 4c illustrates the over-neutralized case. Here, positive feedback occurs at frequencies slightly above the resonant frequency of the collector circuit because of an increase in capacitive feedback over that required for neutralization.

In order to obtain correct adjustment of C_f and R_f , small-signal operation of the transistor must be maintained. Nonlinearity of the transistor input impedance due to excessive signal input is illustrated in Figure 4d; for this case the signal input was increased several times over that used in Figure 4b. (A sweep signal of about one millivolt peak-to-peak at the base is usually small enough to prevent this type of nonlinearity.)

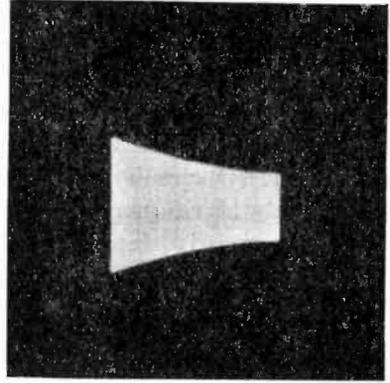
CIRCUIT DESCRIPTION AND GAIN-MEASUREMENT TECHNIQUE

A schematic diagram of the test set is shown in Figure 5. Two tapped 4.5-volt "C" batteries comprise the power supply. Suitable bias polarities for either p-n-p or n-p-n transistors may be selected by means of switch S5. The collector voltage is adjustable from 0 to 7.5 volts, in 1.5-volt steps, by means of switch S6. A constant-emitter-current biasing arrangement, variable over the range from 0.5 to 2.5 milli-amperes (for germanium transistors), is utilized, and consists of R3 and R4 in combination with one 1.5-volt cell of the battery. This bias supply is bypassed to ground for signal frequency by C11; the base circuit is returned to chassis ground.

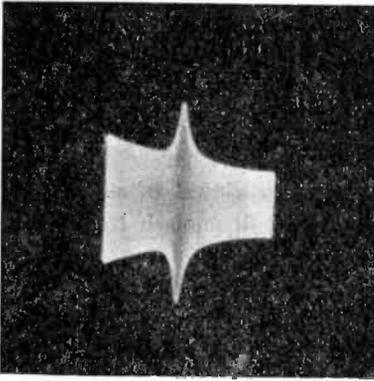
Input and output transformers, designated T1 and T2 respectively, are wound on toroidal ferrite cores. Winding and electrical data are given on the schematic. Taps 1-10 of T1 permit matching the generator resistance to a range of transistor input resistances from 150 to 2000 ohms, while a like number of taps on T2 affords matching of the load to transistor output resistances ranging from 4000 to 100,000 ohms. (The generator and load are described in a subsequent paragraph.)



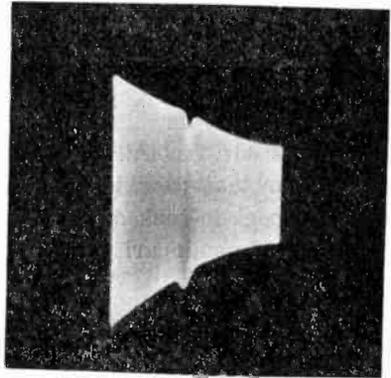
(a) Unneutralized.



(b) Neutralized.



(c) Over neutralized.



(d) Excessive input signal when neutralized.

Fig. 4—Oscillograms of transistor input impedance as a function of frequency.

The output transformer is provided with a separate secondary winding which drives a neutralizing circuit made up of R6 and C14. The neutralizing circuit is calibrated with respect to the collector tap of the output transformer (there is a 10.6 to 1 turns ratio between the collector tap and neutralizing winding), C14 reading 1-27 micro-microfarads, and R6 0-1 megohm. Capacitor C13, shunting the neutralizing winding, serves to suppress parasitic oscillations at frequencies in the megacycle range, and does not significantly affect circuit operation at 455 kilocycles.

Tuning of the input transformer is accomplished by means of S2, which is employed to select an appropriate value from the bank of fixed

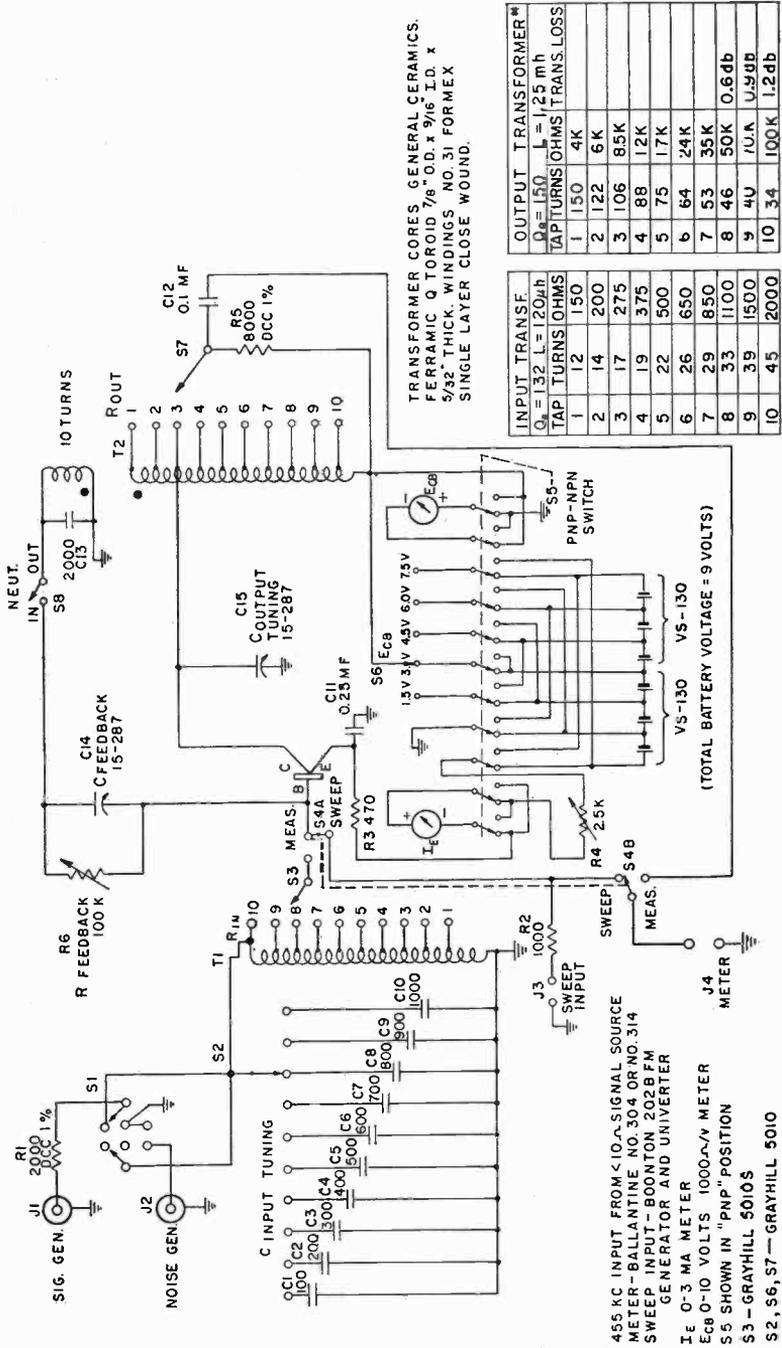


Fig. 5—Schematic diagram of test set.

capacitors C1 to C10. Output circuit tuning is provided by variable capacitor C15.

In making gain measurements, a 455-kilocycle signal source having an internal impedance of less than ten ohms (for example, a Measurements Corporation Standard Signal Generator Model 65-B) is connected to J1. The 2000-ohm resistor, R1, serves as the generator resistance, R_g . Resistor R5 is utilized as the output circuit load; this resistor is assigned the value of $4R_g$ (8000 ohms) to facilitate calibration of the test set, as mentioned above. A Ballantine type 304 or 314 vacuum-tube voltmeter is employed as an output indicator and is connected to J4. (A voltmeter with an input resistance greater than 0.1 megohm and response to 1 megacycle is required.)

For neutralizing purposes, a 200-700 kilocycle sweep signal, derived from a Boonton 202-B Generator-and-Univerter combination, is applied at J3. (Any sweeper which covers approximately this frequency range and has at least 50 millivolts output may be used.) Switch S4, when placed in the "Sweep" position, feeds the sweep signal to the transistor base via R2, and simultaneously connects the Ballantine voltmeter to the base. The voltmeter is then employed as an amplifier, the output of which feeds the vertical deflection input of an oscilloscope. (Any oscilloscope having relatively uniform response up to 1 megacycle and a vertical deflection sensitivity of about 0.1 volt peak-to-peak per inch is satisfactory.) A high-pass RC filter is employed between the voltmeter-amplifier output and the oscilloscope to remove residual hum in the amplifier output. The "Neut" switch, S8, is then switched "In", and R6 and C14 are adjusted to provide an oscilloscope pattern similar to that of Figure 4b (the neutralized condition).

Switch S4 is then moved to the "Measure" position, which connects the transistor base to the input transformer tap-selecting switch, S3, and the voltmeter to R5 (via C12) where it is employed to indicate output voltage. Initial calibration to provide a direct reading of gain is accomplished by placing the meter across the 455-kilocycle signal generator output and setting the output level at 1 millivolt (0 decibels on the 0.01-volt voltmeter scale). The meter, when returned to J4, will read 0-20 decibels gain on the 0.01-volt scale, 20-40 decibels gain on the 0.1-volt scale, etc. Switches S2, S3, and S7, and capacitor C15 are adjusted to provide maximum meter deflection (conjugate matching at input and output). Any loss introduced by the output transformer (which depends upon the tap at which matching occurs and is listed on the schematic) must be added to the meter reading.

METHOD OF NOISE MEASUREMENT

Transistor noise factor is measured here by the noise diode method. The noise factor is given by

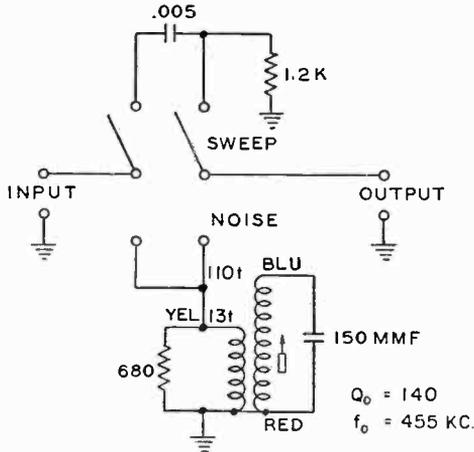
$$F = \frac{20I_d}{M - 1} R_g \quad \text{at } 290^\circ\text{K},$$

where I_d is the plate current of the diode, R_g is the generator resistance, and M is the ratio of noise power output with the diode turned on to the noise power output (transistor noise plus thermal noise) with the diode turned off. For convenience, M is chosen to be 10 decibels. The above formula was used to calibrate the diode current meter directly in decibels noise factor. The calibration is valid for any condition of input or output circuit mismatch.

Noise Generator Circuit and Noise Measurement Technique

A schematic diagram of the noise generator is shown in Figure 6. The type 5722 noise diode serves as the noise source. The diode filament current (and hence the plate current) is varied by means of a 6-ohm rheostat in the filament circuit. Separate ground return leads are employed for the filament and plate circuits so that filament switching transients are confined to the filament circuit and are not transmitted to the transistor. The diode plate is connected to the plate-supply battery via a 2000-ohm load resistor and a milliammeter. The milliammeter is provided with two shunts, which may be selected by means of switch S2. When S2 is placed in the "Direct" position, the meter reads directly 6.5 decibels noise factor full scale, corresponding to a maximum of 1.0 milliamperes diode plate current. When S2 is switched to the "+5 db" position, the meter range is extended to 3.17 milliamperes, and 5 decibels must be added to the meter-scale noise factor reading. Similarly, if S2 is placed in the "+10 db" position, the meter range is extended to 10 milliamperes, and 10 decibels must be added to the meter-scale noise factor. Thus, measurement of noise factors in the 0-16.5 decibels range is accommodated.

Before measuring noise factor, a transistor to be tested should be properly biased, and adjustment of neutralization and input- and output-circuit tuning should be made. (For this reason, it is convenient to make a noise measurement following a measurement of gain.) Switch S1 is then moved to the "Noise Generator" position. This substitutes the noise diode for the 455-kilocycle signal source. Each of these signal sources has an effective internal resistance of 2000 ohms so that switching from one to the other leaves input circuit



TRANSFORMER: # 36 SSE ON 0.125 I.D. x 0.145 O.D. FIBRE FORM RANDOM WOUND ON 0.230 WIDE x 0.125 D. ARBOR. SECONDARY WOUND UNDER PRIMARY — DISTRIBUTED OVER LENGTH OF COIL. FINISHED COIL CUMAR DOPED, FERRITE TRANSFORMER CORE AND MECHANICAL ASSEMBLY ARE EXPERIMENTAL.

Fig. 7—Schematic diagram of test set filters.

PHYSICAL DETAILS

Test Set

A front view of the test set is shown in Figure 8. An aluminum chassis is employed, descriptive lettering and calibration information

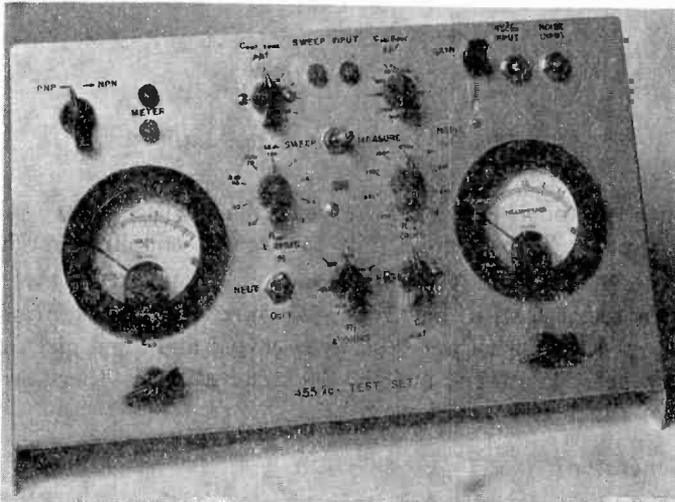


Fig. 8—Front view of test set.

being silk-screened directly on the front panel. Note that the output transformer losses (0.6, 0.9, and 1.2 decibels, corresponding to the 50,000, 70,000 and 100,000 ohms positions of the " R_{out} " switch) are indicated on the front panel. The output transformer losses corresponding to the other tap positions are less than 0.5 decibel and are neglected.

A rear view of the test set appears in Figure 9. The toroidal input and output transformers are mounted directly over their respective tap switches, and are supported by their leads. An aluminum shield, which bridges the transistor socket, serves to isolate input and output circuits. Pin 4 of the transistor socket (a vacant pin between collector and base pins) is grounded to provide further isolation.

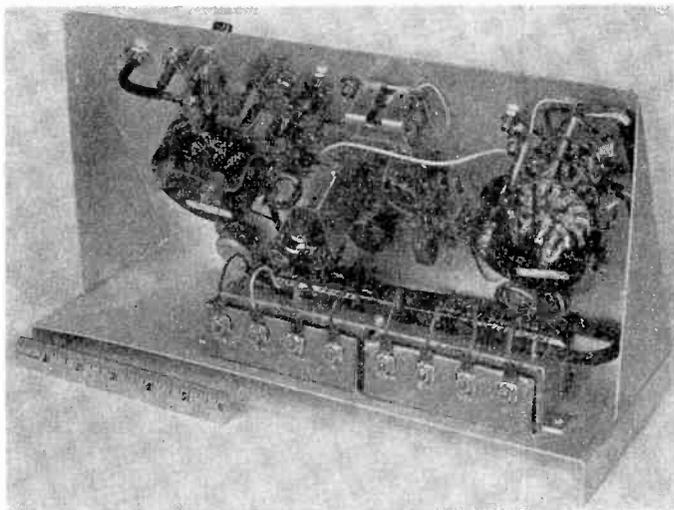


Fig. 9—Rear view of test set.

A shielded lead is employed between the "Noise Input" jack, J2, and the input selector switch, S1, to prevent coupling of the 455-kilocycle signal into the noise input.

Leads associated with the "Sweep-Measure" switch, S4, are dressed so that equal stray capacitances between output and input circuits exist for either position of the switch. Neutralization is therefore independent of the switch position.

Noise Generator

A view of the completed noise generator is shown in Figure 10; Figure 11 shows the unit with both front panel and cover open.

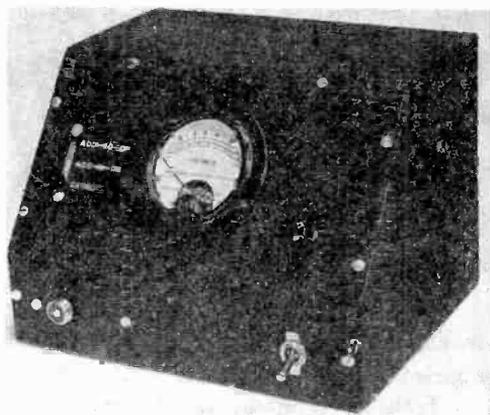


Fig. 10—Noise generator, front view.

455-Kilocycle Signal Source

As mentioned above, a Measurements Corporation Standard Signal Generator, Model 65-B, is suitable for use as a 455-kilocycle signal source. An alternative generator consists of the simple battery-operated transistor oscillator shown in Figure 12. A schematic diagram

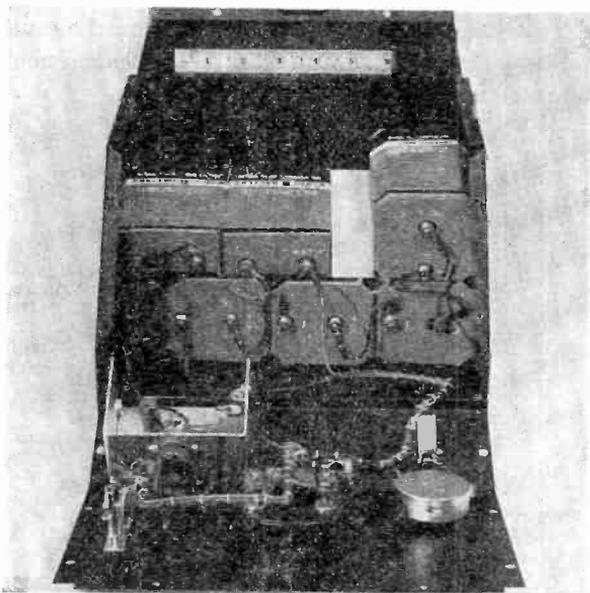


Fig. 11—Noise generator, front view with both front panel and cover open.

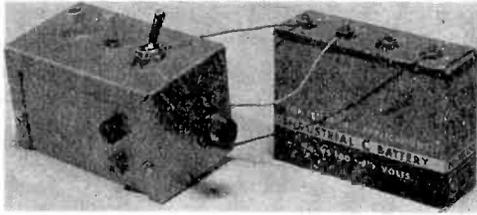


Fig. 12—Transistor signal generator.

of the oscillator appears in Figure 13. The oscillator transformer embodies an experimental arrangement and is ferrite cored. It is approximately $\frac{3}{8}$ inch in diameter by $\frac{1}{2}$ inch high. Winding and electrical data are given on the schematic. A variety of transistors may be used interchangeably in the oscillator circuit; these include the type described by Mueller and Pankove,¹ the Raytheon CK-760 series, and the Philco L5106, using the battery polarity shown on the schematic, or Sylvania 2N94A, Germanium Products RD2517A and the Texas Instruments 225 series, using reversed battery polarity.

Hum and Bandwidth-Limiting Filters

The RC filter, employed to attenuate residual hum at the output of the Ballantine vacuum-tube voltmeter amplifier in making neutralizing adjustments, and the bandwidth-limiting tuned circuit utilized at the voltmeter amplifier output in making noise measurements are contained in the small chassis shown in Figure 14.

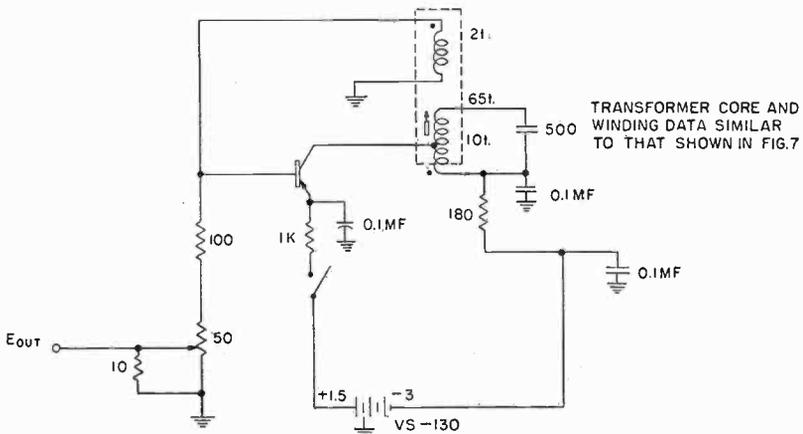


Fig. 13—Schematic of transistor signal generator.

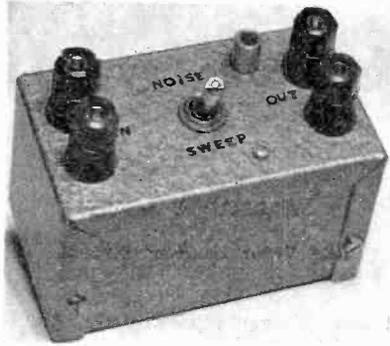


Fig. 14—Filter unit.

Complete Setup

The complete measuring setup is shown in Figure 15. The equipment, from left to right, includes the oscilloscope, the first Ballantine meter and associated output filters, the Boonton FM Generator and Univerter, the second Ballantine meter (atop the FM generator), the test set proper, the noise generator, and the transistor 455-kc oscillator.

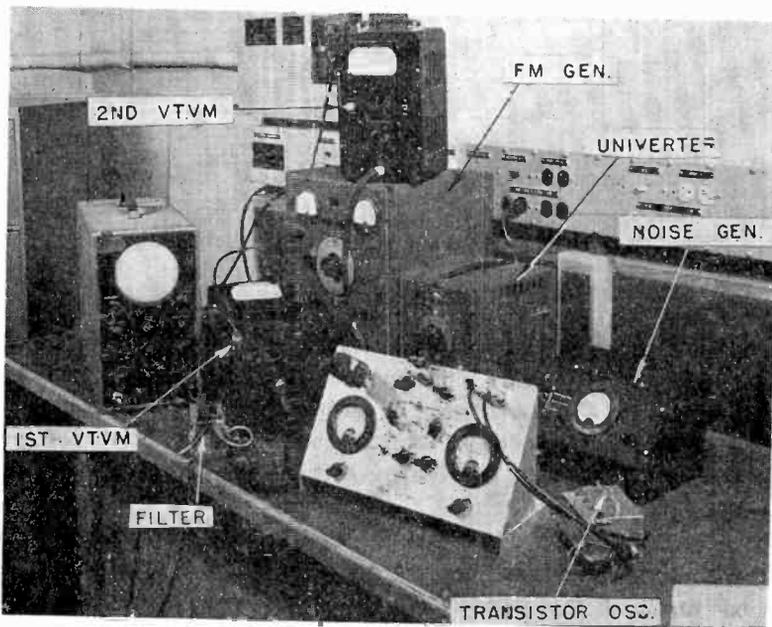


Fig. 15—Complete measuring setup.

A HIGH-FREQUENCY MEASURING EQUIPMENT FOR TRANSISTORS

BY

WILLIAM F. SANDS

RCA Victor Television Division,
Camden, New Jersey

Summary—A transistor-measuring equipment has been developed which will measure power gain, terminal input and output resistance and capacitance values, and feedback parameters over the frequency range from 400 kilocycles to at least 50 megacycles. The power-gain values are estimated to be accurate to within ± 1 decibel. Biasing and meter facilities as well as socket connection are provided for any known high-frequency transistor type. In addition, either base #1, base #2, or the emitter may be utilized as the input signal electrode. The investigation provided knowledge whereby the equipment may be further improved when a supply of special high-frequency resistors and a miniature turret assembly become available.

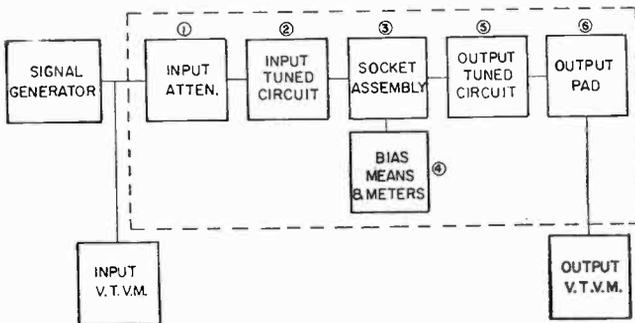


Fig. 1—Block diagram of measuring equipment.

INTRODUCTION

THE measuring instrument described in this paper is capable of determining effective input and output resistance and capacitance values and the power gain of all presently available transistors over the frequency range of 0.4 to 50 megacycles. In addition, the feedback parameters may be secured by suitable neutralization and unilateralization means.

The block diagram of Figure 1 presents the basic units of the instrument and the required auxiliary equipment. The input attenuator (1) has an input impedance of 50 ohms, a constant attenuation of 20 decibels (when matched), and an output impedance range of 10 to 1000 ohms. The input tuned circuit (2) consists of a 100-micromicrofarad

variable capacitor and various plug-in toroidal coil assemblies which provide a multiplicity of test frequencies in the 0.4 to 50 megacycle range.

A group of socket assemblies (3) have been provided whereby various input electrode connections, transistor terminal configurations, neutralization, and biasing means may be effected. The in-line sockets have been mounted in octal tube bases which in turn are inserted in the octal socket mounted in the instrument. The d-c portion of the measuring equipment (4) consists, in the main, of the biasing means (i.e., batteries, switches, and variable resistor controls), the collector voltmeter, and the current meters for base #1, base #2, and the collector.

The output tuned circuit (5) consists of a 50 micromicrofarad variable capacitor and a group of toroidal coil assemblies mounted in 5-pin plug-in assemblies. In addition, feedback windings have been provided on the lower frequency coils whereby neutralization measurements may be conducted. The output pad (6) consists of a simple shunt array of output load resistors mounted on a switch structure. The details of units (1) to (6) inclusive are described fully in the body of the report which follows.

INPUT RESISTIVE ATTENUATOR

The desired matching network to be used between a low-impedance (50 ohms) signal generator with unbalanced resistive termination and the wide range of transistor input-terminal resistive values may be in the form of a multistep array of "T" or " π " resistive pads. It is noted that the effect of the input terminal reactance value will be obviated by the use of a tuned input circuit and that many r-f signal generators are designed for a resistive termination of the order of 50 ohms, (R_1). The desired matching resistance (R_2) range was arbitrarily considered to be 10 to 1000 ohms in 11 steps with substantially uniform logarithmic intervals, i.e., 10, 16, 25, 40, 63, 100, 160, 250, 400, 630, and 1000 ohms. The available total range for R_2 using 20-decibel attenuation pads is more than sufficient for the arbitrarily selected 10 to 1000 ohm limitation in the input matching for available transistor types. The use of 20-decibel units is particularly desirable when used in conjunction with an electronic log voltmeter which has 20-decibel scale calibrations. After preliminary measurements with T structures, it was found that they possess three disadvantages which may be avoided by use of a π pad array, namely:

1. The path to "ground" from either input or output terminal is through two resistors and will, therefore, always exhibit a higher inductive reactance than in the case of a single shunt resistor.

2. The T configuration does not permit a simple in-line assembly of the resistors if mounted on strips of a turret type mechanism or on the wafers of a rotary selector switch.

3. If the input and output branches of the pads are encased in brass block assemblies so as to partially cancel the inductive reactance, the resulting capacitance at the midpoint of the T pad will not be effectively tuned out by the input circuit. Therefore, the voltage division of the higher resistance steps at very high frequencies may not be correct.

The following equations were used to calculate the values for the resistors of the π type steps, Figure 2, of an improved attenuator.

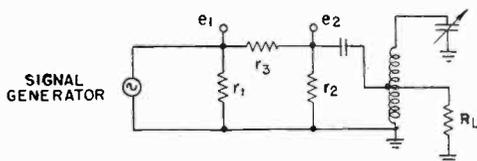


Fig. 2—Tuned setup for input π pad.

$$\frac{1}{r_1} = \frac{1}{R_1} \left(\frac{N+1}{N-1} \right) - \frac{1}{r_3}, \quad (1)$$

$$\frac{1}{r_2} = \frac{1}{R_2} \left(\frac{N+1}{N-1} \right) - \frac{1}{r_3}, \quad (2)$$

$$r_3 = \frac{N-1}{2} \sqrt{\frac{R_1 R_2}{N}}, \quad (3)$$

where N is the ratio of the power output of the pad to the power input.

The calculated values for the 11 π pad steps are given in Table 1. Great care was taken to obtain final resistor elements which would be within 1 per cent of the tabulated values after soldering.

The r_1 and r_2 resistors were mounted in silver-plated brass rings which were separated from their associated switch wafers by thin Teflon* sheets to partially cancel out the inductance of the switch assembly.

* Trademark of the E. I. duPont Co., Wilmington, Del.

Table 1— π Pads ($R_1 = 50$ ohms, attenuation = 20 decibels)

Step	R_2 (ohms)	r_1 (ohms)	r_2 (ohms)	r_3 (ohms)
1	10	88.1	10.8	111
2	16	75.4	17.7	140
3	25	68.1	28.5	175
4	40	63.1	47.8	221
5	63	59.5	79.4	278
6	100	57.0	136	350
7	160	55.1	243	443
8	250	53.8	440	553
9	400	52.7	890	700
10	630	51.9	2070	878
11	1000	51.3	8560	1100

Effect of Soldering on the Pad Elements

An approximate soldering test was made upon a series of small composition type and metallized glass resistors in accordance with the following procedure:

T-1. The units were initially measured by use of a Wheatstone bridge.

T-2. The leads at a point about one-half inch from the body were soldered. The units were permitted to cool and then measured on the bridge.

T-3. The entire procedure was then repeated with the same resistors.

The results of this study are given in Table 2 where T-1, T-2, and T-3 represent the respective conditions of the soldering test. It is understood that no attempt was made to use precise timing of the tinning operations, to use a freshly clean surface of the soldering iron for each individual operation, or to block off the heating of the resistor units by means of hemostats or other clamping devices.

It is evident that precise resistive networks are difficult to fabricate using the commonly available resistors unless great care is observed in blocking off the heat during soldering or unless a correction is made in selecting the resistors.

Table 2 — Effect of Soldering on Resistor Values
(Per Cent Resistance Change)

Initial R (ohms) (T-1)	Composition $\frac{1}{2}$ Watt (T-2) (T-3)		Metallized $\frac{1}{2}$ Watt (T-2) (T-3)		Metallized $\frac{1}{2}$ Watt (T-2) (T-3)	
	100	1.78	2.15	0.18	1.20	5.48
330	1.46	3.03	1.48	1.57	8.90	9.44
1 k	2.98	4.05	0.69	0.79	5.53	4.91
3.3k	4.18	7.63	3.76	5.65	2.29	1.95
10 k	5.37	5.37	0.55	0.55	0.84	-0.19
33 k	2.67	5.93	2.12	2.12	-0.09	-0.20

Effect of Frequency on Pad Elements

To evaluate the effect of frequency on pad elements, the attenuator was mounted on the Boonton type 250-A RX meter, the input terminated by 50 ohms, and the effective parallel resistance and capacitance measured at 0.5, 4.5, 25, and 45 megacycles. The resistance curves are shown in Figure 3. The 0.5-megacycle curve is not shown as it was essentially unity throughout. The resistance ratios somewhat greater than unity for the lower value of resistors are due to lead inductance. The 45-megacycle behavior of the higher resistance steps could possibly be improved by the use of special resistors discussed later.

The terminal parallel capacitance of the attenuator is shown in Figure 4. The 25-megacycle curve is not shown, but would lie intermediate between the 4.5- and 45-megacycle curves. This data indicates

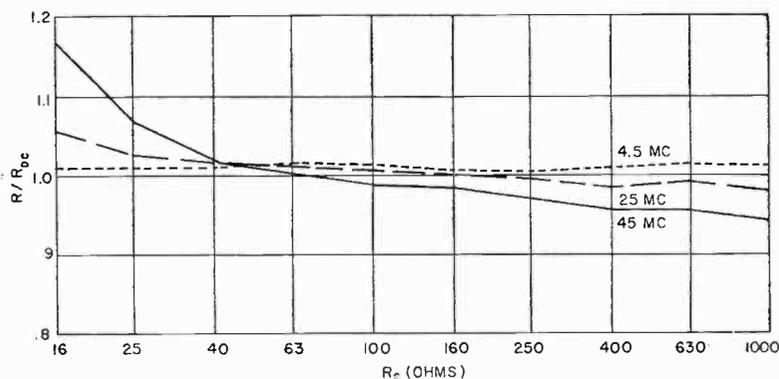


Fig. 3—Ratio of high-frequency parallel resistance to d-c resistance of terminated π attenuator assembly.

that it may be necessary to retune the input circuit slightly when the attenuator is adjusted through the 10 to 63 ohm steps.

Attenuation Measurements

The attenuator was mounted in the instrument with no input tuned circuit, and measurements were made of insertion loss. A series of carefully selected load resistors R_L were secured in substitution for the transistor which would normally be under test. The attenuation measurements were made in the frequency range from 10 to 55 megacycles using a Hewlett-Packard model 608A signal generator. The resulting insertion-loss data is shown in Table 3.

The results contained in Table 3 are quite interesting. The attenuation at the higher frequencies of the lowest steps where the inductive reactance of the resistors is of importance, and the highest steps where

Table 3 — R-F Insertion Loss, in decibels, of " π " Input Attenuator
(No Tuned Circuits)

R_L (ohms)	10 Mc	16 Mc	25 Mc	35 Mc	45 Mc	55 Mc
10	19.94	20.24	20.61	20.85	21.17	21.67
25	20.03	20.00	20.30	20.39	20.58	20.72
63	19.63	19.68	19.72	20.06	20.17	20.37
160	19.50	19.61	19.95	20.24	20.50	20.82
400	19.40	19.71	20.32	21.02	21.76	22.60
1000	19.42	20.32	22.08	23.82

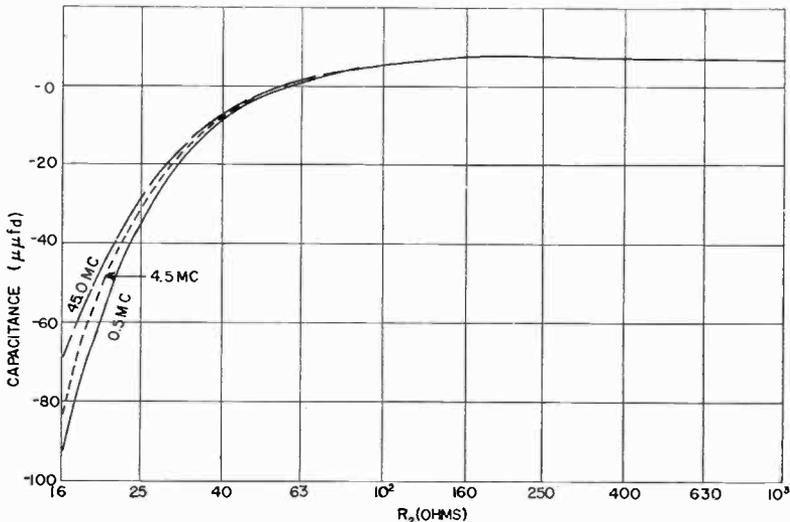


Fig. 4—Effective reactance of terminated π attenuator assembly.

the shunt capacitance and the limited r-f performance of the resistors are of real importance, departed materially from the desired 20-decibel value. It may be concluded that inaccurate data would result, from use of semi-conjugate matching at high frequencies, for transistors with fairly high input resistance values. Such inaccuracies may be corrected by the use of a tuned input circuit.

INPUT COIL

The setup was therefore modified to include a suitable tuned circuit as shown by Figure 2. Measurements were again made of the attenuation in decibels with several of the load resistors and the variation of attenuation in the frequency range from 10 to 55 megacycles was less than 1.0 decibel for any step.

The requirements for the various input coils are:

1. The tap point must be sufficiently high that the tuned impedance at the point will be at least ten times the highest measured input resistance value (i.e., 10×1000 ohms). However, the tap must not be so high that the transformed effective input capacitance of the transistor prevents satisfactory tuning by means of the variable capacitor.

2. The effective field must be small so that the coupling between the input and output coils will be negligible.

To fulfill the above requirements, commercial toroidal cores using Ferramic-Q¹ or experimental Lithium-Zinc-Ferrite² toroidal cores, depending upon the operating frequency, were employed in the input coils for the equipment. Ten input coils, L-1 to L-10 inclusive, have been designed for frequencies from 400 kilocycles to 50 megacycles. The constructional details together with the center frequency for each coil are given in Table 4.

Table 4 — Input Coils

Coil No.	Turns		Wire Size	Toroidal Rings		Center Freq. Mc
	Tap	Total		Material	Diam."	
L-1	20	140	30	2 stacked F421-2	0.50	0.5
L-2	14	78	30	2 stacked F421-2	0.50	0.8
L-3	12	72	LITZ	F421-3	0.50	1.0
L-4	8	32	25	F421-3	0.50	2.0
L-5	11	22	26	F421-2	0.50	4.0
L-6	5	20	23	F421-2	0.50	4.5
L-7	9	33	25	Li-Zn-Fe (High)	0.67	8.0
L-8	8	20	20	Li-Zn-Fe (High)	0.67	15.0
L-9	10	13	26	Li-Zn-Fe (Low)	0.725	20.0
L-10	5	13	20	Li-Zn-Fe (Low)	0.725	30.0

TRANSISTOR SOCKETS AND METERS

A ceramic octal socket is mounted in the instrument and various 5-pin in-line transistor sockets are plugged in by means of the intermediacy of octal type tube bases as shown. Plug-in assemblies are provided for base #1, base #2, and emitter input types of triode or tetrode connection, and with or without connection for neutralization.

Four meters with suitable switching means are used in conjunction with the biasing supplies and controls to measure collector voltage, collector current, base #1 current, and base #2 current.

¹ Trademark of the General Ceramic Corporation, Keasbey, N. J.

² R. S. Weisz, "Ferromagnetic Spinel Using Lithium," *Ceramic Industry*, Vol. 58, #4 & #5. The Li-Zn-Fe (High Temp.) provides $u_0=37$, $Q=90$ at 25 Mc, and Li-Zn-Fe (Low Temp.) provides $u_0=20.5$, $Q=115$ at 50 Mc.

OUTPUT COILS

The output coils were developed in a manner similar to that for the input toroids. However, the effective output capacitance values of junction transistors are relatively low (5 to 20 micromicrofarads). Therefore, to secure the highest possible tuned impedance at the collector tap, the output coils were tapped relatively high or at the top in some cases. Table 5 gives information regarding winding, core material, and center frequency of the 10 output coils designed for use with this equipment.

Table 5 — Output Coils

Coil No.	Turns		Neut.	Wire Size	Toroidal Rings		Center Freq. Mc
	Tap	Total			Material	Diam."	
L'-1	100	200	20	36	F109-2	0.87	0.5
L'-2	40	70	8	30	2 stacked F421-2	0.50	1.0
L'-3	15	38	3	32	F262	0.375	2.0
L'-4	15	20	3	23	F421-3	0.50	4.0
L'-5	10	15	2	—	F421-2	0.50	6.0
L'-6	—	30	—	25	Li-Zn-Fe (High)	0.67	10.0
L'-7	—	16	—	20	Li-Zn-Fe (High)	0.67	18.0
L'-8	—	13/13	—	20	Li-Zn-Fe (Low)	0.725	24.0
L'-9	—	10	—	23	Li-Zn-Fe (Low)	0.725	30.0
L'-10	—	9	—	20	Li-Zn-Fe (Low)	0.725	37.0

The tuned impedance values of the L'-8, L'-9, and L'-10 coils are not sufficiently high largely because of the 21-micromicrofarad built-in capacitance of the measuring equipment at the output test point. By replacing the switch structure of the output pad with a small turret assembly and also substituting a simple transistor socket adapter arrangement in place of the present three-stage socket assembly, the over-all capacitance should be reduced to approximately 10 micromicrofarads. The coil turns could then be increased with consequent improvement in impedance values.

OUTPUT LOAD ASSEMBLY

Design Considerations of L Pad Array

It would be advantageous in an output resistive load network to be able to secure an r-f voltmeter indication of the correct matching of the network and the transistor under test. One means of obtaining the desired result involves the utilization of an array of L pads, each similar to the r_4 and r_5 portion of Figure 5.

Consider the r-f power available at the collector circuit of the transistor to be constant. Then e_3 will remain the same for any given transistor output impedance which is matched by $R_L = r_4 + r_5$, provided

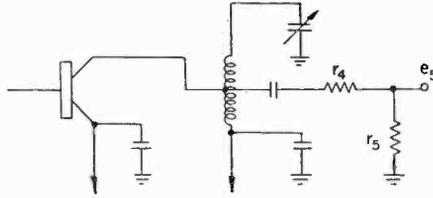


Fig. 5—Measuring equipment output circuit.

that the values of r_4 and r_5 are properly selected. Let two L pads $r_4 + r_5$ and $r'_4 + r'_5$ be considered to constitute any two steps of an array wherein each of the steps properly terminate transistors having output resistance values R_L and R'_L respectively. To determine r_4 , r_5 , r'_4 and r'_5 , we may write

$$e_5 = \frac{e_0 r_5}{r_4 + r_5} = \frac{e'_0 r'_5}{r'_4 + r'_5}, \quad (4)$$

$$r'_5 = r_5 \sqrt{\frac{R'_L}{R_L}}. \quad (5)$$

It was considered that the lowest possible output resistance for which the L pad array should be designed would be 400 ohms. Further, to secure the maximum sensitivity in reading e_5 , the first r_4 element was set equal to zero. The resulting calculations for the 11 steps of a 400 ohm to 40,000 ohm array are given in Table 6.

Because of the high values of resistors required by the output pad array, the choice of resistor types becomes extremely important for high frequency measuring equipment.

Table 6 — Output L Pad Array

Step	R_L (ohms)	$\sqrt{\frac{R'_L}{R_L}}$	r_4 (ohms)	r_5 (ohms)
1	400	1.00	0	400
2	630	1.26	128	502
3	1,000	1.58	368	632
4	1,600	2.00	800	800
5	2,500	2.50	1,500	1,000
6	4,000	3.16	2,730	1,270
7	6,300	3.97	4,710	1,590
8	10,000	5.00	8,000	2,000
9	16,000	6.32	13,500	2,530
10	25,000	7.91	21,800	3,160
11	40,000	10.00	36,000	4,000

High-Frequency Properties of Fixed Resistors

Three types of small fixed resistors were evaluated at high frequencies, namely, (A) composition $\frac{1}{2}$ watt, (B) metallized glass $\frac{1}{2}$ watt and $\frac{1}{3}$ watt units, and (C) deposited carbon units. Samples of these resistor types were measured on the RX meter. The results are presented in Figure 6 in a normalized form first proposed by Pontecorvo who based his work on Howe's transmission-line equivalent of a resistor.³⁻⁶ These curves represent the optimum performance obtainable from these three types of resistors. The deposited carbon resistors

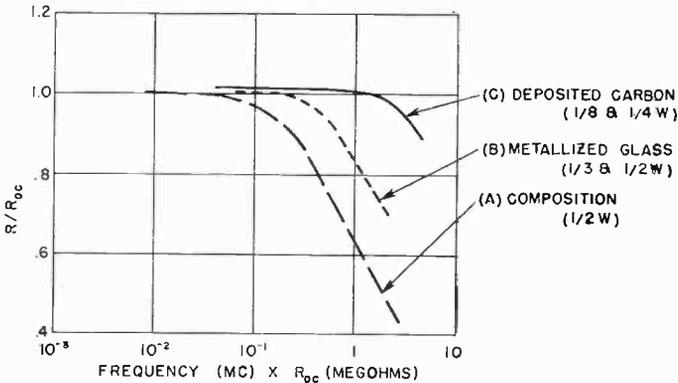


Fig. 6—Radio-frequency performance of fixed resistors.

exhibit superior performance, but these were not used as they were available only in limited ohmic values.

A series of tests was made to determine whether the high-frequency performance of metallized resistors could be improved by forming a desired element from a series of small resistors so as to reduce the net shunting capacitance. An approximate d-c resistance of 33,000 ohms was formed. The results at 25 megacycles for one, two, three, and four resistors in series are given in Table 7.

From the above tabulation the merit of using two resistors in series for each high resistance arm of the L pads is seen. The attainment of

³ P. Pontecorvo, "Behavior of Resistors at High Frequencies and Its Dependence on the Distributed Capacitance," *Alta Frequenza*, Vol. 7, p. 570, August-September, 1938.

⁴ P. Pontecorvo, "On the Behavior of Resistors at High Frequencies," *Wireless Engineer*, Vol. 15, p. 500, September, 1938.

⁵ G. W. O. Howe, "Behavior of High Resistances at High Frequencies," *Wireless Engineer*, Vol. 12, p. 291, June, 1935.

⁶ T. J. F. Pavlasek and F. S. Howes, "Resistors at Radio Frequencies," *Wireless Engineer*, Vol. 29, p. 31, February, 1952.

Table 7 — 25-Megacycle Performance of Series Groups of Metallized Resistors

Rating Grouping	$\frac{1}{3}$ Watt				$\frac{1}{2}$ Watt			
	1	2	3	4	1	2	3	4
R_{dc} (K ohms)	32.44	34.33	33.21	33.52	35.08	33.80	32.40	33.71
R (K ohms)	29.8	33.5	30.0	28.2	30.1	32.5	28.2	27.7
C ($\mu\mu\text{f}$)	0.28	0.23	0.28	0.33	0.34	0.28	0.33	0.50
R/R_{dc}	0.92	0.98	0.90	0.84	0.86	0.96	0.87	0.82

a 96-98 per cent ratio for the $\frac{1}{3}$ watt and $\frac{1}{2}$ watt units is most valuable in the output pad development. It must be understood that the foregoing does not apply to low value resistors less than 100 ohms for which the reactance at high frequencies is essentially inductive.

IMPROVED L PAD ARRAY

In tests with two adjacent L pad steps and an adjoining "open" position on a two-wafer rotary switch, it was determined that all inactive pads were capacitively coupled to the step which was under test. The multiple loading was effectively removed at the high point of the L-pad array by placing a shorting switch wafer in parallel with the high selector wafer. By this means all pads except the one in use are shorted to ground. The increase in capacitance at the high point may be effectively cancelled by readjustment of the output tuned circuit. No shorting switch wafer was used at the tap point of the L pads because the added capacitance at the tap point would cause the voltage division of the high-resistance steps at high frequencies to be incorrect. Further small improvements in the assembly were secured by use of a bakelite switch shaft and side support rods.

Suitable r_4 and r_5 elements each consisting of two metallized glass $\frac{1}{2}$ watt resistors in series were secured for the complete 11-step L pad array in accordance with Table 6 values. Allowance was made for the change in resistance due to the soldering operations and the d-c resistance of all units rechecked three days after assembling. The complete L pad assembly was mounted on the RX meter and measured at 25 megacycles. The ratios of high frequency to d-c resistance for the various steps are given in Table 8.

The low ratio of R/R_{dc} , particularly on the higher steps, represents a deviation from the values shown by curve (b) of Figure 6. Therefore, it may be concluded that the L pad structure is not satisfactory for use in the high-frequency measuring equipment without further improvement.

Table 8 — 25 Megacycle Data for Improved L Pad Assembly

Step	R_{dc} (ohms)	R (ohms)	C (μmf)	R/R_{dc}
1	403	396	7.7	0.985
2	630	612	7.4	0.970
3	998	950	6.5	0.952
4	1,583	1,470	5.8	0.929
5	2,514	2,290	5.4	0.912
6	3,999	3,540	5.0	0.886
7	6,311	5,480	4.9	0.868
8	9,995	8,600	5.0	0.861
9	16,000	12,900	4.9	0.807
10	24,900	19,100	4.9	0.768
11	40,000	29,600	4.0	0.740

SHUNT OUTPUT PAD

The final equipment used a ten-step simple shunt array of resistors (630 to 40,000 ohms) on an improved rotary switch. The first position, normally 400 ohms, was made open circuit to permit $\frac{1}{2}$ -voltage checks on output resistance values. As previously described, a shorting wafer was connected in parallel with the high connections of the various steps. Each resistor step of the array was secured by using two metallized glass ($\frac{1}{3}$ watt) units in series.

The assembly was measured by use of the RX meter at 0.5, 4.5, 25 and 45 megacycles. The normalized ratio of high-frequency to d-c resistance is presented by the curves of Figure 7. The 0.5, 4.5 and 25 megacycle data may all be fitted on a single curve α . However, the results at 45 megacycles definitely result in a separate curve β . It has been found that all available resistors when measured at VHF from 30 to 250 megacycles on the RX meter do not produce data for a single type which may be plotted as a simple normalized curve. Curves α and β of Figure 7 compare favorably with the theoretical values represented by curve (b) of Figure 6. Therefore, the shunt pad structure does not degrade the high-frequency performance of fixed resistors.

OVER-ALL INSTRUMENT ASSEMBLY

General Description

The schematic diagram of the instrument is given in Figure 8. The 20-decibel π type input attenuator is connected to a tap on the input tuned circuit and to the selected transistor input terminal. The collector circuit is connected to the shunt output pad and to a point on the output tuned circuit. An auxiliary winding in reverse phase is provided on the various lower frequency coils, i.e., 0.5 to 4.5 megacycles. This winding may be used to provide neutralization or unilateralization when used in conjunction with the components C_n and R_n .

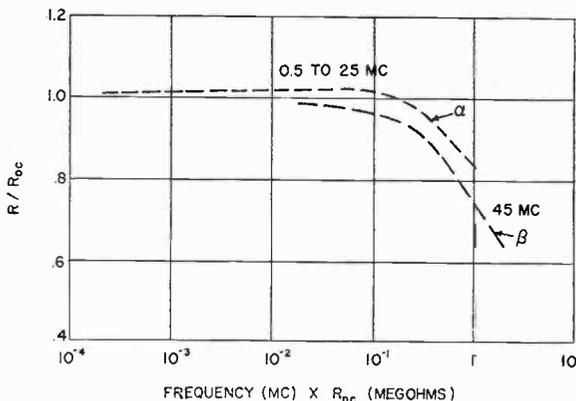


Fig. 7—RX meter tests of shunt output pad.

Batteries are provided for constant-voltage operation of the collector circuit and for constant-current operation of the emitter and the biased base electrode. The front view of the instrument is shown in Figure 9. The four meters, I_c , I_{b1} , I_{b2} , and E_c , the associated switches and current control resistors, and the n-p-n/p-n-p polarity switch are located on the upper sloping panel.

The flat horizontal plane of the instrument contains the input and output tuned circuits, the transistor socket means, the unilateralization controls C_n and R_n , and the small feedthrough test points TP_1 , TP_2 , and TP_3 . Steatite sockets are provided for the coil and transistor

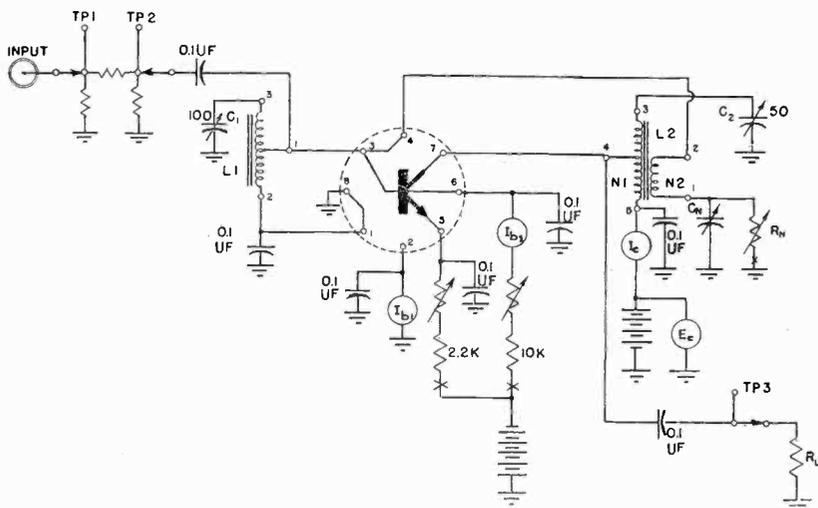


Fig. 8—Instrument schematic diagram.

plug assemblies, four-prong for the input, five-prong for the output coil, and eight-prong octal for the transistor mounting plugs. The various input and output toroidal coils are mounted on Amphenol* type CP mica-filled bakelite plug assemblies. The in-line transistor sockets (various assemblies available for all normal triode and tetrode connections) are set into brass caps which are mounted on octal tube

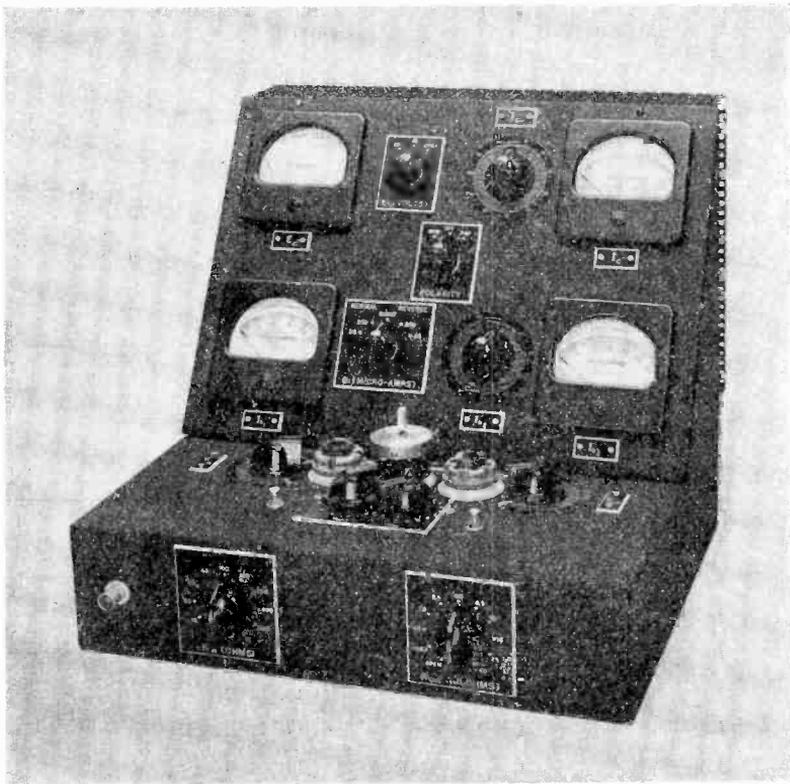


Fig. 9.—Transistor measuring equipment.

bases. A representative set of toroidal coils and a socket assembly may be seen in the photograph. If desired, metal plug caps may be snapped over the coil assemblies to protect the toroids and to facilitate withdrawal from the sockets. The lower vertical plane contains the 50-ohm type BNC input connector and the dial assemblies for the input π pad and shunt output pad.

* Trademark of the American Phenolic Corp., Chicago, Ill.

Measurement Procedures

After the transistor d-c operation has been adjusted, a high-impedance r-f voltmeter is connected to the input test point and the signal generator adjusted to 0.1 volt or other desired output. The voltmeter is then connected to the collector output test point and the shunt output pad switched to "open" and the value of e_0 noted. The pad is then adjusted to the step which causes e_0 to decrease to $\frac{1}{2}$ of the former value. The power gain may now be found by use of the Figure 10 family of curves which are secured from the expression

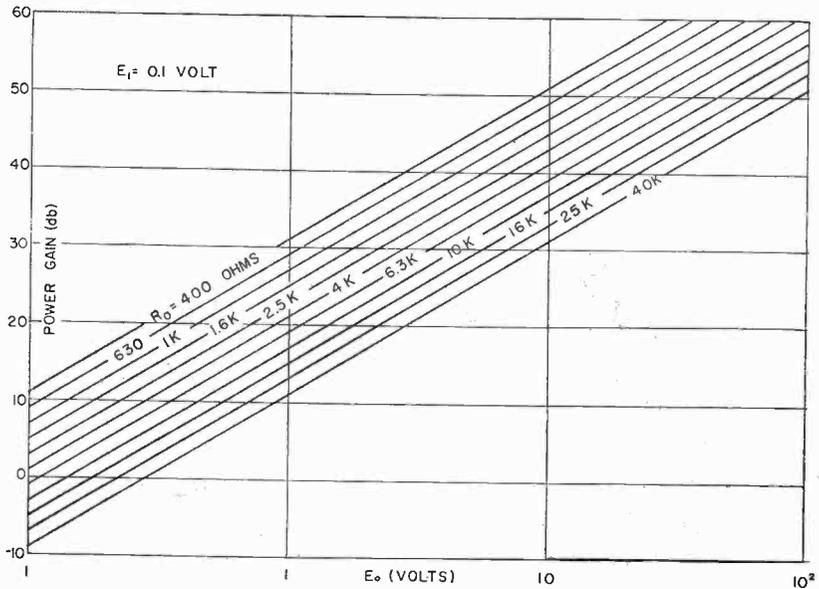
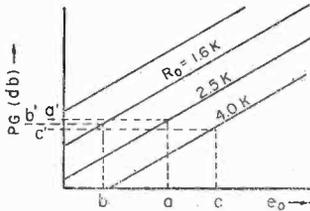


Fig. 10—Transistor power gain versus output voltage and output resistance.

$$PG = 10 \log_{10} \left[\left(\frac{e_0}{e_1} \right)^2 \frac{R_o \times 10^2}{R_L} \right] \text{ decibels.} \quad (6)$$

If the effect of the output tuned circuit is to be considered, more accurate data may on occasion be secured by switching the shunt pad to positions below and above the initial setting. The additional power gain and output resistance values are then noted. As shown by Figure 11, the optimum results will be found by use of the curves of Figure 10. In the representative measurement shown in Figure 11, the $\frac{1}{2}$ -voltage value (a) was secured by use of the 2,500-ohm shunt pad

Fig. 11—Accurate PG determination.

step and the $PG = (a')$. Upon switching to the 1,600-ohm step and then to the 4,000-ohm step, e_0 values of (b) and (c) were secured with resulting PG readings of (b') and (c') respectively. For this particular determination the $\frac{1}{2}$ -voltage reading was sufficient to provide the correct results.

If desired, calibration curves may be obtained for the variable input capacitor, C_1 , and output capacitor, C_2 , shown by Figure 8 for each coil and frequency measurement. The various curves may be secured by mounting the instrument on the RX meter or a Q meter and checking the displacement of the standard capacitor versus the C_1 or C_2 rotation. As an alternative, fixed known capacitors may be inserted in the transistor socket and the change in C_1 or C_2 dial settings noted.

When relatively low-frequency measurements are made on transistors having significant feedback parameters, the setup of Figure 12 has been found to be particularly useful. In addition to the normal measurement equipment, a sweep generator which may be adjusted to the desired center frequency (such as 455 kilocycles) is coupled to the collector circuit. A cathode-ray oscilloscope is connected through a decoupling resistor to the base input. With C_n and R_n out of the circuit, a scope pattern similar to Figure 13 will be obtained after adjustment of the tuned circuits and the input and output pads.

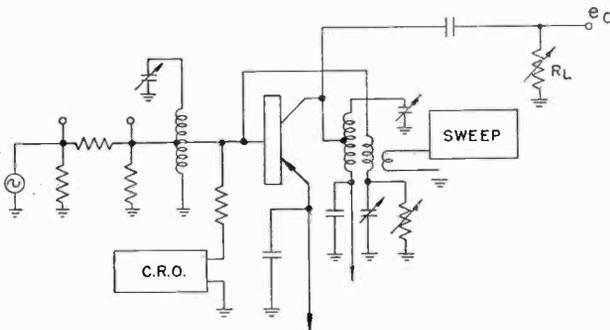


Fig. 12—Neutralization setup.

In the cathode-ray-oscilloscope pattern shown in Figure 13 a frequency-locating beat can be secured by injecting signal from the input carrier generator. The bulge in the pattern is due to the positive feedback at or close to the desired carrier frequency. The C_n and R_n controls are next adjusted so as to secure a flat pattern as shown by the dotted lines. The circuits and pads are then readjusted as required. After finally rechecking the neutralization or unilateralization, the sweep generator signal is removed and the desired measurements obtained.

The values of the parallel terminal feedback quantities C_{bcc} and r_{bcc} may be obtained by the use of the following equations for each frequency and output coil, provided that C_n and R_n have been calibrated by use of the RX meter or a Q meter.

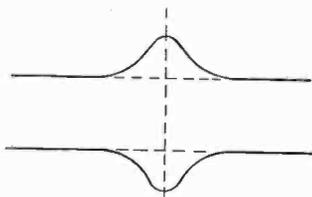


Fig. 13—Oscilloscope pattern (unneutralized).

$$C_{bcc} = C_n \left(\frac{n_2}{n_1} \right) \quad (7)$$

$$r_{bcc} = R_n \left(\frac{n_1}{n_2} \right) \quad (8)$$

where

n_1 = output coil turns at tap,

n_2 = feedback winding turns.

CONCLUSIONS

A transistor-measuring equipment has been developed which will provide power gain, terminal input and output resistance and capacitance values, and feedback parameters over the frequency range from 0.4 to at least 50 megacycles. The resulting power gain values are estimated to be accurate to within ± 1 decibel.

EQUIPMENT FOR MEASURING JUNCTION TEMPERATURE OF AN OPERATING TRANSISTOR

BY

K. E. LOOFBOURROW AND J. OLLENDORF

RCA Semiconductor Division,
Harrison, N. J.

Summary—This paper describes a method of measuring the junction-temperature rise of a transistor due to power dissipation. The method utilizes a temperature-sensitive electrical parameter of the transistor and consequently does not require alteration of the unit. Two equipments which can make these measurements conveniently are described in detail.

INTRODUCTION

EQUIPMENT which will measure junction-temperature rise due to power dissipation is required by both transistor development engineers and circuit design engineers. The transistor development engineer must have information on junction-temperature rise to improve transistor designs and to establish proper maximum ratings. The circuit design engineer requires the information to insure that maximum ratings are not exceeded in equipment designs.

Junction temperature can be measured by a thermocouple placed near the junction. However, because this method involves physical alteration of the transistor, the heat flow geometry may be altered sufficiently to produce an appreciable error.

Several of the electrical characteristics of a transistor change radically with temperature, and can be used for temperature measurement. In particular, the reverse current of the collector with the emitter open, I_{co} , is a convenient parameter. The rise of junction temperature due to dissipation is determined as follows: (1) the transistor is operated for a sufficient time to stabilize the junction temperature; (2) the emitter circuit is opened, and I_{co} is measured before the transistor cools; and (3) the transistor is then heated in an oven until I_{co} has the same value measured in step (2), and the temperature is noted. By utilizing the theoretical expression for I_{co} and special equipment the procedure is considerably simplified as will be described.

CALCULATION OF JUNCTION TEMPERATURE FROM I_{co}

The measured reverse current, I_{co} , is composed of the junction saturation current and a leakage current generally presumed to be due

to conduction along the surface of the semiconductor. It may be expressed mathematically by adding a leakage term to the conventional diode equation:

$$I_{co} = I_s (e^{-qV/kT} - 1) + GV, \quad (1)$$

where,

I_s is the junction saturation current,

q is the electronic charge,

V is the applied reverse voltage,

k is Boltzmann's constant,

T is the absolute temperature,

G is the leakage conductance shunting the junction.

When the applied reverse voltage, V , is greater than 0.2 volt, the quantity $(e^{-qV/kT})$ is much less than unity and may be neglected. In germanium transistors at room temperature, the leakage current, GV , is usually small compared to the saturation current, I_s , provided V is well below the breakdown voltage of the transistor. Consequently, measurements made at voltages below one volt are generally unaffected by the leakage conductance, G . In silicon devices, however, I_s is much smaller and the method is limited to some extent. This limitation will be discussed later. Thus for measurements on germanium transistors good results can be obtained by assuming that I_{co} is essentially equal to I_s . The value of I_s is given in semiconductor theory approximately as

$$I_s = A e^{-qV_g/kT}, \quad (2)$$

where A depends on the physical constants of the semiconductor and the geometry of the device and V_g is the energy gap of the semiconductor expressed in volts.^{1, 2} Figure 1 shows the normalized collector saturation current of a typical germanium transistor as a function of temperature. Since this curve correlates well with Equation (2), it is generally applicable to germanium transistors, and the junction-temperature rise can be obtained directly from saturation current measurements without the time-consuming step involving an oven.

An expression for the junction-temperature rise may be obtained as

¹ W. Shockley, *Holes and Electrons in Semiconductors*, D. Van Nostrand Co., New York, N. Y., 1950; p. 314.

² W. M. Webster, "Saturation Current in Alloy Junctions," *Proc. I.R.E.*, Vol. 43, p. 277, March, 1955.

follows. The ratio of saturation currents measured at two different temperatures, T_1 and T_2 , may be written

$$\frac{I_{s2}}{I_{s1}} = e^{-\frac{qV_g}{kT_2} + \frac{qV_g}{kT_1}} = e^{\frac{qV_g}{kT_1 T_2} [T_2 - T_1]}$$

Therefore,

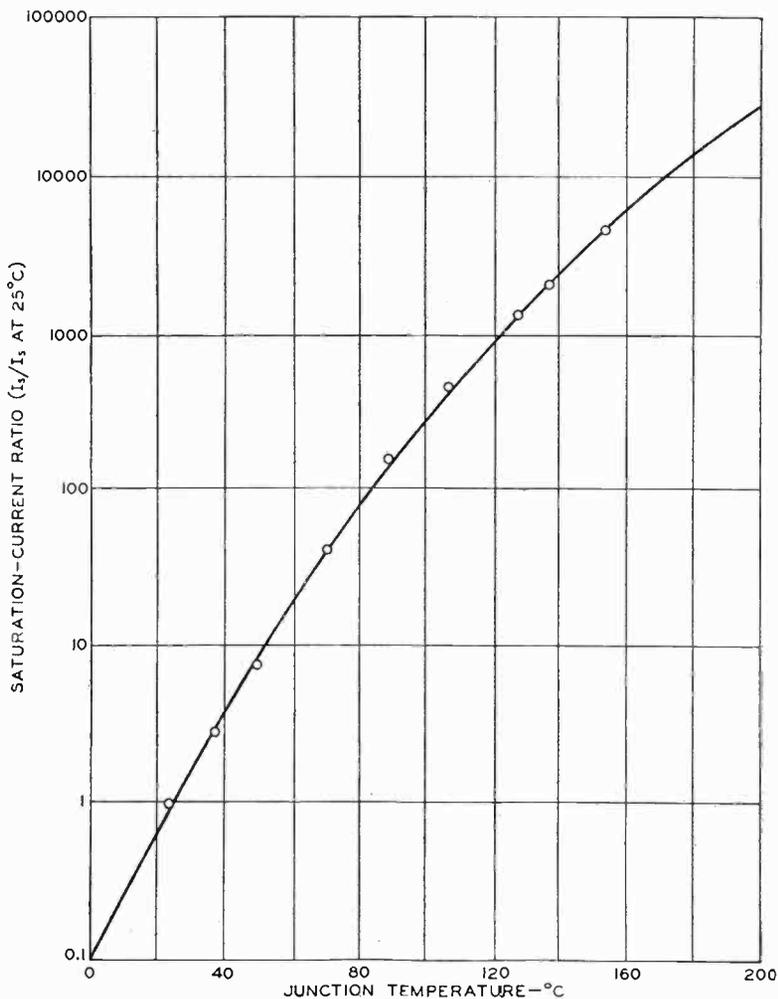


Fig. 1—Collector saturation current as a function of temperature for a germanium transistor. Solid line is plotted in accordance with Equation (2); points indicate measured values.

$$T_2 - T_1 = \frac{kT_1T_2}{qV_g} \ln \frac{I_{s2}}{I_{s1}}. \quad (3)$$

If the temperature range is assumed to be from 20° to 30°C, then $kT_1T_2/qV_g = 10.6$. For a temperature range from 25° to 75°C, this value increases to 12.6.

The first step in the procedure, therefore, is to measure I_{s1} at ambient temperature T_1 . I_{s2} is then measured immediately after the transistor has been dissipating a known amount of power, and the temperature rise is determined from the curve given in Figure 1 or approximately by substitution in Equation (3). The temperature rise for a given dissipation expressed in degrees centigrade per watt is an important parameter of the transistor and is called the thermal resistance.

For transistors in which I_s is not the only significant part of I_{co} (e.g., when leakage is high, or in silicon), calibration using an oven or other external heater is still feasible. However, the shape of the I_{co} -versus- T curve of one unit should not be generalized to all units of a given type because the dependence of leakage current on temperature is usually quite variable.

TEST EQUIPMENT

As mentioned previously, it is important to measure I_{co} , or I_{s2} , before the transistor cools appreciably. The cooling effect can be reduced to a negligible degree by making the measurement in a cyclic fashion. During one part of the cycle, power is dissipated in the transistor. During the other part of the cycle, the saturation current is measured. The period of the cycle must be short enough so that the cooling effect is negligible. The rate of cooling depends on the thermal time constant of the transistor. Although the thermal time constant varies considerably with design, a measuring system operating at a frequency of 30 or more cycles per second is sufficiently fast for most transistors. The switching may be accomplished mechanically by relays or rotating commutators, or electronically by means of diodes. Examples of both types of equipment will be described.

Figure 2 is a photograph of a mechanical switching test equipment. The motor-driven, commutator-type, rotating switch is readily identified. The commutator is composed of six slip rings to provide the proper sequence of switching in the various circuits. The motor turns at 1,800 revolutions per minute, providing a switching rate of 30 cycles per second.

Figure 3 is a circuit diagram of the equipment pictured in Figure 2.

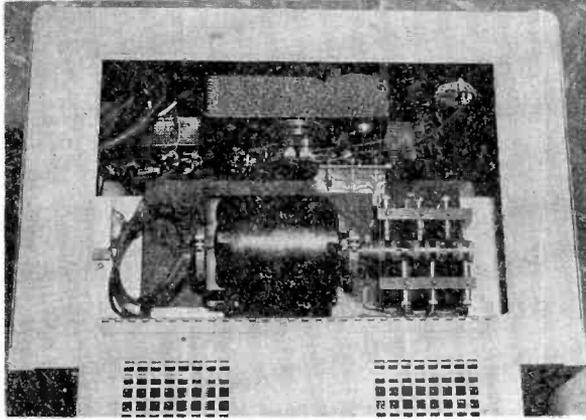


Fig. 2—Motor-driven commutator used in mechanical switching method.

When the commutator is in such a position that the contacts of S-3 are in the DOWN position, the transistor operates at a level of dissipation determined by the applied voltages. The switch remains in this position for a period of 14 milliseconds. Then, during a switching interval of 2 milliseconds, the contacts move to the UP position, the emitter circuit is opened, and the saturation current is measured. The time in the UP position is also 14 milliseconds. When the transistor is inserted in this circuit the motor should be running and switch S-2

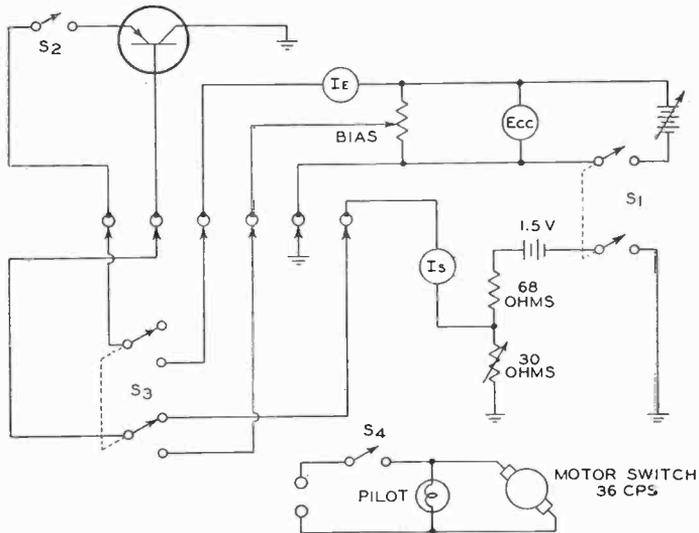


Fig. 3—Circuit of equipment for measuring temperature rise with mechanical switching method.

should be in the open position. With S-2 open, the emitter circuit cannot be closed by the commutator and essentially zero initial dissipation is insured. With S-2 open, the average value of I_{e1} is obtained at ambient temperature. S-2 is now closed and the emitter current is adjusted by means of the bias potentiometer to the desired value of power dissipation. This dissipation is the product of the emitter current and the supply voltage. When equilibrium is reached the saturation current, I_{e2} , is read and the junction temperature may then be obtained from Figure 1 or approximately by substitution into Equation (3).

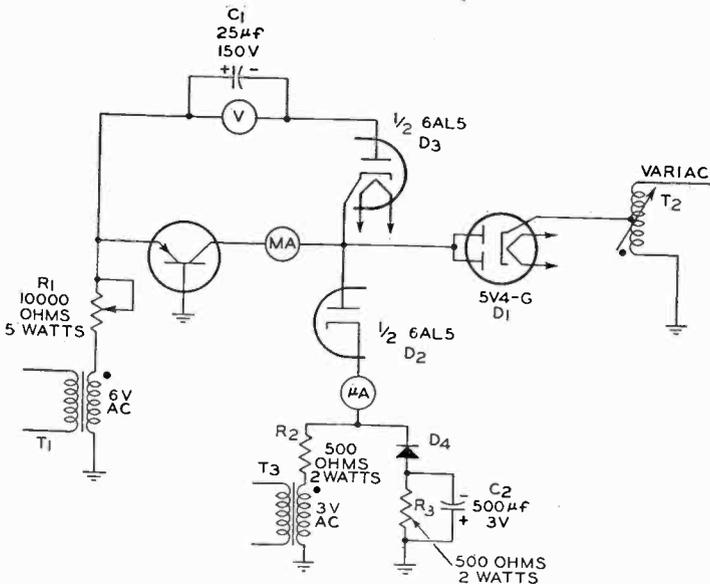


Fig. 4—Circuit of equipment for measuring temperature rise with electronic switching.

Figure 4 is a circuit diagram of an all-electronic equipment which uses vacuum-tube diodes and a-c voltages to provide switching at a rate of 60 cycles per second. The circuit shown is connected for measurements on p-n-p transistors. Operation of this circuit is as follows: When the tap on the transformer T_2 is negative, diode D_1 starts conducting and the negative voltage appears on the collector of the transistor under test. Transformer T_1 is phased so that the emitter becomes positive with respect to ground when the collector becomes negative and power dissipation results. Resistor R_1 is a variable current-limiting resistor. The negative collector-voltage swing also causes diode D_3 to conduct, and charges capacitor C_1 to the peak negative voltage between

the collector and the emitter. As a result, the peak collector voltage V_p appears across the voltmeter. The negative voltage swing at T_2 is accompanied by a positive voltage swing on transformer T_3 which keeps diode D_2 cut off. When the tap on T_2 becomes positive, D_1 is cut off, T_3 swings negative and causes D_2 to conduct, T_1 swings negative and cuts off the emitter current. The negative voltage on the plate of D_3 is greater than the peak negative voltage at T_3 so that D_3 does not conduct during this half cycle. The current flowing through D_2 is the reverse collector current, I_{co} . The circuit consisting of R_2 , D_4 , R_3 , and C_2 is a diode clipper circuit which provides a flat topped negative pulse having an amplitude of about 1 volt for sampling I_{co} .

The average collector current is measured with a d-c ammeter and the peak collector voltage with a d-c voltmeter across C_1 . If pure, rectified sinusoidal waveforms are assumed, the power dissipation in the collector circuit is given by

$$P_c = \frac{\pi}{4} V_p I_{co} \quad (4)$$

The average reverse collector current is read on a microammeter connected in series with T_3 , or it may be estimated by observing the waveforms of the voltage across a resistor connected in series with T_3 . Ideally, this waveform approximates a square wave, but it may be distorted by the charge and discharge of various capacitances in the collector circuit during the reverse current measurement period. Although this distortion makes a waveform measurement more difficult, it does not introduce an error when I_{co} is measured with a meter because the sum of the charge and discharge currents is zero.

When this equipment is constructed care should be taken to keep to a minimum any phase shift between transformers T_1 , T_2 , and T_3 which would result in spikes at the leading or trailing edge of the reverse current waveform. Phase shift between T_1 and T_2 would also affect the power dissipation measurement. It is important that the voltage appearing at T_1 be large compared to the emitter to base voltage. Otherwise, the emitter current will be nonsinusoidal and dissipation calculations will be in error. Care should also be taken to minimize hum pickup because the collector circuit sees a very high impedance to ground during the saturation-current reading.

Measurements made on the equipments of Figures 3 and 4 show reasonably good agreement as is evident from Figure 5. The technique of measurement with either equipment is exceedingly important. Sufficient time must be allowed for the transistor to reach equilibrium

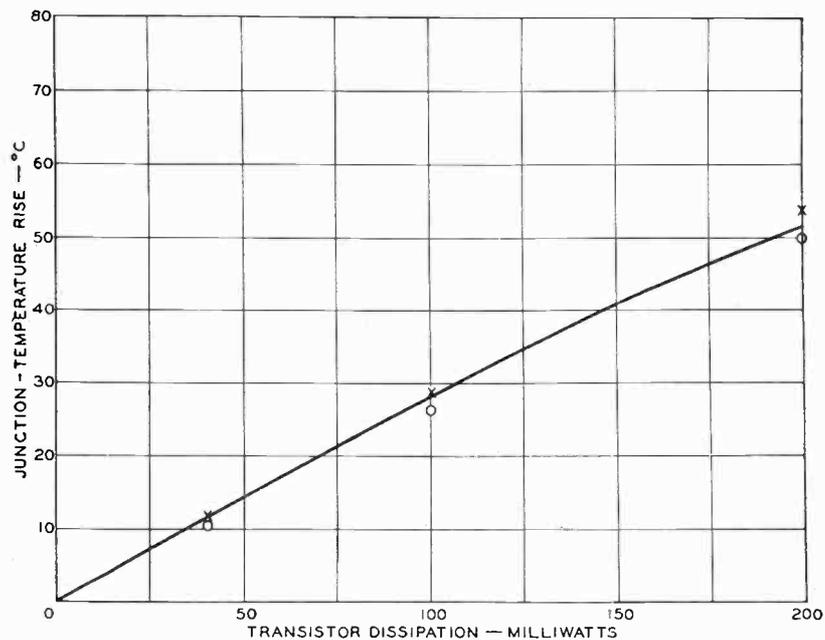


Fig. 5—Junction-temperature rise as a function of dissipation measured on one unit by both methods.

before each reading. In some cases, this time may be as long as five or ten minutes. The heating effect of the operator's hands while connection of the transistor is made to the test set is a source of error which can easily become appreciable. Because many transistor packages have very small external surface areas, the difference between the ambient temperature and the temperature of the case represents a large proportion of the total temperature rise. Consequently, small air currents can produce wide variations in results. With proper care, consistent, useful measurements can be obtained with either equipment.

EVALUATION OF TRANSISTORS FOR CLASS-B POWER OUTPUT AMPLIFIERS

BY

WERNER HASENBERG

RCA Victor Television Division,
Camden, N. J.

Summary—Transistor considerations involved in the design of a class-B amplifier stage are discussed and a simple test circuit to display the dynamic characteristics of interest in this application is described. The large-signal current-transfer characteristics for two transistors in a push-pull circuit and the input characteristic of one transistor are simultaneously displayed on an oscilloscope. The simultaneous display permits accurate matching of a pair of transistors as well as inspection of the linearity of their characteristics. Either *n-p-n* or *p-n-p* transistors can be used and the method can be applied to complementary pairs. The test circuit is simple and easy to build and operate.

INTRODUCTION

IN this paper the transistor considerations involved in the design of a class-B amplifier stage are discussed and a simple test circuit to display the dynamic transistor characteristics of interest in this application is described.

In the design of class-B push-pull amplifiers, the following transistor characteristics are of importance to the circuit designer:

- Maximum collector dissipation.
- Maximum peak collector current.
- Maximum collector voltage.
- Input characteristics.
- Base-to-collector current transfer characteristics.

The maximum ratings are generally given by the transistor manufacturer. This paper is concerned primarily with the input and transfer characteristics of the transistor. These operating characteristics may be conveniently determined from the static characteristics of the transistor as shown in Figure 1. The output and the input characteristics and the current transfer curve are shown in this figure.

Consider first the output circuit. The output load resistance may be determined from the collector characteristics in the usual fashion. The maximum collector voltage must be at least twice the applied collector supply voltage, E_{CC} , to accommodate the output voltage swing during the time the transistor is cut off. The minimum collector voltage

is reached at maximum signal input voltage and maximum collector current. This minimum collector voltage will be taken as one volt as a practical value for illustrative purposes. Therefore, the output load resistance of the single transistor unit is determined by

$$R_L = \frac{E_{CC} - 1}{I_C'} \quad (1)$$

where I_C' is the peak collector current. R_L is shown by the load line in Figure 1.

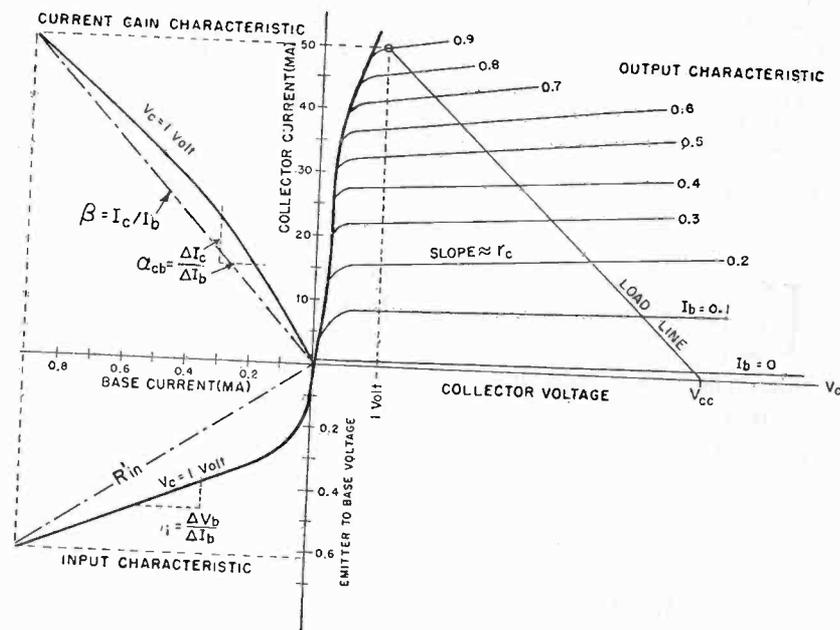


Fig. 1—Output, current gain, and input characteristics taken from data sheets for 2N109's.

Referring again to Figure 1, both the current transfer curve (upper left) and the input characteristic (lower left) are related to the output characteristic and are shown for a collector voltage of 1 volt. In the design of class-B push-pull amplifiers, large-signal design parameters must be used rather than the small-signal parameters. Of particular concern is the d-c current amplification factor, β , at the peak current and the input d-c resistance, R_{IN}' , also at the peak current. As shown in Figure 1, the d-c current amplification factor, β , is defined as the ratio of collector current to base current, and the ratio of base-to-

emitter voltage to the maximum base current is defined as the d-c input resistance, R_{IN}' .

In class-B operation, the importance of the d-c amplification factor is twofold. Since at moderate currents the power gain of the amplifier is directly proportional to the d-c current amplification factor, it should be as high as possible. Furthermore, the variation of the d-c current amplification factor with collector current should be a minimum in order to minimize nonlinear distortion effects, i.e., the current transfer characteristic should be as linear as possible. Also, it is required that, for best performance of a pair of transistors in a push-pull output stage, their current transfer curves should coincide as closely as possible over the entire range of operation up to the peak collector current. In other words, their current transfer characteristics should be well matched.

Both the d-c current amplification factor and the d-c input resistance of a transistor can be measured by d-c methods. Experience, however, has shown that this method is inconvenient. The d-c power applied to the transistor for a direct measurement may, and usually does, give rise to thermal effects that are detrimental not only to the transistor itself, but also to accurate measurement of the desired parameters. Furthermore, no single-point measurement can give information about the shape of the curve or of how well a pair of transistors is matched. These disadvantages are overcome by the dynamic test method to be described. Both the d-c current amplification factor curve and the input characteristic are displayed on an oscilloscope. These characteristics are displayed simultaneously and, furthermore, a simultaneous display can also be made of the characteristics of a pair of transistors. Thus, this method provides for convenient observation of the shapes of the characteristics, permits accurate evaluation of the current amplification factor and the input resistance of each transistor, compares the match between a pair of transistors, and does not incur detrimental thermal effects.

DESCRIPTION OF TEST CIRCUIT

An oscilloscope, a Variac,* and a curve tracer unit are needed for the display of the transistor characteristics specified above. Figure 2 shows the circuit diagram of the curve tracer unit. The transistors are driven in push-pull by a 60-cycle sine wave from transformer T_1 . The composite input current (I_b) is measured as the voltage drop across resistor R_1 . A voltage source is applied to the collectors and the com-

* Trademark of the General Radio Company, Cambridge, Mass.

posite collector current is measured as the voltage drop across resistor R_2 . For normal conditions, the composite currents are essentially currents of a single transistor and will be considered as such hereafter.

For collector currents up to 0.5 ampere, a collector voltage source of 1.5 volts is recommended. A 3-volt battery is necessary for collector currents in excess of 0.5 ampere. A polarity reversing switch, not shown in Figure 2, may be provided to facilitate proper biasing of either p-n-p or n-p-n transistors.

The base current is applied to the horizontal input terminal of the oscilloscope while its vertical input is connected to either the collector

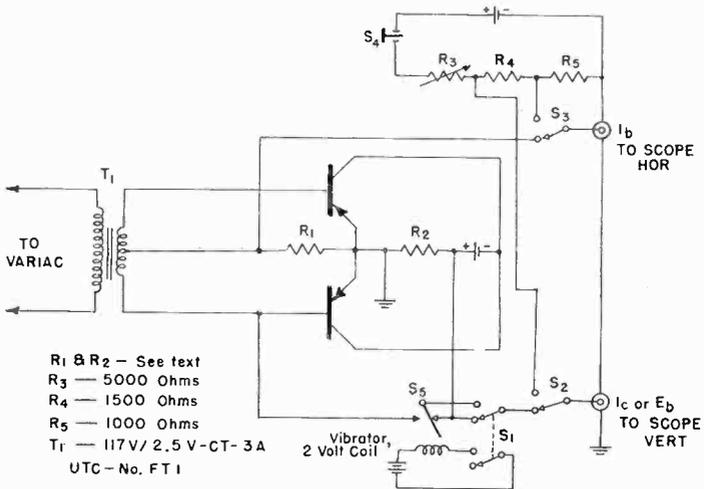


Fig. 2—Circuit for displaying the characteristics of a pair of transistors.

current (I_c) or the base-to-emitter voltage (E_b). If switch S_1 is in the position as shown in Figure 2, the current transfer curves of both transistors will be consecutively displayed on the same quadrant of the screen. Turning switch S_1 into the other position will cause the vibrator to operate with the result that now the input voltage as well as the collector current will be displayed consecutively on the vertical axis of the oscilloscope. Therefore, two current transfer curves and one input characteristic will appear on the screen. The second input characteristic falls in the fourth quadrant and could be displayed if required.

CALIBRATION OF THE OSCILLOSCOPE

A calibration circuit shown in Figure 2 consists of the voltage divider R_3 , R_4 , and R_5 , fed from a 1.5-volt battery. With switches S_2 and S_3 in the calibrating position and push-button switch S_4 closed,

0.5 volt will be applied to the vertical input of the oscilloscope and 0.2 volt to its horizontal input.

The oscilloscope is now calibrated so that equivalent deflections resulting from signals from the test circuit will represent the following currents and voltages:

Horizontal Axis		Vertical Axis	
R_1 ohms	I_b ma	R_2 ohms	I_c ma
10	20	1	500
100	2	10	50

The vertical-axis calibration for E_b is 0.5 volt.

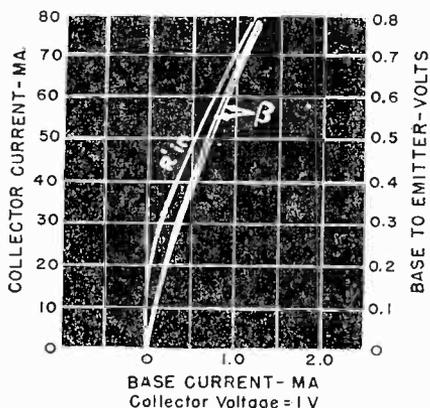


Fig. 3—Current gain and input characteristics of a pair of 2N109's.

SAMPLE CURVES AND EXAMPLE OF EVALUATION

Figure 3 shows a photograph of the transfer and the input characteristics taken on a developmental transistor similar to the 2N109. A pair was chosen whose current transfer curves are not quite matched so that the purpose of this test method can better be made evident. As explained earlier, only one input characteristic appears on the picture. This is not a handicap since, if the transistors are matched, all the data needed for the design of the input circuit of the amplifier is contained in the photograph.

The following procedure may be used to determine the operating

¹ K. E. Loofbourrow, "Class-B Operation of Transistors," *Electronic Design*, pp. 28-31, July, and pp. 34-39, August, 1955.

characteristics of a class-B push-pull output amplifier.¹ With the transistors biased to cutoff and a sinusoidal input voltage applied to their bases, the following values are found:

Average total power output and the peak collector current are related by

$$P_0 = \frac{\text{Peak Power}}{2} = \frac{(E_{CC} - 1)I_C'}{2}, \quad (2)$$

$$\text{Average current, each unit } I_{AVE} = \frac{I_C'}{\pi} = 0.318 I_C', \quad (3)$$

$$\text{Battery power } P_B = 2(I_{AVE} \times E_{CC}), \quad (4)$$

$$\text{Percent efficiency } \% E_{FF} = (P_0/P_B) \times 100, \quad (5)$$

$$\text{Collector dissipation, each unit } P_D = \frac{P_B - P_0}{2}, \quad (6)$$

$$\text{Collector-to-collector load resistance } R_{CC} = 4 R_L, \quad (7)$$

$$\text{Base-to-base resistance } R_{ii} = 4 R'_{IN}. \quad (8)$$

It will be assumed that the collector supply is 9 volts and that an output power of 250 milliwatts is desired. From Equation (2)

$$I_C' = \frac{2 \times 250}{9 - 1} = 63 \text{ milliamperes.}$$

For this value of peak current, a base current, I_B' , of 1.0 milliamperere and a corresponding base-to-emitter voltage, E_B' , of 0.7 volt is required according to Figure 3. Then

$$\beta = 63/1.0 = 63,$$

$$R'_{IN} = 0.7/1.0 \times 10^{-3} = 700 \text{ ohms for each unit,}$$

$$I_{AVE} = 0.318 \times 63 \times 10^{-3} = 20 \text{ milliamperes for each unit,}$$

$$P_B = 2(20 \times 9) = 360 \text{ milliwatts,}$$

$$\% E_{FF} = (250/360) \times 100 = 69.5,$$

$$P_D = (360-250)/2 = 55 \text{ milliwatts for each unit.}$$

If it is assumed that the transistor is rated for a dissipation of 55 milliwatts, the output load, input resistance, and power gain can be

found. From Equation (1), $R_L = \frac{9 - 1}{63 \times 10^{-3}} = 127$ ohms, and according to Equation (7), $R_{CC} = 4 \times 127 = 508$ ohms. $R_{ii} = 4 \times 700 = 2800$ ohms (Equation (8)) at maximum signal voltage.

The power gain of a class-B push-pull amplifier is proportional to the square of the current amplification factor of the transistors. Since the current amplification factor decreases somewhat with increase in collector current, the power gain is commonly calculated at peak output power and is

$$P_G' = 10 \log_{10} \frac{P_0'}{P_{IN}'} = 10 \log_{10} \frac{(E_{CC} - 1) I_C'}{E_B' I_B'} \text{ decibels.} \quad (9)$$

For this example the power gain is 28.5 decibels.

The principles of the described test method can be extended to the evaluation of a pair of transistors in a complementary symmetry circuit. A slight modification of the circuit of Figure 2 and the addition of a second collector voltage source are required. However, since currents and voltages of p-n-p and n-p-n transistors are of opposite polarities, the current transfer curve of the p-n-p transistor would appear on the third quadrant. Such a display makes the evaluation of the match between the two units difficult if not impossible. It becomes necessary, therefore, to modify the sweep circuit so that a simultaneous display of both current transfer curves on the same quadrant can be obtained. An arrangement which accomplishes this is illustrated schematically in Figure 4.

The voltages taken from the current sampling resistors R_1 and R_2 , at points A and B, respectively, are applied to the inputs A' and B' of amplifier-rectifier units before they are fed into the proper input terminals of the oscilloscope. The rectification action insures that the oscilloscope will be swept in one direction regardless of the polarity of the voltage developed across the current sampling resistors of the sweep circuit.

Since the input resistance of the transistor amplifier is low, of the order of 1,000 ohms, the use of 10- and 1-ohm resistors for R_1 and R_2 , respectively, is recommended for the entire range of currents to be measured. The remaining circuit is straightforward and needs no detailed explanation.

The calibration of the oscilloscope can be accomplished by the application of a half sine wave voltage of known amplitude to terminals A' and B' .

CONCLUSIONS

A method has been developed to determine both the d-c current amplification factor and the input d-c resistance of class-B audio output transistors without causing thermal effects. A pair of output transistors to be tested is driven in a push-pull circuit by a 60-cycle sine wave

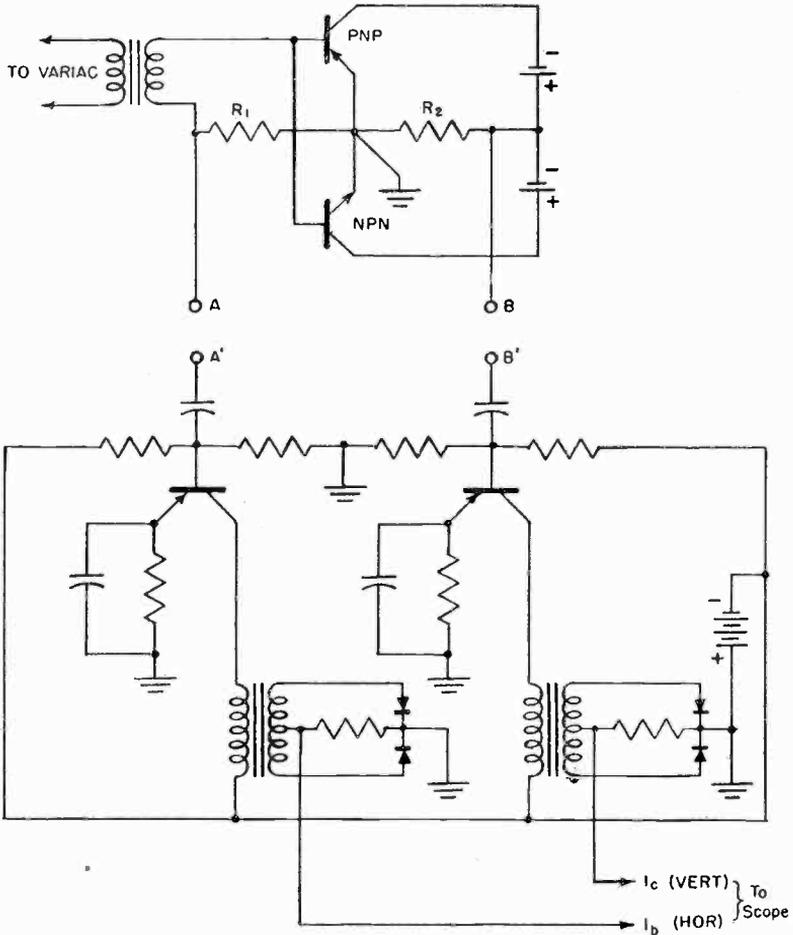


Fig. 4—Circuit for display, on the same quadrant, of curves of a pair of p-n-p/n-p-n transistors.

input signal. Input and output currents, as well as the input voltage, are displayed on an oscilloscope where they can be conveniently measured. A simultaneous display of the current transfer curve of both transistors together with the input characteristic of one of them is accomplished. Either p-n-p or n-p-n transistors can be used and the method can also be applied to the measurement of p-n-p and n-p-n transistors in a complementary symmetry circuit.

The described test method has given excellent results not only in the evaluation of characteristic data of output transistors but also in the matching of transistor pairs.

The test equipment is inexpensive and easy to build and operate.

TEMPERATURE EFFECTS IN CIRCUITS USING JUNCTION TRANSISTORS*

BY

H. C. LIN AND A. A. BARCO

RCA Laboratories,
Princeton, N. J.

Summary—The effects of varying ambient temperature upon the performance of transistor circuits have been evident since the beginning of the transistor art. However, the nature of these effects, their causes, and the methods by which they may be minimized have not generally been appreciated by the circuit-design engineer. This paper is concerned with these matters, principally as they relate to germanium alloy-junction transistors and the circuits in which they are used.

The behavior of the transistor with respect to temperature is most adversely affected by changes in the saturation current and the d-c input conductance. The effects of variations in these parameters upon the transfer characteristics are shown experimentally and an appendix indicates their basis in theory.

When transistors are used in some of the more familiar and frequently employed circuits, these two causes may affect performance by shifting the d-c operating point, i.e., the bias. Such circuits are inherently sensitive to temperature and some may be severely affected by relatively small temperature changes.

For many circuit applications, from audio frequencies up through broadcast frequencies, satisfactory operation is obtained if the quiescent emitter current is held constant with changing temperature. Several means for achieving this condition, including circuits employing d-c degeneration or feedback, are described. Such circuits are inherently stable with temperature and may be made to operate satisfactorily over a wide range of temperature.

Circuits such as those of detectors or class B amplifiers present special biasing problems which may be solved by means of techniques which use diodes, transistors, or thermistors to provide bias compensation.

An analytical approach to evaluating the relative stability of various circuits is indicated in an appendix.

When due circuit consideration is given to the effects both of saturation current changes and d-c input conductance changes, germanium alloy-junction transistor circuits operate satisfactorily over a wide range of ambient temperature.

EFFECTS AND CAUSES

A BASIC d-c circuit of a transistor stage is shown in Figure 1. R_B is the resistance in the base circuit, R_E is the resistance in the emitter circuit and R_C is the resistance in the collector circuit. The base-to-emitter bias is furnished by V_1 , and collector po-

* Part of this material will be incorporated in a thesis for presentation to the E. E. Dept., Polytechnic Institute of Brooklyn.

tential is supplied by V_2 . The manner in which the d-c collector current, i.e., the operating point, is shifted by temperature effects depends greatly upon the values of the various resistors.

In the following discussions attention is directed principally toward the problems of d-c stability. It should be noted that in considering specific applications of various circuits to working apparatus, provision must be made for coupling signals into and out of the circuit, and suitable bypassing must be provided. The values of blocking capacitors and of emitter bypass capacitors have no effect on the d-c stability, but should be commensurate with the signal frequency to be employed.

Consider the following:

Case (a): $R_B = 0$, $R_E = 0$, $R_C \neq 0$.

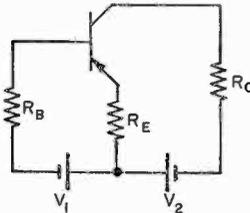


Fig. 1—D-C circuit of a transistor stage.

The value of V_1 is chosen so as to cause a certain desired collector current I_C to flow at a particular temperature, e.g., room temperature. The value of collector current is substantially independent of both R_C and collector supply voltage if the collector potential always exceeds a few tenths of a volt, but is less than the collector breakdown voltage, i.e., the operating point is above the knee of the collector characteristic.

Curve (a) of Figure 2 shows the variation of d-c collector current of a 2N34 type transistor for the resistor values defined as case (a) as temperature is varied. Note that the transistor is practically cut off at low temperature. At high temperature, the collector current becomes exceedingly high. When the current becomes too high the transistor may "run away." A runaway condition is due to thermal regeneration in the transistor.¹

The variation of collector current with temperature for this circuit arrangement has been found to be principally caused by two effects: (1) one which acts throughout the entire temperature range, and (2) a second which is substantial only at the higher temperatures.

¹ L. D. Armstrong and D. A. Jenny, "Behavior of Germanium-Junction Transistors at Elevated Temperatures and Power-Transistor Design," *Proc. I.R.E.*, Vol. 42, p. 527, March, 1954.

Effect (1) exists alone at lower temperatures. It may be demonstrated by plotting the transfer characteristic of the transistor as shown in Figure 3. Note that the effect of increasing temperature is to move the transfer characteristic to the left. This is equivalent to shifting the bias scale to the right, so that the shift may be expressed

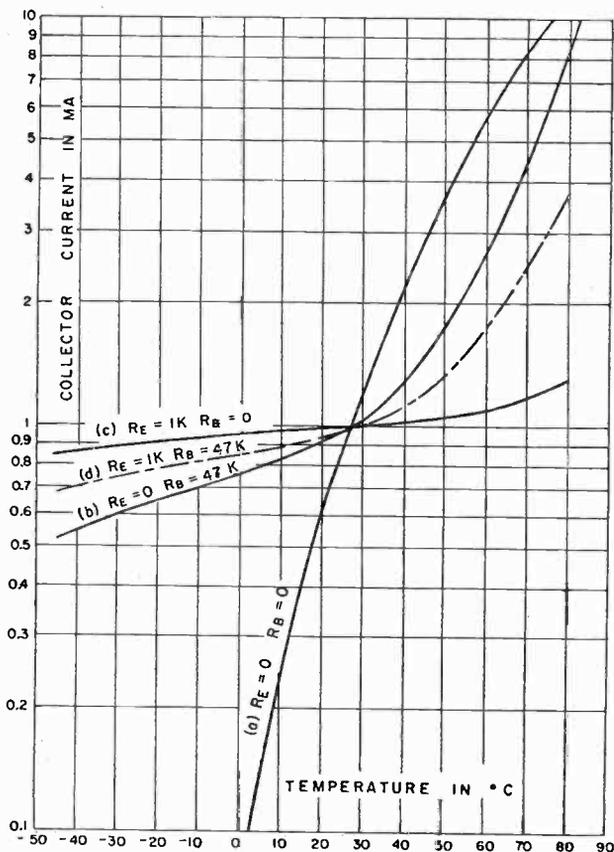


Fig. 2—Collector current versus temperature for (a) $R_E=0$; $R_B=0$; (b) $R_E=0$, $R_B \neq 0$; (c) $R_E \neq 0$, $R_B=0$.

in terms of equivalent bias volts. The average rate of shift for this transistor is -2.5 millivolts per degree centigrade. This shift is caused by the variation of the d-c input conductance of the p-n emitter junction^{2, 3} of the transistor with temperature, and is an inherent

² W. Shockley, M. Sparks, and G. K. Teal, "p-n Junction Transistors," *Phys. Rev.*, Vol. 83, p. 151, July, 1951.

³ L. J. Giacoletto, "Power Transistors for Audio Output Circuits," *Electronics*, Vol. 27, p. 144, January, 1954.

characteristic. The rate of shift is predominantly a function of the energy gap of the semiconductor material, and may be determined from theoretical considerations as indicated in Appendix I. It will be substantially the same for all germanium transistors. It will be slightly higher for silicon or other higher-energy-gap materials.

Some insight as regards the influence of this effect upon circuit behavior may be derived from a consideration of a similar circumstance in the vacuum-tube art. There, shifting of the bias scale occurs due to what are commonly termed "contact potential" effects. The tube acts as though there were a "built-in" bias which is affected by the filament temperature (as well as by electrode materials, etc.). This situation, along with variations from tube to tube, has been in part

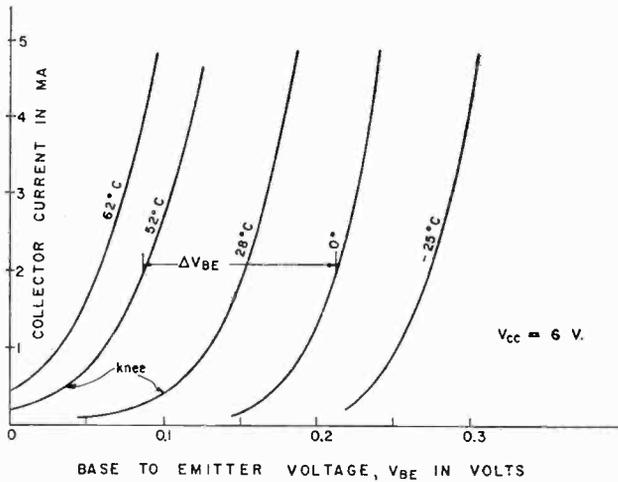


Fig. 3—Transistor characteristics of a 2N34 p-n-p transistor. $\Delta V_{BE}/\Delta T = -2.5$ millivolts per degree centigrade.

responsible for the generally accepted usage of self-bias resistors in cathode circuits, especially for high g_m tubes. Similarly, the use of a resistor (generally bypassed) in the emitter circuit of a transistor will be shown to be beneficial.

The shifting characteristic requires that if a transistor is to be operated at constant collector current in spite of temperature changes (in the lower temperature range where effect (1) is dominant), the circuit must be arranged so that either (a) the emitter current must be held constant while V_{BE} is free to adjust itself as necessary, or (b) that V_{BE} is forced to follow the -2.5 millivolts per degree centigrade rule, approximately. Obviously, these are statements of boundary conditions; other intermediate conditions are possible and are some-

times employed. However, for the sake of simplicity it is convenient to consider these two limiting conditions first. Both are useful; the particular condition (or combination) which is chosen for use in a specific circuit is determined in part by the function to be performed by the transistor. Circuits in which the average (d-c) collector and emitter currents must be permitted to vary with signal, e.g., detectors and class B amplifiers, generally cannot employ constant current biasing. Further, the choice must be determined in part by consideration which must be given to effect (2).

This second temperature effect is important at higher temperatures, but does not exist alone at higher temperatures since effect (1) continues. Effect (2) may be examined by selecting an operating point such that the controllable part of the collector current is zero. The emitter current which is supplied should be zero and the base-to-emitter voltage should be free to adjust itself as necessary. This is accomplished by connecting nothing to the emitter. Of course, this is not a useful arrangement since the transistor can not amplify, but it does permit an examination of effect (2). Under these conditions the variation of collector current with temperature (for the same transistor as used in Figure 3) is found to be as shown in Figure 4. The collector current which flows under these conditions is uncontrolled collector current; i.e., it is not controlled by the input elements of the transistor and hence does not participate in the signal amplification process.

This current is thermally generated in the base region of the transistor and may be shown, in accordance with semi-conductor theory (see Appendix I), to be a function of the materials and geometry.^{2, 3, 4}

This effect, (2), is dominantly influenced by the energy gap of the semiconductor material which is employed, higher-energy-gap materials⁵ giving greatly reduced current. For example, at temperatures of the order of 100°C these currents in germanium transistors are expressed in milliamperes, whereas for silicon transistors⁶ they are expressed in microamperes. It is not easy to grasp intuitively the manner in which such small differences in energy gap (germanium = 0.7 electron volt and silicon = 1.1 electron volts) operate to produce such a striking effect. However, by noting that the effect is a function of absolute temperature, it is possible to evaluate materials in terms of the absolute temperature required to produce a stipulated current.

⁴ W. M. Webster, "Saturation Current in Alloy Junctions," *Proc. I.R.E.*, Vol. 43, pp. 277-280, March, 1955.

⁵ E. R. Johnson and S. M. Christian, "Some Properties of Germanium-Silicon Alloys," *Phys. Rev.*, Vol. 95, p. 560, July, 1954.

⁶ H. Nelson, "A Silicon N-P-N Junction Transistor by the Alloy Process," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 172-181.

Thus, for example, if a germanium transistor has a saturation current of 1 milliampere at 100°C , a comparable silicon transistor should have that same current at roughly $(100 + 273) 1.1/0.7 - 273 = 312^{\circ}\text{C}$.

Since this uncontrollable collector current saturates with respect to collector voltage, i.e., it is substantially unaffected by collector voltage when the collector voltage is greater than a few tenths of a volt, it is often referred to as a saturation current. The effect of this current on the transfer characteristic, Figure 3, is to shift the curves upward

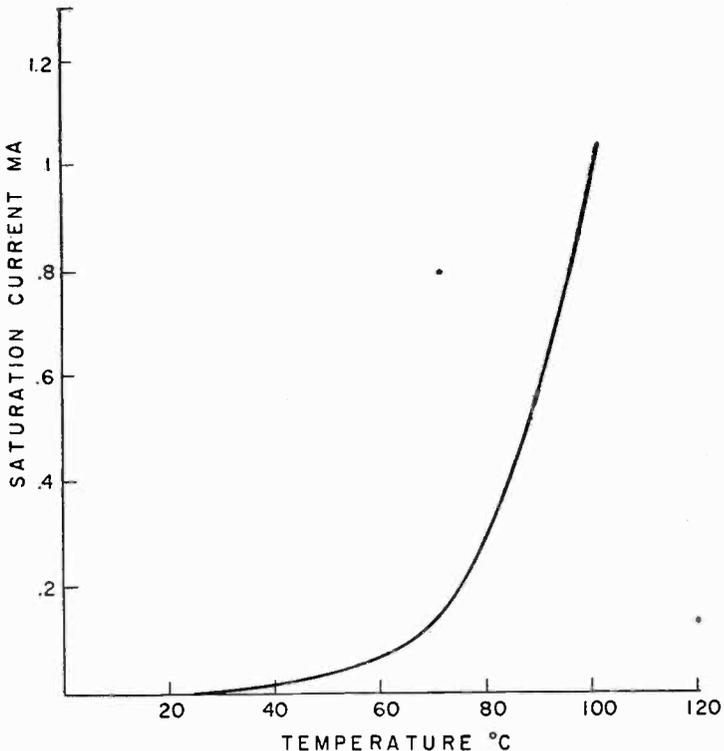


Fig. 4—Saturation current versus temperature.

by an amount equal to the value of this uncontrolled saturation current at the temperature corresponding to each curve. Note that from Figure 4 the value at 62°C is of the order of 0.1 milliampere. Thus, the 62°C curve of Figure 3 is shifted upward by this amount. The effect becomes progressively greater as the temperature is increased. The lower temperature curves are substantially unaffected, except at very low currents. If this were the limit of this effect it would not be especially serious. However, as will be apparent subsequently, this

upward-shift effect may be magnified by the current gain of the transistor for certain circuit configurations, so that it may be increased in importance by several orders of magnitude.

The effect of increased saturation current at increased temperature is most pronounced in circuits which have appreciable resistance in the base circuit. In such circuits, any increase in saturation current is amplified by the transistor to produce a further increase in the collector current. In the discussion thus far, R_B of Figure 1 has been considered as an external resistor connected to the base of a somewhat idealized transistor. In reality, the transistor contains internal base-lead resistance; the value of this internal base-lead resistance must be considered as a lower practical limit for the R_B of Figure 1.

Here again, consideration of a vacuum-tube effect may provide insight which is useful in circuit application work. It may be noted that, in a very crude manner, there is some similarity between the effects of grid current in a vacuum tube and saturation current in a transistor. In the case of the vacuum tube, grid-current effects may be serious in circuits where the value of the d-c grid-return-path resistance is high because the grid current flow through this resistor produces an IR drop which modifies the net bias on the grid. This has led to the familiar specification for maximum permissible grid circuit resistance for the various types of tubes. Similarly, transistors require that due consideration be given to the value of resistance employed in the base circuit.

BASIC CIRCUIT CONSIDERATIONS

To provide further understanding of the manner in which circuit arrangement affects behavior as a function of temperature, consider the circuit of Figure 1 for the following case:

$$\text{Case (b): } R_B \neq 0, \quad R_E = 0, \quad R_C \cong 0.$$

When a high resistance is used in the base circuit, the net current flowing into the base lead remains nearly constant as temperature is varied, being principally determined by V_1/R_B . The net base lead current is the difference between the intrinsic base current and the saturation current. The intrinsic base current is amplified by the current amplification factor, α_{CB} , to produce the intrinsic collector current. The net collector current is the sum of the intrinsic collector current and the saturation current. At low temperatures, the saturation current is negligible and the net base current is equal to the intrinsic base current. As the collector-to-base current amplification factor does not

vary appreciably with temperature, the variation of collector current with temperature is slow. However, at higher temperatures the saturation current is no longer negligible, but increases as shown in Figure 4. The intrinsic base current must necessarily increase the same amount as I_{BS} so as to maintain a constant net base current. This results in a net collector current increase which is approximately α_{CB} times the increase in the saturation current. Curve (b) in Figure 2 shows the variation of collector current of a 2N34 transistor with temperature when $R_B = 47,000$ ohms. Note that the current changes slowly below room temperature but increases rapidly as temperature is raised. For smaller values of R_B the current variation would lie somewhere between the $R_B = 0$ and $R_B = 47,000$ ohms curves.

The upturn in the curve occurs when the saturation current becomes comparable to the intrinsic base current. Thus, if a lower collector current had been chosen as the initial operating point at room temperature, the upturn would have occurred at a lower temperature. Similarly, a higher initial current would delay the upturn until a higher temperature had been reached.

For this case, the operating point is dependent upon the current gain of the particular transistor which is employed. Thus, the substitution of different transistors having a spread of current gains or a spread of saturation currents might require readjustment of R_B or V_1 to achieve uniformity of operating point.

Consider next the following case:

$$\text{Case (c): } R_B = 0, \quad R_E \neq 0, \quad R_C \cong 0.$$

When R_E is connected in series with the emitter, the bias V_{BE} appearing between the base and the emitter terminals is the difference between V_1 and the voltage drop $I_E R_E$. If R_E is high, such that $I_E R_E$ is much greater than V_{BE} , I_E is practically equal to V_1/R_E . The higher the value of R_E , the greater is the ratio of the voltage drop across R_E to the voltage between the base and the emitter, and the greater is the ability of the circuit to hold the emitter current constant. However, for the same emitter current, a high R_E requires a high bias voltage V_1 . This may be a disadvantage when operating from a fixed value of V_2 because as V_1 becomes higher the available collector-to-base voltage becomes smaller.

Although this circuit arrangement operates directly to maintain constant emitter current, its net effect is to maintain substantially constant collector current also; i.e., the controllable part of the collector current is substantially equal to the emitter current and the net col-

lector current is greater only by an amount equal to the saturation current. The variation of collector current for $R_E = 1,000$ ohms using a typical 2N34 transistor is as shown in Figure 2c. Note that the variation in collector current is less than 20 per cent of room temperature value throughout the range of temperature plotted. The gradual upturn at high temperature is due to the increase in base saturation current. This arrangement is, in general, very stable and the operating point is substantially unaffected by the current gain of the transistor.

Consider now a combination of the two preceding cases:

$$\text{Case (d): } R_E \neq 0, \quad R_B \neq 0, \quad R_C \approx 0.$$

Although case (c) is more stable than case (b), there are instances such as resistance-coupled stages or single-battery operated stages where external base resistance can not be avoided. This, then, is an intermediate case between (b) and (c). Shea⁷ has shown that the temperature stability can be improved by increasing the ratio of the external emitter resistance, R_E , to the external base resistance, R_B . In circuits with single-battery operation, such as the one shown in Figure 5, the d-c base-to-emitter bias is derived from a voltage divider, R_2 and R_3 . In this case the external base resistance is equal to the parallel resistance of R_2 and R_3 . This resistance should be much greater than the input impedance at signal frequencies lest some useful signal power be absorbed. The performance of this circuit is shown in Figure 2d.

Cases (c) and (d), which employ a resistor in the emitter circuit, constitute a form of d-c degeneration or feedback which is analogous to the use of a cathode bias resistor in vacuum-tube circuits. Other feedback arrangements are possible in which d-c information from the collector circuit is fed back into the input circuit. One such arrangement is shown in Figure 6. A resistor, R_F , is connected between the base and the collector. A portion of the collector current is fed back to the base through R_F . Since the feedback is degenerative it reduces the changes in d-c collector current with temperature. If the R_C used in Figure 6 is the same as the R_E used in Figure 1, then, at room or reduced temperature where the collector current is nearly equal to the emitter current, the variation in voltage drop across the resistors with changes in temperature is nearly the same in both cases. Since the voltage variation is all fed back between the base and emitter in case (d), Figure 1, whereas only a fraction of the voltage variation is fed

⁷ R. F. Shea, *Principles of Transistor Circuits*, John Wiley & Sons, Inc., New York, N. Y., 1953.

back in case (e), Figure 6, the former is usually more effective than the latter. The performance of these circuits is shown in curves (d) and (e) in Figure 7. Note that below 30°C , there is a marked advantage in using (d). However at elevated temperature where I_C is appreciably greater than I_B , due to saturation current, the variation in voltage drop in R_C with temperature is considerably greater than that in R_B , thus the voltage fed back to the base by means of R_C may become comparable with that fed back by using R_B . As shown in Figure 7, at temperatures above 30°C curve (d) and curve (e) merge together. It is possible to further improve the temperature stability of case (d) by using d-c collector feedback in addition to the emitter degeneration. This

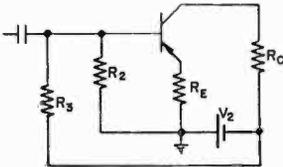


Fig. 5—Transistor stage with resistance in the base and emitter circuits.

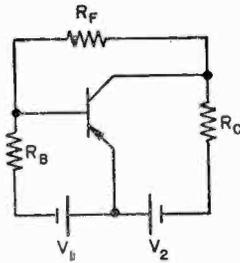


Fig. 6—Transistor stage with d-c feedback.

arrangement is shown in Figure 8. The performance is shown in curve (f) Figure 7.

An analytical comparison of the stability with respect to temperature of various cases is given in Appendix II.

Many arrangements are possible which provide for d-c feedback over a number of stages,^{7, 8} as contrasted with d-c feedback "internally," i.e., within a single stage. One example of such an arrangement has been described.⁸

BIAS-COMPENSATING METHODS

Compensation can be applied in the form of a temperature-dependent base-to-emitter bias. As temperature is increased the bias for germanium transistors should decrease approximately -2.5 millivolts per degree centigrade in order to hold the emitter current constant.

⁸ D. D. Holmes, L. A. Freedman, and T. O. Stanley, "A Developmental Pocket-Size Broadcast Receiver Employing Transistors," *Proc. I.R.E.*, Vol. 43, p. 662, June, 1955; also *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 477-496.

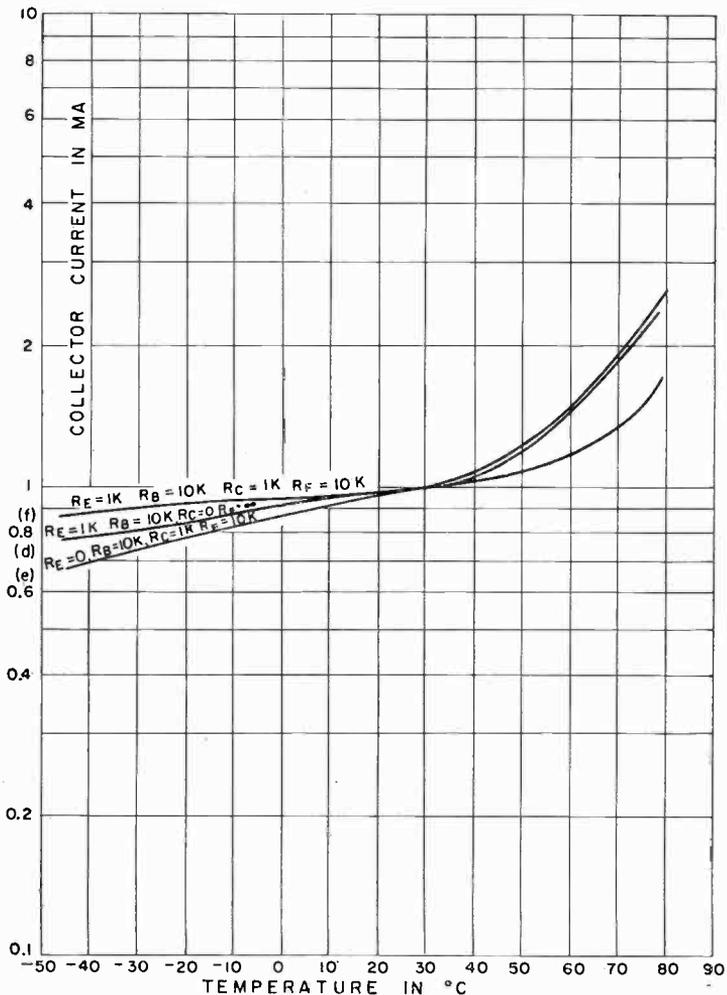


Fig. 7—Collector current versus temperature for a transistor stage with d-c feedback.

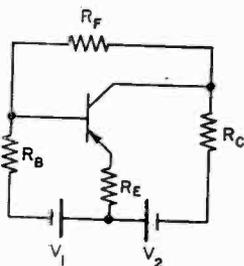


Fig. 8—Composite circuit employing emitter resistance and feedback.

For example, in a transistor detector circuit the operating point might be chosen at the point of maximum curvature of the transfer characteristic, as this is the point which yields maximum small-signal detection. However, the transistor can not be constant-current biased because the current is expected to change with changes in the level of the incoming signal. Therefore, the bias must be set to the proper operation point, and this point must remain fixed despite signal changes. This optimum bias has been found by measurement to be about 0.12 volt at 25°C for the transistors of the type described by Mueller and Pankove.⁹

Figure 9 shows how the bias of such a transistor must be adjusted to provide maximum detection sensitivity with varying temperature. Note that the rate of change of bias with temperature is approximately -2.5 millivolts per degree centigrade. The collector current for optimum bias remains substantially constant at low and room temperatures; it increases slightly at elevated temperature, due to flow of saturation current.

A number of different methods can be used for deriving a compensating bias which varies in the desired manner. Several bias-compensating methods will be shown as applied to detectors. However, since the function of the bias is to hold the emitter current constant with varying temperature, methods used for stabilizing detector circuits may be applied also to other modes of operation.

Single-Diode Compensation

In the single-diode compensation method,¹⁰ the base-to-emitter bias is derived from a junction diode which is constructed of the same kind of material as the transistor to be compensated and which is designed for low base-lead resistance so that the current-voltage relationship in the diode is controlled principally by the junction. The rate of change of voltage across the diode as a function of temperature for constant-current operation is the same as the change of base-to-emitter bias required for constant emitter current. As shown in Figure 10, the diode is fed a constant current through the high resistor R_1 . The diode has low forward resistance in comparison with the base input resistance of the transistor so that the voltage across the diode is relatively unaffected by the base current of the transistor. The variation of voltage across the diode is then suitable for compensating the effect of variation of the d-c input conductance of the transistor with tempera-

⁹ C. W. Mueller and J. I. Pankove, "A P-N-P Triode Alloy Junction Transistor for Radio Frequency Amplification," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 189-201.

¹⁰ L. E. Barton, "An Experimental Transistor Personal Broadcast Receiver," *Proc. I.R.E.*, Vol. 42, p. 1062, July, 1954.

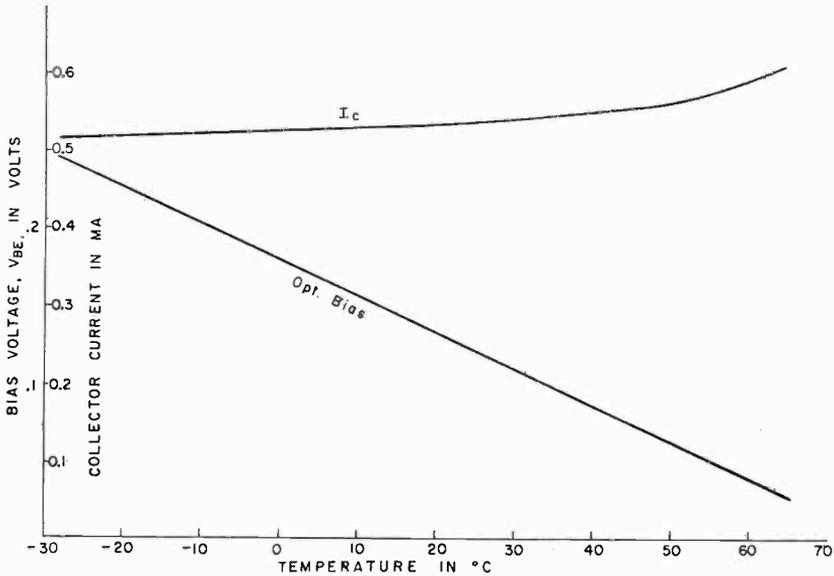


Fig. 9—Optimum base-to-emitter bias and collector current of a grounded emitter transistor detector.

ture. However, at elevated temperature the saturation current of a transistor becomes appreciable. Thus the change in bias caused by the saturation current flowing in the internal base resistance of the transistor is not compensated by the diode. This is evidenced experimentally in Figure 11. At low temperature, where the effect of base saturation current is negligible, the collector current remains practically constant as desired, but at elevated temperature, where saturation current becomes large, the collector current departs from the desired value at a rapid rate due to the flow of saturation current.

Two-Diode Compensation

In order to compensate for the bias change due to base saturation current, a second diode may be incorporated into the single-diode

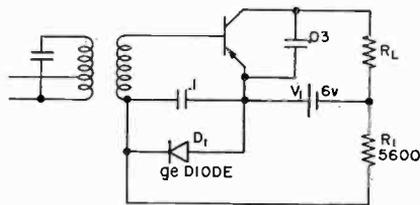


Fig. 10—Transistor detector with single-diode compensation.

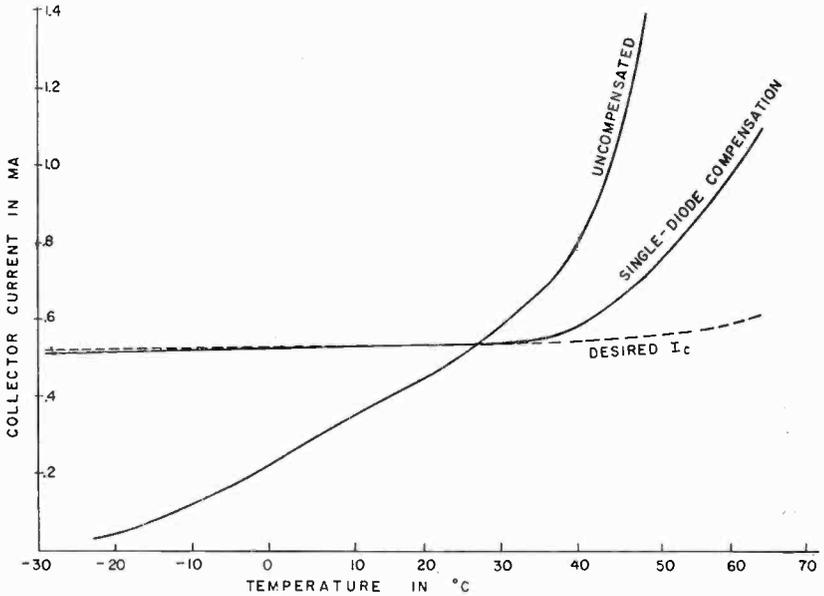


Fig. 11—Quiescent collector current of a transistor detector with and without single-diode compensation.

circuit. Figure 12 shows one way to accomplish the purpose, the added part being drawn in heavy line. D_1 , R_1 , and V_1 function as described under Single-Diode Compensation. D_2 is poled in the reverse direction to a voltage V_2 . At low temperature, D_2 is practically an open circuit. As temperature is increased, the current flowing in the mesh including D_2 , V_2 , and R_2 increases, because of increased saturation current in D_2 , causing a voltage drop across R_2 which compensates the effect due to saturation current in the transistor. The current through D_2 should be larger than I_{BS} . R_2 is adjusted to obtain optimum compensation. The small collector current variation obtained by this method is shown in Figure 13.

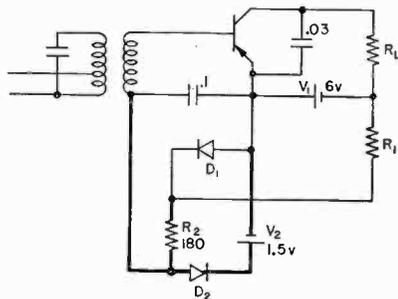


Fig. 12—Transistor detector with double-diode compensation.

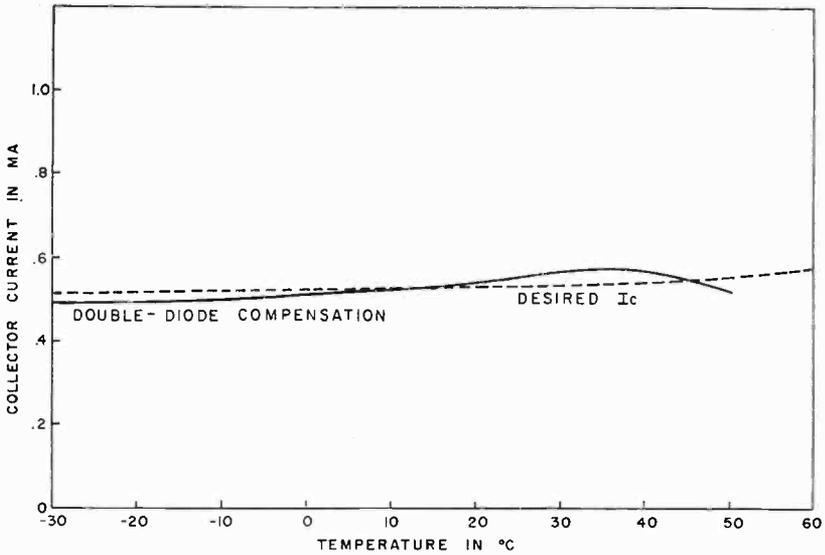


Fig. 13—Quiescent collector current of a transistor detector with double-diode compensation.

Transistor Compensation

In this method, the base-to-emitter voltage of a transistor, operated at constant emitter current, serves as a bias source for the transistors which are to be compensated.

In the circuit shown in Figure 14, the p-n-p transistor detector derives its base bias from the base-to-emitter voltage of n-p-n transistor amplifier which is constant-emitter-current biased by means of a high-

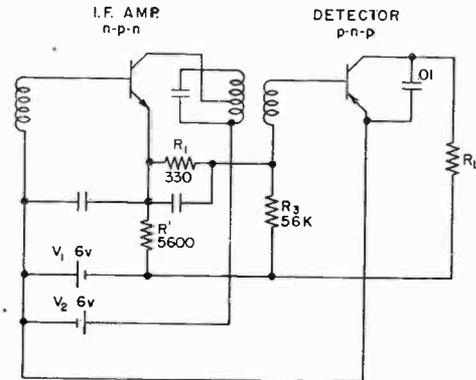


Fig. 14—Transistor detector with compensation derived from another transistor.

voltage source V_1 and a high resistance R' . Because the n-p-n transistor which was used in the test had a higher $I_{BS}r_{BB}$ drop in its lead than the p-n-p transistor, R_1 was inserted in series with the base lead of the p-n-p transistor to equalize the voltage drop due to saturation current. To bring the detector to its proper operating point, i.e., the knee of the transfer characteristic, a large resistor, R_3 , is connected to the base of the detector and fed from a high-potential point. All

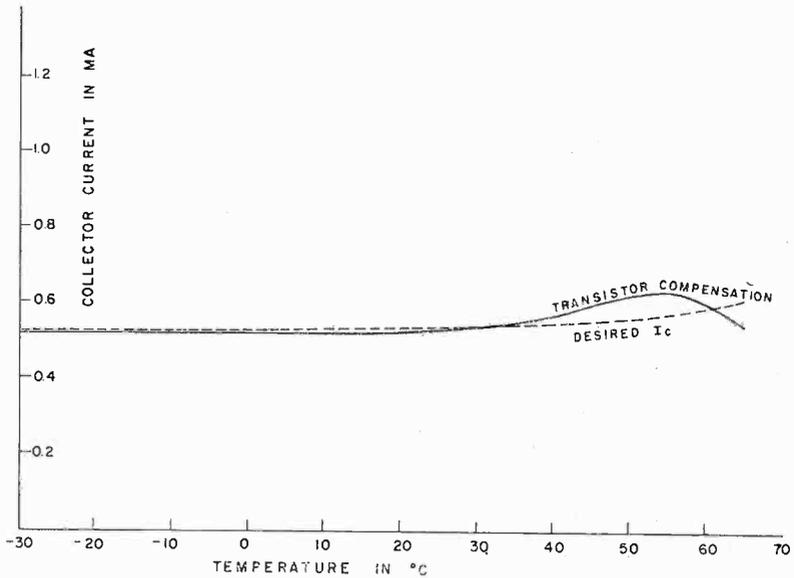


Fig. 15—Quiescent collector current of a detector with transistor compensation.

the resistors mentioned are bypassed to avoid signal degeneration. This compensation maintains the collector current, relatively constant over a wide temperature range, as shown in Figure 15.

Thermistor Compensation

Thermistors can also be used to derive the desired temperature-dependent bias.¹¹ Because commercially available thermistors have a higher temperature coefficient than germanium, they should be used in conjunction with resistors for this purpose. A circuit using thermistor compensation is shown in Figure 16. The degree of compensation is shown in Figure 17; note that "tracking" occurs at only three points.

¹¹ L. A. Freedman, T. O. Stanley, and D. D. Holmes, "An Experimental Automobile Receiver Employing Transistors," *Proc. I.R.E.*, Vol. 43, p. 671, June, 1955; also *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 458-476.

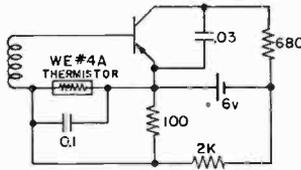


Fig. 16—Transistor detector with thermistor compensation.

APPARATUS-DESIGN ASPECTS OF BIAS COMPENSATION

In the present treatment only the effects of temperature on the transistor and compensating elements, if any, are considered. All the experiments were performed with only the transistor and the compensating elements heated or refrigerated. Other components associated with the transistor circuits remained at room temperature. Actually, most components such as resistors, capacitors, inductors, batteries, etc. have small temperature coefficients. Whereas they are insignificant in most instances, their importance may not be ignored in certain applications.

In applying any of the compensation methods described, it should be remembered that the temperature of the transistor to be compensated may be appreciably higher than the ambient temperature, due to

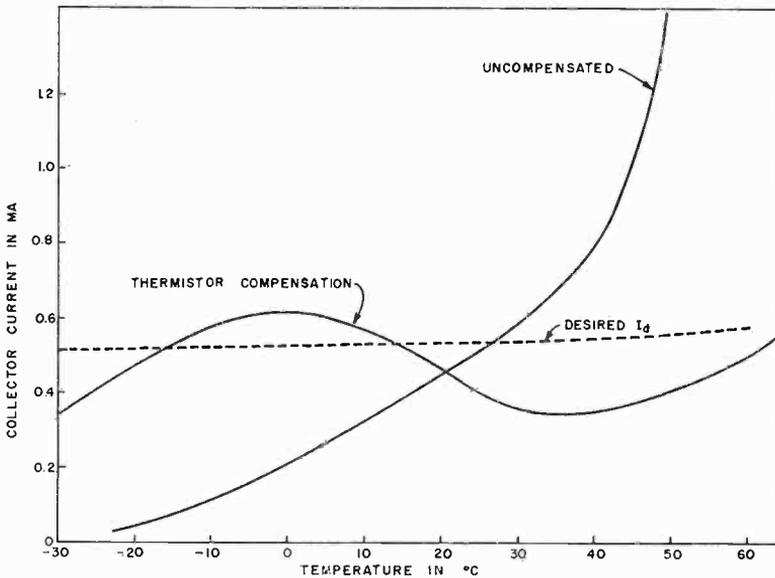


Fig. 17—Quiescent collector current of a detector with thermistor compensation.

internal heating. It is therefore important that the compensating device be thermally coupled as tightly as possible to the transistor to be compensated. Since the compensating element in the biasing circuit usually has one terminal connected electrically to either the emitter or the base of the transistor to be compensated, it is possible to provide good thermal coupling by physically locating the compensating element within the transistor. The single-diode compensating circuit of Figure 10 may be readily adapted to accommodate a transistor which contains a built-in compensating diode. First, the circuit of Figure 10 is rearranged as shown in Figure 18a, where the diode has been moved to

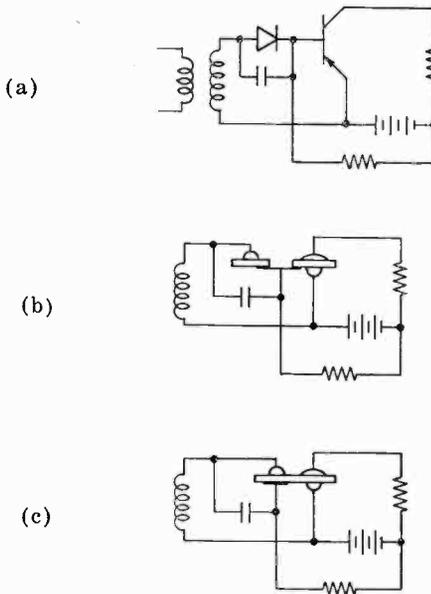


Fig. 18—Evolution of single-diode compensating circuit into a self-compensating transistor arrangement.

the upper side of the input circuit. With the circuit thus arranged, the transistor and diode can be constructed, as in Figure 18b, with an electrical connection between the base of the diode and the base of the transistor. The interconnecting lead may be designed also to provide a path of low thermal resistance to insure that the temperatures of the diode and of the transistor are substantially the same. Figure 18c shows the manner in which such a combination has been constructed as a single unit by alloying an additional dot to a conventional triode transistor structure to provide the required diode.

In a similar manner, the circuit may be rearranged with the com-

compensating diode shifted adjacent to the emitter of the transistor. In this case, an additional base wafer alloyed to the conventional emitter dot may be used to form a single self-compensating unit.

One advantage of locating the compensating element inside the transistor is that compensation is provided with respect to internal heating of the transistor. This is especially important where internal heating effects are severe, such as in high-power applications. In some instances, particularly in low-power apparatus which is intended to meet severe ambient temperature specifications, internal heating effects may be small as compared to the effects of ambient temperature variation. In such cases it is possible to employ a single compensator, responsive to ambient temperature, to provide compensation to a number of transistors, perhaps even an entire apparatus.

In some applications the possible complexity or cost of compensation may be avoided. For example, note that if a push-pull amplifier is biased for class B operation at one temperature, increasing the temperature biases the amplifier first to class AB and then to class A operation under which conditions satisfactory performance is usually achieved, but at the cost of increased power drain and increased dissipation in the transistors. Alternatively, decreasing the temperature biases the amplifier to class C operation, which results in intolerable distortion. Thus, the amplifier bias can be adjusted to provide class B operation at the *lowest* temperature of interest, in which event, if the dissipation rating of the transistor is not exceeded at the upper temperature limit, and if the increased power drain can be tolerated, the amplifier can be used without compensation.

CAPACITIVE EFFECTS

The a-c equivalent circuit of a junction transistor can be represented approximately as a π network,⁹ shown in Figure 19. Both the "diffusion" capacitance, $C_{b'e}$, and the "feedback" capacitance $C_{b'c}$, are of importance in the r-f operation of a transistor.

The diffusion capacitance varies linearly with emitter current.⁹ When I_E is kept constant, it is found by measurement that the change in $C_{b'e}$ with temperature is small.

The feedback capacitance, $C_{b'c}$ consists of two parts. The major part is usually the transition junction capacitance due to potential distribution in the space charge layer. The other part is an effect of space charge widening. If the emitter current varies, then $C_{b'c}$ also varies, since the portion of this capacitance which is due to space charge widening is directly proportional to the collector current, which is nearly equal to the emitter current. The data shown in Figure 20,

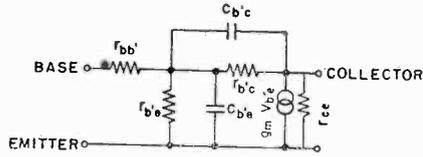


Fig. 19— π -Equivalent circuit of a junction transistor.

taken on apparatus described by Giacoletto,¹² indicates that the feedback capacitance is substantially constant with temperature variation, provided the d-c emitter current is held constant.

A typical circuit in which such capacitive effects might be of interest is shown in Figure 21. The circuit has adjustments for input and output r-f impedances, d-c emitter current, collector voltage, and neutralization. The input and output impedances were conjugate matched and the circuit was neutralized at room temperature for a frequency of 455 kilocycles. As the temperature of the transistor was varied, the emitter current was adjusted to one milliampere at every temperature. The effect on gain is shown in Figure 22. The variation in power gain is about 3 decibels as temperature varies from -30°C

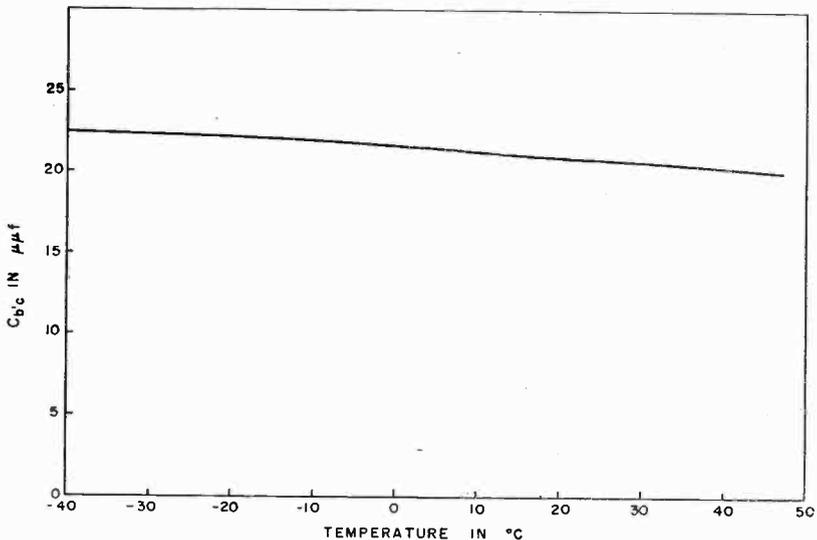


Fig. 20—Collector-to-intrinsic-base capacitance, $C_{b'e}$, versus temperature.

¹² L. J. Giacoletto, "Equipments for Measurement of Junction Transistor Small-Signal Parameters for a Wide Range of Frequencies," *RCA Review*, Vol. XIV, p. 269, June, 1953.

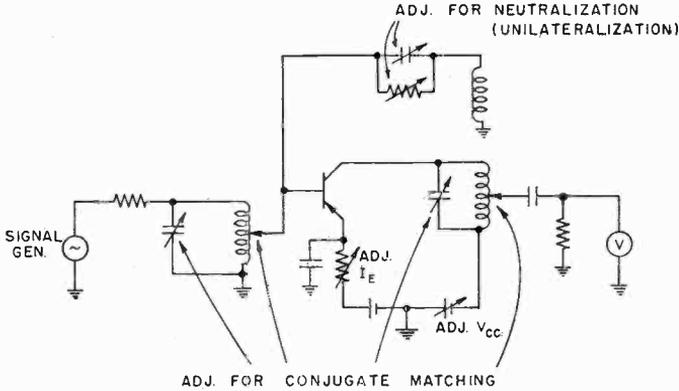


Fig. 21—Circuit for testing transistor at i-f.

to 70°C. The small increase in gain as temperature is reduced is due mainly to a decrease in base-lead resistance. The reduction in gain at the higher temperatures is accompanied by a reduction in $r_{b'c}$. When the output impedance is readjusted for matching at each different temperature, the variation in gain at any temperature is only a fraction of a decibel different from the curve taken above. Between -30°C and 70°C, the output impedance drops from about 25,000 ohms to about 15,000 ohms and the input impedance remains substantially constant at 200 ohms.

By holding the emitter current constant, the change in feedback

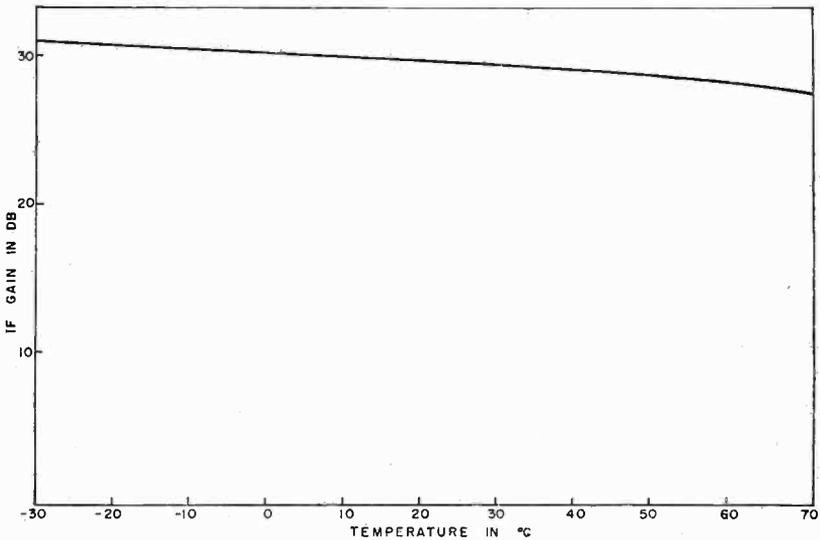


Fig. 22—I-f gain versus temperature.

capacitance is negligible. In r-f circuits where neutralization is used, such as the i-f stages described by Holmes, Stanley, and Freedman,⁸ a constant feedback capacitance means that the neutralizing capacitor need not be changed at different temperatures. In a test similar to that just described, except that neutralization was readjusted at each different temperature, it was found that the value of the neutralizing capacitor remained practically unchanged from -30°C to 70°C . However, if a neutralizing resistance is used in conjunction with the neutralizing capacitance (unilateralization), the proper value of the resistance changes considerably with temperature because it varies as the intrinsic base-to-collector resistance $r_{b,c}$ of the transistor. However, in transistors such as those described by Mueller and Pankove,⁹ the resistive component of the feedback is relatively small, and resistance neutralization need not be employed.

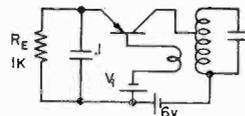


Fig. 23—A transistor oscillator.

Thus from the standpoint of temperature stabilization, constant-emitter-current operation not only stabilizes the d-c characteristics but also gives satisfactory r-f operation for most purposes.

However, constant-emitter-current operation may not be the solution to all temperature problems. For instance, if a very strict requirement on frequency stability is imposed on a transistor oscillator, then an oscillator such as that shown in Figure 23 may not serve the purpose. Note in Figure 24 that the frequency varies by 3 kilocycles over a temperature range of 70°C when no compensation bias is used, and the frequency varies by 1 kilocycle when constant emitter current bias is used. The latter frequency variation appears to be due mainly to the small change in gain, which affects the frequency.

The rate at which the base-to-emitter bias of this transistor oscillator should change in order to maintain a constant frequency, as shown in Figure 24, is -2.89 millivolts per degree centigrade, greater than the -2.5 millivolts per degree centigrade needed for maintaining a constant emitter current. Another method for maintaining the frequency constant as temperature is varied is to change the collector voltage with temperature so as to change the collector junction capacitance. For this oscillator, the collector voltage should be made to increase as temperature is increased. For instance, a bypassed network of thermistor and resistor in series with the collector may serve the purpose.

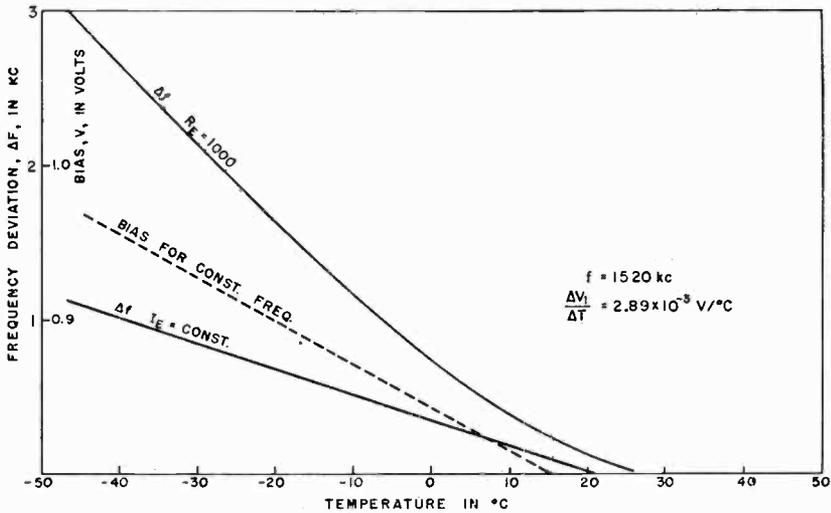


Fig. 24—Transistor oscillator characteristics versus temperature.

APPENDIX I — ANALYSIS OF THE EFFECT OF TEMPERATURE ON D-C TRANSFER CHARACTERISTICS

The d-c characteristics of a junction transistor have been analyzed by Shockley² and extended to include the effect of base-lead resistance by Giacoletto.³ By rearranging Giacoletto's expression for the transfer characteristic and assuming the collector current to be much greater than the collector saturation current,

$$I_C = \frac{\alpha_{CB} g_{BE}}{\Delta} e^{\Lambda} \left[V_{BE} - \left(\frac{I_C}{\alpha_{CB}} - I_{BS} \right) r_{BB'} \right], \quad (1)$$

where

- I_C = d-c collector current,
- I_{BS} = d-c base saturation current,
- g_{BE} = base-to-emitter d-c conductance coefficient,
- $\Delta = q/kT$,
- q = charge of carrier,
- k = Boltzmann's constant,
- T = absolute temperature,

V_{BE} = d-c base-to-emitter voltage,

$r_{BB'}$ = internal base-lead resistance,

α_{OB} = collector-to-base current amplification factor,

$$= \frac{\text{collector-to-emitter d-c conductance coefficient}}{\text{base-to-emitter d-c conductance coefficient}},$$

$$= \frac{I_C - I_{CS}}{I_B + I_{BS}},$$

I_B = d-c base current,

I_{CS} = d-c collector saturation current.

The d-c conductance coefficient, g_{BE} , is highly sensitive to temperature variations. This is essentially due to the change in conductivity of pure semiconductor material, σ_i , and that of impure semiconductor material, σ_b , used in the base of the transistor. The d-c conductance coefficient is proportional to σ_i^2/σ_b . With respect to temperature,

$$\sigma_i^2 \propto e^{-\epsilon_G/kT},$$

$$\sigma_b \propto T^{-3/2},$$

where ϵ_G is the energy gap of the semiconductor material. Then at a temperature ΔT degrees higher than room temperature T_0 , the d-c conductance becomes

$$(g)_{T_0+\Delta T} = (g)_{T_0} e^{(\epsilon_G/kT_0^2 + 3/2T_0)\Delta T} \quad (2)$$

where $(g)_{T_0}$ and $(g)_{T_0+\Delta T}$ are the values of g_{BE} at temperatures T_0 and $T_0 + \Delta T$ respectively. Figure 25 shows the normalized variation with temperature of g_{BE} for n-type germanium; the variation for p-type germanium is essentially the same. Note that at room temperature, g_{BE} increases 10 per cent as temperature is raised one degree centigrade. This rate of change, as seen from Equation (2), is essentially a function only of the energy gap of the semiconductor material, and is therefore approximately the same for all transistors using the same kind of material.

The base saturation current, I_{BS} , which is due to thermally generated carriers in the base, is also very sensitive to temperature. It is proportional to $\sigma_i^2 T \sigma_b$. Thus, if $(I_{BS})_{T_0}$ and $(I_{BS})_{T_0+\Delta T}$ are the base saturation currents at T_0 and $T_0 + \Delta T$ respectively, then

$$(I_{BS})_{T_0+\Delta T} = (I_{BS})_{T_0} e^{(\epsilon_0/kT_0^2 + 5/2T_0)\Delta T} \tag{3}$$

With the saturation current at 25° C taken as 4 microamperes, the theoretical variation of I_{BS} with temperature for a typical germanium

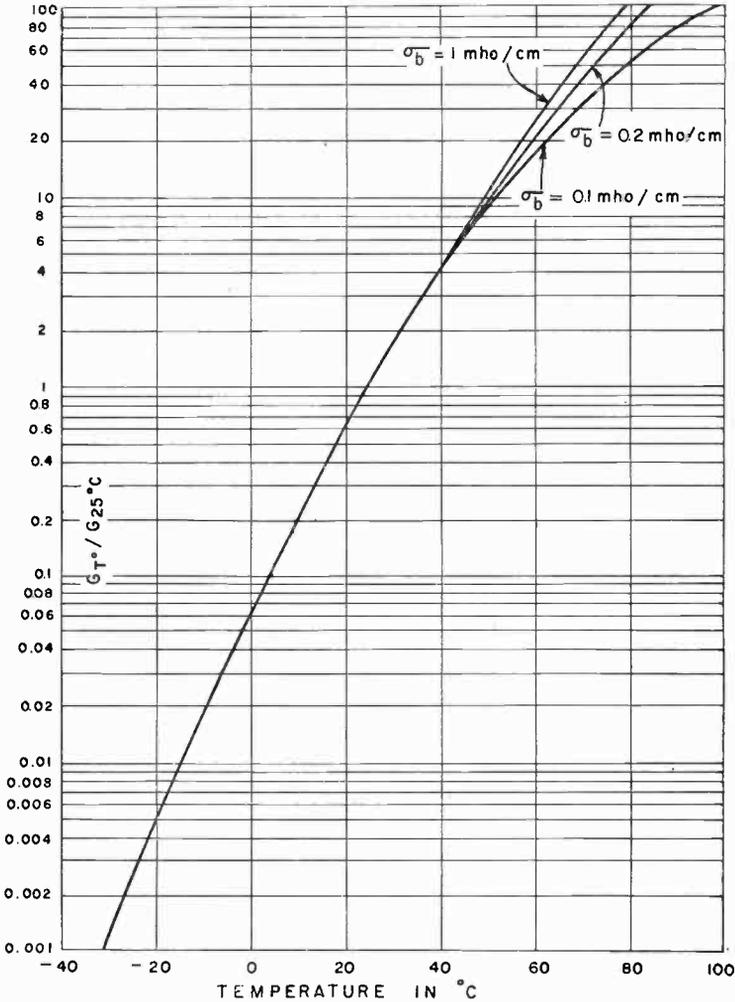


Fig. 25—Variation of d-c conductance coefficient with temperature (n-type germanium).

alloy junction transistor is plotted in Figure 26. Note that the saturation current also increases 10 per cent as temperature is raised one degree centigrade.

The saturation current should be approximately equal in magnitude to the open-emitter collector current. However, in some transistors,

there is a considerable leakage current component. The leakage current is the part of the open-emitter collector current which varies with the collector voltage for values greater than 0.1 volt. It has been found

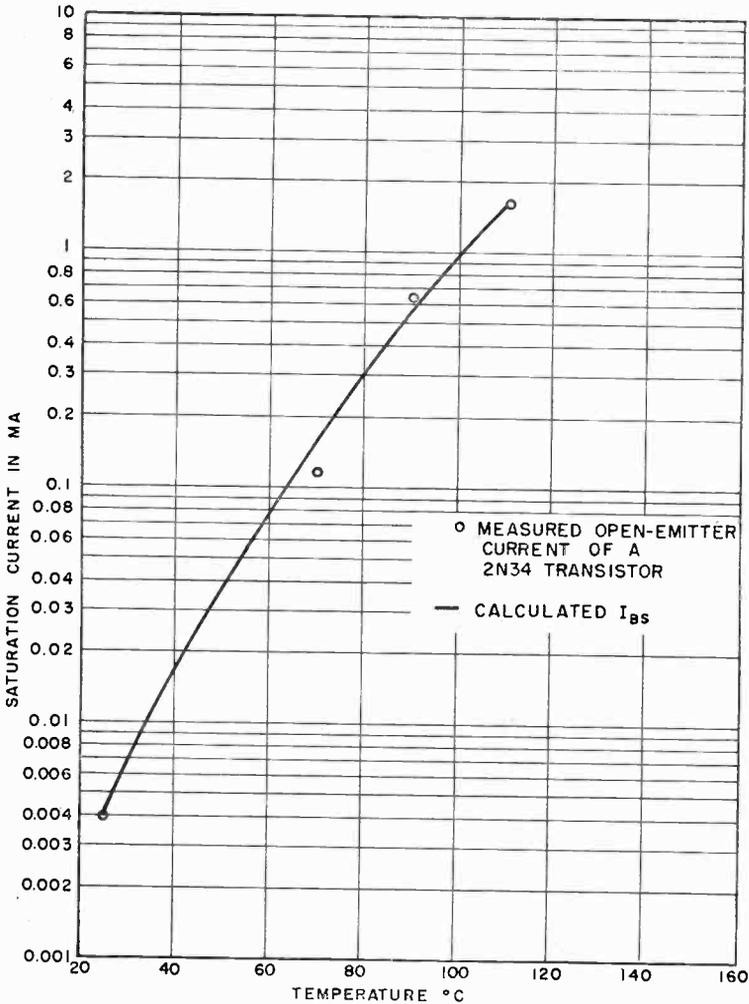


Fig. 26—Saturation current versus temperature.

experimentally that the leakage current of type 2N34 and 2N35 transistors is insensitive to temperature variations.

The base-lead resistance, r_{BB} , varies with temperature in approximately the same manner as the resistivity of the material used in the transistor base. The theoretical variation with temperature of the

resistivity¹³ of n-type germanium is shown in Figure 27. In general variations due to this cause are small compared to the variations of other parameters.

The current amplification factor, α_{CB} , varies little with temperature. This amplification factor is equal to the ratio of the d-c collector-to-emitter conductance coefficient to the d-c base-to-emitter conductance coefficient. Since these coefficients have essentially the same variation with temperature, their ratio is theoretically independent of temperature. Experiments indicate that the variation of α_{CB} with temperature is relatively small. For most practical purposes, the current amplification factor may be considered as constant.

The value of Λ varies inversely as the absolute temperature. Its change near room temperature is therefore quite slow.

From the foregoing discussion it can be seen that the most temperature-sensitive quantities affecting the transistor characteristics are the d-c input conductance coefficient and the base saturation current. For analysis purposes, it is sufficient to consider these two quantities as the only variables with temperature and treat all others as constant.

From Equation (1), if the bias voltage V_{BE} is held constant, a decrease in temperature reduces the collector current quite rapidly as a result of the approximately exponential decrease in g_{BE} . On the other hand, if the collector current is to be held constant, V_{BE} must be increased as temperature is decreased in order to counterbalance the decrease of the d-c conductance coefficient.

The rate at which the bias should change with temperature in order to maintain the collector current constant will next be examined. Rewriting Equation (1) in terms of the intrinsic bias, $V_{B'E}$, at a temperature T_0 ,

$$I_C = \frac{\alpha_{CB} g_{BE} k T_0}{q} e^{qV_{B'E}/kT_0}, \quad (4)$$

where

$$V_{B'E} = V_{BE} - (I_C/\alpha_{CB} - I_{BS}) r_{BB'}$$

Neglecting $I_{BS} r_{BB'}$,

$$V_{B'E} \cong V_{BE} - I_C r_{BB'}/\alpha_{CB}$$

¹³ P. G. Herkart and J. Kurshan, "Theoretical Resistivity and Hall Coefficient of Impure Germanium at Room Temperature," *RCA Review*, Vol. XIV, p. 427, September, 1953.

It can be shown that $g_{BE}T$ has a temperature coefficient equal to $e^{(\epsilon_G/kT_0^2 + 5/2T_0)\Delta T}$.

If the collector current at some temperature $T_0 + \Delta T$ is to be equal to the collector current at T_0 , the forward bias must be changed in order to offset the change in $g_{BE}T$. Let this change in bias be ΔV_{BE} .

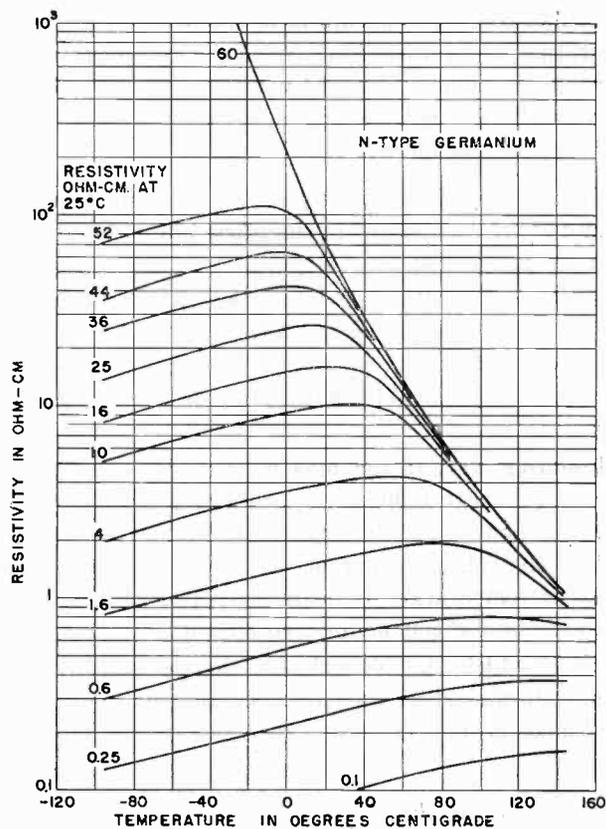


Fig. 27—Resistivity of germanium versus temperature.

The expression for collector current at $T_0 + \Delta T$ becomes

$$I_C = \left(\frac{\alpha_{CB} g_{BE} k T_0}{q} \right)_{T_0} e^{(\epsilon_G/kT_0^2 + 5/2T_0)\Delta T} e^{q(V_{BE} + \Delta V_{BE})/k(T_0 + \Delta T)} \quad (5)$$

where the subscript T_0 indicates that the quantity within the bracket is evaluated at T_0 . Equating (4) and (5),

$$\frac{q\Delta V_{BE}}{k(T_0 + \Delta T)} = \frac{qV_{B'E}\Delta T}{kT_0(T_0 + \Delta T)} - \left(\frac{\epsilon_G}{kT_0^2} + \frac{5}{2T_0} \right) \Delta T. \quad (6)$$

As ΔT approaches zero, Equation (6) becomes

$$\left(\frac{\Delta V_{BE}}{\Delta T} \right)_{\Delta T \rightarrow 0} = - \left(\frac{\epsilon_G/q - V_{B'E}}{T_0} + \frac{5k}{2q} \right). \quad (7)$$

For germanium transistors operated at normal current densities, ϵ_G/q is much greater than $V_{B'E}$, so that $\Delta V_{BE}/\Delta T$ varies only slightly with operating bias. Evaluating Equation (7) at $T_0 = 25^\circ\text{C}$ and $V_{B'E} = 0.1$ volt, $\Delta V_{BE}/\Delta T$ is approximately -2.3 millivolts per degree centigrade. As the temperature is lowered, the bias required for maintaining the collector current fixed increases, so that the fraction $(\epsilon_G/q - V_{B'E})/T_0$ in Equation (7) changes very little with temperature. If the change in internal bias due to $I_{BS}r_{BB'}$ is taken into account, it is found that $\Delta V_{BE}/\Delta T$ changes even less.

However, when the temperature is sufficiently high, the $I_{BS}r_{BB'}$ drop increases at an accelerated rate and may become the dominating factor in the change in bias. In these elevated temperature ranges, Equation (7) no longer holds. The required rate of change in bias then becomes a function of I_{BS} and $r_{BB'}$.

APPENDIX II—STABILITY FACTOR

This appendix presents an analytical comparison of the stability of several basic transistor circuits and derives the quantitative relationships which define the conditions for most effective stabilization for each circuit configuration. The stability of a transistor circuit may be measured by the sensitivity of the transistor collector current to temperature, i.e., dI_C/dT . As was pointed out previously, the dominant effects of temperature on collector current are variations in I_{BS} and g_{BE} . The latter, in turn, is a function of saturation current,

$$g_{BE} = \frac{\Delta I_{ES}}{\alpha_{CB}} = \frac{\Delta I_{BS}}{1 + (I_{CS}/I_{ES})}, \quad (8)$$

where I_{CS} and I_{ES} are the collector and emitter saturation currents. dI_C/dT can be expressed as

$$\frac{dI_C}{dI_{BS}} \frac{dI_{BS}}{dT}$$

The change in I_{BS} with respect to temperature dI_{BS}/dT ($\cong 0.1 I_{BS}$ per degree centigrade) is a property of the transistor, as explained in Appendix I. Therefore, the change in collector current with respect to change in I_{BS} , dI_C/dI_{BS} , is an indication of the stability of the transistor circuit. This derivative is defined as the stability factor, S .

It should be noted in the following derivations that at any particular temperature, a smaller value of S signifies greater circuit stability. Direct comparison of the stability factors for the following circuits may be made only at the same temperature, since S may be a rapidly changing function of temperature.

Case (a): $R_E = 0, R_B = 0$.

Substituting Equation (8) in Equation (1),

$$I_C = \frac{\alpha_{CB} I_{BS}}{1 + I_{CS}/I_{ES}} e^{\Lambda(V_{BE} - r_{BB'} I_C)} \quad (I_C/\alpha_{CB} - I_{BS}) \quad (9)$$

Differentiating and transposing

$$S_a = \frac{dI_C}{dI_{BS}} = \frac{I_C/I_{BS} + \Lambda r_{BB'} I_C}{1 + \Lambda r_{BB'} I_C/\alpha_{CB}} \quad (10)$$

When $r_{BB'}$ is small, the second terms of both the numerator and the denominator are negligible. The stability factor becomes

$$S'_a = \frac{I_C}{I_{BS}} \quad (11)$$

At very high temperatures, when $I_{BS} > 1/\Lambda r_{BB'}$ and I_C is increased to a value such that $\Lambda r_{BB'} I_C/\alpha_{CB} > 1$, the second terms of both the numerator and the denominator of Equation (10) become dominant

$$S'_a = \alpha_{CB} \quad (12)$$

Case (b): $R_B \neq 0, R_E = 0$.

An external resistance in the base circuit is equivalent to increasing $r_{BB'}$. The stability factor is

$$S_b = \frac{I_C/I_{BS} + \Lambda(r_{BB'} + R_B) I_C}{1 + \Lambda(r_{BB'} + R_B) I_C/\alpha_{CB}} \quad (13)$$

Note that at moderate and low temperatures, where $I_C/I_{BS} \gg \alpha_{CB}$, an increase in R_B increases the denominator faster than the numerator, hence reducing the stability factor. In this temperature range, if R_B is of such value that

$$\frac{1}{\Delta I_{BS}} \gg R_B + r_{BB'} \gg \frac{\alpha_{CB}}{\Delta I_C}.$$

Equation (13) becomes

$$S'_b = \alpha_{CB}/\Delta R_B I_{BS}. \quad (14)$$

The improvement over case (a) depends on the factor $I_C \Delta R_B / \alpha_{CB}$.

At elevated temperatures where $R_B + r_{BB'} \gg 1/\Delta I_{BS}$, the rate at which the numerator and the denominator increase becomes nearly the same. Equation (13) reduces to

$$S'_b = \alpha_{CB}. \quad (15)$$

Then there is no improvement over case (a).

Thus this circuit is effective in improving the stability only in the temperature range where the collector current $I_C > \alpha_{CB} I_{BS}$ and the value of R_B should be such that

$$\frac{1}{\Delta I_{BS}} > R_B + r_{BB'} > \frac{\alpha_{CB}}{\Delta I_C}. \quad (16)$$

Case (c): $R_B = 0, R_E \neq 0$.

An external resistance in the emitter circuit reduces the base-to-emitter bias by an amount equal to

$$I_E R_E \cong (I_C + I_B) R_E = (I_C - I_C/\alpha_{CB} + I_{BS}) R_E. \quad (17)$$

Equation (1) becomes

$$I_C = \frac{\alpha_{CB} I_{BS}}{1 + I_{CS}/I_{BS}} e^{\Lambda[V_{BE} + I_{BS}(r_{BB'} + R_E) - \frac{I_C}{\alpha_{CB}}(r_{BB'} + R_E + \alpha_{CB} R_E)]} \quad (18)$$

The stability factor is then

$$S_c = \frac{I_C/I_{BS} + \Lambda(r_{BB'} + R_E) I_C}{1 + \Lambda(r_{BB'} + R_E + \alpha_{CB} R_E) I_C/\alpha_{CB}}. \quad (19)$$

An examination of Equations (10) and (13), for S_a and S_b , reveals that any reduction in α_{CB} increases the denominators and, hence, the stability.

When $R_E \gg (R_E + r_{BB'})/\alpha_{CB}$, S_c can be simplified as

$$S_c = \frac{I_C/I_{BS} + \Lambda(r_{BB'} + R_E) I_C}{1 + \Lambda R_E I_C}. \quad (20)$$

If, furthermore, $R_E \gg 1/\Lambda I_C$ and $I_C/I_{BS} \gg \Lambda(r_{BB'} + R_E) I_C$ as is usually the case at moderate and low temperatures, Equation (20) approximately equals

$$S'_c = \frac{1}{\Lambda R_E I_{BS}}. \quad (21)$$

Compared with Equation (11), the improvement of S'_c over S'_a depends on the product $\Lambda R_E I_C$.

At very high temperatures when $I_{BS} \gg 1/\Lambda(r_{BB'} + R_E)$, Equation (19) becomes

$$S'_c = 1 + r_{BB'}/R_E. \quad (22)$$

Note that so long as $R_E > r_{BB'}/\alpha_{CB}$, S'_c is smaller than the corresponding stability factors in either case (a) or case (b). When R_E is greater than $r_{BB'}$, the circuit can be made nearly α_{CB} times more stable. For stability considerations, R_E is most effective when

$$\frac{1}{I_C} < R_E > \frac{r_{BB'}}{\alpha_{CB}}. \quad (23)$$

Case (d): $R_B \neq 0$, $R_E \neq 0$.

By adding R_B to $r_{BB'}$ in Equation (19) the stability factor can be written as

$$S_d = \frac{I_C/I_{BS} + \Lambda(R_B + r_{BB'} + R_E) I_C}{1 + \Lambda(R_B + r_{BB'} + \alpha_{CB} R_E) I_C/\alpha_{CB}}. \quad (24)$$

The result is as though the base-lead resistance became $R_B + r_{BB'} + R_E$ and the current amplification factor,

$$\left(\frac{R_B + r_{BB'} + R_E}{R_B + r_{BB'} + R_E + \alpha_{CB} R_E} \right) \alpha_{CB}.$$

The effect of increased resistance in the base circuit was discussed in case (b). The effect of reduced α_{CB} was discussed in case (c).

Improvement in stability over case (a) at low temperatures can be effected by making the second term of the denominator in Equation (24) much greater than unity. Improvement at high temperatures can be accomplished by reducing the effective current amplification factor. The value of R_B satisfying these conditions should be

$$\frac{1}{\Delta I_C} < R_B > \frac{R_B + r_{BB'}}{\alpha_{CB}}. \tag{25}$$

Case (e): D-C Feedback Circuit.

In this type of d-c feedback circuit as shown in Figure 6, the base-to-emitter voltage can be determined by using loop equations.

$$V_{BE} = V_2 - \frac{R_C + R_F}{R_C + R_F + R_B} (V_2 - V_1) - \frac{R_B R_C + (R_B R_C + R_B R_F) / \alpha_{CB}}{R_B + R_F + R_C} I_C + \frac{R_B R_C + R_B R_F}{R_B + R_F + R_C} I_{BS}. \tag{26}$$

Substituting Equation (26) into Equation (1),

$$I_C = \frac{\alpha_{CB} I_{BS}}{1 + I_{CS} / I_{BS}} \exp \left[\Lambda \left\{ V' - \left(\frac{R_B R_C + (R_B R_C + R_B R_F) / \alpha_{CB}}{R_C + R_F + R_B} + \frac{r_{BB'}}{\alpha_{CB}} \right) I_C + \left(r_{BB'} + \frac{R_B R_C + R_B R_F}{R_B + R_F + R_C} \right) I_{BS} \right\} \right], \tag{27}$$

where
$$V' = V_2 - \frac{R_C + R_F}{R_C + R_F + R_B} (V_2 - V_1).$$

Differentiating with respect to I_{BS} ,

$$S_e = \frac{\frac{I_C}{I_{BS}} + \Lambda \left(r_{BB'} + \frac{R_B R_C + R_B R_F}{R_B + R_F + R_C} \right) I_C}{1 + \frac{\Lambda [(R_B + r_{BB'}) (R_F + R_C) + R_B r_{BB'} + R_B R_C \alpha_{CB}] I_C}{(R_B + R_F + R_C) \alpha_{CB}}}. \tag{28}$$

The effective base resistance now becomes

$$r_{BB'} + \frac{R_B R_C + R_B R_F}{R_B + R_F + R_C};$$

the effective current amplification factor,

$$\frac{\alpha_{CB}}{1 + \frac{R_B R_C \alpha_{CB}}{(R_B + r_{BB'}) (R_F + R_C) + R_B r_{BB'}}$$

Stability can be improved by making

$$\frac{R_B R_C \alpha_{CB}}{(R_B + r_{BB'}) (R_F + R_C) + R_B r_{BB'}} > 1,$$

so that the effective α_{CB} is reduced substantially. In the extreme case where $R_B = \infty$, the condition becomes that of case (d).

At high temperatures when the second terms in both the numerator and the denominator of Equation (28) become dominant,

$$S'_e = \frac{[(R_B + r_{BB'}) (R_C + R_F) + R_B r_{BB'}] \alpha_{CB}}{(R_B + r_{BB'}) (R_C + R_F) + R_B r_{BB'} + R_B R_C \alpha_{CB}}. \quad (29)$$

Effective stabilization is provided if

$$\frac{R_F}{\alpha_{CB}} < R_C,$$

$$\frac{1}{\Delta I_C} < \frac{R_B R_C}{R_B + R_F + R_C}. \quad (30)$$

For other circuit configurations, a similar method of attack may be used to estimate the relative stability and the approximate circuit resistances.

STABILITY CONSIDERATIONS IN TRANSISTOR INTERMEDIATE-FREQUENCY AMPLIFIERS

BY

DAVID D. HOLMES AND T. O. STANLEY

RCA Laboratories,
Princeton, N. J.

Summary—Transistors which are capable of providing 30 to 40 decibels gain at 455 kilocycles have become relatively commonplace; the development of such units has led to the need for a better understanding of the factors which determine the stability of transistor i-f amplifiers. An understanding of these factors permits intelligent amplifier design, and facilitates comparison of various transistor types in terms of maximum usable gain.

This paper presents procedures for the design of single- and double-tuned interstages to be employed in tuned transistor amplifiers; these procedures are based on stability considerations. Examples of single- and double-tuned interstage designs are included, and the design considerations for amplifiers having many or different stages are outlined.

GENERAL DISCUSSION

THE problem of stability in i-f amplifiers is not a new one; stability in vacuum-tube amplifiers has been discussed in the literature.¹ The basic factors which govern stability in transistor i-f amplifiers (feedback capacitance, driving and load resistances, gain, etc.) are the same as those which apply to vacuum tubes. Transistor terminal impedances are generally different from those of tubes, with the result that the magnitudes of stability-determining parameters are different.

Some of the important factors which determine the stability of an amplifier may be derived with reference to Figure 1, which shows a common-emitter tuned transistor amplifier (transistor designated as a "black box") having finite input and output resistances, R_{in} and R_{out} , and transconductance, g_m . The circuits at the amplifier input and output are synchronously tuned, their tuned resistances being designated R_y and R_L respectively. The amplifier g_m is assumed to be in phase with the input voltage. (Generally, the transistor g_m will have a somewhat lagging phase angle at 455 kilocycles; for practical purposes, using transistors intended for 455-kilocycle operation, this is a second-order effect.) It is further assumed that the amplifier itself has been

¹ B. J. Thompson, "Oscillation in Tuned R-F Amplifiers," *Proc. I.R.E.*, Vol. 19, p. 421, March, 1931; see also F. E. Terman, *Radio Engineers' Handbook*, p. 470 First Edition, McGraw-Hill Book Co., Inc., New York, 1943.

rendered unilateral. If a capacitance C is connected between input and output circuits, the amplifier circuit as a whole is no longer unilateral. The added capacitor can have a profound effect upon the amplifier input resistance. At frequencies above the resonant frequency of the output circuit, a positive component of input resistance is contributed by the feedback capacitor, while at frequencies below resonance, a negative component is contributed. The negative input resistance has a minimum value at the frequency at which the total input- and output-circuit

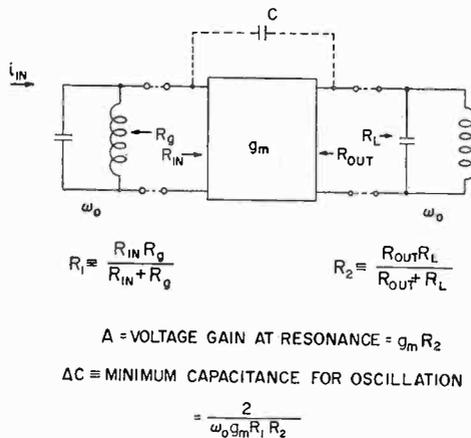


Fig. 1—Stability-determining factors.

phase shift is $+90$ degrees; the magnitude of minimum negative resistance is given by

$$-R = \frac{2X_c}{A}, \quad (1)$$

where $X_c =$ reactance of feedback capacitor C ,

$A =$ voltage gain at resonance.

If the minimum value of negative resistance becomes equal to the net positive resistance shunting the input terminals (the parallel combination of R_g and R_{in} , designated R_1 in Figure 1) the amplifier will oscillate. The minimum value of capacitance required to produce oscillation is designated ΔC ; an expression for the magnitude of ΔC may be determined as follows. The amplifier voltage gain at resonance is equal to $g_m R_2$ (assuming g_m to be independent of output voltage swing), where R_2 is the parallel combination of R_{out} and R_L . The minimum value of negative resistance, from Equation (1), is then

$$-R = \frac{2X_c}{g_m R_2}. \quad (2)$$

For the oscillatory condition, $-R = R_1$. Substituting this in Equation (2) and solving for X_c ,

$$X_c = g_m \frac{R_1 R_2}{2}, \quad (3)$$

and

$$\Delta C = \frac{2}{\omega g_m R_1 R_2}. \quad (4)$$

The value of ΔC determined from Equation (4) applies for a single amplifier stage. If several similar stages are cascaded, the value of ΔC per stage diminishes.¹ The factors by which ΔC is reduced are 2, 2.61, and 3 for two, three, and four stages, respectively.

The determination of ΔC for a given amplifier configuration permits a determination of whether or not neutralization is required, and is an aid to determining the effects of production variation of transistor feedback capacitance and of a neutralizing capacitor, if such an element is employed.

The value of ΔC is the minimum value required to produce oscillation; a smaller capacitor, while not sufficiently large to produce oscillation, may introduce objectionable skew in the selectivity characteristic, and may make alignment difficult. The amount of skew of the selectivity characteristic that can be expected when the feedback capacitor is 20 and 40 per cent of ΔC is shown in Figure 2. It is generally desirable, then, to design the amplifier in such a manner that the maximum net feedback capacitance which can be expected from normal parameter limits is considerably smaller than ΔC .

Inspection of Equation (4) shows that ΔC may be increased by reducing R_1 and/or R_2 , i.e., by reducing the generator and load resistances. The stability obtained in this manner is accompanied by a loss in gain.

In order that the transistor amplifier of Figure 1 provide *maximum* gain, R_g and R_L should be made equal to R_{in} and R_{out} , respectively (conjugate matching). For this condition,

$$R_g = R_{in} = 2R_1,$$

$$R_L = R_{out} = 2R_2.$$

Substituting in Equation (4),

$$\Delta C = \frac{8}{\omega g_m R_{in} R_{out}} \equiv \Delta C_0, \quad (5)$$

where ΔC_0 is defined as the minimum capacitance for oscillation in the *conjugate-matched* condition. The maximum available unilateralized gain (M.A.G.) of the transistor is given by

$$\text{M.A.G.} = (g_m^2/4) R_{in} R_{out}. \quad (6)$$

Making use of Equation (6), Equation (5) may be written

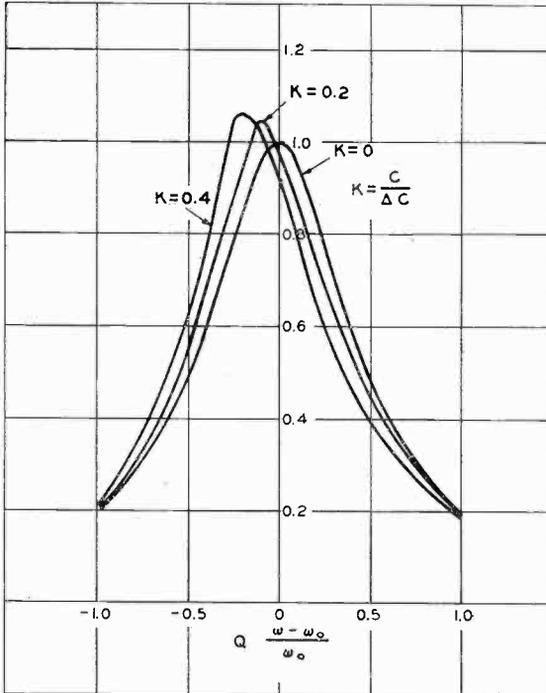


Fig. 2—Selectivity skew resulting from misneutralization.

$$\Delta C_0 = \frac{4}{\omega \sqrt{\text{M.A.G.}} \times R_{in} R_{out}}. \quad (7)$$

The quantities on the right-hand side of Equation (7) may be measured at 455 kilocycles on the test set described by Holmes, Freedman, and Scott.²

² D. D. Holmes, L. A. Freedman, and T. M. Scott, "A Test Set for Transistor Performance Measurement at 455 Kilocycles," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 322-335.

STABILITY FACTOR

As mentioned above, the value of ΔC may be increased by reducing the generator and load impedances. The following is a convenient definition of a stability factor, S , of the amplifier:

$$S = \frac{\Delta C}{\Delta C_0}. \quad (8)$$

This equation describes the factor by which the oscillation-producing feedback capacitance is increased above that applying for the conjugate-matched condition.

A useful expression for the factor S , derived by dividing Equation (4) by Equation (5), is as follows:

$$S = \frac{\Delta C}{\Delta C_0} = \frac{R_{in}}{2R_1} \times \frac{R_{out}}{2R_2}. \quad (9)$$

It is convenient to consider the right-hand side of Equation (9) as comprising the product of an input stability factor (IS) and an output stability factor (OS).

Then

$$IS = \frac{R_{in}}{2R_1}, \quad (10)$$

and

$$OS = \frac{R_{out}}{2R_2}. \quad (11)$$

Equation (10) shows that the IS is equal to the ratio of the net shunt resistance at the input terminals for the matched condition ($R_{in}/2$) to the net shunt resistance which applies for a specific design. A similar statement can be made with respect to the OS .

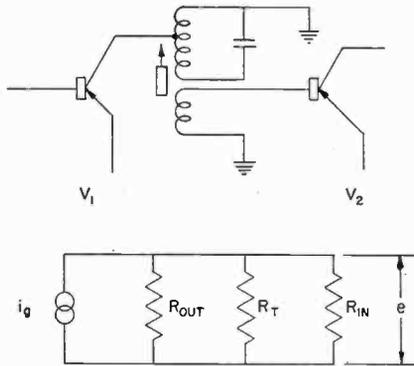
STABILITY FACTOR AND INSERTION LOSS USING
SINGLE-TUNED INTERSTAGES

An equivalent signal circuit for a single-tuned transistor interstage is shown in Figure 3, along with pertinent power relationships. The term "insertion loss" is used here to describe the ratio of the maximum available power from transistor V_1 to the power delivered to the input of transistor V_2 . The resistors are referred to one point in the circuit, e.g., to the point where the collector of the driving transistor is con-

nected to the tuned circuit. Resistors R_{out} , R_T , and R_{in} represent the output resistance of transistor V_1 , the tuned resistance of the transformer, and the reflected input resistance of V_2 . It can be seen that the interstage transformer design determines the *OS* of V_1 and the *IS* of V_2 .

As shown in Figure 3 the transformer insertion loss is given by

$$\text{Insertion loss} = \frac{R_{in} R_{out}}{4R^2}, \quad (12)$$



$R \equiv R_{OUT}, R_T, R_{IN}$ IN PARALLEL

$$\text{MAXIMUM AVAILABLE POWER} = \frac{i_g^2 R_{OUT}}{4}$$

$$\text{LOAD POWER} = \frac{e^2}{R_{IN}} = \frac{i_g^2 R^2}{R_{IN}}$$

$$\text{INSERTION LOSS} = \frac{\text{MAX. AVAIL. P}}{\text{LOAD P}} = \frac{R_{IN} R_{OUT}}{4R^2}$$

Fig. 3—Equivalent circuit of transistor single-tuned interstage with power relations.

which may be written

$$\text{Insertion loss} = \frac{R_{in}}{2R} \times \frac{R_{out}}{2R}. \quad (13)$$

Note that Equation (13) is similar to Equation (9), and may be written:

$$\text{Insertion loss} = OS \times IS. \quad (14)$$

Thus, insertion loss and stability factor are intimately related. For example, if the circuit shown in Figure 3 were repeated iteratively, and if the transformers were designed to provide a stability factor of 2 per

stage, a power loss of 2 times, or 3 decibels, per transformer would be incurred.

A further interesting point may be brought out by writing the expressions for stability factors and insertion loss in terms of circuit Q 's. The following equations are obtained:*

$$OS = \frac{1}{2} \left(\frac{1}{1 - Q/Q''} \right) \quad (15)$$

where Q is the operating circuit Q with transistors V_1 and V_2 connected in the circuit, and Q'' is the circuit Q with only transistor V_2 in the circuit ($R_{out} = \infty$).

$$IS = \frac{1}{2} \left(\frac{1}{1 - Q/Q'} \right) \quad (16)$$

where Q' is the circuit Q with only transistor V_1 in the circuit ($R_{in} = \infty$).

$$\text{Insertion loss} = \frac{1}{4} \left(\frac{1}{(1 - Q/Q') (1 - Q/Q'')} \right). \quad (17)$$

Further,

$$1 - \frac{1}{2} \left(\frac{1}{OS} - \frac{1}{IS} \right) = \frac{Q}{Q_0}, \quad (18)$$

where Q_0 is the unloaded coil Q . Equation (18) shows that if the input and output stability factors are prescribed, the ratio of operating Q to unloaded Q is dictated. If the operating Q is fixed by selectivity considerations the choice of stability factors determines the value of required unloaded coil Q .

EXAMPLE OF 455-KILOCYCLE SINGLE-TUNED INTERSTAGE DESIGN

Consider that the average characteristics of the transistor to be used are as follows:

$$R_{in} = 500 \text{ ohms,}$$

$$R_{out} = 30,000 \text{ ohms,}$$

$$\text{M.A.G.} = 6,000 \text{ (38 decibels), at 455 kilocycles,}$$

$$C_f = \text{base-to-collector capacitance} = 12 \text{ micromicrofarads.}$$

* Derivation in Appendix A.

The range of C_j among transistors is found to be 8 to 16 or ± 4 micromicrofarads. The operating Q of the single-tuned circuit is to be 35. From Equation (7),

$$\begin{aligned}\Delta C_0 &= \frac{4}{\omega \sqrt{\text{M.A.G.} \times R_{in} R_{out}}} \\ &= \frac{4}{6.28 \times 455 \times 10^3 \sqrt{6,000 \times 500 \times 30,000}} \\ &= 4.66 \text{ micromicrofarads.}\end{aligned}$$

Since the collector-to-base capacitance is 12 ± 4 micromicrofarads, the stage must be neutralized to prevent oscillation. If a ± 10 per cent neutralizing capacitor is employed, the net variation from "bogie" feedback capacitance will be $4 + 1.2 = 5.2$ micromicrofarads. On the basis of the curves shown in Figure 2, this value should not be more than 20 per cent (arbitrary) of ΔC in order that excessive skew be avoided. Then,

$$\Delta C = \frac{5.2}{0.2} = 26 \text{ micromicrofarads,}$$

$$S = \frac{\Delta C}{\Delta C_0} = \frac{26}{4.66} = 5.58 \text{ (considering a single stage only).}$$

Make

$$IS = OS = \sqrt{5.58} = 2.36.$$

The insertion loss, equal to S , will be 5.58 times, or 7.47 decibels. This shows that the maximum practical single-stage gain for this transistor type is $38 - 7.47 = 30.53$ decibels.

To find the required unloaded coil Q , from Equation (18),

$$\frac{Q}{Q_0} = 1 - \frac{1}{2} \left(\frac{1}{OS} + \frac{1}{IS} \right) = 0.577,$$

$$Q_0 = \frac{35}{0.577} = 60.7.$$

The problem now is to find the required turns ratios of the transformer. From Equations (12) and (14),

$$S = \frac{R_{in}R_{out}}{4R^2}.$$

Since

$$R_{in} = R_{out} \text{ (equal IS and OS),}$$

$$S = \frac{R_{out}^2}{4R^2}, \text{ and}$$

$$R = \frac{R_{out}}{2\sqrt{S}} = \frac{30,000}{2 \times 2.36} = 6,360 \text{ ohms,}$$

$$Q = \frac{R}{X} = 35,$$

$$\frac{6360}{35} = X = 182 \text{ ohms,}$$

$$C = 1,920 \text{ micromicrofarads.}$$

This is the required circuit capacitance (including transistor capacitances) at the collector of the driving transistor. It is usually desirable to use a smaller tuning capacitor, say 220 micromicrofarads. Then the turns ratio from the collector to the top of the tuned circuit is determined as follows:

$$n_1^2 = \frac{1920}{220} = 8.73,$$

$$n_1 = 2.95.$$

The primary-to-secondary turns ratio is

$$n_2 = \sqrt{\frac{30,000}{500}} = 7.75.$$

The circuit, with turns ratios normalized with respect to the collector tap, is shown in Figure 4.

The total primary inductance should be 0.56 millihenry, and the unloaded primary Q should be 60.7; the primary and secondary should be essentially unity-coupled. The neutralizing capacitance is

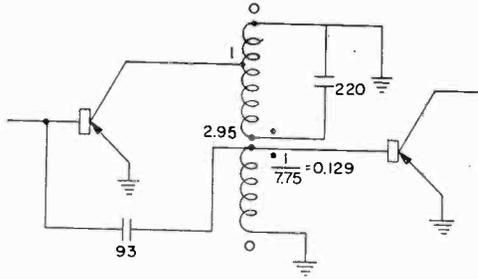


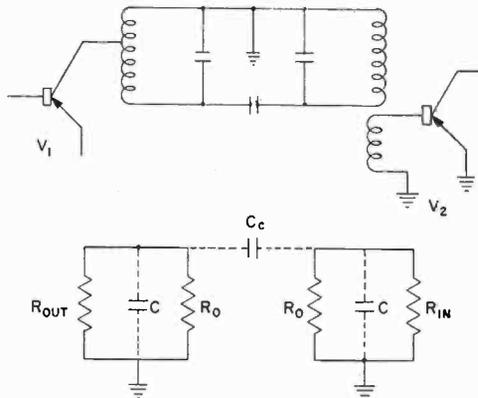
Fig. 4—Example of single-tuned interstage design.

$$C_{fn_2} = 12 \times 7.75 = 93 \text{ micromicrofarads.}$$

STABILITY FACTOR AND INSERTION LOSS USING DOUBLE-TUNED INTERSTAGES

An equivalent signal circuit for two transistors coupled by a double-tuned circuit is shown in Figure 5. The two tuned circuits are assumed to be identical and, further, the loaded primary and secondary Q 's are made equal. These assumptions are reasonable from a practical point of view since it may be desirable production-wise to make the primary and secondary unloaded Q 's the same, while equal loaded Q 's provide maximum adjacent-channel attenuation for a given stability factor.

A general derivation of pertinent equations for the double-tuned



- R_0 = PRIMARY AND SECONDARY TUNED RESISTANCE.
 Q_0 = PRIMARY AND SECONDARY UNLOADED Q .
 Q = PRIMARY AND SECONDARY OPERATING Q .
 R_{out} = TRANSISTOR OUTPUT RESISTANCE
 R_{in} = REFLECTED TRANSISTOR INPUT RESISTANCE
 k = COEFFICIENT OF COUPLING $\approx \frac{C_c}{C}$

Fig. 5—Transistors coupled by a double-tuned circuit.

interstage is given in Appendix B. Substitution of the conditions assumed in the example of Figure 5 in these general equations yields the following:

$$OS = \frac{R_{out}}{2R} (1 + K^2Q^2), \quad (19)$$

$$IS = OS. \quad (20)$$

$$\text{Insertion loss} = \frac{R_{in}R_{out} (1 + K^2Q^2)^2}{4K^2Q^2R^2} \quad (21)$$

$$= OS \times IS \times \frac{1}{K^2Q^2}. \quad (22)$$

Further,

$$OS = \frac{1 + K^2Q^2}{2 \left(1 - \frac{Q}{Q_0} \right)}. \quad (23)$$

The symbols are defined in Figure 5.

The equations describing the stability factor and insertion loss for a double-tuned circuit are similar in form to those governing a single-tuned circuit, with the exception that the factor K^2Q^2 appears in the double-tuned case.

EXAMPLE OF 455-KC DOUBLE-TUNED INTERSTAGE DESIGN

Assume that the transistor type to be employed is the same as that used in the single-tuned interstage example, that the primary and secondary unloaded Q 's are to be equal and that the primary and secondary loaded Q 's are to be equal to 35. The coefficient of coupling is to be 85 per cent of critical. Then $IS = OS = 2.36$ (from single-tuned example).

From Equation (13),

$$OS = \frac{1 + K^2Q^2}{2 \left(1 - \frac{Q}{Q_0} \right)}$$

$$Q_0 = \frac{Q}{1 - \frac{1 + K^2Q^2}{2OS}}$$

$$= \frac{35}{1 - \frac{1 + 0.722}{2(2.36)}} = 55.2.$$

This is the value for primary and secondary unloaded Q .

From Equation (22),

$$\begin{aligned} \text{Insertion loss} &= S \times \frac{1}{K^2 Q^2} \\ &= 5.58 \times \frac{1}{0.722} = 7.72, \text{ or } 8.87 \text{ decibels.} \end{aligned}$$

From Equation (19),

$$OS = \frac{R_{\text{out}}}{2R} (1 + K^2 Q^2),$$

$$R = \frac{30,000}{2(2.36)} (1.722) = 10,950 \text{ ohms,}$$

$$\frac{R}{X} = 35,$$

$$X = \frac{10,950}{35} = 313 \text{ ohms,}$$

$$C = 1117 \text{ micromicrofarads.}$$

This is the value of circuit capacitance at the collector of the driving transistor. Assuming that the tuning capacitance is to be 220 micromicrofarads for both primary and secondary, then

$$n_1^2 = \frac{1117}{220} = 5.07,$$

$$n_1 = 2.25,$$

$$n_2 = 7.75 \text{ (from single-tuned example).}$$

The circuit, with turns ratios normalized with respect to the collector tap of the driving transistor, is shown in Figure 6. The neutralizing capacitance is again 93 micromicrofarads. The value of coupling capacitor is determined as follows:

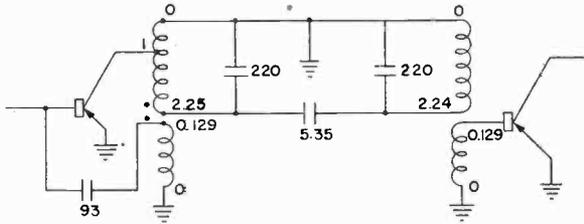


Fig. 6—Example of double-tuned interstage design.

$$K = \frac{0.85}{\sqrt{Q_1 Q_2}} = \frac{0.85}{35} = \frac{C_c}{220},$$

$$C_c = 5.35 \text{ micromicrofarads.}$$

Alternatively, the circuit may be arranged as shown in Figure 7.

Here, the neutralizing capacitance is

$$C_n = \frac{1}{n_1 - 1} C_f = \frac{1}{1.25} \times 12 = 9.68 \text{ micromicrofarads.}$$

Mutual inductance is employed as the coupling element between primary and secondary.

DESIGN OF MULTISTAGE AMPLIFIERS

The design of an amplifier having more than one stage requires examination of the amplifier or system as a whole. Two typical broadcast receiver complements will be investigated as examples of multi-stage amplifiers.

Figure 8 shows a block diagram of a three-transistor converter-i-f amplifier-second-detector combination. The coupling network between the converter and the i-f stage is a double-tuned circuit, while that between the i-f amplifier and detector is single tuned. The impedance presented to i-f at the converter input and at the detector output is

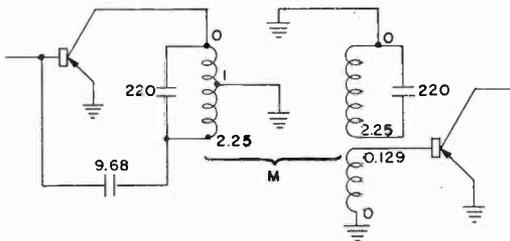


Fig. 7—Alternative to Figure 6.

assumed to be negligible, and therefore the stability factor of these stages is not a practical problem. The S of the i-f stage is important, however. The IS of this stage is principally determined by the secondary of the double-tuned circuit, while the OS is determined by the single-tuned circuit. A fruitful design approach may be followed if it is recalled that the IS is related to the factor by which the Q of the secondary of the double-tuned circuit is reduced when the i-f transistor is connected in the circuit. The OS is similarly related to the Q change of the single-tuned circuit when the i-f transistor is connected in the circuit.

An example of the philosophy of the design is as follows. Suppose that the operating Q of each tuned circuit is set at 35 by selectivity requirements. The OS of the converter is of no concern, and therefore the major requirement to be satisfied by the primary of the double-

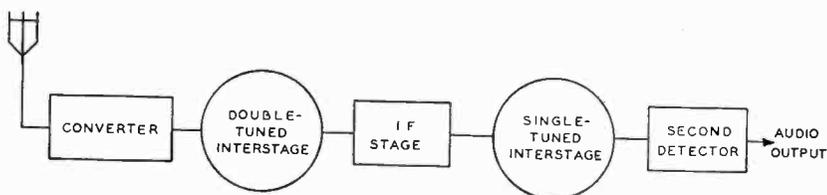


Fig. 8—Three-transistor complement.

tuned circuit is that its operating Q be 35. Minimum loss will be obtained if the primary unloaded Q is made as high as practicable; the output resistance of the converter should then be reflected across the primary in such a manner that the loaded Q becomes 35. The primary operating Q , thus, may be relatively dependent on the output resistance of the converter. The power loss associated with the primary of the double-tuned circuit is determined by (a) the highest attainable primary unloaded Q , and (b) the permissible variation in primary Q due to variation in converter output resistance among transistors. Typically, a primary unloaded Q of about 100 would result in a moderate power loss (2 or 3 decibels), and the primary Q variation incurred by converter output resistance variation among transistors might be in the order of 10 to 20 per cent.

The secondary of the double-tuned circuit should be designed to provide the desired IS for the i-f stage.

The design of the single-tuned circuit involves a combination of the design considerations of the primary and secondary of the double-tuned circuit. The desired OS of the i-f stage, together with the desired operating Q , dictates the value of circuit Q existing when the i-f transistor is removed from the circuit. For example, the desired OS

might be obtained when the circuit Q changed from 60 to 35 when the single-tuned circuit is loaded by the i-f stage output resistance. There are no stability considerations involved with the second detector, or output, side of the single-tuned circuit. Thus, as was pointed out with respect to the primary of the double-tuned circuit, minimum loss will be obtained if the unloaded Q is made as high as practicable. Continuing with the example, then, the unloaded Q of the single-tuned circuit might be 100. The turns ratios would be selected so that the loading provided by the second detector input circuit brought the Q down to 60, while the loading contributed by the i-f stage output resistance brought the Q from 60 to 35.

A second receiver complement is shown in Figure 9. Here, two i-f stages are employed, and therefore the individual stability factors

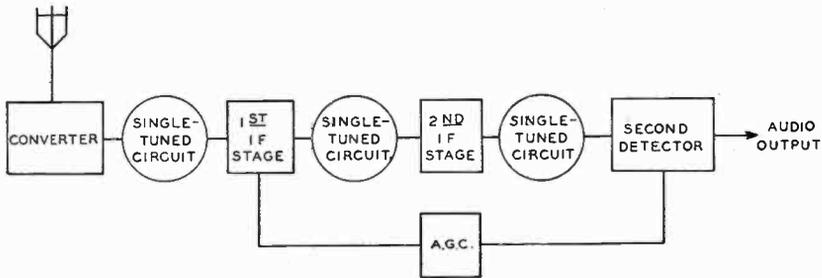


Fig. 9—Four-transistor complement.

of the two stages must be twice that of a single stage in order that the same degree of over-all stability be obtained. The design of the input of the first and the output of the last coupling networks, working from the converter and into the detector, is similar to that described above for the case of the single i-f stage.

It is pertinent to consider the effect of the application of automatic-gain-control (a-g-c) to the first i-f stage; this is generally accomplished by emitter current reduction. As the emitter current is reduced, the input and output resistances of the first i-f stage increase. The increase in output resistance of the first i-f stage reduces the IS of the second i-f stage. Therefore, the contribution to the IS of the second i-f stage by the output resistance of the first i-f stage cannot be counted on under strong-signal conditions. The stability factor of the second stage, which has been made twice that required in a single-stage amplifier, affords a considerable margin of safety when the gain of the first stage is reduced by a-g-c action. Further, in many applications the detector input resistance decreases with increasing signal, increasing the OS of the second stage, so that the overall second stage S remains more nearly constant.

APPENDIX A—DERIVATION OF S AND INSERTION LOSS IN TERMS OF
CIRCUIT Q 'S FOR A SINGLE-TUNED CIRCUIT

Referring to the equivalent circuit of Figure 3,

$$Q_0 = \text{primary unloaded } Q = \frac{R_0}{X},$$

$$Q'' = \text{circuit } Q \text{ with V2 loading, i.e., } R_{\text{out}} = \infty,$$

$$Q' = \text{circuit } Q \text{ with V1 loading, i.e., } R_{\text{in}} = \infty,$$

$$Q = \text{circuit operating } Q, \text{ i.e., } R_{\text{in}} \text{ and } R_{\text{out}} \text{ loading,}$$

$$\frac{1}{R} = G_{\text{out}} + G_0 + G_{\text{in}},$$

$$\frac{1}{R''} = G_0 + G_{\text{in}},$$

$$\frac{1}{R'} = G_{\text{out}} + G_0.$$

To determine OS in terms of Q 's defined above,

$$\frac{R}{R_{\text{out}}} = \frac{G_{\text{out}}}{G_{\text{out}} + G_0 + G_{\text{in}}} = 1 - \frac{G_0 + G_{\text{in}}}{G_{\text{out}} + G_0 + G_{\text{in}}} = 1 - \frac{R}{R''} = 1 - \frac{Q}{Q''},$$

$$OS = \frac{R_{\text{out}}}{2R} = \frac{1}{2} \left(\frac{1}{1 - Q/Q''} \right). \quad (15)$$

To determine IS in terms of Q 's,

$$\frac{R}{R_{\text{in}}} = \frac{G_{\text{in}}}{G_{\text{out}} + G_0 + G_{\text{in}}} = 1 - \frac{G_{\text{out}} + G_0}{G_{\text{out}} + G_0 + G_{\text{in}}} = 1 - \frac{R}{R'} = 1 - \frac{Q}{Q'},$$

$$IS = \frac{R_{\text{in}}}{2R} = \frac{1}{2} \left(\frac{1}{1 - Q/Q'} \right). \quad (16)$$

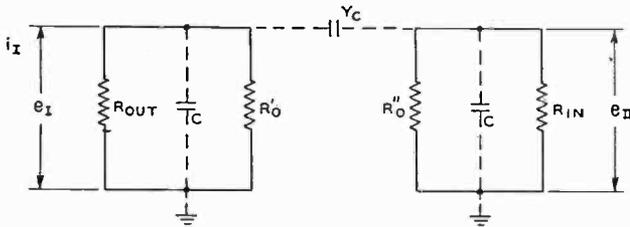
$$\text{Insertion loss} = IS \times OS = \frac{1}{4} \left(\frac{1}{1 - Q/Q'} \right) \left(\frac{1}{1 - Q/Q''} \right). \quad (17)$$

To derive Equation (18),

$$\begin{aligned} \frac{R}{R_{out}} + \frac{R}{R_{in}} &= 2 - \frac{G_{out} + 2G_0 + G_{in}}{G_{out} + G_0 + G_{in}} = 1 - \frac{G_0}{G_{out} + G_0 + G_{in}} \\ &= 1 - \frac{R}{R_0} = 1 - \frac{Q}{Q_0}, \\ \frac{R}{R_{out}} &= \frac{1}{2OS} \quad \text{and} \quad \frac{R}{R_{in}} = \frac{1}{2IS}, \\ \frac{1}{2OS} + \frac{1}{2IS} &= 1 - \frac{Q}{Q_0}, \\ 1 - \frac{1}{2} \left(\frac{1}{OS} + \frac{1}{IS} \right) &= \frac{Q}{Q_0}. \end{aligned} \tag{18}$$

APPENDIX B—DERIVATION OF EXPRESSIONS FOR *S* AND INSERTION LOSS FOR A DOUBLE-TUNED CIRCUIT

Assume that primary and secondary impedances are referred to equal shunt susceptance points, and consider the schematic diagram.



R'_0 = primary unloaded tuned resistance,

Q'_0 = primary unloaded Q .

$$R' = \frac{R'_0 R_{out}}{R'_0 + R_{out}},$$

Q' = primary loaded Q ,

R''_0 = secondary unloaded tuned resistance,

Q''_0 = secondary unloaded Q ,

$$R'' = \frac{R''_0 R_{in}}{R''_0 + R_{in}},$$

Q'' = secondary loaded Q ,

$$K \approx \frac{\text{coupling capacitance}}{\text{tuning capacitance}}.$$

Writing nodal equations,

$$i_I = \frac{e_I}{R'} - e_{II} j Y_c,$$

$$0 = -e_I j Y_c + \frac{e_{II}}{R''},$$

$$e_{II} = \frac{-i_I j Y_c}{1/(R'R'') + Y_c^2} = \frac{-j i_I R' K Q''}{1 + K^2 Q' Q''},$$

$$e_I = \frac{i_I / R''}{1/(R'R'') + Y_c^2} = \frac{i_I R'}{1 + K^2 Q' Q''}.$$

Now,

$$OS = \frac{R_{out}}{2e_I/i_I} = \frac{R_{out}}{R'} \frac{1 + K^2 Q' Q''}{2}, \quad (19a)$$

from which Equation (19) follows, substituting the assumptions of Figure 5.

By symmetry,

$$IS = \frac{R_{in}}{R''} \frac{1 + K^2 Q' Q''}{2}.$$

$$\text{Insertion loss} = \frac{\text{Max available power}}{\text{Load power}}$$

$$= \frac{i_I^2 R_{out}}{4} \cdot \frac{R_{in}}{e_{II}^2} = \frac{R_{in} R_{out} (1 + K^2 Q' Q'')^2}{4K^2 Q' Q''^2 Q'^2}$$

$$= \frac{R_{in} R_{out} (1 + K^2 Q' Q'')^2}{4K^2 Q' Q'' R' R''}, \quad (21a)$$

from which Equation (21) follows, substituting the assumptions of Figure 5.

Further,

$$\text{Insertion loss} = OS \times IS \times \frac{1}{K^2 Q' Q''}, \quad (22a)$$

from which Equation (22) follows similarly.

For a general derivation leading to Equation (23),

$$\frac{1}{R'} = G_{\text{out}} + G_0', \quad \frac{1}{R''} = G_{\text{in}} + G_0'',$$

$$\frac{R'}{R_{\text{out}}} = \frac{G_{\text{out}}}{G_{\text{out}} + G_0} = 1 - \frac{G_0'}{G_{\text{out}} + G_0'} = 1 - \frac{R'}{R_0'} = 1 - \frac{Q'}{Q_0'}.$$

Similarly,

$$\frac{R''}{R_{\text{in}}} = 1 - \frac{Q''}{Q_0''}.$$

Then

$$OS = \frac{R_{\text{out}}}{R'} \frac{1 + K^2 Q' Q''}{2} = \frac{1 + K^2 Q' Q''}{2 (1 - Q'/Q_0')},$$

and

$$IS = \frac{1 + K^2 Q' Q''}{2 (1 - Q''/Q_0'')}.$$

$$\text{Insertion loss} = \frac{(1 + K^2 Q' Q'')^2}{4 (1 - Q'/Q_0') (1 - Q''/Q_0'')} \times \frac{1}{K^2 Q' Q''}.$$

CROSS MODULATION IN TRANSISTOR RADIO-FREQUENCY AMPLIFIERS

By

DAVID D. HOLMES

RCA Laboratories,
Princeton, N. J.

Summary—The performance of gain-controlled transistor r-f and i-f stages with respect to cross modulation has frequently been questioned. This paper presents an analysis which shows the performance that can be expected from alloy-junction transistors. Measurements which substantiate the analysis are given.

Because of the differences in impedance levels between transistors and tubes, normalization of voltage levels is required in order that the performance of the two devices be compared at an equal input power level. Comparative data for a transistor and a "variable- μ " tube, when employed as a gain-controlled r-f stage in an automobile receiver, are shown. This example shows that the transistor can be equal to or better than the tube with respect to cross modulation.

GENERAL DISCUSSION

CROSS modulation in a device, or system, is a term applied to the transfer of modulation of an undesired carrier to a desired carrier. The per cent cross modulation may be defined as the per cent modulation of the desired carrier resulting from the presence of a 100 per cent modulated undesired carrier. The magnitude of the undesired, or interfering, signal required to produce a stated per cent cross modulation is a convenient yardstick by which the interference susceptibility of signal-processing devices may be measured.

An analysis of cross modulation in vacuum tubes has shown¹ that the magnitude of the interfering voltage required to produce a given per cent cross modulation, K , may be derived from a plot of the log of the tube transconductance versus bias voltage. For a constant bias, the per cent cross modulation is independent of the desired signal voltage, and varies as the square of the interfering signal voltage. The equation for determining the interfering voltage, V_2 , for K per cent cross modulation is as follows:

$$V_2 = \frac{dV_{\text{bias}}}{d(\log g_m)} \sqrt{\frac{K}{2.65}} \frac{1}{\sqrt{2}} \text{ volts r-m-s.} \quad (1)$$

¹ Philips Technical Library, Book IV, *Application of the Electronic Valve in Radio Receivers and Amplifiers*, p. 325.

It is usually convenient to draw a tangent to the curve at the operating point of interest, and take a decade change in g_m along the tangent; then

$$\begin{aligned} d(\log g_m) &= \Delta \log g_m = \log g_{m_1} - \log g_{m_2} \\ &= \log \frac{g_{m_1}}{g_{m_2}} = \log 10 = 1. \end{aligned}$$

Equation (1) becomes

$$V_2 = \left(\Delta V_{\text{bias}} \Big|_{g_{m_1}/g_{m_2} = 10} \right) \left(\sqrt{\frac{K}{2.65}} \times \frac{1}{\sqrt{2}} \right) \text{volts r-m-s.}$$

For $K = 1$ per cent,

$$V_2 = 0.0434 \Delta V_{\text{bias}} \Big|_{g_{m_1}/g_{m_2} = 10} \text{volts r-m-s.} \quad (3)$$

(This equation is valid for portions of the $\log g_m$ versus bias voltage characteristic which have small curvature; generally there are large segments of the characteristic which satisfy this condition.)

APPLICATION TO TUBES

For most tubes of the "variable- μ " type, a plot of transconductance (\log scale) versus grid No. 1 volts may be found in the tube manual. Curves for the 6BA6 are shown in Figure 1. To find the interfering signal for 1 per cent cross modulation, a tangent is drawn at the operating point of interest, the ΔV_{G1} for $g_{m_1}/g_{m_2} = 10$ is read from the abscissa, and Equation (3) is then employed to obtain V_2 . For the example shown in Figure 1, $\Delta V_{G1} = 2.7$ volts; $V_2 = 117$ millivolts r-m-s.

Measurements on the 6BA6 r-f stage of a contemporary high-quality auto receiver have reasonably confirmed this value for weak-signal (small a-g-c bias) operation. Under strong-signal conditions, the increasing a-g-c voltage shifts the operating point into the region of a more gradual slope of the $\log g_m$ characteristic, and an improvement of several times (3 to 10) may be obtained, depending upon the extent to which a "sliding screen" arrangement is employed. The use of a separate screen dropping resistor on the r-f stage effects a considerable improvement in cross modulation characteristics, but in recent years full advantage has not been taken of it, possibly for economic reasons.

As a further illustration of the correlation of the slope of the \log

g_m characteristic with the cross modulation characteristic, venerable data on a type 6K7 tube, taken two decades ago, is shown in Figures 2 and 3. In Figure 3, the ordinate is the interfering voltage required to give 12 per cent cross modulation of the desired carrier,² and must be divided by $\sqrt{12}$ to obtain the value for 1 per cent cross modulation. The log g_m slope method outlined above gives good agreement with the values shown in Figure 3. For instance, at -15 volts bias the curve

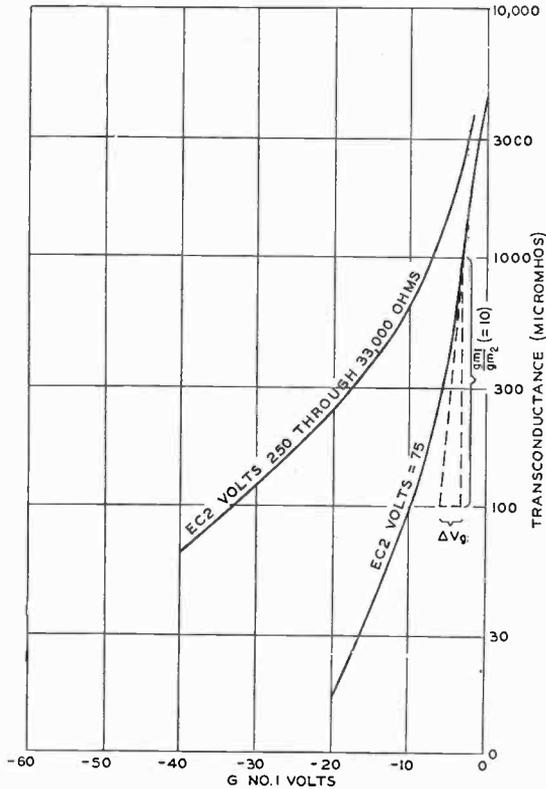


Fig. 1—Transconductance versus grid No. 1 voltage for 6BA6 tube.

of Figure 3 (divided by $\sqrt{12}$) shows 0.6 volt for 1 per cent cross modulation, while the slope method, applied to Figure 2, gives 0.65 volt. Corresponding values at -35 volts bias are 1.3 volts and 1.5 volts.

CROSS MODULATION IN A P-N JUNCTION

An approximate equation (neglecting lead resistances, saturation

² E. W. Herold, "Simple Methods for Checking R-F Distortion or Cross Modulation of Pentode Amplifier Tubes," *Electronics*, Vol. 13, p. 82, April, 1940.

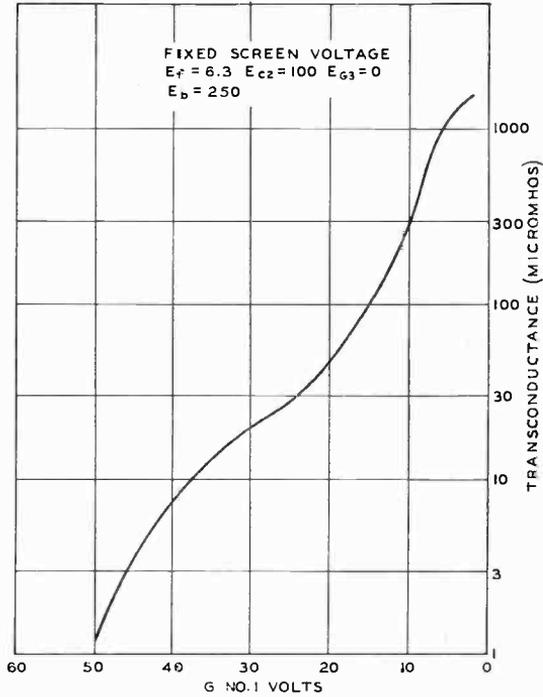


Fig. 2—Transconductance versus grid No. 1 voltage for 6K7 tube.

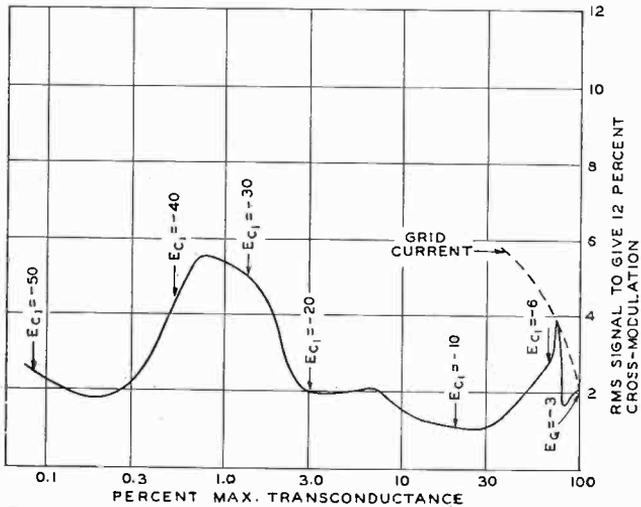


Fig. 3—Interfering voltage for 12 per cent cross modulation; 6K7 tube — fixed screen voltage = 100 volts.

current, and frequency effects) for a forward biased p-n junction is as follows:

$$I \propto e^{kt - \frac{q}{V}} \approx e^{40V} \text{ (at } 25^\circ\text{C)}. \quad (4)$$

In a transistor, $g_m = dI/dV \approx 40 Ke^{40V} \approx 40 I$, so that Equation (4) may also be employed to describe the variation of g_m with bias voltage. If $\log I$ (or $\log g_m$) is now plotted against bias voltage, a straight line is obtained. It is interesting to note that the slope of this line (the important factor in determining cross modulation characteristics) depends on the constant q/kT , and is therefore independent of the semiconductor material, geometry, etc. Any ideal p-n junction, then, when driven from a constant-voltage source, produces constant cross-modulation distortion for a given interfering voltage, independent of the operating point. The significance of this with respect to transistors will now be developed.

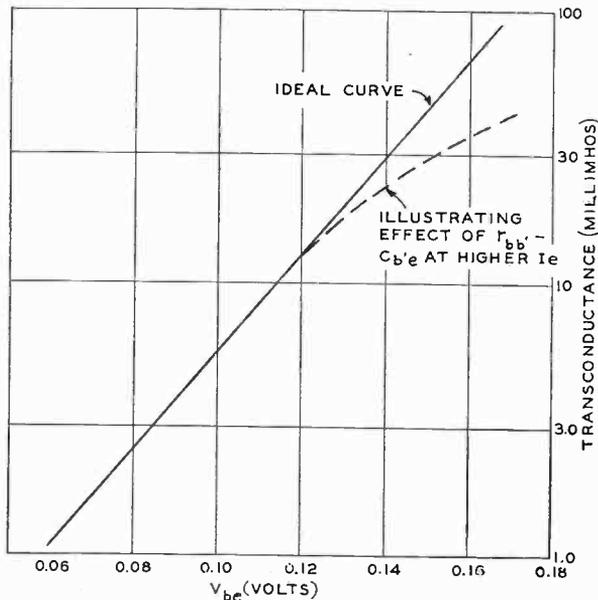


Fig. 4—Transconductance versus base-to-emitter voltage for ideal transistor (solid curve) and illustration of effect of r_{bb} - $C_{b'e}$ combination of an r-f alloy-junction transistor at 1000 kilocycles (dashed curve). $r_{bb} = 100$ ohms; $C_{b'e} = 2000$ micromicrofarads at $I_e = 1.0$ milliamperes.

CROSS MODULATION IN TRANSISTORS

A theoretical plot of $\log g_m$ versus V_{be} for a transistor is shown as the solid line of Figure 4. This curve assumes zero generator resistance

and zero base lead resistance. This assumption is valid for most circuits employing radio-frequency alloy-junction transistors operating in the broadcast band at emitter currents below 0.5 milliamperere, since at low currents the difference between V_{be} and $V_{b'e}$ is small, i.e., the voltage drop across the base lead resistance is small. In a typical r-f or i-f stage, selectivity and stability considerations dictate that the transistor be effectively driven from a generator whose source conductance is considerably greater than the transistor input conductance (see the r-f stage of the receiver described by Freedman, Stanley, and Holmes³). The effective interfering-signal generator impedance is even smaller than that of the desired signal (by an amount depending upon the frequency separation of the desired and undesired signals).

The dashed curve of Figure 4 illustrates the effect of the $r_{bb'}-C_{b'e}$ combination of an r-f alloy-junction transistor⁴ at higher emitter currents ($C_{b'e} = KI_e$). It may be seen that this combination serves to reduce the apparent g_m , reducing the slope of the log g_m characteristic and improving the cross modulation characteristic. Finite signal source resistance produces a similar improvement.

Generally, gain control of a transistor r-f stage is effected by emitter current reduction; the emitter current control range is typically from about 1.0 milliamperere to zero. Thus, the gamut of the dashed curve of Figure 4 is traversed. This curve coincides with the solid curve at low emitter currents (below about 0.5 milliamperere), and here the greatest susceptibility to cross modulation is always encountered.

Computed and measured cross modulation values for an r-f alloy-junction transistor are shown in Figure 5. The curve is drawn through the computed points. (Some difficulty may be encountered in making measurements at very low emitter currents, since the signal levels must be kept low enough that the transistor operating point is not materially shifted by application of signal voltage. Ultimately one is limited by the noise factor of the measuring equipment.) Agreement between measured and computed data is quite good.

When the transistor emitter is biased in the reverse direction ($g_m = 0$), the interfering voltage required for a given cross-modulation distortion increases very rapidly, since the transistor is then a linear passive network until pushed into conduction by the undesired signal.

³ L. A. Freedman, T. O. Stanley, and D. D. Holmes, "An Experimental Automobile Receiver Employing Transistors," *Proc. I.R.E.*, Vol. 43, p. 671, June, 1955; also, *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 458-476.

⁴ C. W. Mueller and J. I. Pankove, "A P-N-P Triode Alloy Junction Transistor for Radio-Frequency Amplification," *RCA Review*, Vol. XIV, p. 586, December, 1953; also, *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 189-201.

This fact can be employed to advantage in receiver design, as will be shown later.

Measured values for an r-f alloy-junction transistor for two frequencies of the undesired signal are shown in Figure 6. These curves demonstrate that the lower the frequency of the *undesired* signal, the higher the emitter current at which r_{bb} begins to have effect. The frequency of the *desired* signal is of itself unimportant.

The above analysis and measurements apply to an alloy-junction transistor, or more specifically to a transistor to which the pi-equivalent circuit of L. J. Giacoletto⁵ applies. The presence of other or different

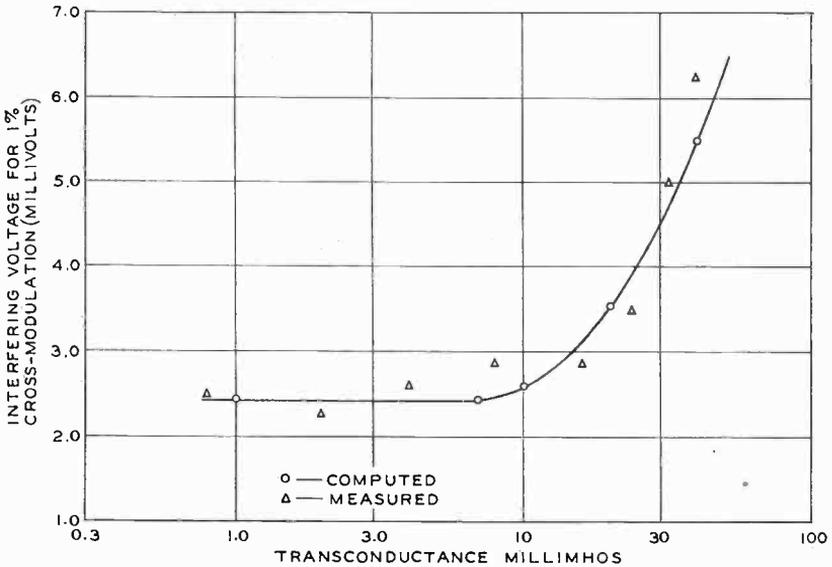


Fig. 5—Computed and measured cross-modulation data for an r-f alloy-junction transistor. Interfering voltage for 1 per cent cross modulation versus transconductance. Frequency of interfering signal = 1000 kilocycles.

elements, such as emitter-lead resistance, or a complex base-lead impedance, must be taken into account. These elements change the proportion of signal voltage which appears *across the junction* relative to the signal voltage impressed on the external transistor terminals, and have effects analogous to that of r_{bb} in the pi circuit.

COMPARISON OF A TRANSISTOR AND A TUBE IN AN AUTOMOBILE RECEIVER R-F STAGE

In comparing the performance of a transistor and a tube, *nor-*

⁵ L. J. Giacoletto, "Study of P-N-P Alloy Junction Transistor From D-C Through Medium Frequencies," *RCA Review*, Vol. XV, p. 506, December, 1954.

malized, interfering voltages must be obtained. An alloy-junction transistor will be compared with a 6BA6 tube; this tube is representative of the type used in automobile radio r-f stages. From the example shown previously, the 6BA6 can be expected to require 117 millivolts at its grid for 1 per cent cross modulation under weak-signal conditions. The transistor, at reduced emitter currents, requires 2.5 millivolts at its base for 1 per cent cross modulation. A typical turns ratio for the transistor input transformer might be 40:1. This means that the transistor, normalized with respect to the tube (assuming equal input circuit operating Q and dummy antenna) would require

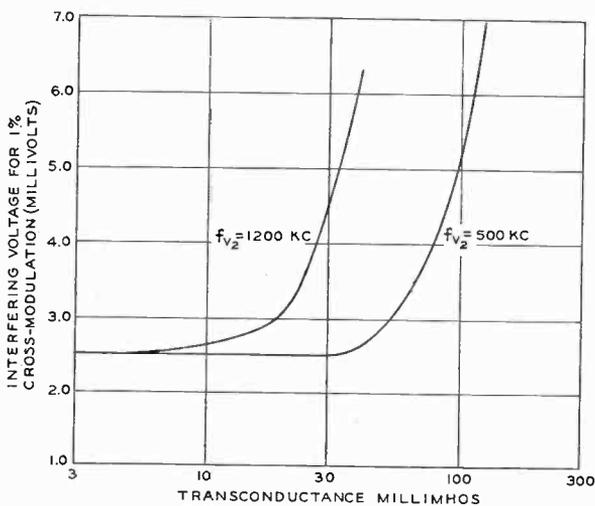


Fig. 6—Measured cross-modulation data; two interfering-signal frequencies for an r-f alloy-junction transistor.

$40 \times 2.5 = 100$ millivolts. This is minimum performance for both tube and transistor. In practice, the tube is biased where a relatively poor cross-modulation characteristic obtains for weak signals, while the transistor operates at a high emitter current and its performance is better than that shown above for the reduced-emitter-current condition. Under this condition, the transistor may be superior to the tube by a factor of 2. As signal increases, producing more a-g-c bias, the tube improves, operating at a more gradual $\log g_m$ slope, while the transistor performance decreases. At some moderate signal level the two are equal. If the transistor is not cut off beyond this point, the tube will excel; if the transistor is sufficiently cut off beyond this point, it will excel. The relative performance obtained when this latter condition is satisfied is shown in Figure 7. These curves show the performance of a high-quality tube receiver and that of the transistor receiver described in Reference (3) in which the a-g-c has been reapportioned

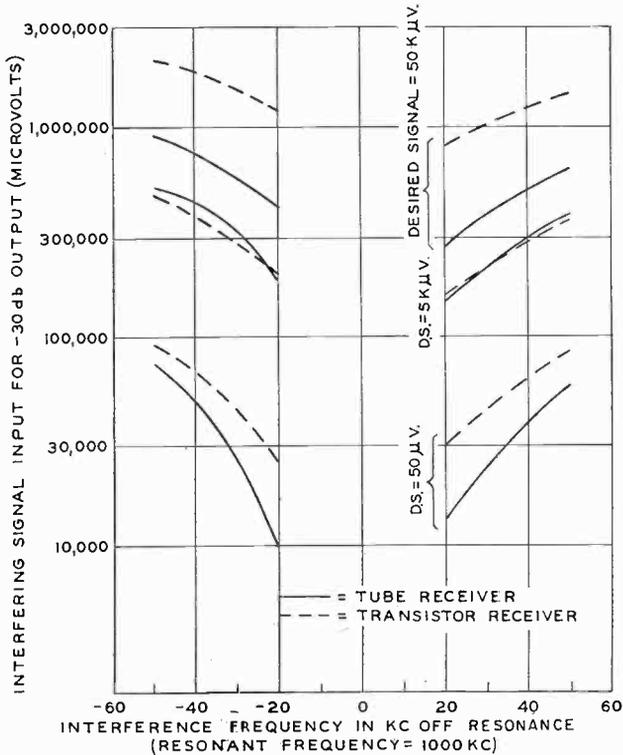


Fig. 7—Two-signal cross-modulation measurements (IRE Standard 4.05.16) for transistor receiver and high-quality tube receiver.

so that the r-f stage is cut off at approximately 5 millivolts input signal, and is reverse-biased beyond this point. The curves show that the transistor receiver is superior by about 2 to 1 at weak desired signal, equivalent at medium signal (transistor r-f stage just beyond cutoff), and superior by about 3 to 1 under strong-signal conditions.

The difference in slope between the transistor and tube receiver curves occurs because of a small difference in antenna circuit Q .

It should be pointed out that all of the cross-modulation distortion in these two receivers takes place in the r-f stage. Although no comparative data is presented here for receivers in which the first stage is a converter, it appears that performance can be expected to be comparable.

Analysis, data, and design experience accumulated to date indicate that the alloy-junction transistor can be employed successfully in a gain-controlled i-f or r-f stage. Performance with respect to a-g-c figure of merit and cross modulation can be as good or better than that which current practice indicates is acceptable.

PERFORMANCE OF A RADIO-FREQUENCY ALLOY JUNCTION TRANSISTOR IN DIFFERENT CIRCUITS

BY

L. J. GIACOLETTO

RCA Laboratories,
Princeton, N. J.

Summary—The small-signal performance of a radio-frequency p-n-p alloy junction transistor is evaluated over a frequency range of 0.01 to 10 megacycles with the aid of a digital computer. The computations are based upon a hybrid- π common-emitter equivalent circuit whose parameters were determined by measurements over a limited frequency range. The evaluation of terminal parameters, current and voltage amplification factors, maximum power amplification, and conjugately matched input and output admittances of the common-emitter, common-base, and common-collector circuit configurations are included. The maximum power amplification and conjugately matched admittances are in each case determined for three different circuit arrangements, unneutralized, partially neutralized, and unilateralized. The common-emitter circuit has the best over-all characteristics followed by the common-base and common-collector in order of increasing internal feedback. The circuit performance obtained will depend to a considerable degree upon the unilateralization circuit employed. A simple method is developed for quickly evaluating the approximate performance of a transistor with the aid of asymptotic low- and high-frequency values of the transistor terminal parameters.

INTRODUCTION

IT IS the purpose of this paper to compare the small-signal performance of a radio-frequency alloy junction transistor under conjugate matched conditions in the three circuit arrangements, common-emitter, common-base, and common-collector. For each of these circuit arrangements the performance will be considered first for the unneutralized circuit (transistor conjugately matched at input and output, but no auxiliary circuit elements), second for a partially neutralized circuit (auxiliary circuit elements added so that the feedback is non-reactive), and third for a unilateralized circuit (auxiliary circuit elements added so that the feedback is zero).

GENERAL CONSIDERATIONS

For the purposes of this paper, several p-n-p radio-frequency transistors¹ were examined briefly and a suitable unit selected. This tran-

¹C. W. Mueller and J. I. Pankove, "A P-N-P Triode Alloy Junction Transistor for Radio Frequency Amplification," *RCA Review*, Vol. 14, pp. 586-598, December, 1953; or *Proc. I.R.E.*, Vol. 42, pp. 386-391, February, 1954.

sistor was then carefully measured² over a limited frequency range, and from these measurements the hybrid- π common-emitter equivalent circuit was deduced.³ In accordance with the results of a previous study³ this equivalent circuit together with common-base and common-collector equivalent circuits derived therefrom were assumed to be correct representations of the transistor over the full frequency range of interest. With the aid of these equivalent circuits the circuit performance was set up in analytical form and computations carried out. The equivalent circuits used are simplified first approximations to exact representations and contain the minimum complexity required for satisfactory engineering results. The accuracy of the analytical results to be given is limited by the approximate nature of the equivalent circuits used.

Throughout this report, the d-c operating point was assumed to be the same: collector-to-emitter voltage, $V_{CE} = -6$ volts; collector current, $I_C = -1$ milliamperes; together with a base current, $I_B = -31.8$ microamperes. In addition to the different circuit arrangements, the variable parameter was the operating frequency ranging from 0.01 to 10 megacycles. The system of terminology employed together with a tabulation of suitable equations for making the circuit performance calculations can be obtained from a published paper.⁴ No attempt will be made to give the details of the calculations; rather, only the results will be presented.

There are numerous circuit arrangements that may be employed to obtain unilateralization, and these circuits will in general be different in their performance. S. J. Mason⁵ has shown that if unilateralization is accomplished with linear, lossless, bidirectional components, the power amplification of the resulting unilateralized circuit is independent of the terminal chosen as common to both input and output. However, the power amplification obtained with the Mason unilateralized circuit will, in some cases, be less than the power amplification obtained with a circuit arrangement where unilateralization is accomplished with lossy circuit components.

² L. J. Giacoletto, "Equipment for Measuring Junction Transistor Admittance Parameters for a Wide Range of Frequencies," *RCA Review*, Vol. 14, pp. 269-296, June, 1953.

³ L. J. Giacoletto, "Study of P-N-P Alloy Junction Transistor from DC through Medium Frequencies," *RCA Review*, Vol. 15, pp. 506-562, December, 1954.

⁴ L. J. Giacoletto, "Terminology and Equations for Linear Active Four-Terminal Networks Including Transistors," *RCA Review*, Vol. 14, pp. 28-46, March, 1953.

⁵ Sam J. Mason, "Power Gain in Feedback Amplifiers," *Transactions of the I.R.E. Professional Group on Circuit Theory*, Vol. CT-1, pp. 20-25, June, 1954.

The general two-generator nodal representation of a linear active device is as shown in Figure 1a. The $y_{12}V_2$ current generator in the input determines the amount of feedback, and y_{12} is often called the feedback admittance parameter. One method of altering the amount of feedback is to connect a feedback element, y_f , between the output, 2, and input, 1, terminals as shown in Figure 1b. This feedback element can be considered a part of a new linear active device whose two-generator nodal parameters as shown in Figure 2b are

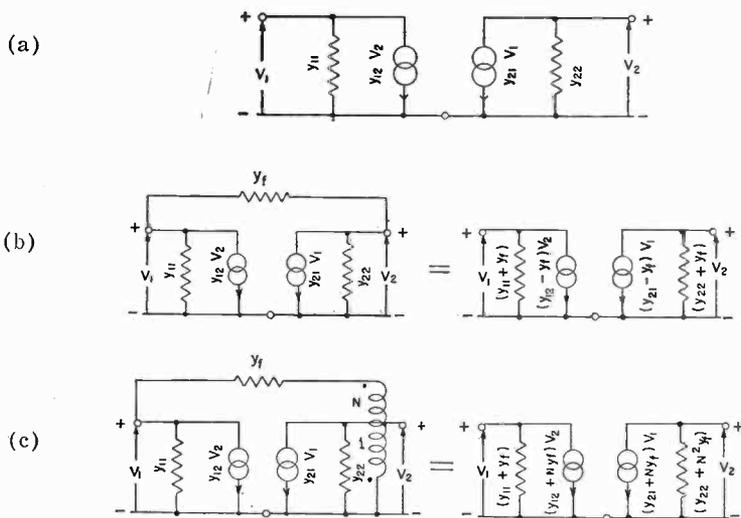


Fig. 1—Circuit arrangements for neutralization and unilateralization: (a) General two-generator nodal representation; (b) Circuit with direct-connected feedback element; (c) Circuit with transformer connected feedback element.

$$\begin{aligned}
 y'_{11} &= y_{11} + y_f, \\
 y'_{12} &= y_{12} - y_f, \\
 y'_{21} &= y_{21} - y_f, \\
 y'_{22} &= y_{22} + y_f.
 \end{aligned}
 \tag{1}$$

For partial neutralization, y_f is chosen so that y'_{12} is a pure conductance, i.e., since $y_{12} = g_{12} + jb_{12}$, $y_f = jb_{12}$ so that $y'_{12} = g_{12}$. For unilateralization, $y_f = y_{12}$ so that $y'_{12} = 0$. However, in order to satisfy this condition, a negative conductance will generally be required in y_f . However, y_f would then be a source of power, and such a circuit arrangement would yield an erroneous evaluation of the power capa-

bilities of the original device. For these reasons, a feedback arrangement as shown in Figure 1c may be preferable. Here, an ideal $N:1$ ratio transformer is used to reverse the phase and alter the amplitude of the output voltage, V_2 . The feedback element and transformer are now considered a part of a new linear active device whose two-generator nodal parameters as shown in Figure 1c are

$$\begin{aligned} y''_{11} &= y_{11} + y_f, \\ y''_{12} &= y_{12} + Ny_f, \\ y''_{21} &= y_{21} + Ny_f, \\ y''_{22} &= y_{22} + N^2y_f. \end{aligned} \tag{2}$$

For unilateralization, $y_f = -y_{12}/N$, and $y''_{12} = 0$. For the transistor to be considered and for all three circuit connections considered, a negative conductance would be required in the circuit of Figure 1b for unilateralization. Therefore, the circuit of Figure 1c with a phase reversing transformer of turns ratio, $N = 1$, and $y_f = -y_{12}$ will be employed when considering unilateralized operation; the circuit of Figure 1b with $y_f = jb_{12}$ will be employed when considering partially neutralized operation; and the circuit of Figure 1a will be employed when considering unneutralized operation. Only the parameters of the original device as shown in Figure 1a will be plotted; the parameters for the circuits of Figures 1b and c can be easily obtained with the aid of Equations (1) or (2).

It is worth noting that the maximum power amplification for the circuit of Figure 1c when adjusted for unilateralization will be the largest when the transformer turns ratio is chosen so that $N = \sqrt{g_{11}/g_{22}}$. This optimum turns ratio assumes zero losses in the transformer and corresponds to adjustment for minimum circuit losses. A unity turns ratio is employed in this report for simplicity.

COMMON-EMITTER CIRCUIT

As a result of terminal measurements and calculations, the common-emitter hybrid- π equivalent circuit is as shown in Figure 2. This circuit is assumed to hold throughout the range of frequencies of interest herein and forms the basis for calculations for this as well as subsequent circuit modifications. An important item in the frequency behavior of this circuit is the validity of the $r_{bb'}$, $g_{b'e}$, and $C_{b'e}$ lumped representation. Measurements to confirm the high-frequency validity of this lumped representation are shown in Figure 3. Here, the input impedance in an $R + jX$ form is plotted. The circuit in Figure 2 yields

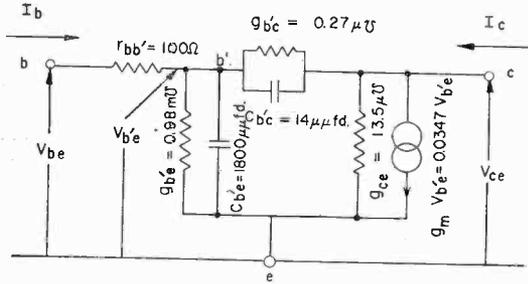


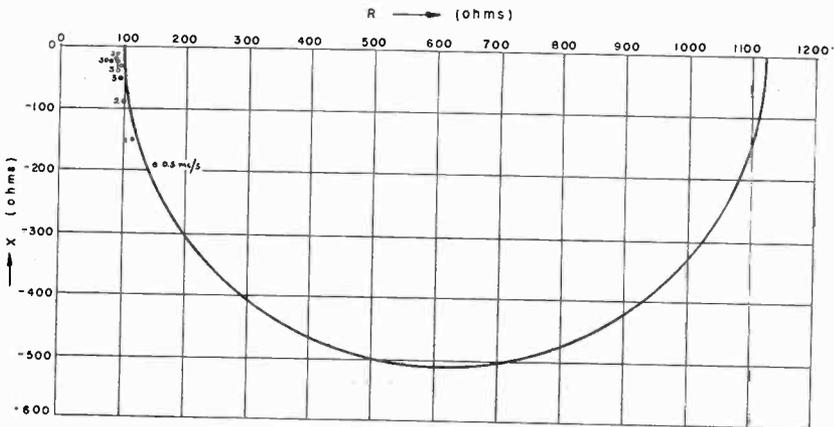
Fig. 2—Common-emitter hybrid- π equivalent circuit.

the solid semicircle. The data shown is measured values for frequencies as designated. The agreement is quite good.

Many of the parameters of Figure 2 can either be computed analytically or can be used to determine auxiliary transistor constants. The formulas used in these computations will not be given here but can be obtained in a published paper.³ The results of the calculations are tabulated in Table I together with associated measured values.

The four-terminal admittance parameters for the circuit of Figure 2 are shown graphed in Figure 4 over a frequency range from 0.01 to 10 megacycles. All curves of Figure 4 have an inflection point at a characteristic frequency, $f_c = 0.97$ megacycle, corresponding approximately to the $r_{bb'} C_{b'e}$ time constant; more exactly,

$$\omega_c = \frac{1 + r_{bb'} g_{b'e}}{r_{bb'} (C_{b'e} + C_{b'c})}$$



INPUT IMPEDANCE CIRCLE DIAGRAM

Fig. 3—Input impedance circle diagram.

Table I—Comparison of Measured and Computed Values

Quantity	Measurement	Computation
$r_{bb'}$	100 Ω
$g_{b'e}$	0.98 $m\Omega^{-1}$	1.31 $m\Omega^{-1}$
$C_{b'e}$	1800 $\mu\mu fd$
$g_{b'o}$	$g_{b'o}$ (intrinsic) = 0.13 $g_1 = 0.14$ <hr/> 0.27 $\mu\Omega^{-1}$	$L_b = 135$ microns $\tau_b = 4.2$ μsec
$C_{b'o}$	14 $\mu\mu fd$	Transition capacitance = 5.7 Transit-time capacitance = .6 <hr/> $C_{b'o} = 6.3$ $\mu\mu fd$
g_{oe}	13.5 $\mu\Omega^{-1}$	11.6 $\mu\Omega^{-1}$
g_m	0.0347 Ω^{-1}	0.0386 Ω^{-1}
α_{ob}	-34.6
μ_{cb}	-2,160
Maximum power amplification at 1 kc	41 db	41.7 db
g_{input}	1.0 $m\Omega^{-1}$	1.13 $m\Omega^{-1}$
g_{output}	33 $\mu\Omega^{-1}$	18.5 $\mu\Omega^{-1}$
Unneutralized maximum power amplification at 0.5 mc	oscillates	oscillates
Partially neutralized maximum power amplification at 0.5 mc	30.2 db	27.2 db
g_{input}	2.9 $m\Omega^{-1}$	4.7 $m\Omega^{-1}$
C_{input}	700 $\mu\mu fd$
g_{output}	67 $\mu\Omega^{-1}$	119 $\mu\Omega^{-1}$
C_{output}	27 $\mu\mu fd$
Unilateralized maximum power amplification at 0.5 mc	34 db	29.0 db
g_{input}	2.4 $m\Omega^{-1}$	2.8 $m\Omega^{-1}$
C_{input}	1192 $\mu\mu fd$
g_{output}	29 $\mu\Omega^{-1}$	88 $\mu\Omega^{-1}$
C_{output}	59 $\mu\mu fd$
Unneutralized maximum power amplification at 1 mc	18.8 db	19.5 db
g_{input}	10 $m\Omega^{-1}$	8.7 $m\Omega^{-1}$
C_{input}	700 $\mu\mu fd$
g_{output}	179 $\mu\Omega^{-1}$	240 $\mu\Omega^{-1}$
C_{output}	35 $\mu\mu fd$
Unneutralized maximum power amplification at 3 mc	7.9 db	9.6 db
g_{input}	10 $m\Omega^{-1}$	9.9 $m\Omega^{-1}$
C_{input}	810 $\mu\mu fd$
g_{output}	1000 $\mu\Omega^{-1}$	280 $\mu\Omega^{-1}$
C_{output}	16 $\mu\mu fd$
Figure of merit = maximum oscillation frequency	9.0 mc	9.3 mc

The computed transistor parameters of Table I were obtained with the aid of the following auxiliary transistor data: base conductivity, $\sigma_b = 77$ mhos/meter, n-type; collector area, $A_c = 1.14 \times 10^{-7}$ square meters (circular); $V_{CB} \approx V_{CE} = -6$ v; $I_C = -1$ ma; $I_B = -31.8 \mu\text{a}$; $I_{BS} = 1.27 \mu\text{a}$; $W_b = 20$ microns (0.80 mils) computed from the measured $C_{b'e}$. The computed circuit performance data is based upon the measured transistor parameters tabulated in Table I shown in Figure 2.

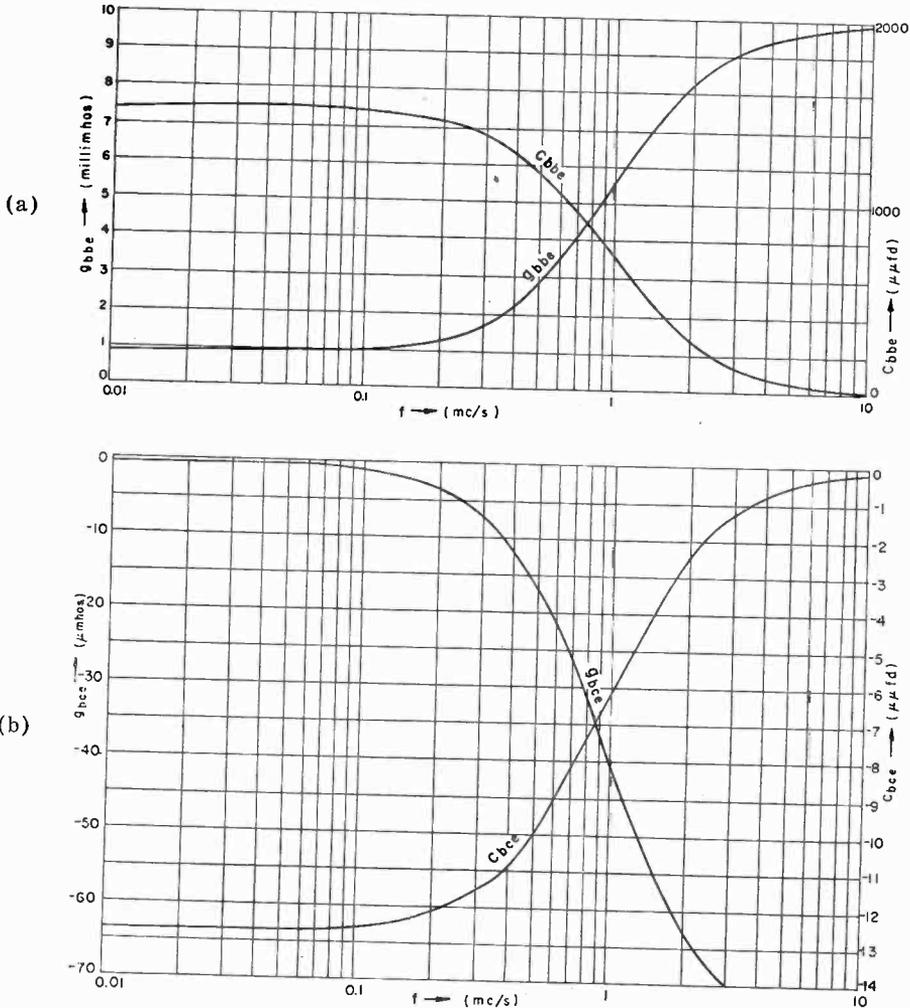


Fig. 4(a) and (b)—Common-emitter terminal admittance parameters.

In terms of this characteristic frequency, the various conductive and capacitive terminal parameter quantities, Q , can be expressed approximately for normal operating conditions in a general form,

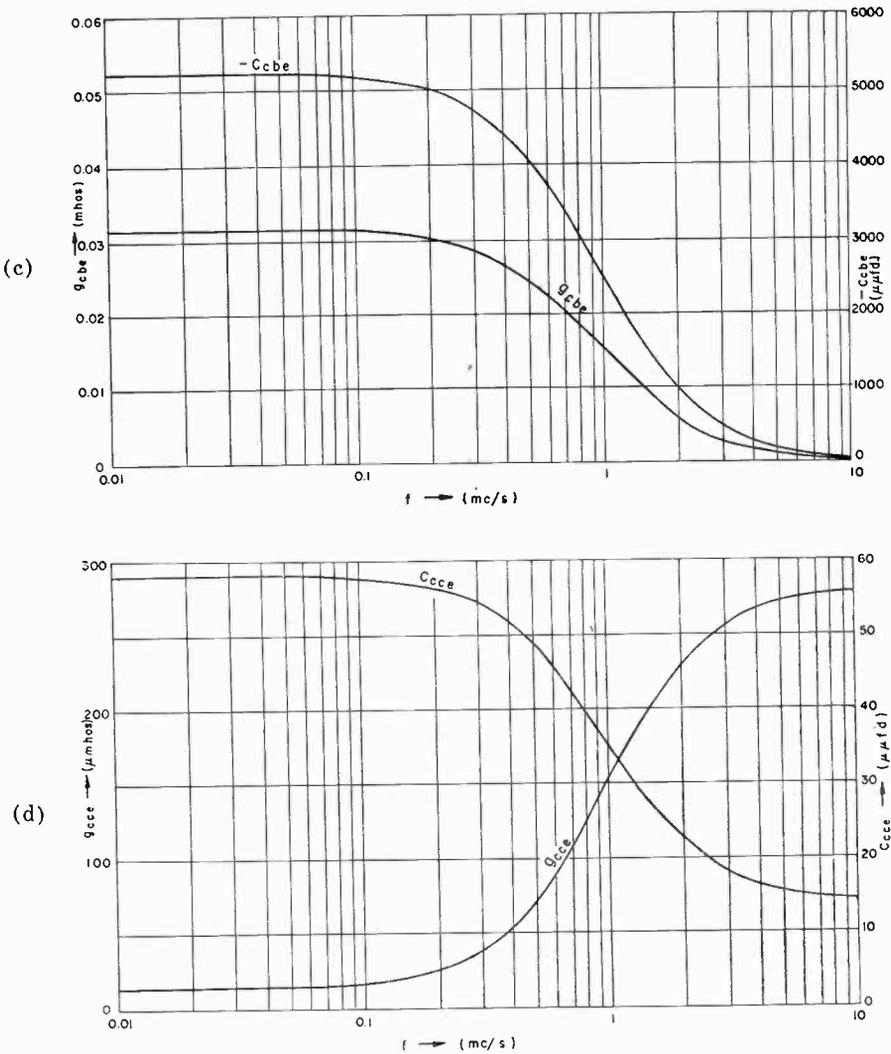


Fig. 4(c) and (d)—Common-emitter terminal admittance parameters.

$$Q = Q_1 + \frac{Q_2}{1 + (f/f_c)^2}.$$

Therefore the graphs shown in Figure 4 can be sketched out quickly by determining the limiting values at low frequencies, $(Q_1 + Q_2)$, and limiting values at high frequencies, (Q_1) . These two sets of limiting values are tabulated in Table II. On a logarithmic frequency abscissa

plot a straight line can be drawn through the $f = f_c$, $Q = Q_1 + Q_2/2$ point so as to intersect the $Q = Q_1 + Q_2$ line at a frequency $f_1 = f_c/2.718$ and the $Q = Q_1$ line at a frequency $f_2 = 2.718 f_c$. This method of approximately constructing the terminal parameters as a function of frequency from the hybrid- π equivalent circuit in Figure 2 is shown in Figure 5.

The current and voltage amplification factors for the common-emitter circuit are shown in Figures 6 (a) and (b) respectively. The vectors corresponding to these amplification factors are on the negative

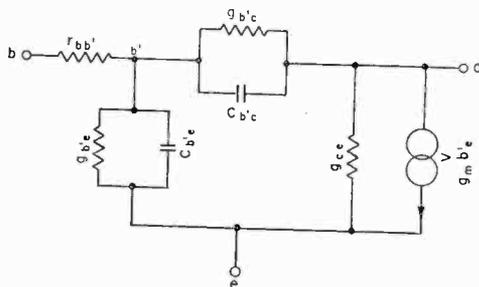


Table II—Approximate Limiting Values of Common-Emitter Circuit Terminal Parameters

Quantity	Low-frequency value	High-frequency value
g_{bbe}	$(g_{b'e}) (1 + r_{bb'} g_{b'e})^{-1}$	$(r_{bb'})^{-1}$
C_{bbe}	$(C_{b'e}) (1 + r_{bb'} g_{b'e})^{-2}$	0
g_{bcc}	$-(g_{b'c}) (1 + r_{bb'} g_{b'e})^{-1}$	$-C_{b'c} (r_{bb'} C_{b'e})^{-1}$
C_{bcc}	$-[C_{b'c} (1 + r_{bb'} g_{b'e}) - r_{bb'} g_{b'c} C_{b'e}] [1 + r_{bb'} g_{b'e}]^{-2}$	0
g_{cbe}	$g_m (1 + r_{bb'} g_{b'e})^{-1}$	$-C_{b'c} (r_{bb'} C_{b'e})^{-1}$
C_{cbe}	$-r_{bb'} g_m C_{b'e} (1 + r_{bb'} g_{b'e})^{-2}$	0
g_{cce}	$g_{ce} + [(1 + r_{bb'} g_m) g_{b'c}] [1 + r_{bb'} g_{b'e}]^{-1}$	$g_{ce} + g_{b'c} + (C_{b'e}) (C_{b'c})^{-1} g_m + (1 + r_{bb'} g_{b'e}) (r_{bb'})^{-1} (C_{b'e})^2 (C_{b'c})^{-2}$
C_{cce}	$[[(1 + r_{bb'} g_m) (1 + r_{bb'} g_{b'e}) - r_{bb'}^2 g_m g_{b'c} C_{b'e}] C_{b'c} [1 + r_{bb'} g_{b'e}]^{-2}$	$C_{b'c}$

Characteristic frequency, $f_c = \frac{1}{2\pi} \left[\frac{1 + r_{bb'} g_{b'e}}{r_{bb'} C_{b'e}} \right]$. The approximations

contained in the above results are those valid for normal transistor operation. For unusual conditions the exact formulation for the terminal parameters should be used.

real axis at low frequencies and rotate towards the positive imaginary axis at high frequencies. The current amplification factor always remains in the second quadrant, but the voltage amplification factor moves into the first quadrant at high frequencies. The current or voltage amplification obtained in a particular circuit will depend upon the load admittance and will differ from the corresponding amplification factors.

Maximum power amplification is obtained when both input and output are conjugately matched. The maximum power amplification together with input and output admittances corresponding to conjugate match conditions are shown in Figure 7 for three different circuit arrangements as discussed under General Considerations. Some values obtained from these curves are tabulated in Table I and can be compared with the corresponding measured values. Because of transistor internal feedback, the input and output conductances become zero, and

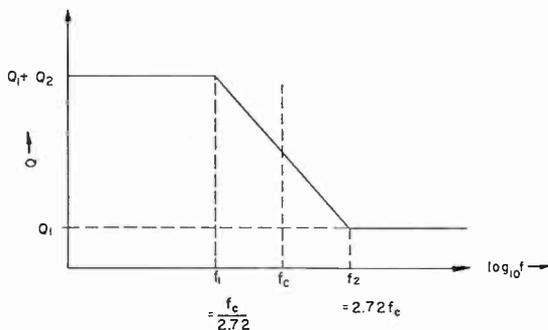


Fig. 5—Approximate construction of terminal parameters.

oscillations can be expected from about 0.016 to 0.38 megacycle if input and output are conjugately matched. Oscillations can be completely removed by partial neutralization as for instance by means of a suitable inductor between collector and base terminals. The amount of inductance required can be determined from Figure 4b since it must resonate with $-C_{bce}$ at the operating frequency. Unilateralization with the circuit of Figure 1c ($N = 1$) yields as much as 3 decibels more power amplification than the partially neutralized circuit at certain mid-frequencies; at higher frequencies the unilateralized power amplification is a little less than that for the partially neutralized or unneutralized circuit. The unilateralized power amplification could have been improved slightly if the optimum transformer turns ratio, $N = \sqrt{g_{bde}/g_{cce}}$, had been used at each frequency. The feedback admittance required to produce unilateralization can be obtained from Figure 4b: $y_f = -(g_{bce} + j\omega C_{bce})$ for $N = 1$. At higher frequencies

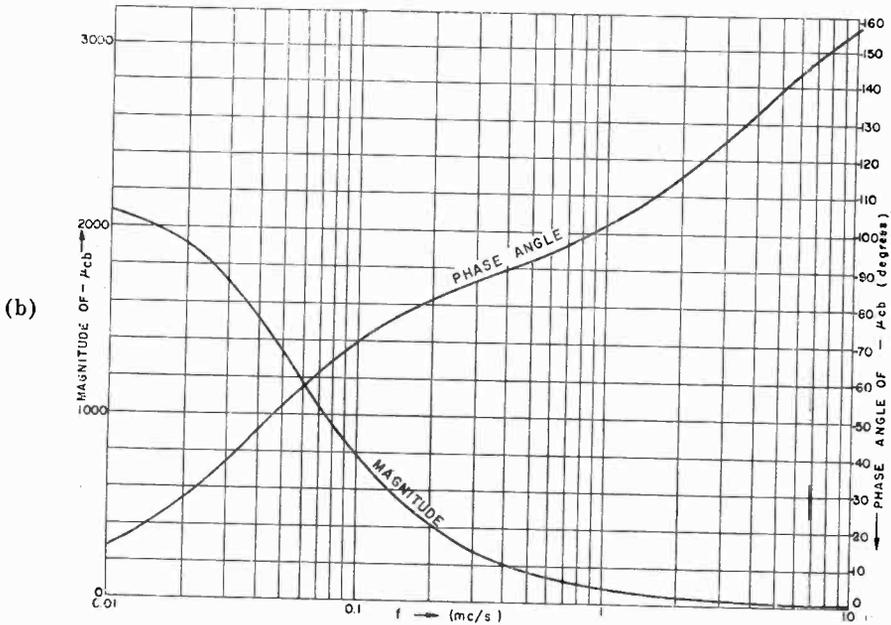
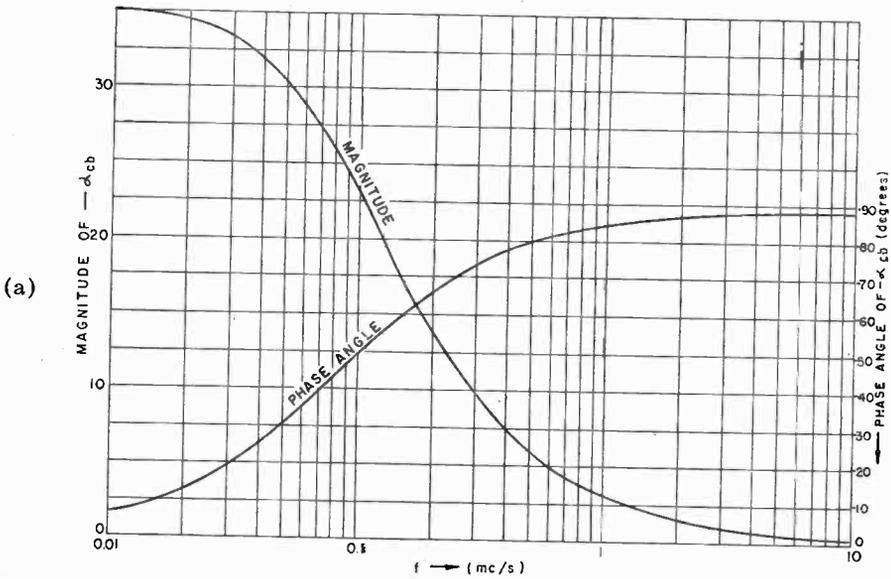


Fig. 6—Common-emitter current and voltage amplification factors.

the maximum power amplification is roughly independent of the circuit arrangement and decreases approximately 6 decibels per octave. The

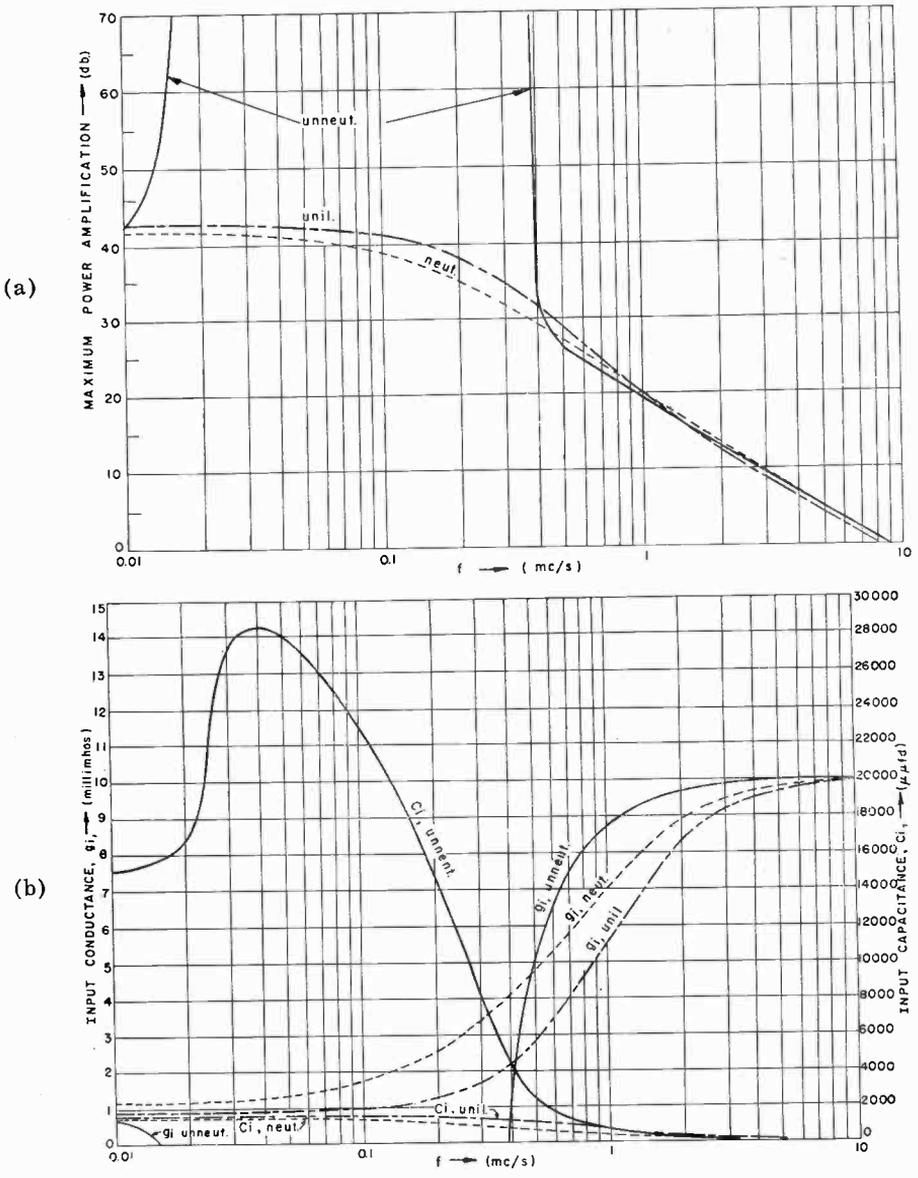


Fig. 7(a) and (b)—Common-emitter maximum power amplification and input admittance.

variation at higher frequencies follows closely the formulation³

$$P.A. \approx \frac{g_m}{4\omega^2 r_{b'b'} C_{b'e} C_{b'e}}$$

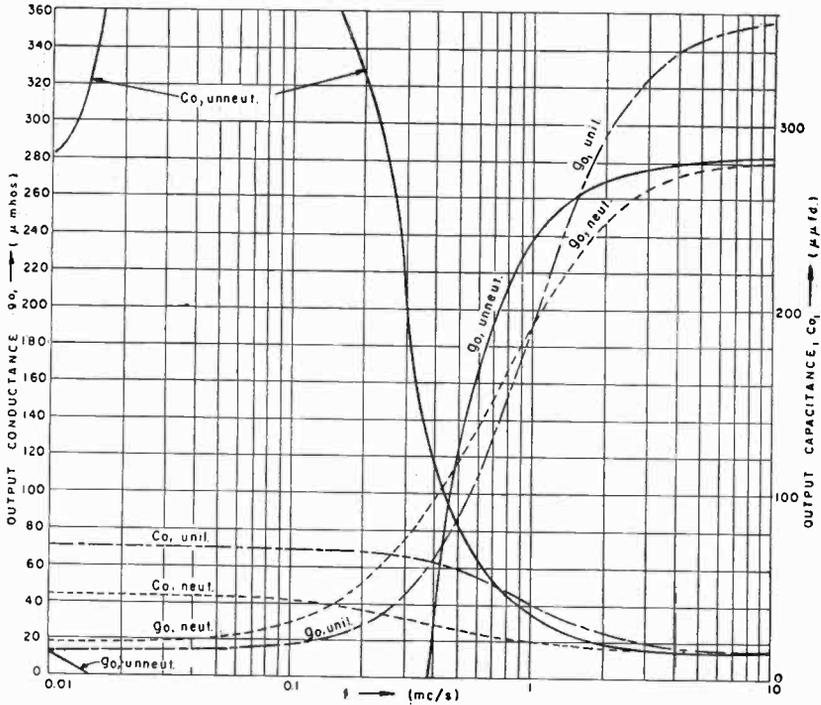


Fig. 7(c)—Common-emitter output admittance.

The approximate power amplification curve can therefore be sketched in quickly on a db-logarithmic frequency graph by determining the figure of merit angular frequency,

$$\omega = \frac{1}{2} \left[\frac{g_m}{r_{bb'} C_{b'e} C_{b'o}} \right]^{1/2}$$

corresponding to a power amplification of unity (0 decibels) and drawing in a straight line from this point with a 6 decibels per octave slope to the low-frequency maximum power amplification,

$$\frac{\alpha_{cb} \mu_{cb}}{(1 + \sqrt{1 - \alpha_{cb} \alpha_{bc}})^2}$$

In terms of the hybrid- π equivalent circuit of Figure 2, the low-frequency maximum power amplification is given approximately by

$$\text{P.A.}_{LF} \approx \frac{g_m^2}{g_{b'e}g_{ce} (1 + r_{bb'}g_{b'e}) \left[1 + \sqrt{1 + \frac{g_{b'e}g_m}{(1+r_{bb'}g_{b'e})g_{ce}g_{b'e}}} \right]^2}$$

This method of sketching in the power amplification curve will check the curve for the partially neutralized circuit reasonably well. The curve for the unilateralized circuit will exhibit somewhat higher power amplification and the difference may be particularly noticeable in the transition region from the constant low-frequency power amplification to the 6 decibels per octave variation at higher frequencies. The limits of the range of oscillations of the unneutralized circuit cannot be simply expressed.

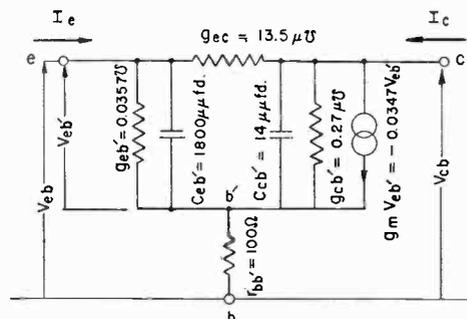


Fig. 8—Common-base hybrid- π equivalent circuit.

COMMON-BASE CIRCUIT

The common-base hybrid- π equivalent circuit shown in Figure 8 can be obtained by simple transformation of the circuit shown in Figure 2. The four-terminal admittance parameters for this circuit as shown in Figure 9 can be obtained by addition of the appropriate parameters shown in Figure 4. The curves therefore have the same general shape and same characteristic frequency as for the common-emitter terminal parameters. For normal operating conditions the approximate method of constructing the terminal parameters as shown in Figure 5 is applicable. In terms of the parameters in Figure 8,

$$\omega_c = \frac{1 + r_{bb'}(g_{eb'} + g_m)}{r_{bb'}C_{eb'}}$$

The asymptotic low-frequency and high-frequency values of the terminal parameters are tabulated in Table III.

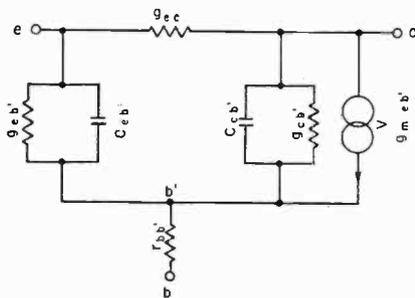


Table III—Approximate Limiting Values of Common-Base Circuit Terminal Parameters

Quantity	Low-frequency value	High-frequency value
g_{ceb}	$g_{eb'} [1 + r_{bb'}(g_{cb'} + g_m)]^{-1}$	$(r_{bb'})^{-1}$
C_{ceb}	$\frac{[(1 + r_{bb'}g_m) C_{eb'}]}{[1 + r_{bb'}(g_{eb'} + g_m)]^{-2}}$	$C_{cb'}$
g_{ccb}	$\frac{-g_{ec} - [r_{bb'}g_{cb'}g_{eb'}]}{[1 + r_{bb'}(g_{eb'} + g_m)]^{-1}}$	$\frac{-g_{ec} - g_{cb'} + [(1 + r_{bb'}g_m) C_{eb'}]}{(C_{eb'}r_{bb'})^{-1}}$
C_{ccb}	$-\frac{[[1 + r_{bb'}(g_{eb'} + g_m)]r_{bb'}g_{eb'}C_{eb'} + (1 + r_{bb'}g_m) r_{bb'}g_{cb'}C_{eb'}]}{[1 + r_{bb'}(g_{eb'} + g_m)]^{-2}}$	$-C_{cb'}$
g_{cbb}	$g_m [1 + r_{bb'}(g_{eb'} + g_m)]^{-1}$	$\frac{-g_{ec} - g_{cb'} + [(1 + r_{bb'}g_m) C_{eb'}]}{[r_{bb'}C_{eb'}]^{-1}}$
C_{cbb}	$r_{bb'}g_mC_{eb'} [1 + r_{bb'}(g_{eb'} + g_m)]^{-2}$	$-C_{cb'}$
g_{cbb}	$\frac{g_{ec} + [(1 + r_{bb'}g_{eb'})g_{cb'}]}{[1 + r_{bb'}(g_{eb'} + g_m)]^{-1}}$	$\frac{g_{ec} + g_{cb'} - (C_{eb'})^{-1}g_m}{(r_{bb'})^{-1} (C_{eb'})^2 (C_{eb'})^{-2}}$
C_{cbb}	$\frac{[[1 + r_{bb'}(g_{eb'} + g_m)] (1 + r_{bb'}g_{eb'})C_{eb'} + r_{bb'}^2g_mg_{cb'}C_{eb'}}{[1 + r_{bb'}(g_{eb'} + g_m)]^{-2}}$	$C_{cb'}$

Characteristic frequency, $f_c = \frac{1}{2\pi} \left[\frac{1 + r_{bb'}(g_{eb'} + g_m)}{r_{bb'}C_{eb'}} \right]$ The approx-

imations contained in the above results are those valid for normal transistor operation. For unusual conditions the exact formulation for the terminal parameters should be used.

The current and voltage amplification factors for the common-base circuit are shown in Figures 10 (a) and (b) respectively. The vectors corresponding to these amplification factors are predominantly in the fourth quadrant, but the voltage amplification factor moves into the third quadrant at high frequencies.

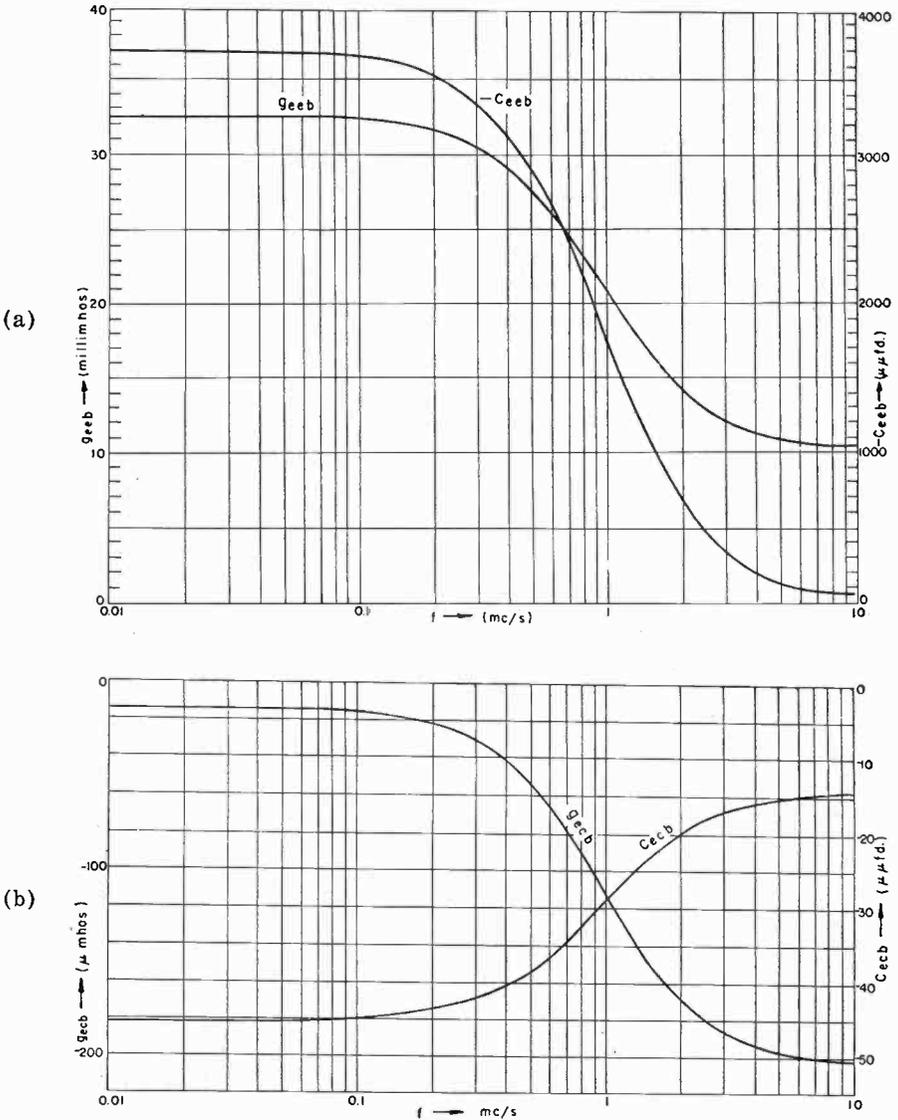


Fig. 9(a) and (b)—Common-base terminal admittance parameters.

The maximum power amplification together with input and output admittances corresponding therewith are shown in Figure 11 for unneutralized, partially neutralized, and unilateralized operation as discussed under General Considerations. The unneutralized circuit conjugately matched exhibits oscillations from 0.022 to 1.7 megacycles.

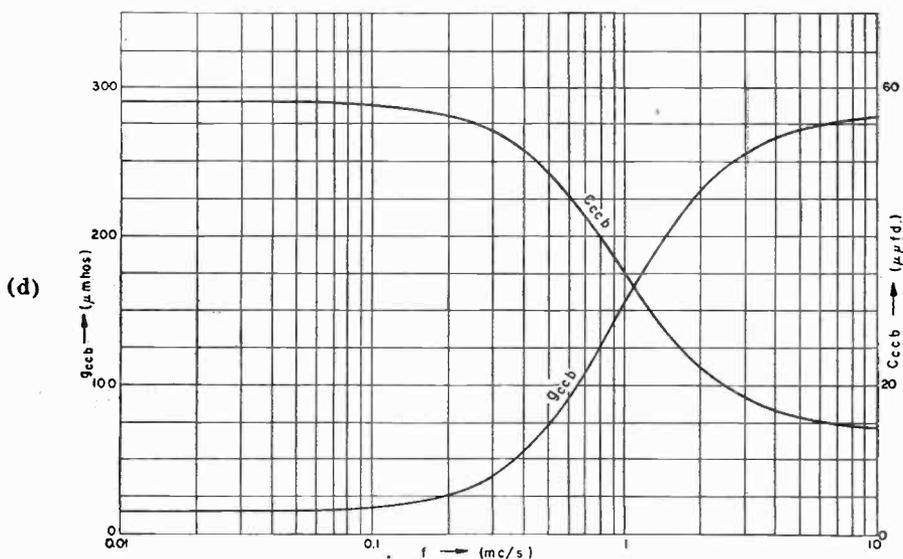
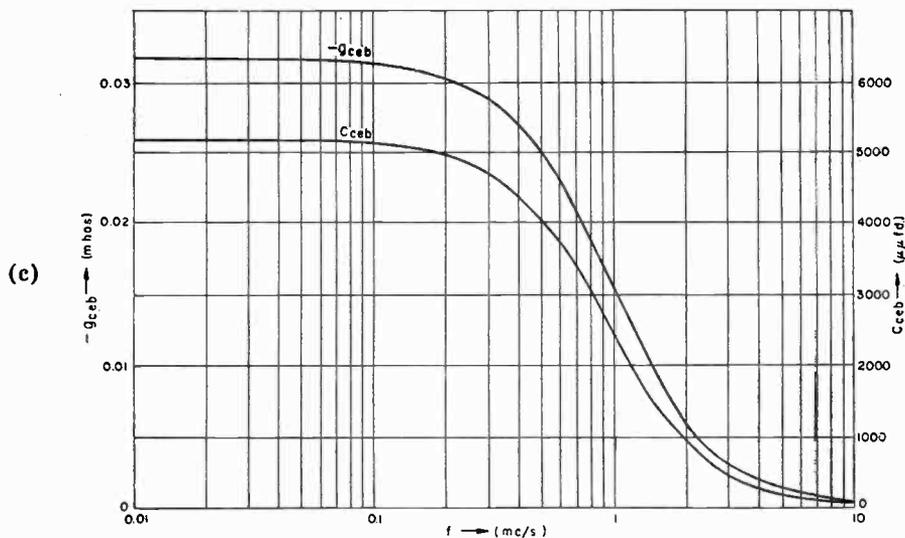


Fig. 9(c) and (d)—Common-base terminal admittance parameters.

Oscillations can be completely eliminated by partial neutralization with, for instance, an inductor connected between the collector and emitter terminals. The value of the inductor required can be ascertained from Figure 9(b) since the inductor must resonate with $-C_{ccb}$ at the operating frequency. In comparison with the common-emitter

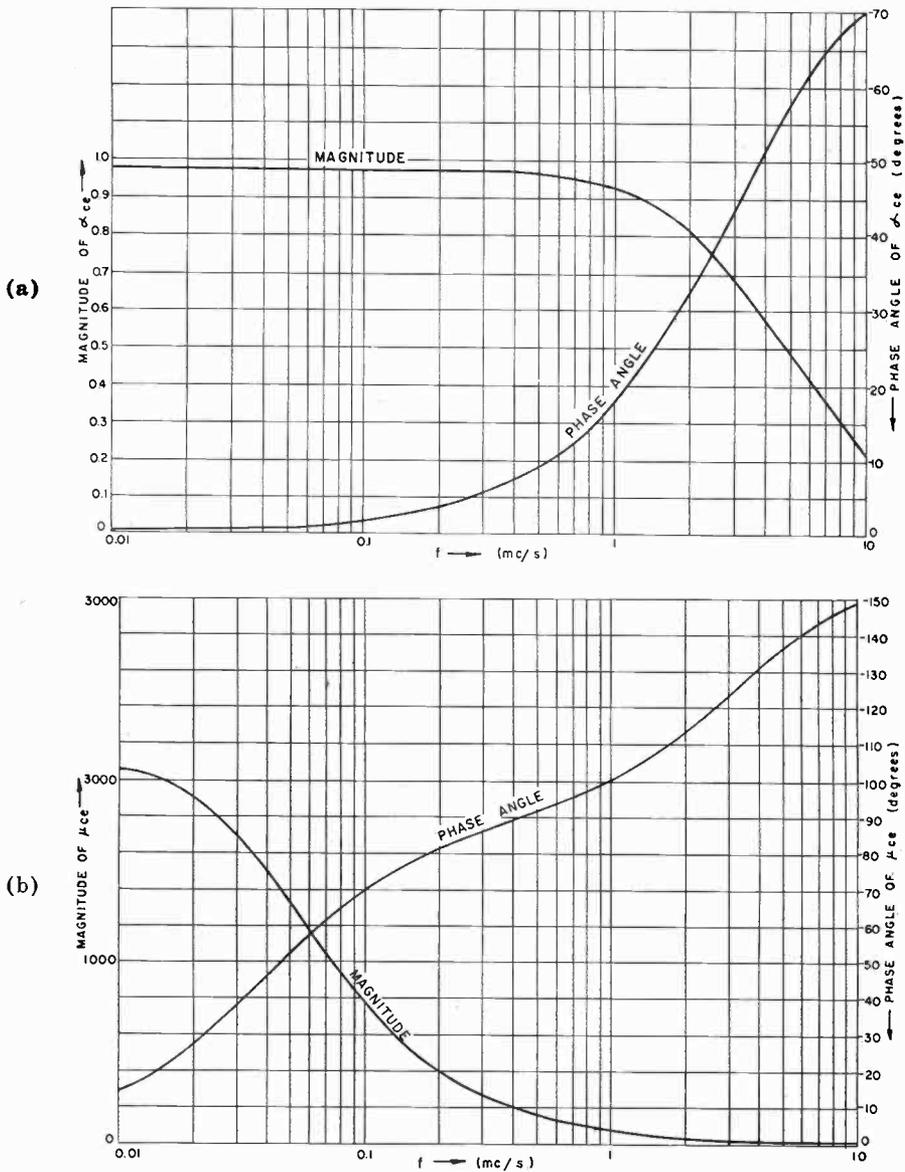


Fig. 10—Common-base current and voltage amplification factors.

power amplification (Figure 7(a)), the common-base power amplification exhibits a wider range of oscillations in the unneutralized operation. In addition, the unilateralized circuit, at least the type considered here, gives considerably poorer performance than the

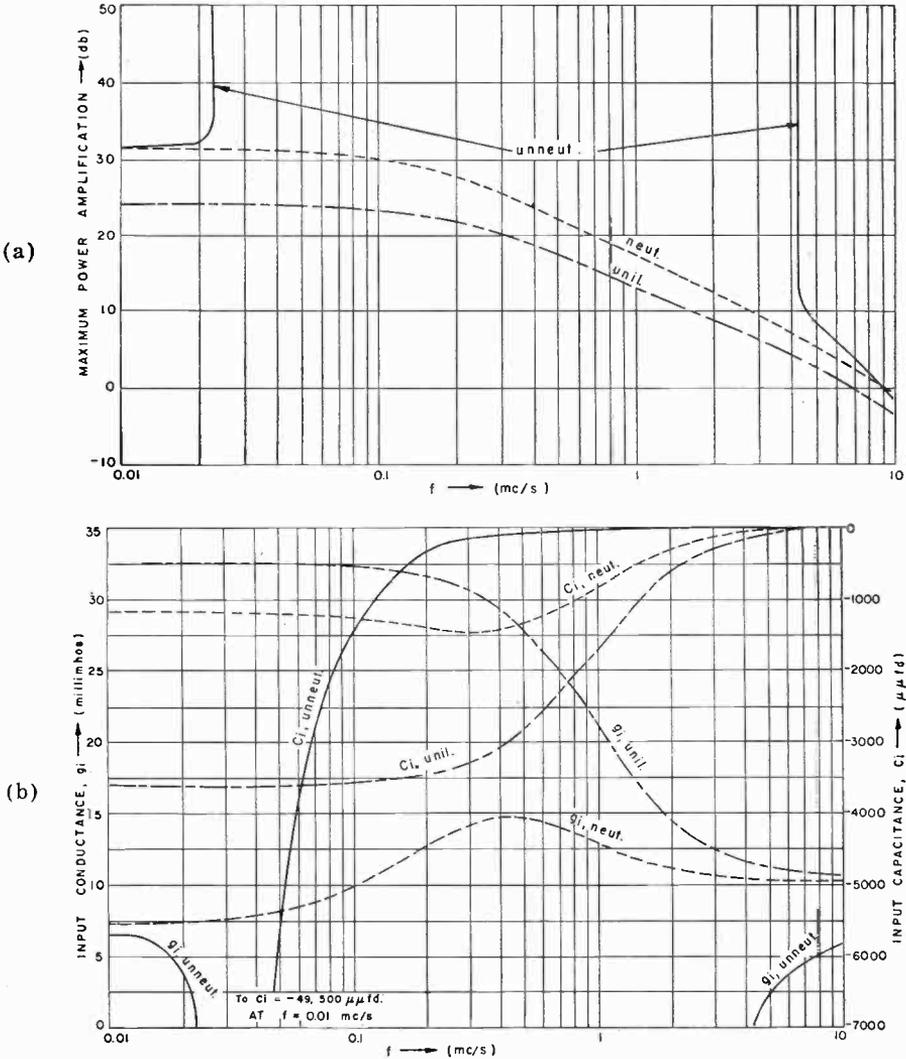


Fig. 11(a) and (b)—Common-base maximum power amplification and input admittance.

partially neutralized circuit. The frequency at which the partially neutralized power amplification is unity (0 decibels) is the same as the corresponding point in Figure 7(a). However, the rate of change of the partially neutralized power amplification with frequency is somewhat less than that of Figure 7(a) although at the highest frequencies the variation of 6 decibels per octave is obtained in both cases.

COMMON-COLLECTOR CIRCUIT

The common-collector hybrid- π equivalent circuit is shown in Figure 12 and can be obtained by simple transformations of the circuits of either Figures 2 or 8. The four-terminal admittance parameters for the common-collector circuit are shown in Figure 13 and can be obtained by addition of appropriate parameters of either the common-emitter or common-base circuits. The same approximate method of constructing the terminal parameters as discussed before (see Figure 5) can be employed. The asymptotic values required for this approximate construction are tabulated in Table IV.

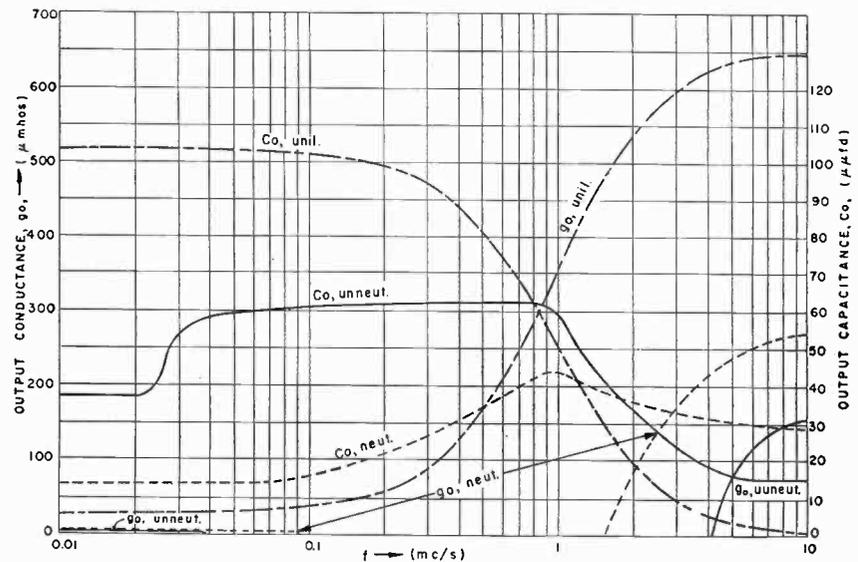


Fig. 11(c)—Common-base output admittance.

The common-collector current and voltage amplification factors are shown in Figure 14. The current amplification factor has a magnitude variation very similar to that of the common-emitter circuit (Figure 6(a)). The phase angle variation is very different however; a maximum phase angle of -71 degrees at 0.5 megacycle is noted for the common-collector circuit. The voltage amplification factor for the common-collector circuit (Figure 14(b)) is very close to unity with a small phase angle throughout the range of frequencies shown. A similar variation will be obtained for the voltage amplification as long as the load admittance is small compared with y_{oec} (Figure 13 (d)). This is the reason the common-collector circuit is often used as a coupling circuit in wide band amplifiers.

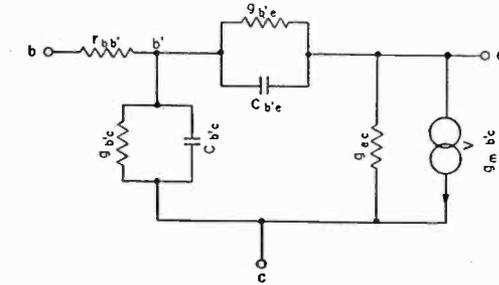


Table IV—Approximate Limiting Values of Common-Collector Circuit Terminal Parameters

Quantity	Low-frequency value	High-frequency value
g_{bbo}	$g_{b'e} (1 + r_{bb'} g_{b'e})^{-1}$	$(r_{bb'})^{-1}$
C_{bbo}	$C_{b'e} (1 + r_{bb'} g_{b'e})^{-2}$	0
g_{bco}	$-g_{b'e} (1 + r_{bb'} g_{b'e})^{-1}$	$-(r_{bb'})^{-1}$
C_{bco}	$-C_{b'e} (1 + r_{bb'} g_{b'e})^{-2}$	0
g_{eoo}	$g_m (1 + r_{bb'} g_{b'e})^{-1}$	$-(r_{bb'})^{-1}$
C_{eoo}	$-[(1 + r_{bb'} g_m) C_{b'e}] [1 + r_{bb'} g_{b'e}]^{-2}$	0
g_{eoc}	$g_{ec} + g_{b'e} (1 + r_{bb'} g_m) (1 + r_{bb'} g_{b'e})^{-1}$	$(r_{bb'})^{-1}$
C_{eoc}	$C_{b'e} + r_{bb'} g_m C_{b'e} (1 + r_{bb'} g_{b'e})^{-2}$	$C_{b'c}$

Characteristic frequency, $f_c = \frac{1}{2\pi} \left[\frac{1 + r_{bb'} g_{b'e}}{r_{bb'} C_{b'e}} \right]$. The approximations contained in the above results are those valid for normal transistor operation. For unusual conditions the exact formulation for the terminal parameters should be used.

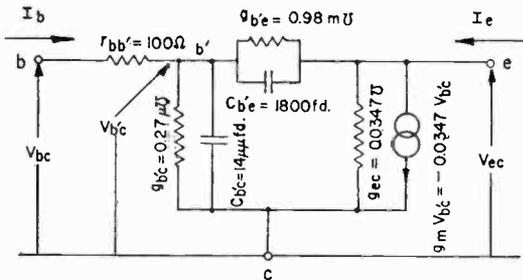


Fig. 12—Common-collector hybrid- π equivalent circuit.

The maximum power amplification together with input and output admittances corresponding therewith are shown in Figure 15 for the

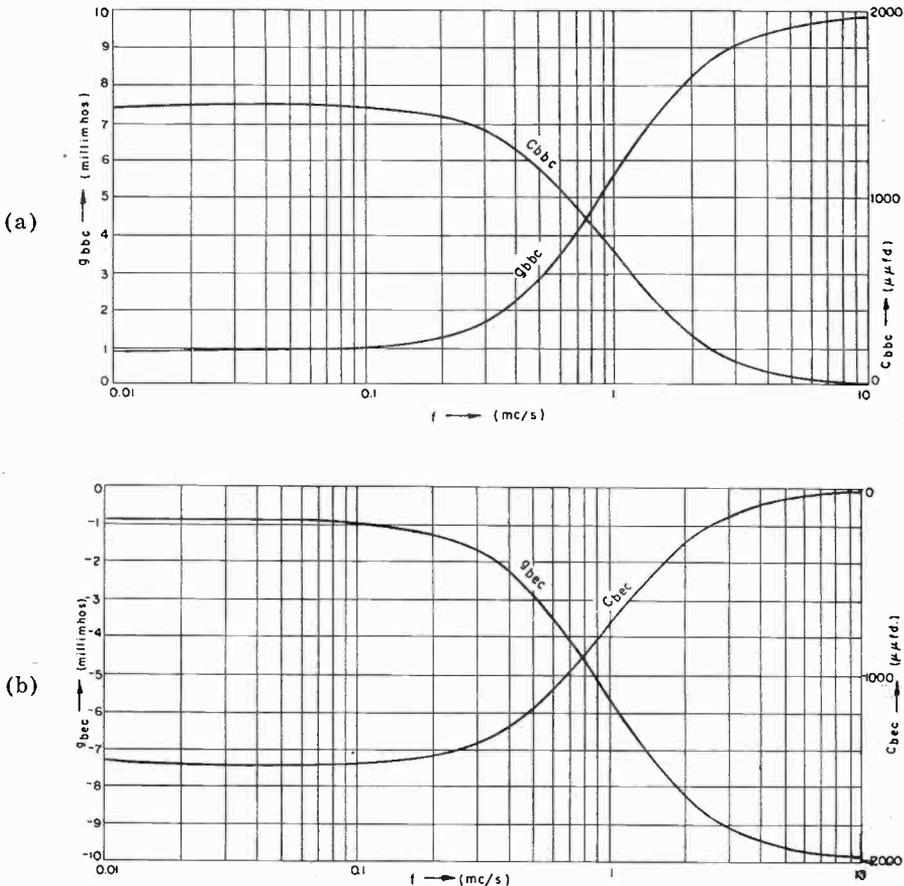


Fig. 13(a) and (b)—Common-collector terminal admittance parameters.

same operating conditions discussed previously. The unneutralized circuit conjugately matched oscillates from about 9.7 kilocycles to 9.0 megacycles. This region of oscillations is not completely suppressed by partial neutralization where oscillation is obtained from about 0.077 to 8.9 megacycles. The frequency at which the partially neutralized power amplification is unity (0 decibels) is 10 megacycles and is slightly higher than 9 megacycles for the common-emitter and common-base circuits. The unilateralized power amplification is free of oscillations and is small in magnitude due to the type of unilateralization employed. The feedback admittance required to produce unilateralization in accordance with the circuit of Figure 1(c) with $N = 1$ can be obtained from Figure 13(b) with $y_f = -(g_{bbc} + j\omega C_{bbc})$. This rela-

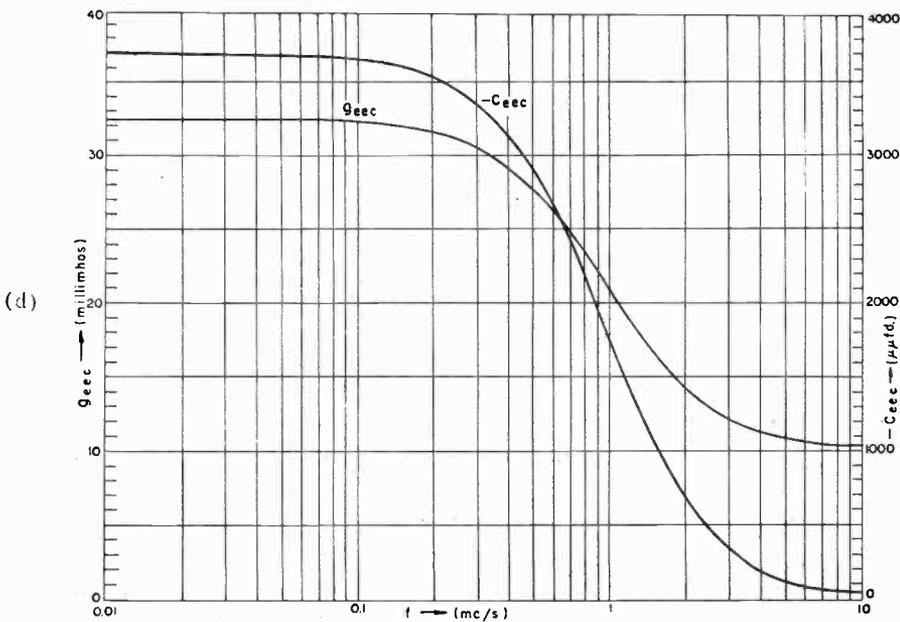
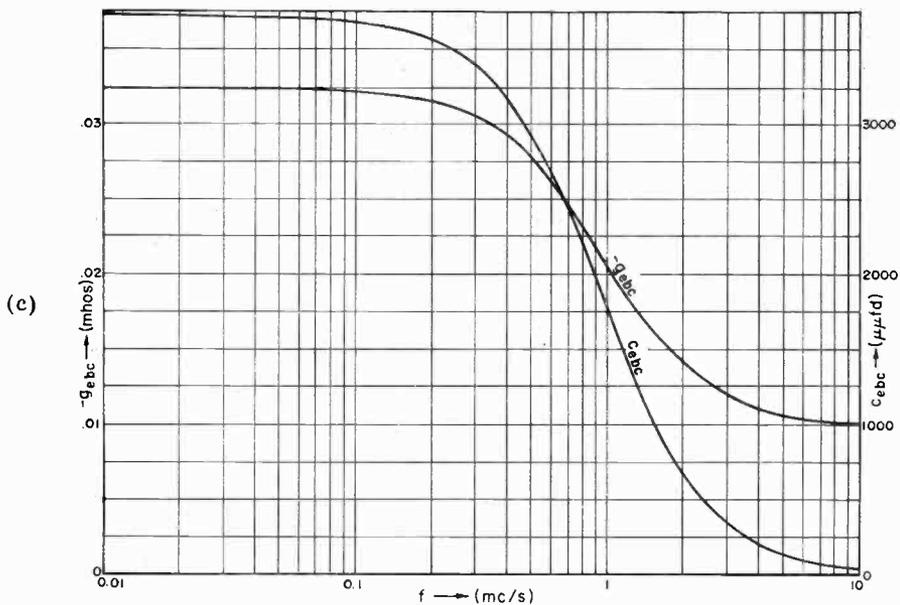


Fig. 13(c) and (d)—Common-collector terminal admittance parameters.

tively large feedback admittance in part accounts for the large input and output conductances for unilateralized operation shown in Figures 15(b) and (c).

CONCLUSIONS

The results given in this paper, although applicable in detail to a single transistor, will indicate in a qualitative manner the variation with frequency to be expected from a transistor in the three different circuit configurations and for three different circuit arrangements.

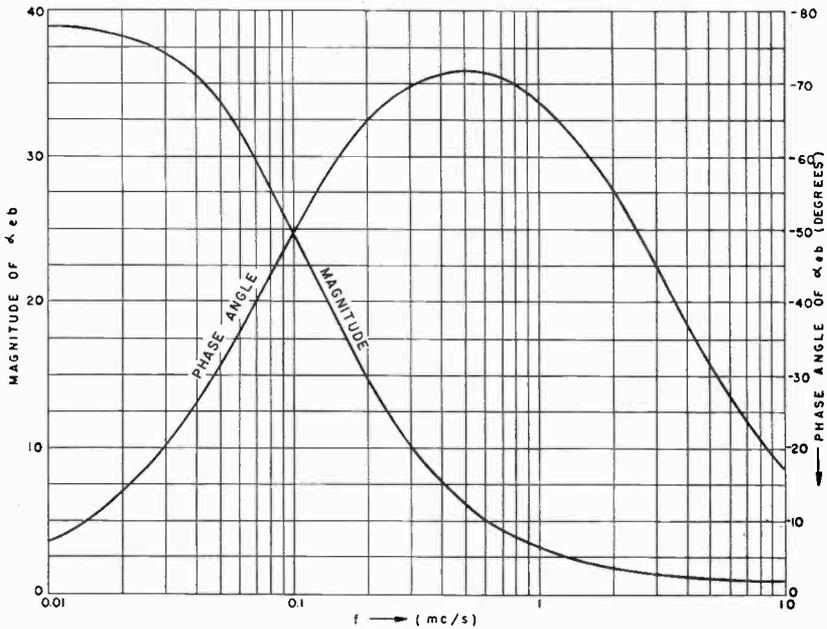


Fig. 14(a)—Common-collector current and voltage amplification factors.

The terminal parameters for the three circuit configurations can be sketched out quickly in an approximate manner using their asymptotic values as given in Tables II, III, and IV.

As may be expected, the circuit configuration that has more internal feedback oscillates over a wider frequency range and has less power amplification when unilateralized. For this reason, the common-emitter circuit has the best over-all operating characteristics, with the common-base circuit next and the common-collector last. This order of rating is based upon small-signal performance as studied herein and upon the circuit arrangements employed herein for partial neutralization

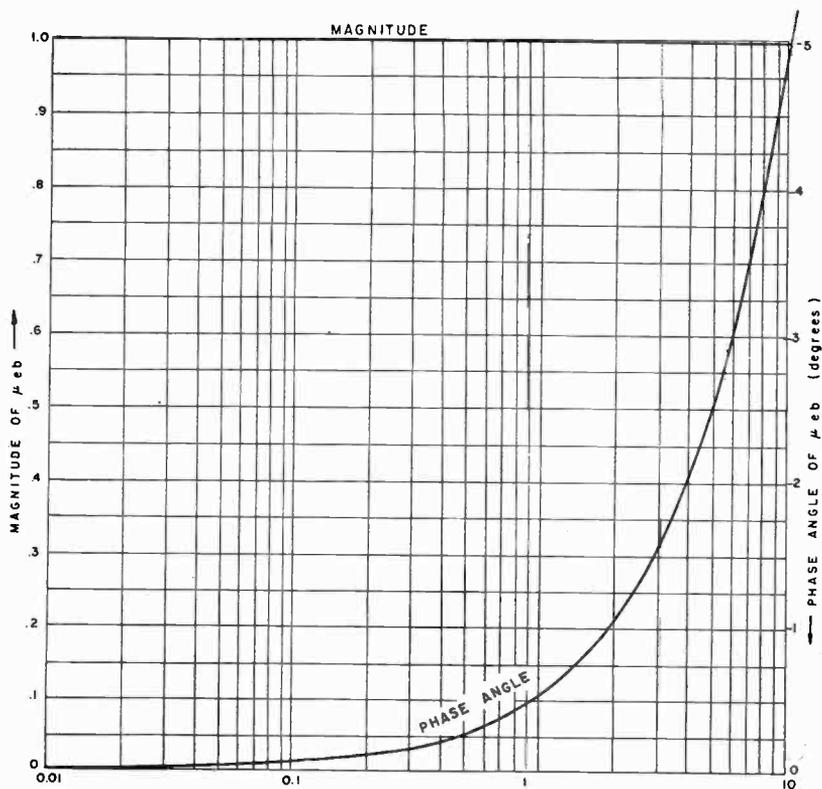


Fig. 14(b)—Common-collector current and voltage amplification factors.

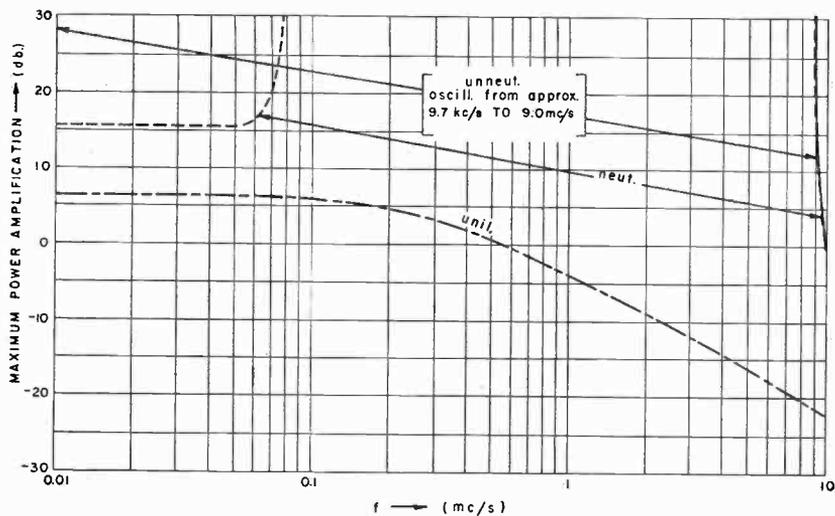


Fig. 15(a)—Common-collector input and output admittances.

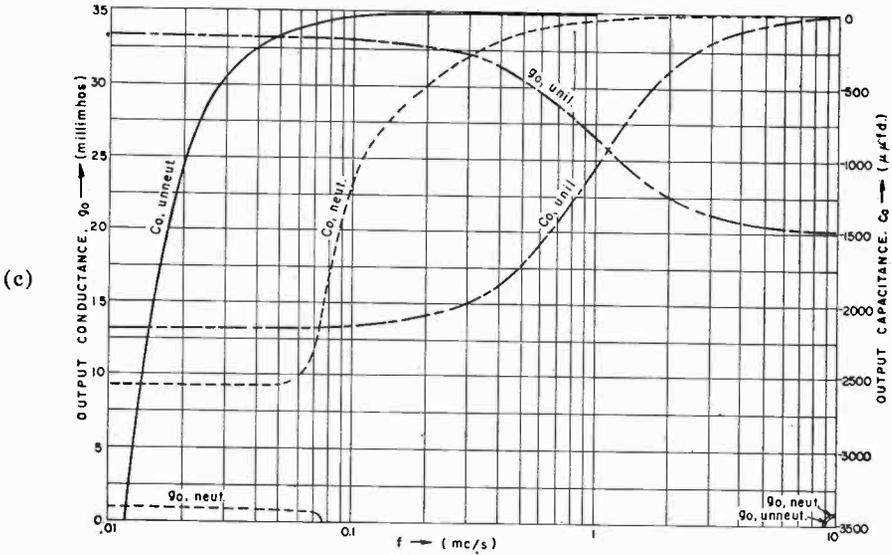
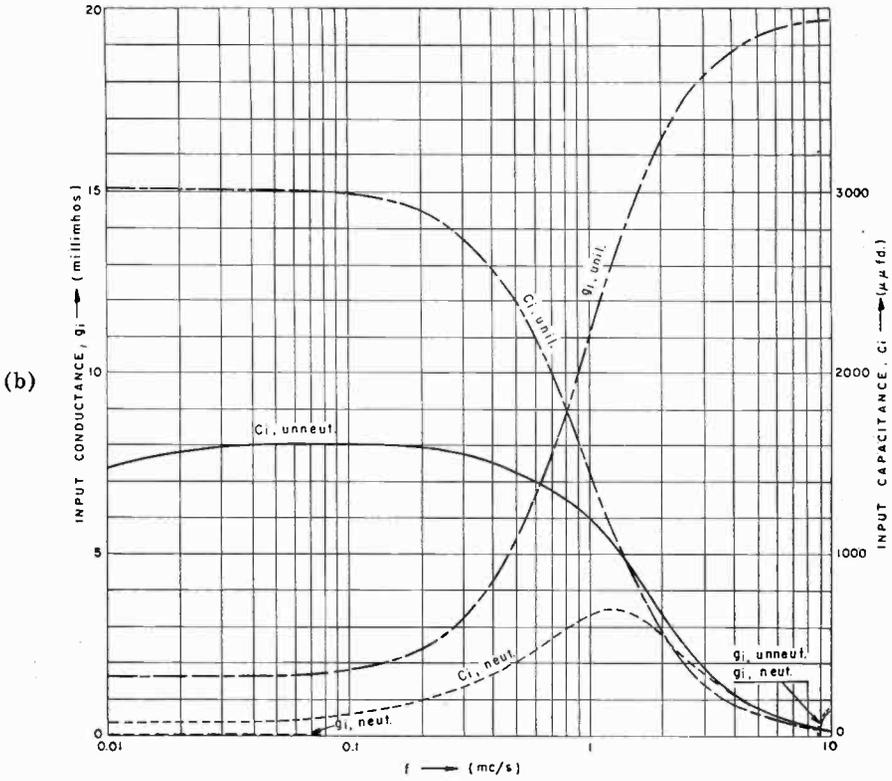


Fig. 15(b) and (c)—Common-collector maximum power amplification.

and unilateralization. It is apparent from the results obtained that the partially neutralized and the unilateralized circuits used, although well suited for the common-emitter circuit, are not well suited for the common-base and common-collector circuits. Thus, the unilateralization circuit proposed by Mason⁵ would at 0.5 megacycle give 25 decibels of power amplification in all three circuit arrangements, whereas, the unilateralization circuit employed herein gives 29 decibels (common-emitter); 17.5 decibels (common-base); and 0.7 decibel (common-collector).

ACKNOWLEDGMENT

The data for the various performance curves given in this paper was evaluated by personnel of the Computation Laboratory using a digital computer. Some of the measurements tabulated in Table I were made by D. D. Holmes and T. Scott.

AN EXPERIMENTAL AUTOMOBILE RECEIVER EMPLOYING TRANSISTORS*

BY

LARRY A. FREEDMAN, THOMAS O. STANLEY AND DAVID D. HOLMES

RCA Laboratories,
Princeton, N. J.

Summary—This paper describes an experimental automobile broadcast receiver utilizing nine experimental junction transistors in a superheterodyne circuit. The receiver operates directly from the six-volt storage battery without vibrator, power transformer, or rectifier. The average current drain, including that for two pilot lights, is approximately one-tenth that of a conventional automobile receiver.

The performance of this receiver is comparable to that of conventional automobile receivers. Particular emphasis has been placed on maintaining performance over a wide range of ambient temperature, both to accommodate the severe requirements specified for automobile service, and to establish the operability over such a temperature range of apparatus employing germanium transistors.

The receiver circuits and performance characteristics, including performance data for the ambient temperature range -40° to $+80^{\circ}$ C, are described in detail. Techniques which render circuit operation insensitive to variation of ambient temperature and which permit interchangeability of transistors are discussed.

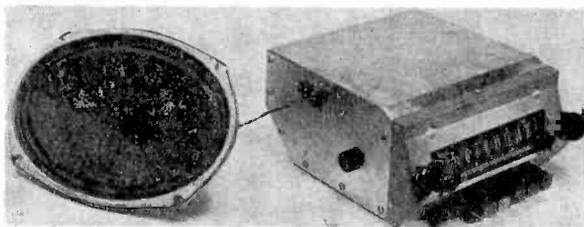


Fig. 1—Experimental transistor automobile receiver.

GENERAL DESCRIPTION

THE automobile receiver shown in Figure 1 uses nine experimental p-n-p alloy junction transistors in a superheterodyne circuit employing a 455-kilocycle intermediate frequency (i-f). A permeability-tuned r-f stage and a class-B transformer-coupled output stage are incorporated. The current drain of the radio is 250

* Reprinted from the *Proceedings of the Institute of Radio Engineers*, June, 1955.

milliamperes for the zero-signal condition; an additional 300 milliamperes are required by the two pilot lights. On a sustained tone, with maximum power output the total drain rises to one ampere.

The performance characteristics of the receiver at 20°C are as follows:¹

Sensitivity	2	microvolts
Noise Performance		
Input for 20 db S/N	12	microvolts
ENSI	0.4	microvolt
Power Output	2	watts
Selectivity (adjacent-channel attenuation)	41	decibels
AGC (Figure of Merit)	63	decibels

The operating temperature range of the receiver is -40° to +80°C. Circuit techniques provide stabilization for the most part, and thermistors are used for temperature compensation in the audio amplifier.

The receiver is constructed in three main sections as shown in Figure 2. These sections are the tuner assembly, in the foreground; the audio amplifier, mounted over the tuner; and the high-frequency part of the receiver, in the background.

CIRCUIT DESCRIPTION

A schematic diagram of the receiver is shown in Figure 3. Transistors V1 through V6 are experimental radio-frequency units² and serve as r-f stage, mixer, oscillator, two i-f stages, and second detector. The audio complement, V7 through V9, are experimental p-n-p transistors electrically similar to those described by Armstrong and Jenny³ but incorporate a mechanically and thermally improved mounting arrangement.

Tuner Transformers

The electrical elements of a push-button permeability tuner manu-

¹This data reflects the characteristics of the experimental transistors used; units of exceptionally high and exceptionally low performance were avoided. Long-term life performance of the transistors was not evaluated.

²C. W. Mueller and J. J. Pankove, "A p-n-p Triode Alloy-Junction Transistor for Radio-Frequency Amplification," *Proc. I.R.E.*, Vol. 42, pp. 386-391, February, 1954; also *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 189-201.

³L. D. Armstrong and D. A. Jenny, "Behavior of Germanium Junction Transistors at Elevated Temperatures and Power Transistor Design," *Proc. I.R.E.*, Vol. 42, pp. 527-530, March, 1954.

factured by Radio Condenser Corporation, Camden, N. J., were revised for operation in the transistor receiver. Three tightly coupled transformers are employed in the antenna-to-r-f input, r-f-to-mixer interstage, and oscillator circuits. Maximum unloaded Q of these transformers is obtained by making the effective diameter of the windings as large as the tuner dimensions will permit. Tracking of the signal circuits is insured by maintaining equal effective diameters of the respective coils. Oscillator tracking is accomplished by approximating a variable-pitch oscillator-transformer winding, the desired variable pitch being obtained by changing the coil-winder gear ratio at intervals

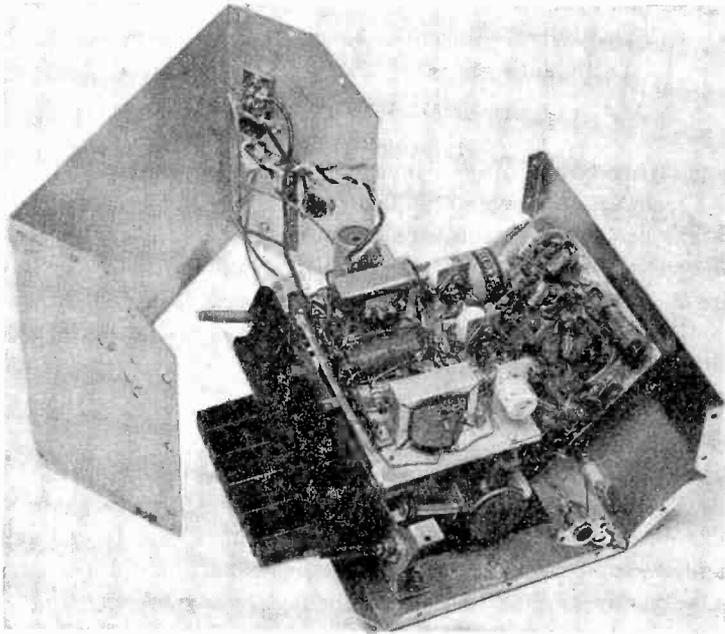


Fig. 2—Receiver with cover removed.

along the coil. The pertinent data for the antenna, interstage, and oscillator transformers is shown on the schematic.

The individual coil assemblies of the tuner are enclosed in shield cans. Magnetite sleeves of $\frac{3}{8}$ -inch inner diameter are provided for the antenna and interstage transformers. The powdered-iron tuning slugs are 1.2 inches long by 0.18 inch diameter; slug travel is about one inch.

Tuner Circuitry

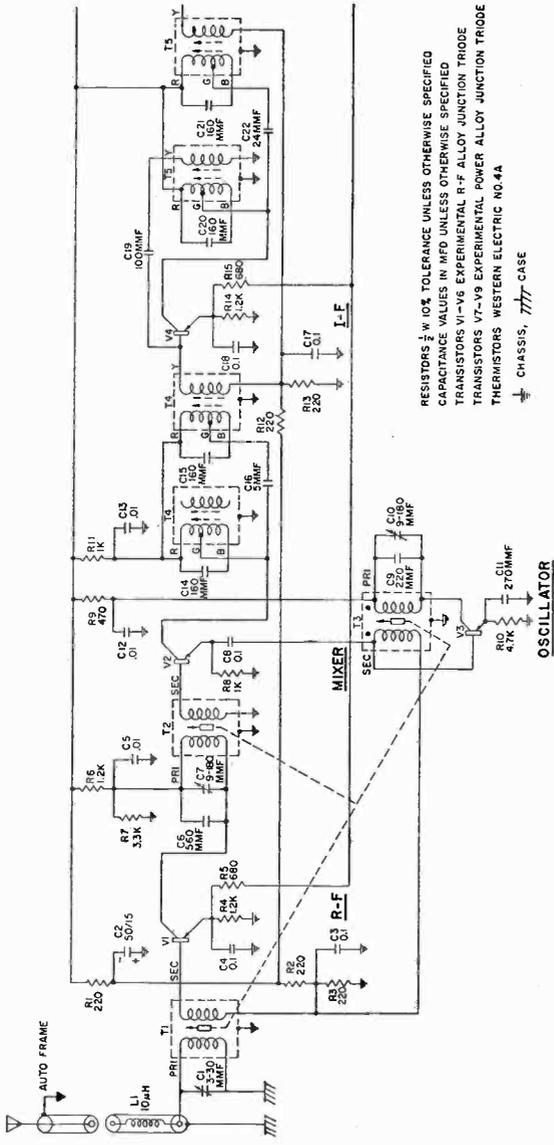
A conventional automobile rod antenna is employed; this antenna feeds the tuned primary of the antenna transformer, *T1*, as shown on

the schematic, Figure 3. Since the antenna is effectively a voltage source in series with a capacitor, the power fed to $T1$ will increase with increasing Q and decreasing shunt capacitance. The shunt capacitance includes the capacitance of the shielded lead from the antenna, the stray capacitance of the wiring, and the reflected r-f stage input capacitance. The coil Q is limited by the winding dimensions to 50-70 across the band. The primary is designed to tune with a total of 75 micromicrofarads, allowing for the above shunt capacitance as well as the antenna capacitance and a trimmer capacitor, $C1$ of 3 to 30 micromicrofarads.

The choice of operating Q for the antenna and interstage transformers represents a compromise between the rejection of image and intermediate frequencies, which increases with operating Q , and insertion loss, which also depends on operating Q . Minimum insertion loss between the antenna and the r-f stage input obtains for the "matched" condition, i.e., when the turns ratio of $T1$ is adjusted so that the transformer tuned impedance, referred to the secondary winding, is equal to the input impedance of the r-f stage. For this condition, the operating Q is one-half the unloaded coil Q . For the compromise used, the operating Q is eight-tenths of the unloaded Q , with an associated increase in insertion loss of 2 decibels above that applying for the matched condition. This compromise is obtained by adjusting the turns ratio so that the secondary tuned-impedance is one-fourth the input impedance of $V1$. A typical value for the input impedance at $V1$ at midband is 75 ohms.

The r-f stage bias arrangement renders the stage relatively insensitive to changes in ambient temperature. The base is returned to a low resistance bias source -1.5 volts at the junction of $R2$ and $R3$. Below automatic-gain-control (a-g-c) threshold, constant-emitter-current bias of 1.3 milliamperes obtains, via the emitter resistor, $R4$, in conjunction with $R5$, $R19$ and $R20$. The emitter is returned to ground for r-f by $C4$. The bias conditions with respect to a-g-c action are discussed in connection with the second detector.

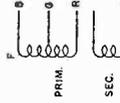
The r-f interstage transformer, $T2$, provides coupling from the collector of $V1$ to the base of the mixer, $V2$. In the mid-frequency range of the broadcast band the output impedance of the r-f transistor is 10,000 to 15,000 ohms, and the mixer input impedance is typically about 500 ohms. The operating Q of $T2$ is 15 to 20 and the transformer insertion loss is 3.7 decibels. This compromise between operating Q and insertion loss is obtained by adjusting the turns ratio of $T2$ so that the tuned primary impedance of $T2$ and the reflected input impedance of the mixer each equal the output impedance of $V1$. Tuning



RESISTORS $\frac{1}{2}$ W 10% TOLERANCE UNLESS OTHERWISE SPECIFIED
 CAPACITANCE VALUES IN MFD UNLESS OTHERWISE SPECIFIED
 TRANSISTORS VI-V6 EXPERIMENTAL R-F ALLOY JUNCTION TRIODE
 TRANSISTORS V7-V9 EXPERIMENTAL POWER ALLOY JUNCTION TRIODE
 THERMISTORS WESTERN ELECTRIC NO.4A

CHASSIS, 7777 CASE

UNIT	SEC	R	G	B
T4	4T	OT	55T	110T
T5	4T	OT	32T	110T
T6	4T	OT	40T	110T



#36 USE ON 0.125 I.D. 0.145 O.D. FIBRE FORM. RANDOM WOUND ON 0.230 WIDE X 0.125 D. ARBOR. SECONDARY WOUND UNDER PRIMARY-DISTRIBUTED OVER LENGTH OF COIL FINISHED COIL CUMAR DOPED.

TRANSFORMER CORES AND MECHANICAL ASSEMBLY MINATURE I-F TRANSFORMERS DESCRIBED (EXPERIMENTAL)

UNIT	PRIMARY	SECONDARY	FORM	GEAR SETTINGS					
	T	SSE	I.D. O.D.	FRONT	CAM	REAR			
T1	322	4-43	5	#35	.167	205	43-40	.25	28-100
T2	107	5-42	28	#35	.167	205	50-40	.25	88-100
T3	102	7-41	14	#35	.167	230	43-40	.25	52-60
	20								46-60
	25								72-60
	20								65-60
	17								71-60

WOUND ON MEISSNER MACHINE-PRIMARY PROGRESSIVE UNIVERSAL ON BAKELITE FORM SECONDARY DISTRIBUTED HAND WOUND OVER PRIMARY. #THE PRIMARY OF T3 IS A STEPPED-PITCH WINDING, THE STEPS ARE AS SHOWN

TUNER TRANSFORMERS

I-F TRANSFORMERS

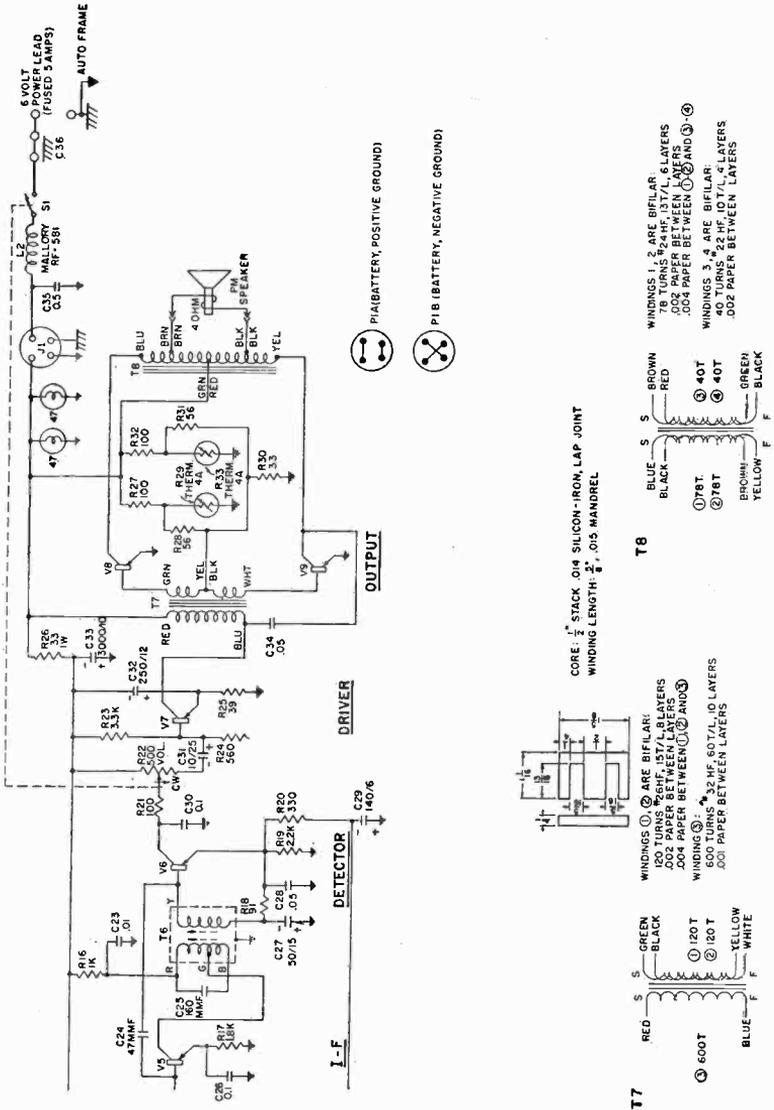


Fig. 3—Schematic of transistor automobile receiver.

of the primary of $T2$ is provided by $C6$ and $C7$. The gain of the r-f stage is about 20 decibels at midband.

The collector circuit of $V1$ returns to the tap on the bleeder formed by $R6$ and $R7$. The bleeder tap is bypassed to ground by $C5$. This arrangement decouples the collector circuit from the common supply, and serves as a voltage divider to reduce the collector voltage of the r-f stage. An improved signal-to-noise ratio is obtained by operation at reduced collector voltage.²

The tuned primary of the oscillator transformer, $T3$, in the collector circuit of the oscillator transistor, $V3$, affords an unloaded Q of from 40 to 60 over the oscillation frequency range. A relatively high tank capacitance is employed for stability. The number of secondary turns was determined experimentally for adequate mixer injection. The secondary applies feedback to the base of $V3$, and is returned to the -1.5 -volt bias source at the junction of $R2$ and $R3$. The $R10$ - $C11$ network in the emitter circuit of $V3$ introduces degeneration which reduces the net positive feedback in the oscillator circuit. The loading of the oscillator tuned circuit by the oscillator transistor input circuit is thus reduced so that oscillator tuning becomes relatively independent of the oscillator transistor input impedance. The reactance of the effective transistor base-to-emitter capacitance is in series with the relatively high reactance of $C11$. Thus, variation of the effective transistor base-to-emitter capacitance with frequency does not deteriorate oscillator tracking. (A decrease of the effective transistor base-to-emitter capacitance with increasing frequency arises from the presence of transistor base-lead resistance, which is in series with the emitter-junction capacitance. In the 1 to 2 megacycle range, the base-lead resistance and the reactance of the emitter-junction capacitance are comparable in magnitude.) The resistor $R10$, in conjunction with the base bias, provides sufficient starting emitter current to initiate oscillation.

The secondary of the interstage transformer, $T2$, is coupled to the base of the mixer, $V2$. The emitter is returned to ground by $R8$, which provides bias stability in a manner analogous to that provided by the emitter-return resistors of the amplifier stages. Approximately 0.4 volt r-m-s of oscillator injection is applied to the emitter through capacitor $C8$; the corresponding average emitter current is 0.4 milliampere. The optimum magnitude of oscillator injection depends in part on the magnitude of the emitter-return resistor; in this instance optimum injection is typically 0.35 volt r-m-s. If injection decreases below this value, the conversion gain falls rapidly, while the conversion gain decreases relatively slowly with increasing injection. Somewhat

greater than optimum injection insures interchangeability of mixer transistors and minimizes variation of conversion gain with small changes in oscillator injection.

Since the coupling capacitor, C_8 , presents a low impedance at the intermediate frequency and the secondary of T_3 has a low impedance at both signal and intermediate frequencies, R_8 is effectively bypassed to ground for both the mixer input and output signals.

The mixer output impedance is typically 60,000 ohms and is relatively independent of signal frequency, while the input impedance is about 500 ohms at midband and decreases somewhat with increasing frequency. The mixer stage conversion gain is about 20 decibels at midband.

I-F Amplifier⁴

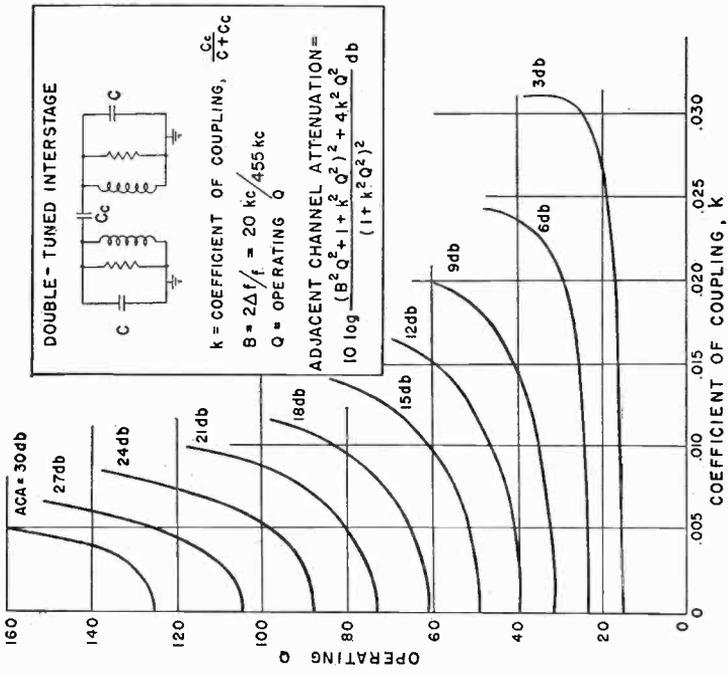
The three i-f interstage coupling networks consist of two capacitively coupled double-tuned transformers, T_4 - T_4 and T_5 - T_5 , and a single-tuned transformer, T_6 . These transformers embody an experimental arrangement and are ferrite cored. They are approximately $\frac{3}{8}$ inch in diameter by $\frac{1}{2}$ inch high. The i-f amplifier contributes 37 decibels of adjacent channel attenuation (ACA); the selectivity provided by the various interstage networks is so apportioned as to minimize over-all insertion loss for this ACA. The choice of turns ratio of a single-tuned transformer for minimum insertion loss at a prescribed ACA is described in the literature.⁴

For the double-tuned circuits, the ACA is determined by the coefficient of coupling, k , as well as the operating Q 's, as shown in Figure 4a. The insertion loss of the double-tuned circuits is determined by k , by the operating Q 's, and by the unloaded coil Q , as shown in Figure 4b. That optimum combination of k and operating Q which results in minimum insertion loss for a prescribed ACA may be determined from these curves. The variation of minimum insertion loss versus ACA is shown in Figure 5; a corresponding curve for a single-tuned circuit is plotted for comparison.

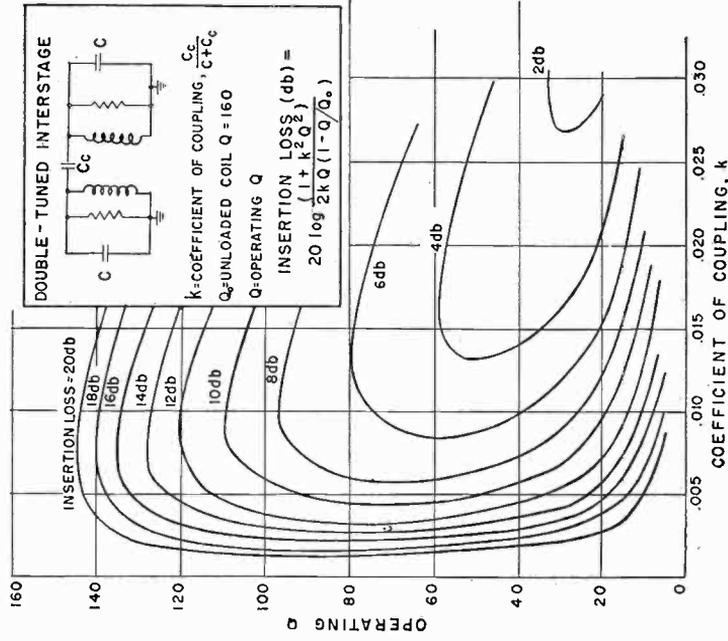
The ACA's provided by T_4 - T_4 , T_5 - T_5 , and T_6 are 15.5, 15.5, and 6 decibels respectively, and the insertion losses are 5, 5, and 2.5 decibels respectively. The amplifier provides about 50 decibels gain from the base of the first i-f amplifier, V_4 , to the base of the second detector, V_6 .⁵

⁴The i-f, detector, and a-g-c circuits are similar to those described in D. D. Holmes, T. O. Stanley, and L. A. Freedman, "A Developmental Pocket-Size Broadcast Receiver Employing Transistors," *Proc. I.R.E.*, Vol. 43, pp. 662-680, June, 1955; also *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 477-496.

⁵By way of example, i-f transistors having $r_{bb} = 75$ ohms, $\alpha_{cb} = 20$, $C_{bc} = 11$ micromicrofarads, and $C_{be} = 0.001$ microfarad ("alpha cutoff" of 6 megacycles) would give typical performance.



(a) Variation of adjacent-channel attenuation with operating Q and coefficient of coupling, for $i-f = 455$ kilocycles, channel separation ≈ 10 kilocycles.



(b) Variation of insertion loss with operating Q and coefficient of coupling, for unloaded $Q \approx 160$.

Fig. 4—Design curves for double-tuned i-f coupling transformers.

Biasing of the first i-f stage, to which a-g-c is applied, is similar to that of the r-f stage. The second i-f stage is constant-emitter-current biased by means of the emitter resistor, $R17$. The i-f stage bases return to the tap on a separate bleeder formed by $R12$ and $R13$. The bleeder tap is bypassed to ground by $C17$. This bleeder circuit, in combination with the collector-circuit decoupling of the second i-f stage provided by $R16-C23$, serves to isolate the i-f amplifier from the r-f and mixer stages. Neutralization of the i-f stages is provided by $C19$ and $C24$.

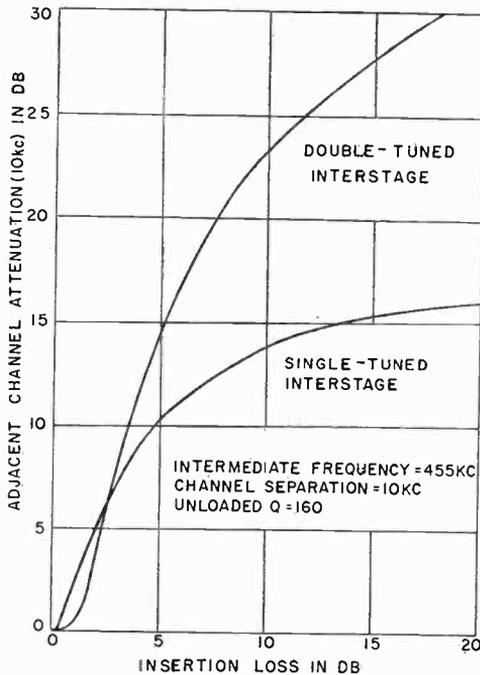


Fig. 5—Minimum insertion loss versus ACA for double- and single-tuned circuits.

Second Detector and A-G-C

As shown in the schematic, a-g-c is applied to the r-f and first i-f stages. Emitter current control is employed, and is obtained from the audio-a-g-c detector, $V6$, via $R20$, $R15$, and $R5$. The bias arrangement of $V1$ and $V4$ holds the currents in $R4$ and $R14$ essentially constant; the controlled stages are constant-emitter-current biased for the zero-signal condition. An increase in the detector collector current produces a proportionate decrease in the emitter currents of the controlled stages. The magnitude of these current at zero signal (about 1.5 milliamperes

per stage) is sufficient that the initial decrease in current results in little change in gain, introducing the desired delay in a-g-c action. The control current at the flat portion of the a-g-c characteristics is 3 milliamperes, an additional 1 milliamperere of detector current flows through the shunt resistor, *R19*. The total detector current of 4 milliamperes is then sufficiently large to minimize the effects of elevated-temperature saturation current in the detector transistor. In the design of an a-g-c circuit of this type, adjustment of the zero-signal emitter currents of the controlled stages provides a convenient level control of the flat portion of the a-g-c characteristic.

The R-C networks in the emitter circuits of the detector, r-f, and first i-f amplifiers provide filtering of the r-f and audio components of a-g-c current generated in the detector.

Detector linearity is improved by degeneration provided by *R18* in the return path for the audio component of detector emitter current. Audio output is taken from the collector through the volume control, *R22*. Decoupling from the collector supply is provided by *R21* and *C30*.

Audio Amplifier

The audio output of the detector is applied, via *C31*, to the base of the audio driver, *V7*. Since the dynamic output impedance of the detector transistor is high, the output impedance of the detector stage is essentially the resistance of the volume control irrespective of volume control setting. The current available for driving the following stage is therefore proportional to the resistance intercepted between the slider and the common supply. Nearly all of this available output current flows into the base of the driver, since the driver presents an input impedance on the order of 75 ohms. A base-to-collector signal current gain of 40 to 50 is realized. For medium audio frequencies, the emitter of the driver is bypassed, by *C32*, to the common supply. The low-frequency response of the amplifier is down 3 decibels at that frequency at which the input impedance at the base of the driver (the product of the driver current gain and the reactance of *C32*) equals the parallel resistance of the volume control, *R22*, and the base-bias bleeder resistors, *R23* and *R24*. The roll-off frequency for operation into a 4-ohm dummy load is approximately 120 cycles. The response into the speaker load remains elevated to below 80 cycles due to the mechanical low-frequency resonance of the loudspeaker.

Control of the driver stage operating current is provided by the base-bias bleeder in conjunction with the emitter resistor *R25*. The driver collector current is approximately 15 milliamperes at moderate temperatures; it increases to 30 milliamperes at 80°C and drops to 10 milliamperes at -40°C.

The driver collector is transformer-coupled to the bases of the push-pull output stage by transformer *T7*. The interstage and output transformer data are given on the schematic.

In addition to the usual transformer design compromises (efficiency, low-frequency response, etc.), the following requirements are met in the design of the interstage transformer:

1. The impedance reflected to the driver collector is low enough that driver overload does not occur before overload of the output stage.
2. The series resistance of the primary winding is low enough that the loss of driver supply voltage is tolerable.
3. The two halves of the secondary winding are sufficiently tightly coupled together to avoid transient voltages when current shifts from one output transistor to the other. This is accomplished by bifilar winding of the secondary.

A current gain of 5 is afforded by the interstage transformer from the collector of the driver to the bases of the output transistors. Non-linearity in the cross-over interval is minimized by providing the output transistors with a small initial threshold bias. Experiment indicates that there is an optimum value of threshold emitter current, about 20 milliamperes, which results in minimum nonlinearity. While this optimum value of current is essentially independent of temperature, the corresponding required base-to-emitter bias voltage varies with temperature at a rate of approximately -0.0025 volt per degree C. The resistor-thermistor networks, *R27* through *R33*, provide a low-impedance bias source exhibiting the proper variation of voltage with temperature. The transistor-thermistor mounting arrangement, shown in Figure 6, provides for close thermal contact between the transistor case and mounting bracket, and thermistor, mounting bracket, and chassis.

The output stage affords large-signal current gain of 20 to 30, and transconductance in excess of 1 mho.⁶ With a 6.6-volt collector supply-voltage, a collector voltage swing of 6 volts peak is obtainable. The corresponding peak collector current, for 2.0 watts output (10 per cent distortion), is 0.67 ampere. The 4-ohm speaker voice-coil impedance is transformed to the required 9-ohm collector load by transformer *T_s*. The use here of a bifilar wound autotransformer provides close coupling and high transformer efficiency. The power gain of the output stage is on the order of 24 decibels.

⁶ These characteristics were obtained with experimental transistors having substantially better high-current performance than that described by Armstrong and Jenny, *loc. cit.* They were the forerunners of those described by B. N. Slade, "Recent Advances in Power Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 153-171.

High-frequency roll-off of the audio amplifier is controlled by capacitor $C34$, which introduces inverse feedback in the output stage at frequencies above 2 kilocycles.

The audio input circuits are decoupled from the power supply by $R26$ and $C33$.

Power and Interference Considerations

Conventional means are utilized to eliminate high-frequency interference on the power lead. A spark plate, $C36$, and an r-f choke and capacitor $L2$, $C35$, remove the VHF and MF components respectively. The receiver chassis is insulated from the receiver case. Power connection for operation in automobiles with positive or negative polarity battery ground is obtained by employing plug P1A or P1B, respectively.

Rejection of high-frequency impulse type of ignition interference appearing on the antenna is accomplished by the choke, L_1 , in series

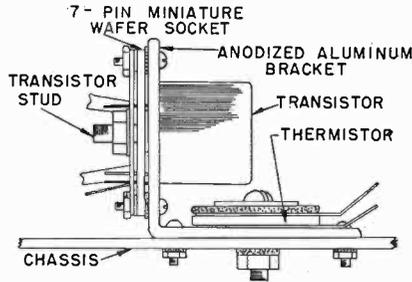


Fig. 6—Power-transistor and thermistor mounting arrangement.

with the antenna lead, which, together with the shunt capacitance across the antenna primary, forms a low-pass filter.

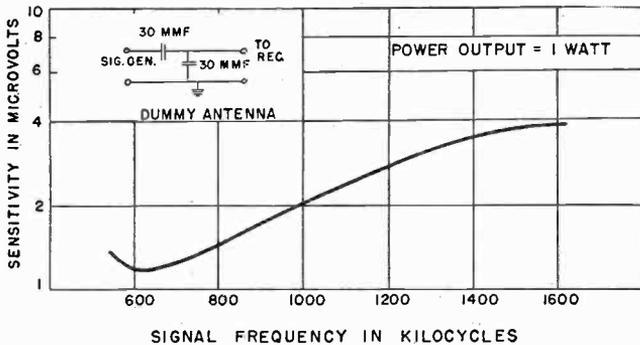
An additional potential source of interference arises from the periodic current drawn by the automobile ignition system, which may produce a low-frequency voltage fluctuation on the power lead. The presence of more than a few millivolts of this type of interference appearing between bases and emitters of the gain-controlled stages would produce objectionable modulation of the received signal. This is avoided by returning the bias bleeders, as well as the collectors of $V1$ through $V6$, to the decoupled voltage available at the junction of $R26$ and $C33$, and by the additional isolation of the base returns provided by $C2$.

PERFORMANCE

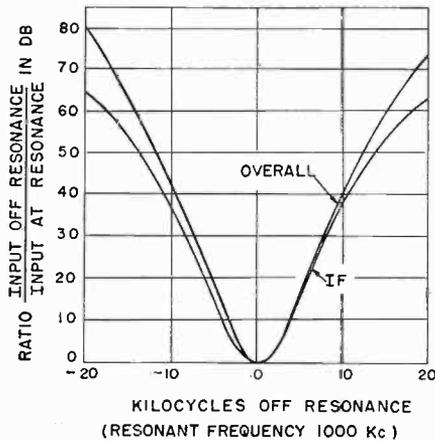
Various receiver performance characteristics are shown in Figure 7.

The receiver sensitivity as a function of signal frequency appears in Figure 7a. The increase in sensitivity at the lower extreme of the

broadcast band is due largely to the increased gain of the r-f and mixer stages over that obtaining at the upper extreme of the band. The shape of this curve is also influenced by the tracking error of the oscillator transformer which is plus or minus 4 kilocycles. This tracking error is due in part to the step approximation to a variable-pitch winding employed in this transformer.



(a) Sensitivity versus signal frequency.

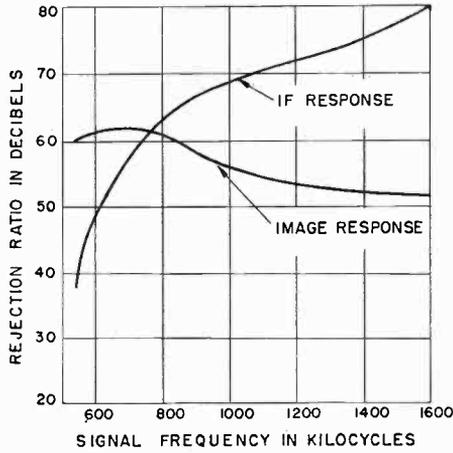


(b) i-f and over-all selectivity.

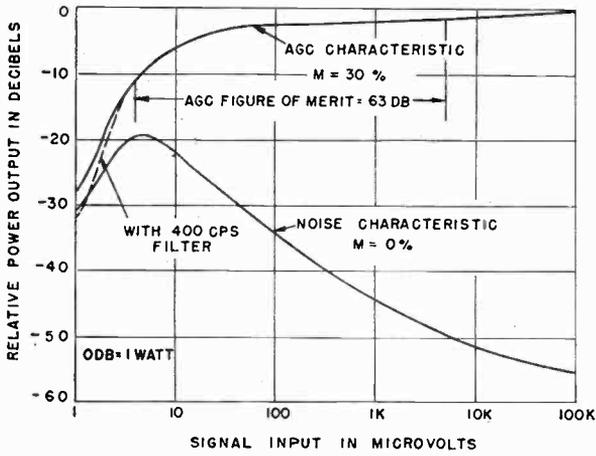
Fig. 7—Receiver performance characteristics.

The i-f and over-all selectivity at 1,000 kilocycles and the i-f and image rejection as a function of receiver tuning are shown in Figures 7b and 7c, respectively. These curves reflect the design compromises in the antenna, interstage, and i-f coupling networks. The over-all adjacent-channel attenuation at 1,000 kilocycles is 41 decibels. Since the over-all selectivity is determined almost entirely by the i-f coupling

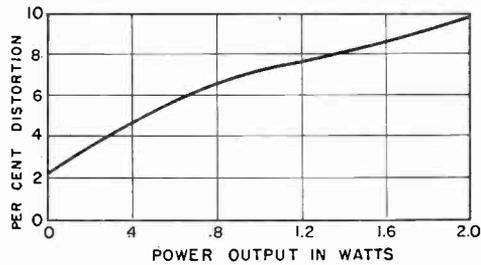
TRANSISTORS I



(c) i-f and image rejection characteristics.

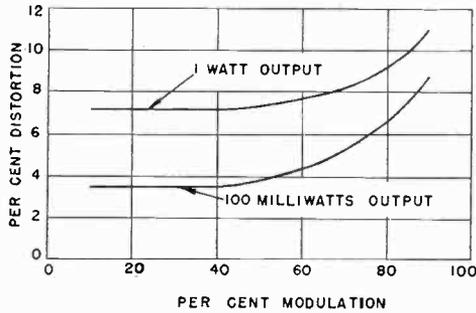


(d) Automatic-gain-control and noise characteristics.

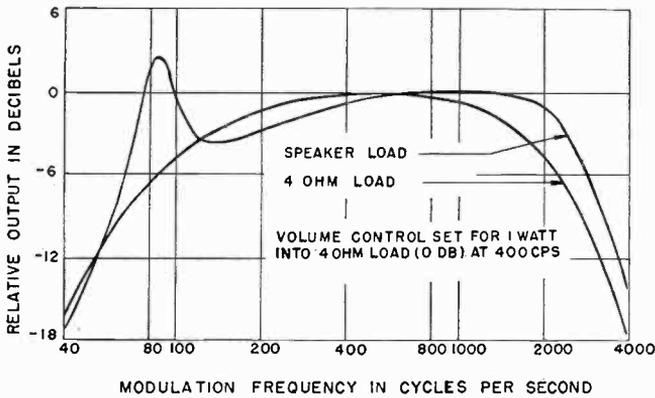


(e) Distortion versus power output.

Fig. 7—Receiver performance characteristics.



(f) Distortion versus per cent modulation.



(g) Electric fidelity.

Fig. 7—Receiver performance characteristics.

networks, the ACA is essentially independent of the signal frequency. The image rejection at 1,600 kilocycles is 52 decibels and increases to about 60 decibels at the low end of the band. The i-f rejection provided at 540 kilocycles is 40 decibels and increases rapidly with increasing signal frequency.

The a-g-c and noise characteristics are shown in Figure 7d. The a-g-c figure of merit is 63 decibels. The signal input required for a 20-decibel signal-to-noise ratio is 12 microvolts; the corresponding equivalent-noise-sideband input is 0.4 microvolt.

The distortion versus power-output characteristic is displayed in Figure 7e. Total harmonic distortion reaches 10 per cent, mostly third harmonic, at approximately 2 watts output. At this level, distortion arises almost entirely from the curvature of the current-gain characteristics of the output transistors. Since the power output is not limited by the dissipation capabilities of the transistors, operation in auto-

mobiles with 12-volt electrical systems should permit a substantial increase of power output, utilizing the same transistors at the same peak currents.

The distortion versus per cent modulation characteristics for 1.0 watt and 100 milliwatts output are shown in Figure 7f. The distortion is substantially constant up to 50 per cent modulation and increases above this level due to an increasing detector contribution.

The electric fidelity into the speaker load and into a 4-ohm dummy load is displayed in Figure 7g. The mechanical resonance of the 6×9 -inch speaker and the series inductance of the voice coil account for the elevated response into the speaker load in the vicinity of 85 and 2,000 cycles, respectively. Volume-control compensation has not been em-

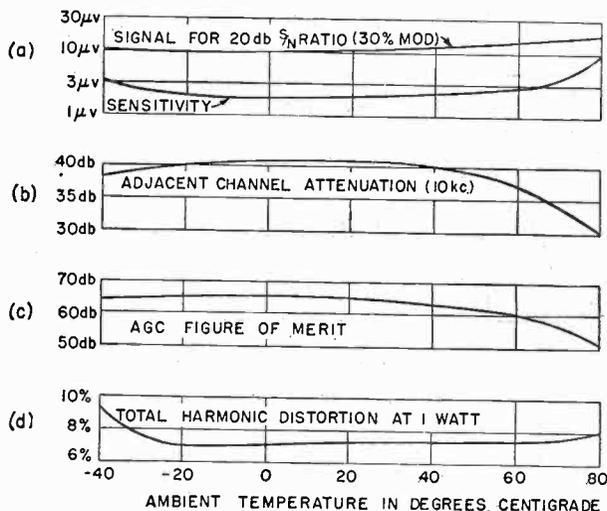


Fig. 8—Receiver performance as a function of ambient temperature.

ployed; the electric fidelity is essentially independent of volume-control setting. The receiver acoustical performance is influenced by the effective baffle which is employed, and thus is different for various makes and models of automobiles.

Curves illustrating the performance of the receiver as a function of ambient temperature are shown in Figure 8.

Figure 8a shows the receiver sensitivity and the signal required for 20-decibel signal-to-noise ratio as a function of ambient temperature. The sensitivity is 2.0 microvolts at 20°C, and is below 10 microvolts over the range from -40° to 80°C. Loss of sensitivity at high temperatures arises from the reduced Q of the core materials in the tuned circuits, from detuning effects, and from a decrease in gain of

the high-frequency transistors of approximately 1 decibel per stage. The receiver noise performance is substantially unaffected over this range of temperature.

Figure 8b shows the receiver ACA as a function of ambient temperature. Differences in tuned-frequency shift of the individual i-f interstages tend to reduce selectivity at the extreme temperatures. This tendency is offset by the increase in Q 's at reduced temperatures, and aggravated by the reduction in Q 's at elevated temperatures, resulting in a reduction in ACA from approximately 40 decibels, at low and moderate temperatures, to 31 decibels at 80°C.

The variation of the a-g-c figure of merit with ambient temperature is shown in Figure 8c. The figure of merit is approximately 65 decibels at low and moderate temperatures, and drops to 51 decibels at 80°C.

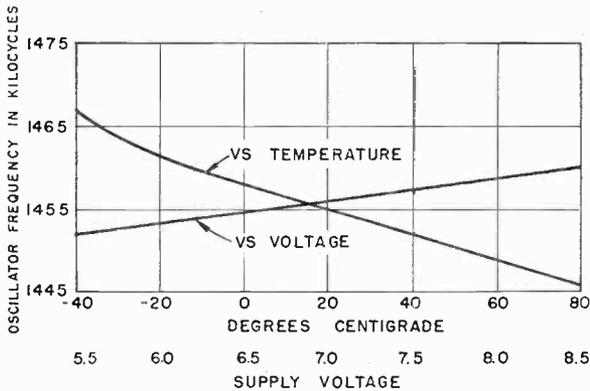


Fig. 9—Oscillator frequency stability.

The shape of the a-g-c characteristic at any temperature in this range is essentially the same as that shown in Figure 7d. The reduction in a-g-c figure of merit at elevated temperatures arises from the corresponding reduction in receiver sensitivity.

The total harmonic distortion at 1.0 watt output, shown in Figure 8d, and the maximum power output (10 per cent distortion), are relatively unaffected over the temperature range from -40° to 80° C. The increase in distortion at extreme temperatures arises for the most part from the imperfect temperature compensation provided by the output-stage biasing network. Additional distortion at -40° C originates in the detector-a-g-c circuitry, and is due mainly to the increased impedance of the sintered tantalum electrolytic capacitor, C29, in the a-g-c filter network. It is the use of this type of capacitor here, notable for good low-temperature performance, that permits operation to

—40° C. Substitution of a conventional aluminum-foil type capacitor restricts the useful lower-temperature limit to approximately —10° C.

The change in oscillation frequency with change in supply voltage, shown in Figure 9, is about 0.2 per cent per volt. The frequency variation of the oscillator with respect to ambient temperature is also shown in Figure 9. The difference in oscillator-frequency shift and in the frequency shift of the signal tuned-circuits is in the same direction and of approximately the same magnitude as the shift in the i-f tuned circuits. The net tracking error as a function of ambient temperature is therefore smaller than either the oscillator- or intermediate-frequency shifts.

A DEVELOPMENTAL POCKET-SIZE BROADCAST RECEIVER EMPLOYING TRANSISTORS*

BY

DAVID D. HOLMES, THOMAS O. STANLEY AND LARRY A. FREEDMAN

RCA Laboratories,
Princeton, N. J.

Summary—This paper describes a pocket-size developmental AM broadcast receiver which utilizes eight junction transistors. Its performance is comparable to that of conventional personal receivers. Emphasis has been given to developments which contribute to stability with respect to temperature, battery voltage, and variations among transistors. The superheterodyne circuit employed uses a single-transistor frequency converter to perform the functions of both mixer and oscillator. Refined detector and automatic-gain-control circuits and an audio amplifier embodying further development of the principle of complementary symmetry are incorporated. Reduction in physical size and battery requirement, as compared to conventional receivers, is substantial.

The circuits are described in detail and certain aspects of components and of physical arrangement, which contribute to the small size, are discussed. Detailed performance data is also included.

GENERAL DESCRIPTION

THE receiver is shown in Figure 1. Operation of tuning and volume-on-off controls is accomplished by means of rim-drive wheels which protrude through slots at the end of the receiver case. The tuning indication is marked on the tuning wheel and is viewed through the window at the lower right. The 2 × 3 inch speaker is located behind the lower portion of the decorative grille, and the back of the case is vented to improve acoustical performance. The 4-cell battery is contained in a compartment which is located at the rounded end of the receiver and provided with a snap-on cover.

The over-all dimensions of the receiver are: height $2\frac{3}{4}$ inches, length $5\frac{1}{8}$ inches, thickness $1\frac{1}{4}$ inches. The total weight of this unit is 17 ounces. The dimensions of the receiver are determined principally by the speaker, tuning condenser, antenna core and battery, i.e., the transistors and small components occupy only a fraction of the total volume of the receiver. For pocket use, the thickness is an important dimension. In this receiver the desired minimum thickness was achieved by employing shallow (approximately 1 inch) versions of a

* Reprinted from the *Proceedings of the Institute of Radio Engineers*, June, 1955.

conventional speaker and a conventional tuning condenser. The magnet structure of the speaker was rearranged to minimize the over-all depth. Approximately the same volumes of magnet and iron are employed in both the original and modified units; no degradation in performance is introduced by the modification. The tuning condenser was modified by removing some of the plates and shortening the shaft and frame.

The salient performance characteristics of the receiver are as follows:¹

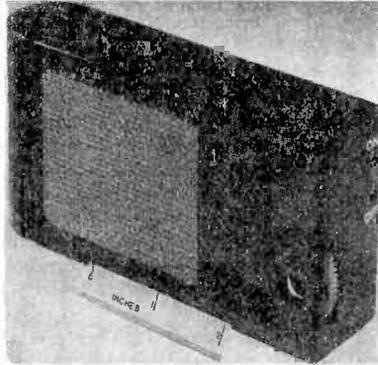


Fig. 1—Pocket-size broadcast receiver.

Sensitivity	100 microvolts per meter
Noise performance	
ENSI	15 microvolts per meter
Input for 20 db S/N	1,300 microvolts per meter
Power output	125 milliwatts
Selectivity (adjacent-channel attenuation)	28 decibels
AGC (figure of merit)	37 decibels

The battery life of the receiver shown in Figure 1, employing RM-1 cells, is approximately 50 hours. Alternative battery cases housing battery complements representing different design compromises among battery life, size, weight, cost, etc., may readily be substituted. For example, the battery case shown in Figure 2 (above) accommodates either RM-502 cells, providing a battery life of 120 hours, or conven-

¹ This data reflects the characteristics of the experimental transistors used; units of exceptionally high and exceptionally low performance were avoided.

tional "pen-lite" cells, providing a 36-hour life. This battery case increases the receiver length by $\frac{1}{2}$ inch.

CIRCUIT DESCRIPTION

A schematic diagram of the complete receiver is shown in Figure 3.

Transistors V1 through V4 are experimental radio-frequency units,² and serve as converter, first and second i-f, (intermediate-frequency) and second detector stages respectively. The intermediate frequency is 455 kilocycles. Transistors V5 through V8 are the audio-amplifier complement. Transistors V5 and V6 are identified by the type numbers 2N34 and 2N35 respectively. Transistors V7 and V8 are experimental high-current p-n-p and n-p-n junction units, respectively, connected as

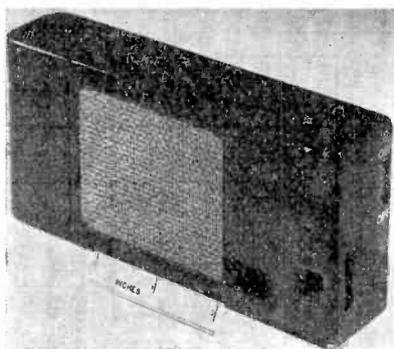


Fig. 2—Receiver case with alternative battery case attached.

a class B complementary symmetry output pair working directly into the voice coil of the speaker.

The 5-volt battery consists of four RM-1 Mallory Mercury cells. The positive side of the battery is set at +1 volt with respect to ground by the bleeder combination of R7 and R8. The battery is centertapped to provide the symmetrical supply required by the output stage.

Antenna

The receiver antenna consists of a ferrite-cored loop. The core is made up of four 4-inch sections of one-quarter inch diameter ferrite rod placed side-by-side to form a flat structure. This arrangement provides a large antenna volume in a shape which is compatible with the remainder of the receiver. As large a volume as was considered

² C. W. Mueller and J. I. Pankove, "A p-n-p Triode Alloy-Junction Transistor for Radio-Frequency Amplification," *Proc. I.R.E.*, Vol. 42, pp. 386-391, February, 1954; also *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 189-201.

practicable was utilized since the power extracted from a given radiated field by such an antenna is proportional to the volume of the ferrite core.

The antenna loop and secondary winding serve as an impedance transformer; the turns ratio is designed to match the high tuned impedance of the antenna to the low input impedance of the converter at approximately midband. This entails no appreciable sacrifice in performance at the extremes of the band.

A grounded copper shield is located beneath the antenna between the antenna and the remainder of the receiver to reduce feedback of i-f and its harmonics from the detector circuit to the antenna.

Converter

Oscillator and mixer functions are performed by V1, the converter transistor. The oscillator transformer, T1, provides tickler feedback from collector to base. The i-f take-off transformer, T2, is located in the collector circuit in series with the oscillator transformer primary. Signal is applied to the converter base via the antenna secondary winding in series with the oscillator feedback winding. The input circuit is returned to chassis ground; the collector voltage is -4 volts. The converter is constant-emitter-current biased at approximately 1 milliamperes by means of the 1000-ohm emitter resistor, R1, which returns to the +1 volt bus. The 1-volt bias supply is several times the emitter-to-base operating potential of the transistor. Thus, variations in emitter-to-base operating potential among transistors, or with temperature for a given transistor, have a negligible effect upon emitter current. The bleeder which fixes ground potential is relatively stiff (2 milliamperes) so that base current variation with temperature and among transistors does not shift the operating point. Constant-emitter-current bias provides stability of operating point over a wide temperature range and affords transistor interchangeability.

The converter emitter is bypassed to ground by C5. Bypassing to ground in this manner (rather than to the +1-volt bus) provides effective r-f isolation of the input circuit from the positive side of the battery.

The turns ratios of the oscillator transformer have been determined experimentally to provide high operating Q and near-optimum oscillator injection over the band. A representative value of optimum injection is 0.1 volt r-m-s at the converter base. The oscillator frequency is substantially unaffected by a 50 per cent reduction in battery voltage or a variation in ambient temperature from 0° to 50° C.

The input impedance of the converter varies approximately 2 to 1 over the broadcast band. An average value is 300 ohms. The converter

output impedance at i-f is approximately twice that of the same transistor as an i-f amplifier, or about 40,000 ohms. A typical value for conversion gain is 22 decibels.

Intermediate-Frequency Amplifier

The three i-f transformers, T2, T3 and T4 comprise three single-tuned circuits which serve as interstage coupling networks, having essentially unity coupling between primary and secondary windings. These transformers embody an experimental arrangement and are ferrite cored. They are approximately $\frac{3}{8}$ inch in diameter by $\frac{1}{2}$ inch high. For any given operating Q , minimum insertion loss is incurred by choosing the transformer turns ratios so that the reflected load impedance is equal to the transistor driving impedance. In this receiver, each transformer is designed for an operating Q of 35; the resulting insertion loss is 2.5 decibels per transformer. The different operating conditions of the various stages give rise to differences in input and output impedances among stages; the various transformer turns ratios are chosen accordingly.

The first i-f transformer, T2, feeds the base of the first i-f transistor, V2, the input impedance of which is 150 ohms. A neutralized common-emitter connection is employed. The operation of the neutralizing circuit may be visualized by referring to the single-generator common-emitter π -equivalent circuit of the transistor, shown in Figure 4. Feedback from the output to point b' occurs via the parallel combination of $C_{b'c}$ and $r_{b'c}$. The transistor feedback capacitance and resistance are in the order of 15 micromicrofarads and 0.1 megohm respectively. This feedback may be neutralized by deriving a reverse-phase voltage from the output which is then fed back through a suitable parallel r-c circuit to the base connection, point b . The presence of $r_{bb'}$ renders the neutralization slightly dependent upon frequency, and influences the requisite parameters in the neutralizing circuit. This effect is second order at 455 kilocycles, however, and can be neglected.

The first r-f transistor is provided with an initial constant-emitter-current bias of 0.5 milliamperes; choice of bias for this stage is dictated in part by automatic-gain-control (a-g-c) considerations. The operation of the a-g-c circuit is described in conjunction with the second detector.

Neutralization is accomplished by utilizing feedback from the secondary of the second r-f transformer, T3, through the 150-micromicrofarad neutralizing capacitor, C6, to the base. The proper primary-to-secondary polarity is as indicated on the schematic. Because the resistive component of transistor feedback is relatively small, resistor neutralization need not be employed. Since the primary-to-secondary turns ratio of T3 is approximately 10 to 1, a neutralizing capacitor

which is ten times the effective transistor feedback capacitance is used. T3 provides interstage coupling between the 20,000-ohm output impedance of the first i-f stage and the 150-ohm input impedance of the second i-f stage. An i-f stage gain of approximately 28 decibels is realized.³

The second i-f transistor, V3, is constant-emitter-current biased at 1.0 milliamperes. The third i-f transformer, T4, operates between the 20,000-ohm output impedance of the second i-f stage and the 20,000-ohm input impedance of the second detector. This value of detector input impedance applies at receiver-sensitivity level; the detector input impedance decreases with increasing signal level. The primary-to-secondary turns ratio of this transformer is approximately 1 to 1 so that a 15-micromicrofarad neutralizing capacitor, C9, is required.

Automatic gain control of the first i-f stage is accomplished by

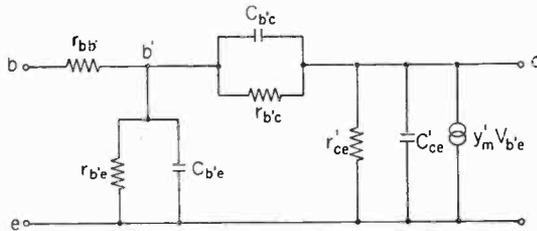


Fig. 4—Single-generator common-emitter π -equivalent circuit of experimental radio-frequency alloy junction transistor.

varying its d-c emitter current as a function of the signal level at the detector. The manner in which 455-kilocycle gain, input impedance, and output impedance vary with emitter current is shown in Figure 5. The transistor gain decreases rapidly as the d-c emitter current is reduced below 0.5 milliamperes. The input and output impedances increase with decreasing emitter current so that the stage gain is further decreased by input and output circuit mismatches. A single-stage control range of approximately 45 decibels is obtained. It is evident that automatic gain control will result in a change in operating Q ; the effect of this change on over-all bandwidth is discussed below.

Second Detector and Automatic Gain Control

The second detector transistor, V4, is driven by the third i-f transformer secondary. This stage operates with the base and emitter at

³ By way of example, i-f transistors having $r_{bb'} = 75$ ohms, $\alpha_{cb} = 20$, $C_{b'e} = 11$ micromicrofarads, and $C_{bc} = 0.001$ microfarad ("alpha cutoff" of 6 megacycles) would give typical performance.

the same d-c potential for the zero-signal condition. As signal is applied to the detector, substantially three components of collector current are developed; these components increase with signal. Included are a high-frequency component (i-f and its harmonics), an audio-frequency component, and a d-c component. The return path for the high-frequency component is from the collector through the 910-kilocycle series resonant circuit formed by C13 and L1, and thence through the series combination of R4 and R5 to the emitter. The 910-kilocycle series-resonant circuit confines the i-f second harmonic current to one return path: reduction of i-f second harmonic feedback to the antenna is simplified since the inductive field of the resulting single loop may

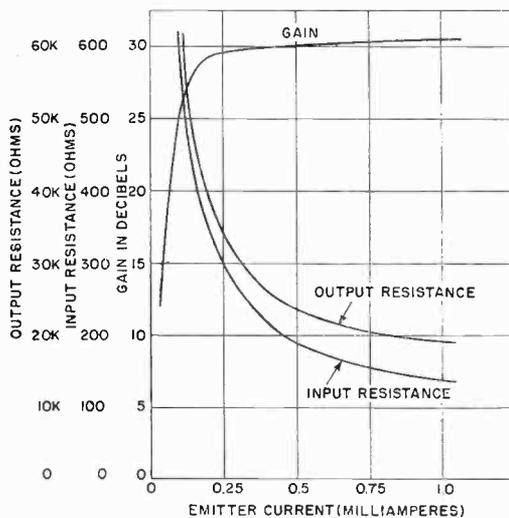


Fig. 5—Typical variation of r-f transistor characteristics with variation of emitter current.

then be readily shielded from the antenna. Resistor R10 blocks high-frequency components of collector current from the volume control and associated wiring. Audio-frequency components of collector current return via C12 through the series combination of R4 and R5 to the emitter. The degeneration provided by R4 and R5 serves to reduce detector distortion.

The detector audio output is developed across the upper portion of the 5,000-ohm volume control, R11. The collector is returned to the volume control slider to provide a detector-circuit output impedance which is substantially independent of volume control setting; the collector impedance is high compared to the volume control resistance.

The desirability of a constant audio driving impedance for the particular audio amplifier utilized is discussed later.

Curves of detector distortion and output voltage versus detector input are shown in Figure 6. Distortion at approximately receiver-sensitivity level is 11 per cent at 80 per cent modulation, and falls to less than 2 per cent at a signal level corresponding to the knee of the a-g-c characteristic.

The d-c component of detector emitter current is employed for a-g-c purposes. The d-c emitter current return path is from the positive side of the battery through the resistor, R2, in the emitter circuit of the first i-f stage, thence through R6 and R5 to the detector

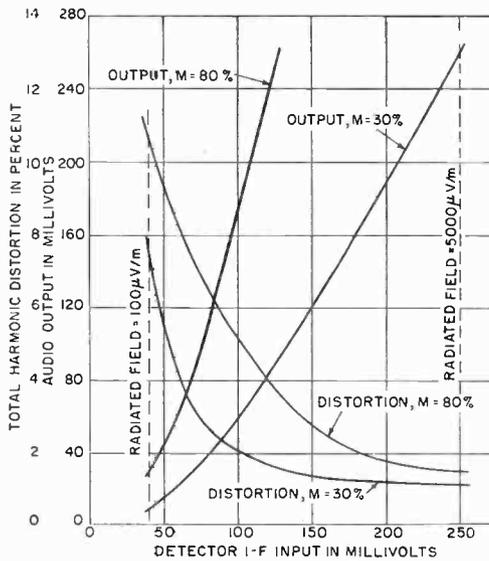


Fig. 6—Detector characteristics.

emitter. Negative feedback for d-c in the detector circuit is provided by R5, introducing an effective delay in the a-g-c action of the detector. The 30-microfarad emitter bypassing capacitor of the first i-f stage, C7, in combination with R6, serves to filter out audio-frequency components of detector emitter current.

The operation of the a-g-c circuit is as follows. For the zero-signal condition, the first i-f stage is constant-emitter-current biased at 0.5 milliamperes, and substantially no d-c emitter current flows in the detector circuit. As signal is increased, detector d-c emitter current is developed, and flows through the resistor R2, in the emitter circuit of the first i-f stage. Since the bias arrangement of the first i-f stage

holds the current in R2 essentially constant, the detector may be considered to "rob" the i-f stage of emitter current, i.e., the d-c emitter current shifts from the i-f stage to the detector. As the detector emitter current approaches 0.5 milliamperes, the i-f stage emitter current (and stage gain) approaches zero. Thus, the detector d-c emitter current at the flat portion of the a-g-c characteristic is equal to the first i-f stage d-c emitter current at zero signal. The value of 0.5 milliamperes is chosen as being high enough to insure operation of the first i-f stage near maximum gain for zero-signal condition and low enough to prevent detector overload at maximum volume control setting under strong signal conditions.

The detector circuit input impedance varies with signal level, being about 20,000 ohms at receiver sensitivity level, and dropping to about 5,000 ohms at the knee of the a-g-c characteristic. This produces a lowering of the operating Q of the last i-f transformer with increasing signal level. At the same time, however, a-g-c action increases the input and output impedances of the first i-f stage; the operating Q 's of the first and second i-f transformers are increased so that a negligible change in over-all bandwidth occurs.

The first four stages of the receiver are decoupled from the battery by R9 and C15 which prevent audio-frequency components developed across the battery impedance from affecting the oscillator and a-g-c circuits. The +1 volt bus is bypassed to ground for audio and radio frequencies by C14.

Audio

The detector is followed by a three-stage audio amplifier employing cascade complementary symmetry⁴ in the first two stages, which operate class A, and push-pull complementary symmetry in the output stage, which operates class B. The voice coil is driven directly from the emitters of the output stage. Signal feedback from the voice coil to the first audio transistor reduces distortion, relaxes the degree of match required in the output transistors, and enhances interchangeability of transistors. Stabilization of operating biases over a wide temperature range is achieved by over-all d-c feedback.

The signal feedback from the voice coil to the emitter of the first stage, V5, attenuated by the R19 to C19-R14 voltage divider, increases the input impedance of the audio section to about 10,000 ohms. For this type of feedback, the degree of degeneration is a function of the relative magnitudes of this input impedance and the signal source impedance. The particular arrangement of detector and volume control

⁴G. C. Sziklai, "Symmetrical Properties of Transistors and their Applications," *Proc. I.R.E.*, Vol. 41, pp. 717-724, June, 1953.

used here was chosen to provide constant source impedance so that the degeneration (10 decibels) would be independent of volume control setting. The increasing impedance of C19 at decreasing frequencies determines the low-frequency roll-off of the amplifier. High-frequency roll-off is controlled at the input to the audio section by the shunt 0.1-microfarad capacitor, C16.

The collector of the first audio transistor feeds the base of the second, or driver transistor, V6. Effectively, the collector of the driver works directly into the bases of both output transistors. Actually these bases are separated only by a 22-ohm resistor, R16, which develops a small initial bias. The major portion of the signal current of the driver is made to flow into the output stage bases by returning the coupling resistor, R17, to the common output stage emitters, via C20. Relatively little signal current flows through R17, since only the output stage base-to-emitter voltage need be developed across this resistor, rather than the full output voltage. The d-c return path for the driver collector current is R17 and R18 (no appreciable direct current flows into the output bases). The quiescent driver current is made sufficiently large to permit the output bases to be driven to a peak-to-peak voltage equal to the battery supply voltage. The total resistance of resistors R16, R17, and R18 is chosen so that the quiescent voltage developed at the driver collector by this current is approximately battery centertap voltage.

The common emitters of the output stage directly drive the 12-ohm voice coil, which returns to the battery centertap. When the bases of the output stage are driven negative with respect to the battery centertap, the p-n-p transistor, V7, conducts, current being fed from the upper half of the battery to the voice coil. The n-p-n transistor, V8, and the lower half of the battery function similarly for positive excursions. The peak-to-peak voltage available across the voice coil is less than the battery voltage by the required peak base-to-emitter voltages in the output transistors and by signal voltage developed in the dynamic battery impedance. The experimental output transistors employed are electrically similar to those described by Armstrong and Jenny,⁵ requiring approximately 0.5 volt between base and emitter for 145 milliamperes collector current. Thus, with a nominal supply voltage of 5 volts, a maximum peak-to-peak voltage across the voice coil of about 3.5 volts is realized, corresponding to a power output of about 125 milliwatts.

The three audio stages are d-c coupled to permit the use of over-all

⁵ L. D. Armstrong and D. A. Jenny, "Behavior of Germanium Junction Transistors at Elevated Temperatures and Power Transistor Design," *Proc. I.R.E.*, Vol. 42, pp. 527-530, March, 1954.

d-c feedback for stabilization purposes. Bias for the base of the first transistor is developed by a bleeder, R12 and R13, between the negative side of the battery and the battery centertap. The emitter of this transistor is returned to the battery centertap through R19 and the voice coil. Thus, any d-c voltage developed across the voice coil by unbalanced currents in the output transistors is subtracted from this bias, constituting d-c feedback. The collector current of the p-n-p first stage develops a voltage from base to emitter of the n-p-n driver stage, controlling the driver collector current. The magnitude of the driver collector current determines the voltages applied to the output stage bases. When these voltages bracket the battery centertrap voltage, balanced currents flow in the output transistors. Any departure from this balance, due to variations in characteristics among transistors or ambient temperature variation, is of the proper polarity to be self-correcting, through the previously mentioned d-c feedback to the first stage. The unbalance is held within ± 10 milliamperes from 0° to 50° C. For example, the replacement of a normal driver transistor by one having three times the current gain results in an unbalance of only 3 milliamperes. The maintenance of this balance avoids excessive quiescent current drain on either half of the battery supply, and prevents loss of dynamic range due to asymmetrical overload.

The collector current of the first stage is shared by the base of the driver and the shunting resistor, R15. The d-c feedback functions to adjust this current to that value which will bias the driver stage to 8 milliamperes collector current. The component of current flowing into the base of the driver stage (for 8 milliamperes collector current) diminishes with increasing temperature, ultimately reversing. The component flowing into R15 remains relatively unchanged. This latter component constitutes essentially the total operating collector current of the first stage near the upper temperature limit of the amplifier. The value of R15 is low enough to insure adequate operating current for the first stage to at least 50° C.

Threshold bias for the output transistors, which operate class B, is developed across R16. The choice of the magnitude of this bias represents a compromise between (a) the appearance of "cross-over" distortion (insufficient bias), and (b) excessive quiescent battery drain (more than sufficient bias). The compromise chosen, good at room temperature, tends toward (a) at temperatures below 10° C, and toward (b) at temperatures above 40° C.

The volume control and the bleeder for the first audio stage are returned to the junction of R20 and C18, decoupling them from the negative supply with respect to the battery centertap, to which the

emitter of the input stage ultimately returns. The base and emitter of the driver are both returned to the negative supply, and so require no decoupling. Voltage developed across the internal impedance of the lower half of the battery, appearing at the positive battery terminal and at ground with respect to the centertap, is not directly applied to the audio section. Paths for current do exist, however, through the 910-kilocycle filter in the detector, and through the dynamic detector collector resistance. The audio impedance of these paths is sufficiently high (more than 50,000 ohms) that negligible distortion is introduced.

Power Supply

The battery consists of four separate cells, grouped into two pairs of two cells each to provide a centertapped supply for the output stage. The three-point power switch is located at the battery centertap, where it simultaneously connects the two halves of the battery to each other and to the centertap lead. The use of a three-point switch, instead of a double-pole switch, facilitated the modification of a conventional hearing-aid volume control and *single-pole switch* combination.

The average battery drain on program material at typical listening level is about 20 milliamperes. The first four stages require 5 milliamperes; the first audio and driver require 8 milliamperes; the output stage draws about 7 milliamperes. The zero-signal current of the output stage is a few milliamperes; the current increases with signal to as much as 50 milliamperes (average) for maximum output on a continuous tone. By the use of the class B output stage rather than a conventional class A output stage the over-all battery requirement has been reduced by a factor of more than three.

As the battery voltage decreases and battery internal impedance increases with operating life, the maximum power output capability of the receiver decreases, ultimately determining the useful battery life. Appreciable degradation of receiver performance in other respects does not occur within this life. The battery life for different types of cells and various maximum undistorted output-power end points is shown in Table I.

Table I—Battery Life Characteristics

Cell Type	RM-1	RM-502	"Pen-Lite"
Initial maximum undistorted output power	125 mw	125 mw	150 mw
Life to 75 mw maximum undistorted output power	50 hours	120 hours	20 hours
Life to 50 mw maximum undistorted output power	50 hours	120 hours	36 hours

PERFORMANCE

The major receiver performance characteristics are shown in Figure 7. These characteristics are generally comparable to those of conventional personal receivers employing vacuum tubes.

The image rejection and sensitivity characteristics are shown in Figure 7(a). The operation of the antenna circuit at approximately one-half its unloaded Q (matched condition) results in a steepening of the slope of the image rejection characteristic. Approximately 30 decibels rejection is provided at the upper extreme of the band; rejection increases to 60 decibels at the lower extreme of the band. The i-f rejection (not shown) is 30 decibels at 600 kilocycles.

The a-g-c characteristic, and the signal-to-noise ratio as a function of signal level, are shown in Figure 7(b). The a-g-c figure of merit is 37 decibels. The receiver noise performance, depending upon many factors (antenna effectiveness, receiver bandwidth, converter noise, etc.), is not readily compared with a conventional receiver on the basis of any one contributing factor. The signal-to-noise ratio for a given signal level is in the order of 4 decibels lower than that of a conventional personal receiver.

The i-f selectivity, as shown in Figure 7(c), represents a design compromise between selectivity and gain for the three single-tuned circuits employed. The adjacent channel attenuation is 20 decibels. A requirement of more selectivity might justify the substitution of one or more double-tuned circuits.

The over-all selectivity characteristic is displayed in Figure 7(d). Over-all selectivity curves for two typical signal levels are included, demonstrating that the selectivity change due to detector and a-g-c action is small.

The electrical fidelity of the receiver is shown in Figure 7(e). The high-frequency roll-off has been adjusted in listening tests to provide a pleasing tonal balance.

The distortion-versus-power output characteristic is shown in Figure 7(f). This characteristic is influenced by the battery complement employed; the maximum power output obtained increases approximately as the square of the battery voltage, and decreases with increased battery internal impedance. For a battery complement, 4 RM-1 cells, (5.0 terminal volts under load) as shown, the maximum power output (10 per cent distortion) is approximately 125 milliwatts. Maximum power output achieved with 4 penlite cells (5.6 terminal volts under load) for example, is approximately 160 milliwatts.

The over-all acoustical performance of the receiver appears in Figure 8. The curve shown was obtained in a free field sound room

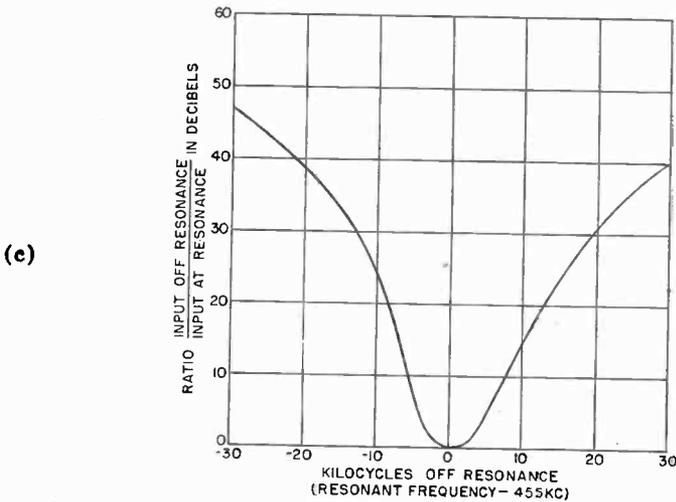
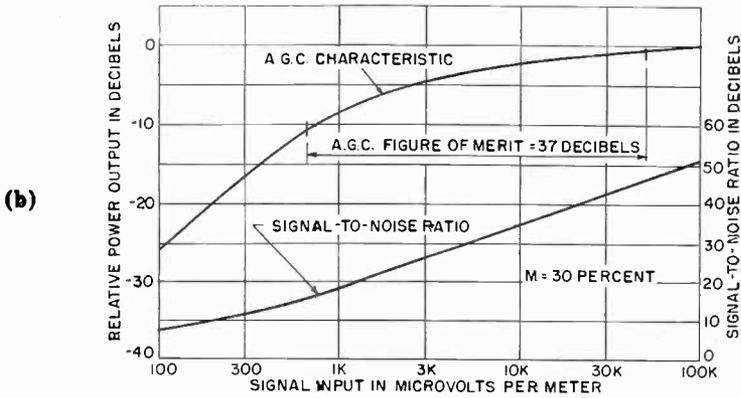
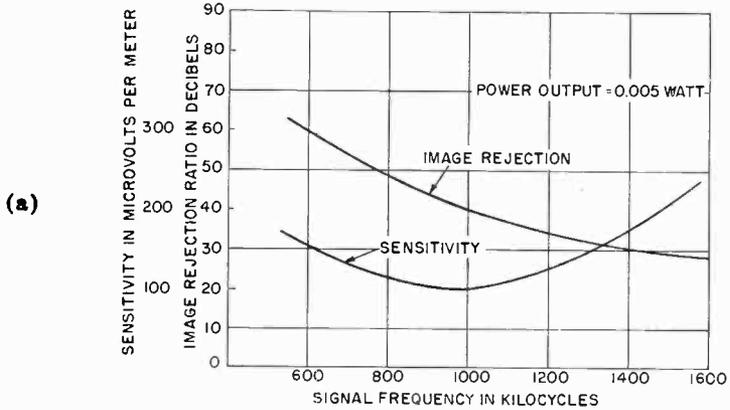


Fig. 7—Major receiver performance characteristics. (a) Image rejection and sensitivity. (b) Signal-to-noise ratio and a-g-c characteristic. (c) I-f selectivity.

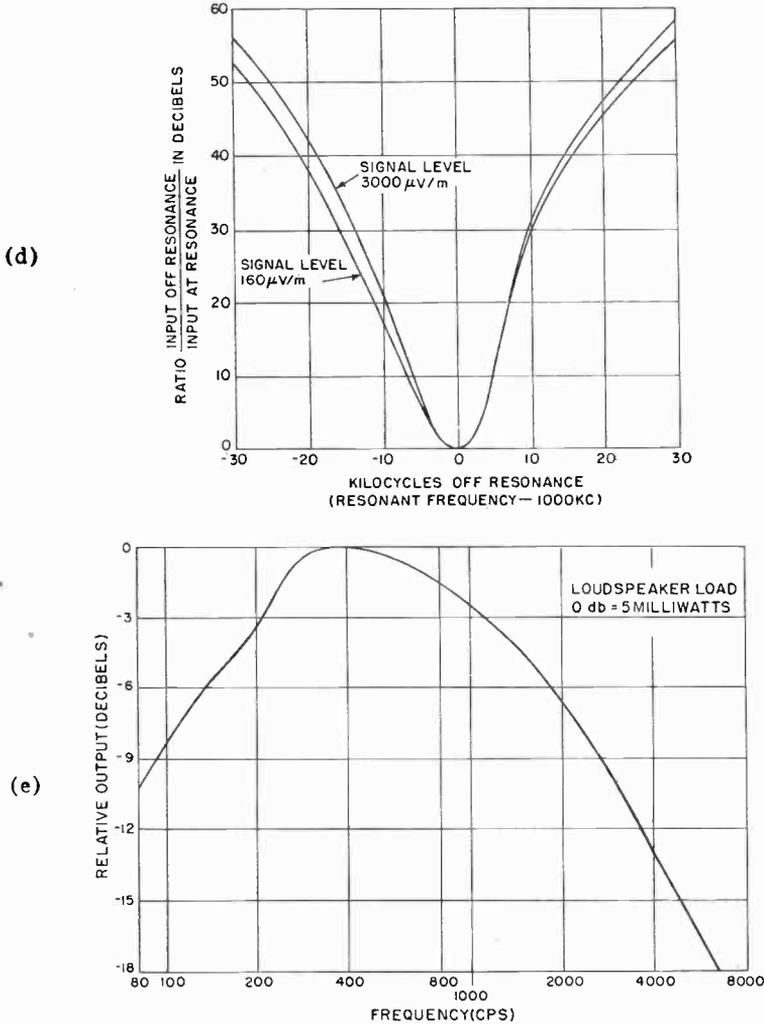


Fig. 7—Major receiver performance characteristics. (d) Over-all selectivity. (e) Electrical fidelity.

with the receiver placed on a table, as indicated in the sketch.

The performance of the converter, i-f amplifier and a-g-c circuits is substantially unaffected by a variation in ambient temperature over the range from 0° to 50° C. The detector (and thus the receiver) sensitivity varies approximately 2 to 1 over this range, being higher at elevated temperatures. The performance of the audio amplifier as a function of ambient temperature is shown in Figure 9. This data was taken on an experimental "breadboard" while the receiver was

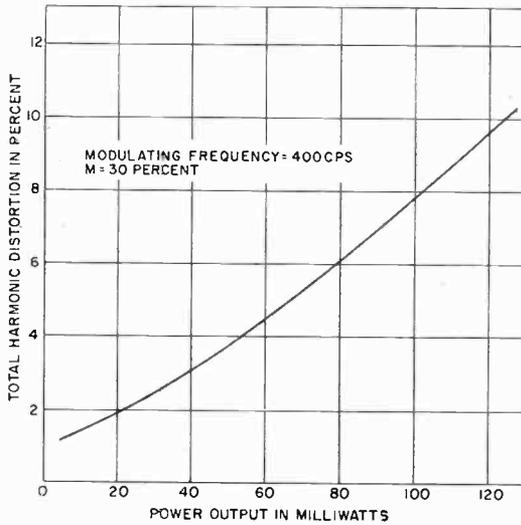


Fig. 7(f)—Distortion characteristic.

under development. The unbalance current remains within limits of approximately ± 10 milliamperes over the temperature range from 0° to 50° C. The distortion, for a constant output of 125 milliwatts, is substantially unaffected over this temperature range.

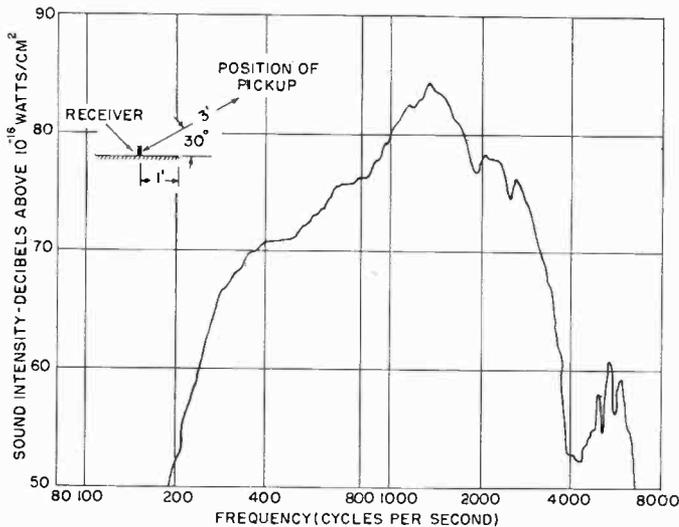


Fig. 8—Acoustical performance.

PHYSICAL DETAILS

Front and rear views of the receiver with the covers removed are shown in Figure 10. The ferrite-cored antenna may be seen at the top of a "chassis" which is essentially vertical and to which are attached, by means of brackets, the tuning condenser, speaker, and battery compartment. The eight transistors, mounted in subminiature sockets, occupy the space immediately below the antenna. The miniature oscillator and i-f transformers are mounted on the chassis, which serves as an electrical ground for the r-f and i-f portions of the receiver. The top of the chassis is bent over to form a horizontal flange. This flange,

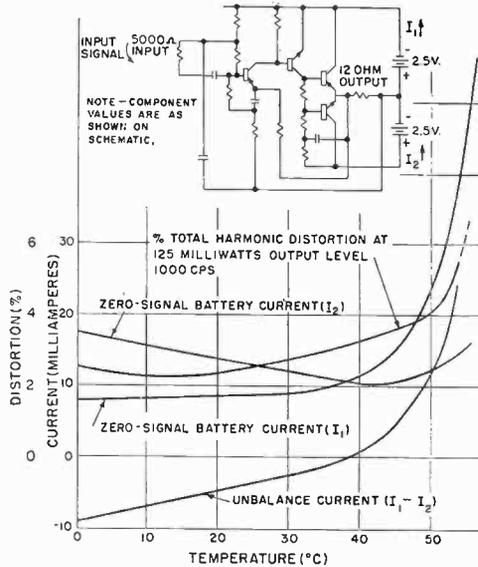


Fig. 9—Audio amplifier performance versus ambient temperature.

the subminiature sockets, a copper shield, fibre spacers, and the antenna are cemented together to form an integral sandwich.

The tuning capacitor, supplied by Radio Condenser Corporation, Camden, N. J., is a modification of a capacitor employed in a conventional personal-portable receiver design. The oscillator and r-f sections employ four stator plates each instead of five and seven respectively, providing the requisite reduction in capacitor depth. The resultant reduction in the ratio of maximum to minimum capacitance can be tolerated because of the relatively low shunt capacitance reflected across the antenna and oscillator-tuned circuits by the converter transistor. The tuning wheel is mounted on the rotor shaft, within the frame of the condenser.

The speaker is a modified version of the RCA 214S1 unit. Standard

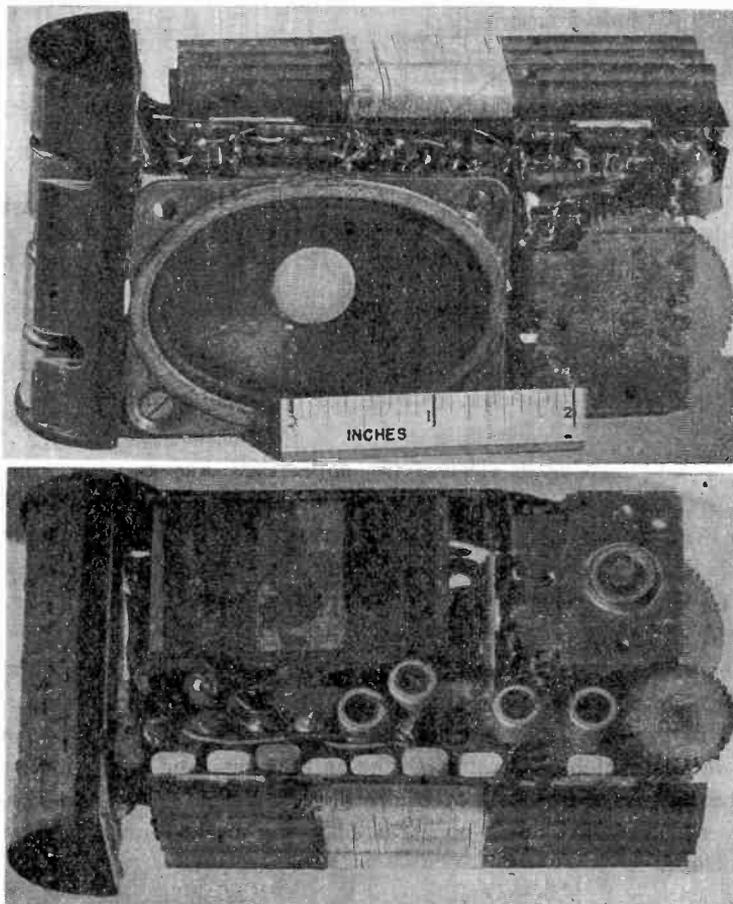


Fig. 10—Front and rear views of the receiver with covers removed.

parts were used except for the field structure, arranged as shown in Figure 11. The structure was magnetized, after assembly, by the impulse method, using a ten-turn coil on each magnet (series-opposing) and a 4,000-ampere discharge.

The volume control as acquired (Centralab Model 1 Radiohm, with switch) is equipped with a single-pole single-throw switch, modified as shown in Figure 12 to provide desired 3-point switch. Switch action is such that the movable contact is forced between the two stationary contacts.

Three small 0.1-microfarad ceramic capacitors of an experimental type are employed. The small size ($0.2 \times 0.50 \times 0.03$ inch) is achieved through the use of partially-reduced barium-strontium titanate as the dielectric.

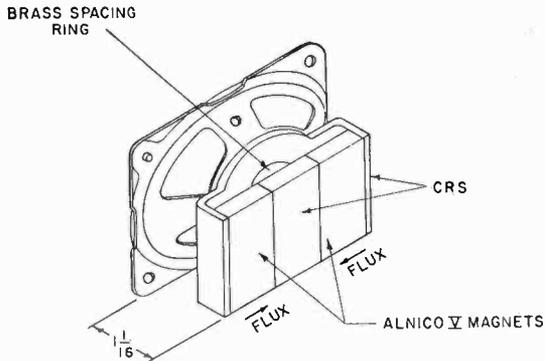


Fig. 11—Speaker, showing modified field structure.

The 30-microfarad electrolytic capacitors are of the sintered-tantalum type which is notable for its high capacitance-to-volume ratio. These units were mounted by pressing them into holes in the chassis plate. In those instances where it was necessary that the outer case of the capacitor be insulated from ground, the unit was first covered with a plastic sleeve.

The battery compartment and receiver case shown here were milled from linen-base phenolic stock. For convenience, the receiver case was milled in two sections which were then cemented together to form a slide-on unit which fits flush with the battery case and is screw-fastened to a bracket on the bottom of the receiver chassis. Other experimental receiver cases have been constructed using resin-impregnated fibre-glass cloth.

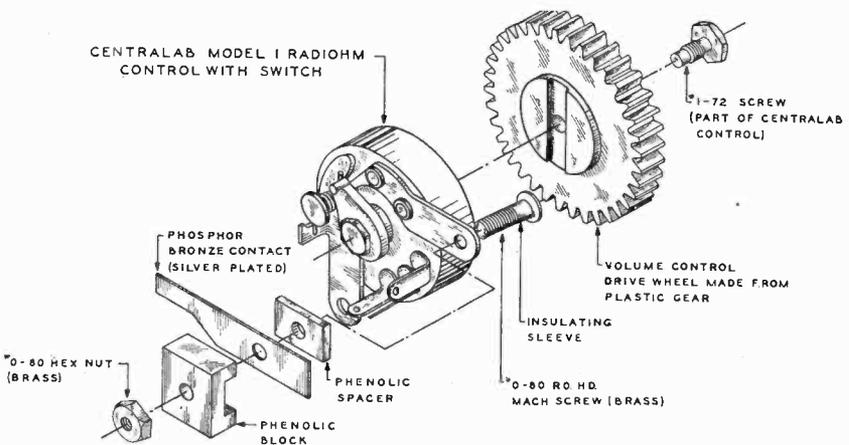


Fig. 12—Exploded view of modified volume control and switch.

DESIGN CONSIDERATIONS IN CLASS-B COMPLEMENTARY SYMMETRY CIRCUITS

BY

T. O. STANLEY AND T. M. SCOTT

RCA Laboratories,
Princeton, N. J.

Summary—The potentialities of the complementary symmetry exhibited by p-n-p and n-p-n transistors were first explored by Sziklai.¹ The present paper discusses design considerations for transformerless complementary-symmetry circuits providing class B push-pull audio output at medium power levels. Two three-transistor amplifiers, one of which provides 75 milliwatts and the other 135 milliwatts output (suitable for small portable receivers) and a four-transistor amplifier which provides 2 watts output (suitable for a portable phonograph or table-model radio) are described in detail. Performance characteristics of these amplifiers, including that for one thermistor temperature-compensated circuit, are shown for the ambient temperature range 0°-50°C. Comparison of these amplifiers with transformer-coupled amplifiers indicates that complementary symmetry affords the circuit designer the option of eliminating two audio transformers while maintaining the same output performance, but with some reduction in gain.

INTRODUCTION

IT was shown by Sziklai¹ that p-n-p and n-p-n transistors have complementary properties which permit circuits not possible with vacuum tubes. Of particular value are class-B push-pull audio output systems which require no transformers; these are discussed in the present paper.

TRANSISTORS FOR COMPLEMENTARY SYMMETRY

Experimental transistors which might be considered for developmental complementary-symmetry circuits have been described.² Many transistors which might be considered for low-level stages, and a number of alloy-junction transistors intended for high-current medium-power applications, may be obtained. In addition, several available "radio-frequency" alloy-junction transistors, while not intended for

¹ George C. Sziklai, "Symmetrical Properties of Transistors and Their Application," *Proc. I.R.E.*, Vol. 41, p. 717, June, 1953.

² For example, see B. N. Slade, "Recent Advances in Power Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 153-171.

high-current service, do exhibit respectable performance. (This is not entirely fortuitous, since several of the factors that enhance high-frequency operation, e.g., close junction spacing and low base-lead resistance, enhance high-current performance as well.) The circuits herein described employ transistors which were obtained in small quantities without consideration of cost or future availability.

The transistor parameters that most directly influence performance in complementary-symmetry output circuits are current gain and transconductance. Loosely speaking, current gain describes the relation between base or emitter current and short-circuit collector current. The base-to-collector current gain, α_{cb} (also sometimes designated β), is the ratio of an increment in collector current to an increment in base current for some specified operating conditions. Since α_{cb} diminishes at higher collector currents, a second parameter, "large-signal current gain," or "d-c alpha," may be specified. This refers to a collector-current to base-current ratio at some relatively high value of collector current. One of the factors that influences α_{cb} and large-signal current gain is minority carrier type; that is, whether the transistor is n-p-n or p-n-p. Other things being equal, one might expect the n-p-n to excel both in α_{cb} and in large-signal current gain, but even with mechanically similar transistors, many other factors enter, and either may excel. The factors that influence current gain, and its dependence on current level are summarized elsewhere.³ The differences in current gain between n-p-n and p-n-p units that can be tolerated in complementary-symmetry applications depend upon the circuits, and will be discussed later.

Transistor transconductance is infrequently specified directly. At low current levels, the incremental emitter current change with incremental base-to-emitter voltage usually follows the relationship for an idealized transistor, $\partial I_e / \partial V_{be} = - (kT/q) I_e$, where $kT/q = 38.9$ volts⁻¹ at 25°C. At higher currents, the incremental voltage drop produced by base current flow in the base-lead resistance becomes appreciable, and the transconductance no longer increases linearly with emitter current; in fact, it ultimately decreases with increasing emitter current. The presence of any emitter-lead resistance also reduces the transconductance from the ideal value. A "large-signal transconductance" may be inferred when the required base-to-emitter voltage is specified for some relatively large value of collector current.

Generally, the impedance levels in complementary-symmetry output

³ E. W. Herold, "New Advances in the Junction Transistor," *British Journal of Applied Sciences*, Vol. 5, pp. 115-126, April, 1954; also *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 27-58; see also L. D. Armstrong, C. L. Carlson, and M. Bentivegna, "P-N-P Transistors Using High-Emitter Alloy Materials," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 144-152.

circuits are relatively low, and signal voltage developed at the collector has little direct effect on collector current. An important exception occurs when the collector "bottoms," that is, when the reverse bias at the collector junction becomes too small for collection to take place. A parameter which describes the minimum collector-to-emitter voltage for collection at some relatively large collector current is then of interest. The ratio of this minimum voltage to the collector current may be thought of as an effective series resistance which the transistor exhibits.

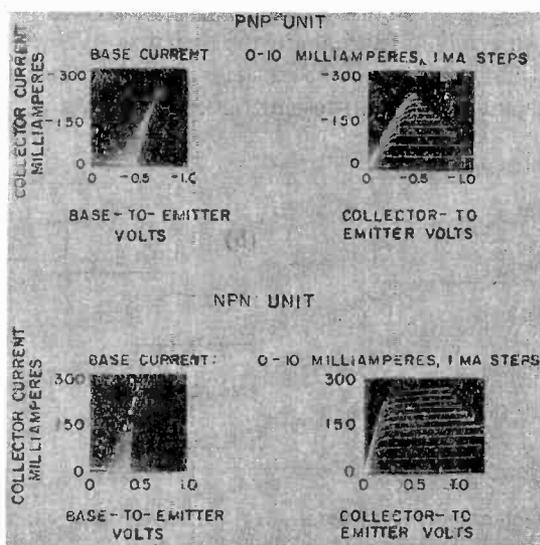


Fig. 1—Representative transistor characteristic curves.

All of these parameters may be derived from sets of characteristics such as those shown in Figure 1. Representative curves for any transistor type are usually available from the manufacturer.

The transistor voltage and dissipation limits specified by the manufacturer must be taken into account in circuit design. The range of ambient temperature over which performance is satisfactory is intimately related to circuit design, as discussed elsewhere,⁴ and will be considered here in conjunction with some representative circuits.

BASIC CIRCUITS

When a single-ended amplifier employing a p-n-p transistor is con-

⁴ H. C. Lin and A. A. Barco, "Temperature Effects in Circuits Using Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 369-402.

nected in parallel (for signal frequencies) with the same type of single-ended amplifier employing an n-p-n transistor, a rudimentary push-pull complementary-symmetry amplifier is born. The circuit becomes something more than rudimentary when advantages in biasing and coupling are exploited, e.g., in class-B operation without transformers. Basic complementary-symmetry circuit types reflect the single-ended circuit types from which they are derived. One of the simplest is the common-collector or emitter-follower circuit, shown diagrammatically in Figure 2a. When the paralleled bases swing negative, the p-n-p conducts and the paralleled emitters follow; for positive swing the n-p-n conducts, and again the emitters follow the bases. The voltage gain of this circuit is necessarily smaller than unity; the current gain is equal to the base-to-emitter current gain ($\alpha_{eb} + 1$) of the transistors.

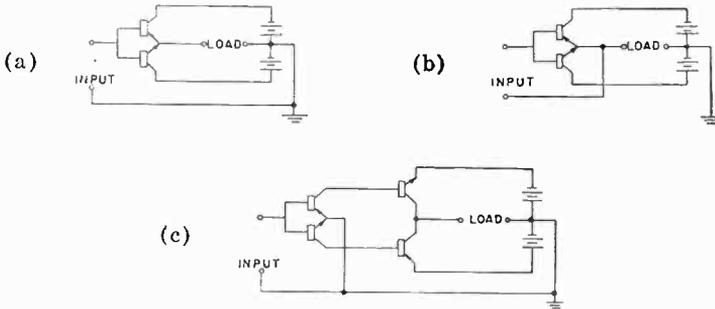


Fig. 2—Basic complementary-symmetry circuits: (a) common collector; (b) emitter loaded; (c) cascade common emitter (in complete circuits, the battery center tap may in some cases be eliminated).

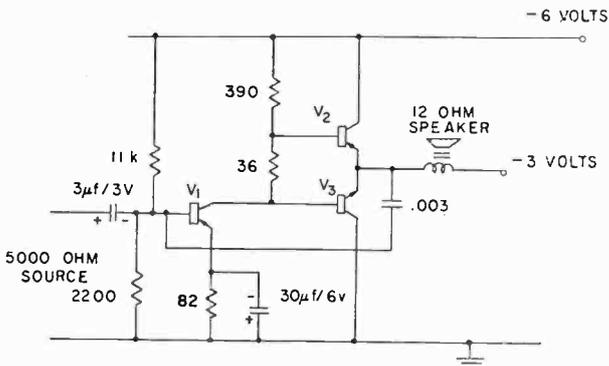
A modification of this circuit, the *emitter-loaded* amplifier, is shown in Figure 2b. In a sense, this is a common-emitter circuit, since the signal source and load are both returned to the paralleled emitters. However, note that in this circuit the *common* point does not correspond to the amplifier *ground*. The voltage gain of this circuit (the product of the transistor transconductance and the load impedance) may be quite large; the current gain is equal to the base-to-collector current gain, α_{cb} , of the transistors.

A two-stage cascaded common-emitter amplifier is shown in Figure 2c. Here the signal source, the load, and the various emitters are all returned to a-c ground. The separate collectors of the first stage drive the separate bases of the second stage, and the paralleled collectors of the second stage are connected to the load. The voltage and current gains of this amplifier are greater than single-stage common-emitter gains by a factor equal to the second-stage current-gain.

SOME ILLUSTRATIVE CIRCUITS

Common Collector

An audio amplifier employing a common-collector output stage and a single-ended class-A driver stage is illustrated in Figure 3. This amplifier, which might be employed in a portable receiver, provides a nominal maximum output of 75 milliwatts. The driver is direct-coupled to the output stage. A bypassed emitter-return resistor stabilizes driver current and the voltage developed by driver collector current flowing through a 36-ohm resistor connected between the out-



V_1 : LOW LEVEL P-N-P

V_2 : HIGH CURRENT P-N-P*

V_3 : HIGH CURRENT N-P-N*

ALL RESISTORS CARBON, 5%

* SEE FIG.1 FOR REPRESENTATIVE CHARACTERISTICS

Fig. 3—Amplifier employing common-collector output stage.

put-stage bases provides threshold bias for the output transistors. The 12-ohm voice coil is direct-driven by the output stage emitters, and returns to the center-tap of the -6-volt supply. A .003-microfarad capacitor controls high-frequency response by introducing negative feedback at frequencies above 3 kilocycles. The sensitivity of the amplifier is 80 microamperes r-m-s for 75 milliwatts output.

The driver stage, which is biased at 7.5 milliamperes, is direct-coupled to the output stage. With no signal present, the driver collector current develops about -2.88 and -3.12 volts at the n-p-n and p-n-p bases respectively. Since the paralleled emitters of the output stage are returned, via the voice coil, to the -3-volt tap, this means that each of the output transistors is provided with a small forward bias, so that

about 2 milliamperes emitter current flows serially through the emitters. This threshold bias minimizes nonlinearity in the crossover interval, when current shifts from one output transistor to the other. The total quiescent current drain of the amplifier, then, is slightly more than 9.5 milliamperes.

The driver stage is constant-emitter-current biased by means of the bypassed emitter-return resistor. The efficacy of this type of bias with respect to changes in ambient temperature may best be understood by reference to Lin and Barco.⁴ The most important consequence of the shift in driver collector current with change in ambient temperature is the unbalancing of the quiescent currents in the output transistors. For example, at reduced ambient temperatures driver collector current is reduced and the bases of each of the output stage transistors becomes more negative. The p-n-p quiescent current increases and the n-p-n quiescent current diminishes, resulting in unbalanced battery drain and direct-current flow through the speaker. A second important effect of ambient temperature change occurs in the output transistors. While the quiescent voltage developed between the output stage bases does not change significantly with temperature change, the corresponding threshold currents change rather rapidly. This results in the appearance of cross-over distortion at low ambient temperatures, and in increased battery drain at high temperatures. The substitution of a suitable thermistor-resistor combination for the 36-ohm threshold-bias resistor essentially eliminates this second effect by developing a threshold-bias voltage which exhibits the proper variation with ambient temperature. An example of thermistor compensation is shown for a subsequent circuit.

Signal current, after amplification by the driver stage, is shared by the bases of the output stage and the collector-return resistance (coupling resistance) of the driver, while the collector-return resistance alone carries essentially all the static driver current. Hence, the dynamic impedance at the driver collector is smaller than the static resistance; this means that a 7.5-milliamper peak signal swing produces a voltage swing at this point which is necessarily smaller than 3 volts peak. Maximum output is limited by the onset of clipping, which occurs due to current limitation in the driver. In the amplifier illustrated, a 1.8-volt maximum peak signal is developed at the driver collector, about 4.5 milliamperes peak flowing in the coupling resistance, and about 3 milliamperes peak into the output bases. The corresponding peak base-to-emitter voltage is about 0.45 volt and peak emitter current is about 110 milliamperes, producing 75 milliwatts maximum in a 12-ohm load. The voltage at the driver collector for peak maximum col-

lector current is about -1.2 volts. This margin is exploited by the driver-stage bias arrangement, which requires that the driver emitter be somewhat negative with respect to ground.

The common-collector connection is degenerative since signal voltage developed across the load diverts driver signal current into the coupling resistance. Increased negative feedback for the same load may be obtained by reducing the coupling resistance. However, since appreciable signal voltage appears from base-to-emitter of the output transistors, it may be shown that the stage gain decreases more rapidly than the negative feedback increases. For instance, the negative feedback for the circuit values shown is 5.2 decibels; increasing the feedback to 8.2 decibels would require reducing the coupling resistance to 137 ohms with a resultant gain reduction of 6.7 decibels, and an increase in quiescent driver current to about 22 milliamperes. The negative feedback may be increased much more efficiently by introducing current feedback from the load to the base of the driver. No increase in driver current is required, and the gain reduction is essentially equal to the increase in negative feedback. Capacitive current feedback, which is employed to control the high-frequency response of this amplifier, ameliorates any residual cross-over distortion. Low-frequency response is controlled by the magnitude of the driver emitter-bypass capacitor.

Increasing the magnitude of the coupling resistance would increase the gain, reduce the negative feedback and the driver quiescent current, but also reduce the maximum power output capability, since the peak driver current, and hence the peak voltage applied at the output bases, necessarily would be smaller. For instance, increasing the coupling resistance to 780 ohms increases the output stage gain by 3 decibels and requires a reduction in driver current to about 3.8 milliamperes but reduces the negative feedback to 3.4 decibels and the maximum power output capability to 40 milliwatts. The performance data of this circuit over the ambient temperature range 0° - 50°C is illustrated in Figure 4.

The data shown applies for output transistors that are fairly well matched. Considerable unbalance can be tolerated, especially if the unbalance is in the direction of increased current gain or transconductance. For example, if the "bogey" n-p-n is replaced by one which exhibits twice the large-signal current gain, the theoretical total harmonic distortion introduced from the mismatch alone is something less than 7 per cent, mostly second harmonic. In practice, less than the theoretical distortion is introduced, due largely to the smooth current transition provided by the threshold bias. Asymmetrical clipping reduces the maximum power out to about 60 milliwatts, and an un-

balance current of about 12 milliamperes d-c flows through the speaker at maximum power out.

The substitution of an output transistor of higher or lower large-signal transconductance increases or decreases, respectively, the output signal level required for clipping in the substituted transistor. In addition, an increase in transconductance acts like an increase in current

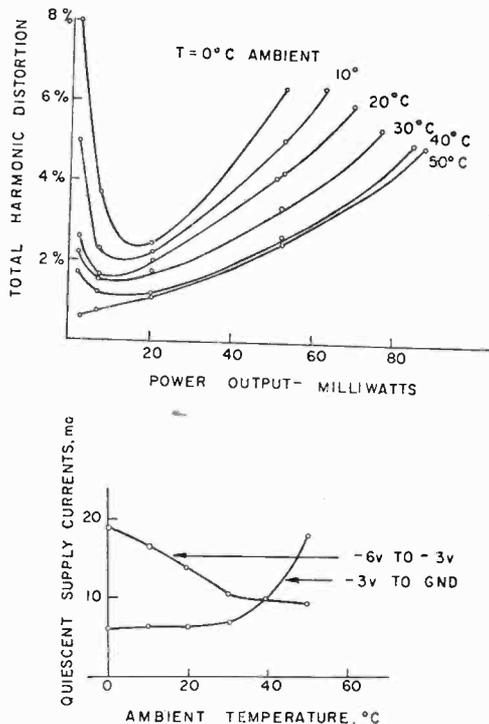


Fig. 4—Common-collector amplifier performance as a function of ambient temperature.

gain, in that the *stage* current gain increases somewhat. In this circuit, collector bottoming does not normally occur, so unbalance in this parameter is of no significance. In extreme cases, however, such as that for the transistor of Figure 5, collector bottoming limits the maximum output signal.

Emitter Loaded

An amplifier employing an emitter-loaded output stage and a single-ended class-A driver stage is illustrated in Figure 6. The configuration of this circuit is similar to that of the common-collector circuit, with the following distinctions: the driver stage bias is stabilized by d-c feedback from the output stage, the voice-coil is capacitively coupled

to the output stage and returned to -6 volts rather than to a center-tap, and the driver coupling-resistance is returned to the "high" side of the voice-coil. This circuit offers some advantages over the common-collector amplifier: the output peak-to-peak voltage may approach the d-c voltage without a concomitant large decrease in gain and increase in driver current, negative feedback and gain may be traded equally, the d-c supply is not center-tapped and thus there is no problem of unbalanced battery drain and no appreciable direct current flows in the speaker. At the same time, some signal voltage is developed across the load-coupling capacitor, reducing the output efficiency and limiting the low-frequency response. (The series resistance as well as the capacitive reactance of this capacitor must be low compared to the voice-coil impedance, and the manufacturer's r-m-s current limit must

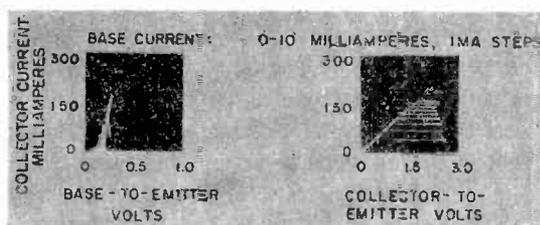


Fig. 5—Transistor characteristic curves for a transistor exhibiting high series collector resistance.

be observed.) The amplifier illustrated, operating from a -6 -volt supply, provides a nominal 130-milliwatt maximum output into a 12-ohm speaker for 100 microamperes r-m-s input.

The driver stage is biased at 5 milliamperes, and provides threshold bias for the output stage in the same manner as in the previously discussed amplifier. However, since the speaker is capacitively coupled to the paralleled emitters here, a shift in driver current produces no unbalanced output current, but only a change in the quiescent voltage at the emitters, with a corresponding asymmetry in clipping and reduction in maximum power output capability. The total quiescent current drain of the amplifier is slightly more than 7 milliamperes.

The driver bias is stabilized by d-c feedback from the paralleled output emitters to the driver base and by the emitter-return resistor. The feedback functions in essentially the same manner as collector-to-base stabilization. The variation of threshold current with ambient temperature for this amplifier does not differ appreciably from the previously discussed circuit. The emitter-loaded amplifier may be modified for greater temperature stability by the incorporation of

thermistor compensation in the output stage, as illustrated in Figure 7.

Operating principles of the circuits, with or without thermistor compensation, are essentially identical with respect to signal. Signal current, after amplification by the driver stage, is shared by the bases of the output stage and the coupling resistance. Since the coupling resistance is returned to the *high* side of the load, the emitter-loaded connection is not degenerative: only the sum of the signal voltage from base-to-emitter and the signal voltage developed across the load-coupling capacitor need be developed across the coupling resistance. The signal voltage at the driver collector is the sum of the above voltages and the output voltage. The magnitude of the dynamic impedance at the driver collector actually somewhat exceeds the coupling resistance: for this condition the coupling resistance is made slightly lower than the product of the output-transistor current gain and the load impedance. Clipping occurs then due to collector bottoming in the driver stage on positive signal peaks, and due to collector bottoming of the p-n-p output transistor on negative signal peaks. The peak-to-peak voltage at the driver collector is approximately equal to the d-c supply voltage; the output voltage is smaller than the supply voltage by the sum of the base-to-emitter voltages and the signal developed across the load-coupling capacitor.

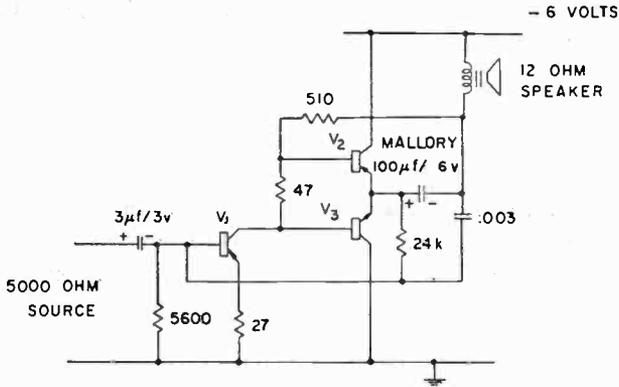
The feedback resistor from the output emitters to the driver base introduces negative feedback for signal (in this case, 6 decibels) as well as for d-c. An R-C network may be employed here if less signal feedback than d-c feedback should be required. High-frequency response is controlled by capacitive-current feedback.

The consequences of unbalanced current gain in the output transistors in this circuit are essentially the same as in the previous circuit, except that large signal unbalance can produce no d-c shift in the load. At the same time, the driver d-c feedback brings about a shift in the driver operating bias which tends to "recenter" the signal, minimizing asymmetrical clipping.

The situation with respect to unbalance in transconductance here does not differ from that in the common-collector circuit, but collector bottoming becomes significant for the emitter-loaded circuit, since the peak-to-peak signal voltage is more nearly equal to the supply voltage.

The performance of the circuits of Figures 6 and 7 over the ambient temperature range 0°-50°C is illustrated in Figures 8 and 9. Performance of the unmodified circuit at 25°C with current-gain unbalance in the output stage is illustrated in Figure 10.

An audio amplifier employing an emitter-loaded complementary



V₁ : LOW LEVEL P-N-P

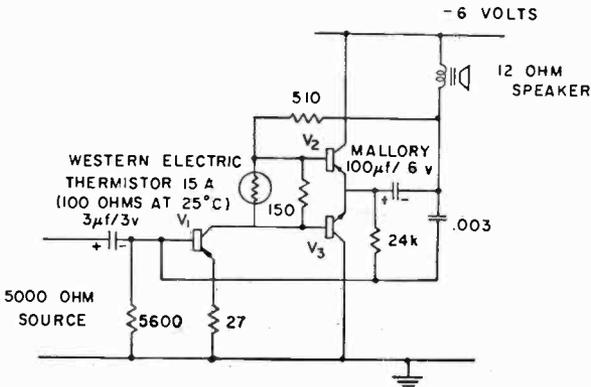
V₂ : HIGH CURRENT P-N-P*

V₃ : HIGH CURRENT N-P-N*

ALL RESISTORS CARBON, 5%

* SEE FIG. 1 FOR REPRESENTATIVE CHARACTERISTICS

Fig. 6—Amplifier employing emitter-loaded output stage.



V₁ : LOW LEVEL P-N-P

V₂ : HIGH CURRENT P-N-P*

V₃ : HIGH CURRENT N-P-N*

ALL RESISTORS CARBON, 5%

* SEE FIG. 1 FOR REPRESENTATIVE CHARACTERISTICS

Fig. 7—Temperature-compensated emitter-loaded amplifier.

symmetry output stage which is direct-coupled to the speaker has been described elsewhere.⁵

Common Emitter

The cascaded common-emitter circuit is employed in the amplifier shown in Figure 11. In this circuit, both driver and output stages operate push-pull, and direct coupling obtains throughout. This amplifier, which might find application in a portable phonograph, table model radio, etc., provides a nominal 2.0-watt maximum output into an 8-ohm speaker for 1.5-milliampere r-m-s input, operating from a 12-volt center-tapped supply.

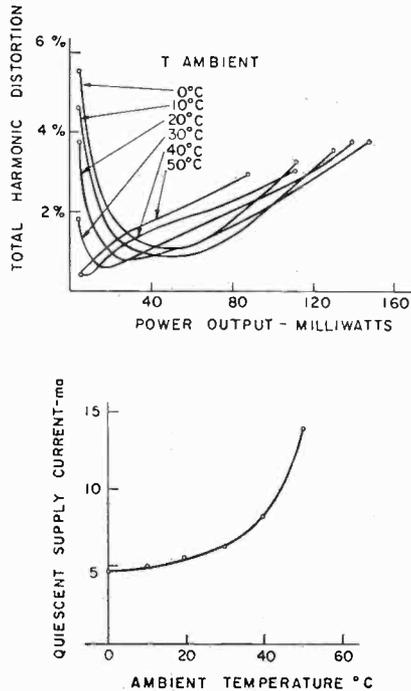


Fig. 8—Emitter-loaded amplifier performance as a function of ambient temperature (circuit of Figure 6).

The sum of the threshold currents of the driver-stage transistors is stabilized by the emitter-return resistor networks in a manner analogous to constant-emitter-current bias, i.e., sufficient emitter current flows out of the junction of the 39-ohm and 1200-ohm resistors so that

⁵D. D. Holmes, T. O. Stanley, and L. A. Freedman, "A Developmental Pocket-Size Broadcast Receiver Employing Transistors," *Proc. I.R.E.*, Vol. 43, p. 662, June, 1955; also *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 477-496.

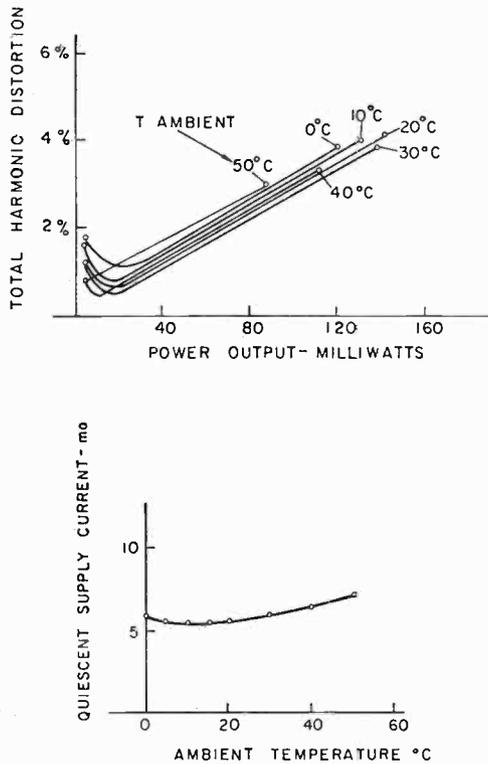


Fig. 9—Temperature-compensated emitter-loaded amplifier performance (circuit of Figure 7).

only the sum of the base-to-emitter voltages of the two driver-stage transistors is developed across the 39-ohm resistors. The ratio of the n-p-n to p-n-p emitter currents is equal to the ratio of the respective current gains, since the static base current that flows out of the p-n-p

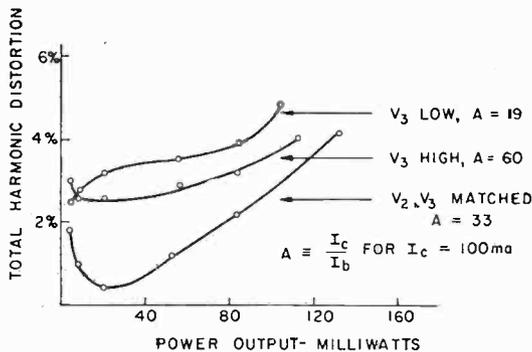
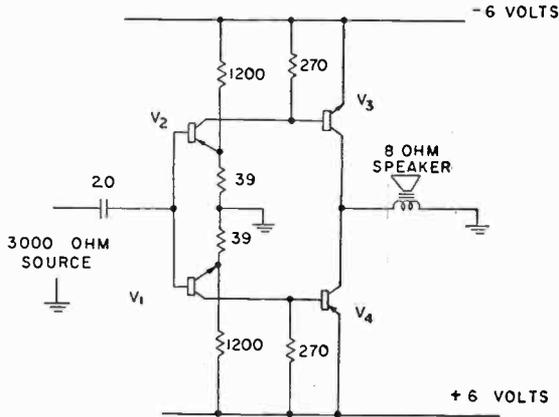


Fig. 10—Emitter-loaded amplifier performance at 25°C for unbalanced output transistor current gains.

flows into the n-p-n. The sensitivity of this bias arrangement to the increase in base-saturation current that occurs at elevated temperatures depends upon the similarity of the n-p-n and p-n-p units with respect to saturation current, any difference in saturation current of the two units producing an unbalance in the two emitter currents. The driver-stage collectors are direct-coupled to the output-stage bases, and returned via 270-ohm resistors to the respective output-stage emitters.

Driver-stage collector current, about 2 milliamperes at 25°C, develops threshold bias for the output-stage transistors, biasing them to



V_1 : LOW LEVEL N-P-N

V_2 : LOW LEVEL P-N-P

V_3 : MEDIUM POWER N-P-N

V_4 MEDIUM POWER P-N-P

ALL RESISTORS $\frac{1}{2}$ W CARBON, 5%

Fig. 11—Cascade common-emitter amplifier.

about 12 milliamperes emitter current at 25°C. Due to the nonlinearity of the base voltage-current characteristic, these resistors do not materially shunt the output bases for large signals.

For output signals larger than a few milliwatts, each cascaded pair of transistors works essentially half the time. Any unbalance of signal peaks arises from differences in the product of the stage current gains for each pair. Since the n-p-n driver works into the p-n-p output transistor and vice versa, any systematic difference in current gain between p-n-p and n-p-n units tends to be balanced. The maximum output voltage obtainable is limited by collector bottoming. This amplifier will tolerate output transistors which exhibit relatively low large-signal

transconductance, since this parameter does not here affect the maximum output power, and since the coupling resistors can be adjusted so that no significant reduction in output-stage current gain is incurred.

The performance of this amplifier over the ambient temperature range 0° to 50°C is illustrated in Figure 12. Negative feedback is not incorporated into this amplifier, so the performance characteristics represent the maximum gain, maximum distortion mode of operation. Low-frequency response is limited only by the input-coupling capacitor, and high-frequency response by the transistors. The response is down 1 decibel at 20 cycles, 3 decibels at 10 kilocycles. Feedback, applied for

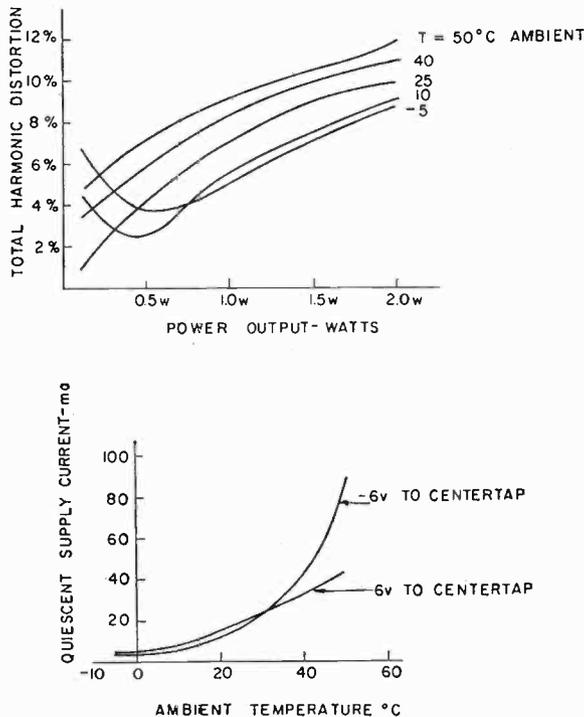


Fig. 12—Cascade common-emitter amplifier performance as a function of ambient temperature.

example by connecting suitable admittances from the output to each of the input stage emitters, could be expected to effect an exchange of gain for improved distortion performance. Additional stabilization of the d-c operating point can be obtained at the same time, by providing a d-c path to ground for the paralleled input-stage bases. An amplifier of this type employing 100 per cent voltage feedback is described by Sziklai.¹

GENERALIZATIONS AND CONCLUSIONS

Among the performance criteria by which an audio amplifier might be judged are gain, linearity, output power capability, stability with respect to ambient temperature changes, and tolerance of variation in transistor parameters. By means of negative feedback, gain and linearity can usually be traded on equal terms. Achieving greater temperature stability, on the other hand, can require that some fraction of the d-c supply voltage be sacrificed for bias purposes, with a resultant reduction in output power capability and gain. The degree of variation in transistor parameters for which the transistors can be said to be interchangeable in the foregoing circuits can generally be increased by the application of increased signal and d-c feedback. On the other hand, the circuits may be revised for increased gain if a wide range in transistor parameters need not be tolerated (for example, it may be possible to eliminate signal feedback). The output power capability of the amplifiers described depends fundamentally on the speaker impedance and the supply voltage. The upper limit on maximum output

power, $\frac{V_{\text{supply}}^2}{8 R_{\text{speaker}}}$, would occur for the peak to peak signal across the

voice coil equal to the supply voltage. The nominal voice-coil impedances of small speakers are in the 10 to 16 ohm range (No. 40 wire); fabrication becomes increasingly difficult at higher impedances where finer wire must be employed. Larger speakers offer a greater latitude in this respect. Thus, for small speakers, power-output requirements usually dictate the supply voltage employed, and emphasis is placed on the output transistors as low-impedance, low-voltage devices.

Comparison of the complementary-symmetry amplifiers with transformer-coupled push-pull amplifiers reveal certain basic differences, the most prominent difference being the elimination, in the complementary-symmetry amplifier, of the interstage and output transformers. This becomes particularly significant where minimizing size and weight, or obtaining a wide frequency response is important. (An intermediate arrangement, permitting the elimination of the output transformer, but not the interstage, is described by Lin.⁶)

Basic differences that affect performance exist in both the output and driver stages. The maximum peak-to-peak signal between collector and emitter of either complementary-symmetry output transistor is necessarily somewhat smaller than the supply voltage; in the trans-

⁶ H. C. Lin, "A 20-Watt Transistor Audio Amplifier," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 536-546.

former-coupled output stage, the peak-to-peak collector-to-emitter voltage is more nearly twice the supply voltage. For the same supply voltage, then, and the same output power, the peak collector currents in the complementary-symmetry stage must be more than twice those in the transformer-coupled stage. This difference is slightly offset by the larger insertion loss of the output transformer, compared to the loss for the capacitively or direct-coupled complementary-symmetry circuits. As a consequence, there is more nearly a two-to-one ratio in collector currents required for the same load power. Due to the curvature of the transistor current-gain characteristic, the distortion (for no inverse feedback) for the complementary-symmetry output stage may be somewhat greater than for the transformer-coupled amplifier, and the required peak currents into the output stage bases are more than twice as large.

The current gain of the driver-stage-and-interstage transformer is larger than the current gain of the complementary-symmetry driver stage by a factor equal to the turns ratio of the interstage transformer. The gain of the complementary-symmetry amplifier, then, from driver-base to speaker is lower than the gain of the transformer-coupled amplifier (for the same supply voltage and output performance) on three counts: (1) the 2:1 difference in output stage peak base currents, (2) the current gain of the interstage transformer, and (3) the likely application of somewhat more negative feedback in the complementary-symmetry amplifier to insure the same output performance. This difference in gain might be on the order of -18 to -24 decibels. An important additional consideration arises upon the incorporation of the amplifier into a receiver. The input presumably would be driven from a detector; the detected audio levels for standard output would be on the order of a few microwatts for the complementary-symmetry amplifier and a few hundredths of a microwatt for the transformer-coupled amplifier. This difference in level should materially affect detector performance. For example, a detector of the general type described by Holmes, Stanley, and Freedman,⁵ adjusted at each level to provide the same distortion, affords a conversion gain approximately proportional to the input power level. The difference in receiver gain effected by the two amplifiers, then, is more nearly -9 to -12 decibels.

There is little fundamental difference in the bias requirements for the two amplifiers. In each case, the output stage requires threshold bias, which can be temperature compensated as required by the application. The driver stages in the complementary-symmetry amplifiers discussed are direct-coupled to the output stages. The operating-current stability required for these driver stages therefore is typically more

stringent than that required for a transformer-coupled stage, the required stability being achieved by d-c feedback.

Complementary-symmetry, then, affords the radio receiver designer the option of eliminating two audio transformers if he can accommodate the resulting 9 to 12 decibel reduction in gain.

TRANSISTOR AUDIO AMPLIFIERS

By

A. ARONSON

RCA Commercial Electronic Products,
Camden, N. J.

Summary—Three audio amplifiers using transistors are described: a 20-dbm (100 milliwatt) preamplifier, and two five-watt power amplifiers. The preamplifier, for use with microphones, tape playback heads, or phonograph pickups, meets broadcast standards with respect to frequency response, distortion, and noise figure. One five-watt power amplifier uses a conventional push-pull output stage with an output transformer; the driver transformer often used with transistor power amplifiers has been eliminated. The other five-watt amplifier which utilizes the complementary-symmetry principle and employs no audio iron-core components, is suitable for high-fidelity sound reproduction.

INTRODUCTION

THE use of transistors in audio circuits has resulted in devices that perform comparably with their tube counterparts at much greater efficiencies, and in some cases have resulted in applications that are not possible with tubes.

The 20-dbm (decibels above 1 milliwatt) preamplifier, the power amplifier, and the complementary-symmetry amplifier described here represent practical applications of transistors to common audio devices. The circuits cover a range of functions from low-noise, low-level input stages to power output stages, and include such elements as volume control, driver, and temperature stabilizing and tone equalizing circuits. Use has been made of the inherent low-impedance characteristics of transistors to eliminate the audio transformers (a source of problems in high-fidelity amplifiers) generally used in the tube counterpart of these amplifiers.

The 20-dbm preamplifier described is characterized by very low distortion, low noise, wide frequency response, and high efficiency. The amplifier is well-suited for use in audio systems requiring a +20-dbm output to a 600-ohm load.

The five-watt power amplifier is characterized by less than 1 per cent distortion at mid-frequencies and a wide frequency response. The driver transformer, generally used in transistor circuits of this type, has been eliminated.

The five-watt complementary symmetry amplifier is characterized by low distortion over the entire audio spectrum. No audio trans-

formers are used and the output is directly coupled to a standard 15-ohm speaker. The amplifier is equalized for use with a variable-reluctance phonograph pickup.

20 DBM PREAMPLIFIER

General Description

The preamplifier shown in Figure 1 uses one n-p-n and five p-n-p transistors. No audio iron-core components are used. The preamplifier will provide a 20-dbm (100 milliwatt) output at low distortion over a range of frequencies from 50 to 15,000 cycles. Using low-noise transistors for the input stages, a noise figure of less than 3 decibels can be obtained.

The performance of the amplifier is sufficiently good for use in broadcast application to operate with microphones, tape playback heads, and low-impedance phonograph pickups.



Fig. 1—A 20-dbm preamplifier.

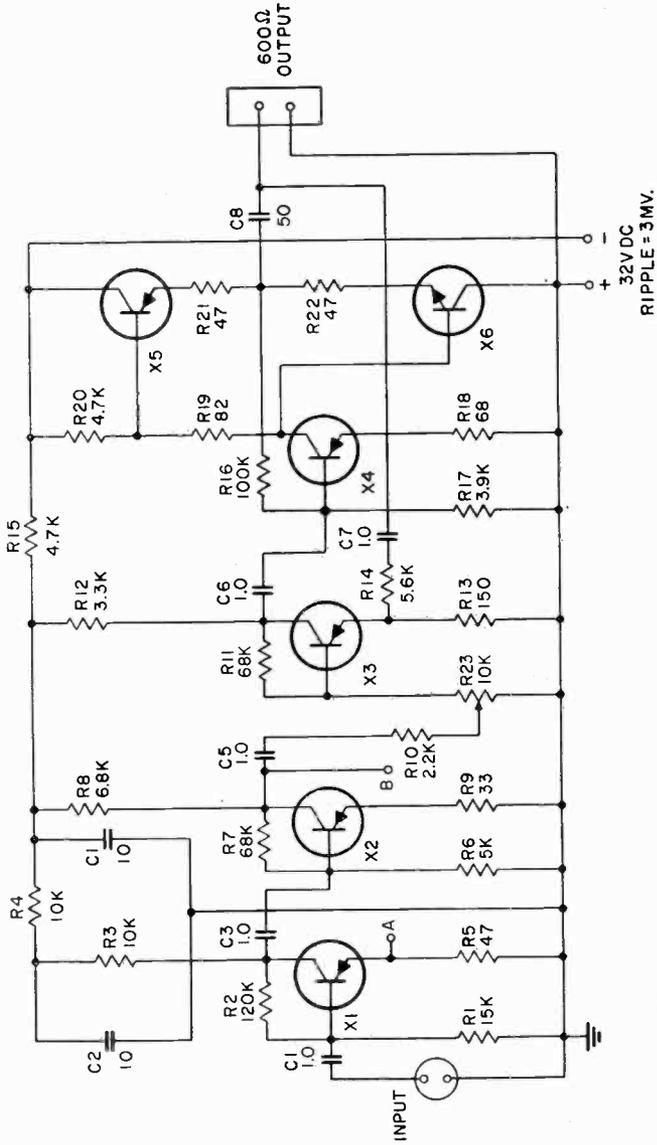
The power drain of the amplifier is very low, making it attractive for a remote amplifier in portable audio equipment. The total power drain is 200 milliwatts, idling, and about 500 milliwatts under full-signal conditions.

Circuit Description

Figure 2 is a complete circuit diagram of the preamplifier. Transistor X-1 is a low noise input stage with X-2 and X-3 serving as common-emitter amplifiers. X-4 is a class A driver amplifier direct-coupled to a complementary-symmetry class-B output stage, X-5 and X-6. The circuit is analyzed by considering the input, medium level and driver-output stages.

Input stage

Noise is the most important consideration in low-level audio-amplifier input stages. Transistors make ideal low-noise amplifiers because the elimination of a heater removes a possible source of hum. The transistor is relatively free from microphonic tendencies, and the low-impedance circuitry associated with transistors makes them less susceptible to electrostatic noise pickup. Finally, the inherent noise of



- NOTE:
 CAPACITIES IN μf
 RESISTANCES IN OHMS
 K=1000
- TRANSISTORS
 X1 2N109
 X2 2N109
 X3 2N109
 X4 2N109
 X5 2N109
 X6 RCA EXP. NPN

Fig. 2—20-dbm preamplifier schematic.

present transistors is low enough to enable audio amplifier designs in some applications to approach the theoretical minimum noise limit with a given source. Figure 3 shows the output signal of the preamplifier with a one-microvolt input signal to illustrate the low noise capability of modern transistors.

To achieve low noise with a transistor amplifier requires a careful selection of source impedance and d-c operating point.¹ Measurements indicate that the minimum noise figure for present junction transistors occurs with a source impedance in the range from 200 to 1,000 ohms. This impedance range matches that of many broadcast microphones so that the microphone can be directly coupled to the input. Transistor noise has been shown to increase with collector current, and to be relatively independent of collector voltage to values as high as 20 volts.

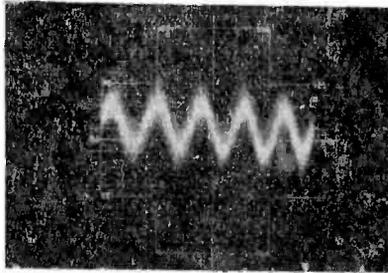


Fig. 3—Output signal of the 20-dbm amplifier with a one microvolt 1,000-cycle input signal. Source impedance is 10 ohms, effective bandwidth is 60 kilocycles.

This is true with transistors having low collector-base leakage. Low-noise operation can be obtained with collector currents in the range from 0.2 to 1.0 milliamperes.

The amplifier has a flat frequency response as shown in Figure 4 and can be used directly with low-impedance microphones with 150- to 500-ohm impedances. At maximum gain-control setting, a 0.3-millivolt signal will drive the amplifier to a 20 dbm output. For use with tape playback heads or magnetic phonograph pickups, a suitable series R-C feedback network can be connected between points A and B in the circuit of Figure 2 to provide for proper low-frequency equalization. The specific R-C value will depend on the particular magnetic transducer used. Typical values with a variable reluctance cartridge are those of the first two stages of the five-watt complementary symmetry amplifier described later.

¹ P. L. Bargellini and M. B. Herscher, "Investigations of Noise in Audio Frequency Amplifiers Using Junction Transistors," *Proc. I.R.E.*, Vol. 43, p. 217, February, 1955.

Medium-Level Amplifiers

Transistors X-2 and X-3 are conventional common-emitter amplifiers. Stabilization of operating point^{2, 3} with ambient temperature change and transistor variation is accomplished by collector-to-base voltage feedback and current feedback in the emitter circuit.

Driver-Output Stage

A class A driver stage, X-4, is direct coupled to a class-B complementary symmetry output stage with transistors X-5 and X-6* connected in a common-collector configuration.

The complementary symmetry configuration was used to eliminate the output transformer and phase inverter generally needed with push-pull stages. The elimination of reactive elements permits a more liberal use of negative feedback for distortion reduction.

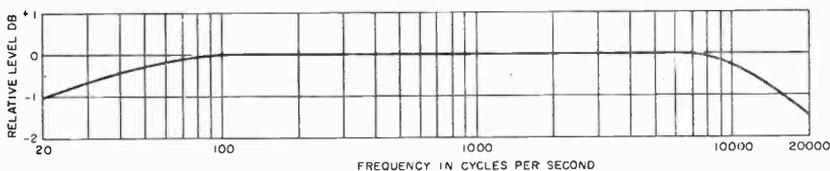


Fig. 4—Frequency response 20-dbm preamplifier.

The transistors in the output stage were connected common collector for the following reasons:

- A. The 100 per cent voltage feedback inherent in this configuration balances the complementary transistors so that low distortion can be obtained with unselected pairs. Transistors with an α_{cb} range of 20 to 100 will work in this circuit provided they have the peak current capability.
- B. Zero-signal biasing can be achieved in a more straightforward manner than in a common-emitter version of this circuit.
- C. Stabilization against ambient temperature change can be achieved more easily because a low resistance can be maintained in the base-emitter circuit.

The main disadvantage of the common-collector connection is the low

² A. W. Lo, R. O. Endres, J. Zawels, F. D. Waldhauer, C. Cheng, *Transistor Electronics*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1955, pp. 134, 152, 165.

³ H. C. Lin and A. A. Barco, "Temperature Effects in Circuits Using Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., pp. 369-402.

* X-6 is an experimental n-p-n transistor with characteristics roughly similar to the 2N109 except for polarity.

voltage gain (unity or less) making drive requirement difficult when a common d-c supply is used for the driver-output stage.

The peak current required from each output transistor when driving a 600-ohm load at 20 dbm output is about 25 milliamperes. The distortion as a function of power output is shown in Figure 5. This freedom from distortion results from the inherent linearity of the output stage and an additional 15 decibels of negative feedback applied from the load to R-13.

The output transistors are connected in series across a single-polarity 32-volt power supply with capacitive coupling to the load. To operate properly in this circuit, the transistors should divide the supply

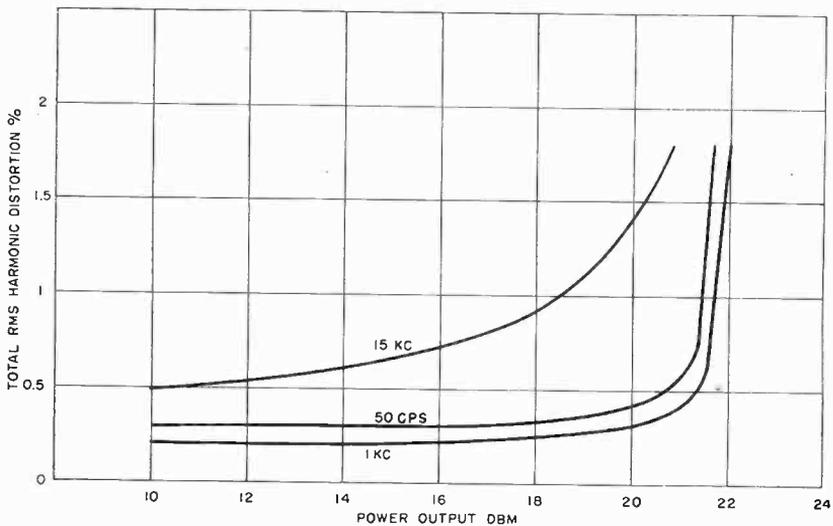


Fig. 5—20-dbm preamplifier power output versus distortion.

voltage equally. This is accomplished by biasing the base circuit with resistor R-19 about 16 volts with respect to ground. The half-supply-voltage point will reflect to the emitter circuit by the emitter follower action of the output stage so that proper collector-to-emitter potential of 16 volts per output transistor will be established. Forward bias voltage, to remove any remaining crossover distortion in the class-B stage, is provided by the collector current flowing in X-4 providing a 300-millivolt drop in R-19. The reference for this bias voltage applied between the bases of X-5 and X-6 will be midpoint of R-19 so that the effective bias for each output transistor will be 150 millivolts.

Stabilization of the output stage for temperature effects is accomplished by emitter feedback due to R-21 and R-22 in addition to the

inherent stability of the common-collector configuration. A portion of the available output power is lost in R-21 and R-22, but good stabilization is achieved. This amplifier is stable up to at least 60°C. The temperature range can be extended, if necessary, by substituting a thermistor for R-19.

5-WATT POWER AMPLIFIER

General Description

Figure 6 is a photograph of the power amplifier showing top and bottom views of the chassis. The two experimental p-n-p power transistors can be seen; the output transformer is also shown. The bottom of the chassis shows the wiring of the subassembly board with the power transformer and filter capacitor representing the two largest components.

A unique feature of this amplifier is the absence of the driver transformer often used in transistor power amplifiers. The elimination of this component helps achieve lower distortion performance, which in this amplifier is limited mainly by the quality of the output transformer.

Circuit Description

Figure 7 is a complete schematic of the power amplifier. The circuit consists of two common-emitter amplifiers, X-1 and X-2, coupled to a split-load phase-inverter. The phase-inverter is capacitance-coupled to a class-B push-pull common-collector driver which is finally direct coupled to the class-B power stage.

Driver-Output Stage

The driver-output stage consists of a push-pull class-B common-collector driver direct-coupled to a push-pull class-B common-emitter output stage.

In order to achieve wide frequency response (about 50 to 15,000 cycles) at low distortion with a practical amplifier, it is desirable to eliminate the driver transformer used to couple signal to the power transistors. Driver transformers tend to become bulky when good low-frequency response below 100 cycles is desired and the added phase shift introduced by this element limits the amount of stable negative feedback that can be employed.

Replacing the driver transformer with a resistance-coupled driver stage requires that the driver source present a low d-c path between the base-emitter circuit of the power transistors. This is necessary to bypass the I_{co} current from the base circuit to avoid amplification of this thermally dependent current. The incremental d-c output resistance of the drivers is 47 ohms per side. The 2N109 transistors

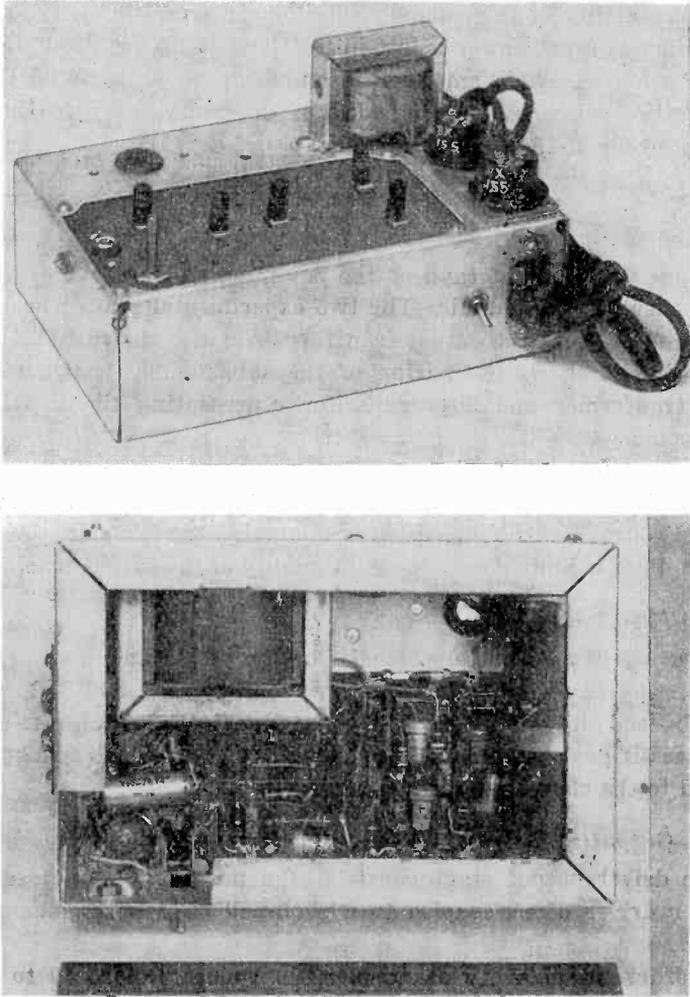


Fig. 6—Five-watt power amplifier.

connected common-collector make ideal drivers for this application. Direct coupling the driver to the output stage also permits one bias supply to bias both the driver and power stage simultaneously. Also, with this direct-coupled arrangement, the coupling capacitors to the power transistor bases are eliminated, thus avoiding a difficult cross-over distortion problem caused by bias changes in class-B circuits due to charging coupling capacitors.

The peak driver output current required will depend on the large-signal current gain of the power transistors. For this particular am-

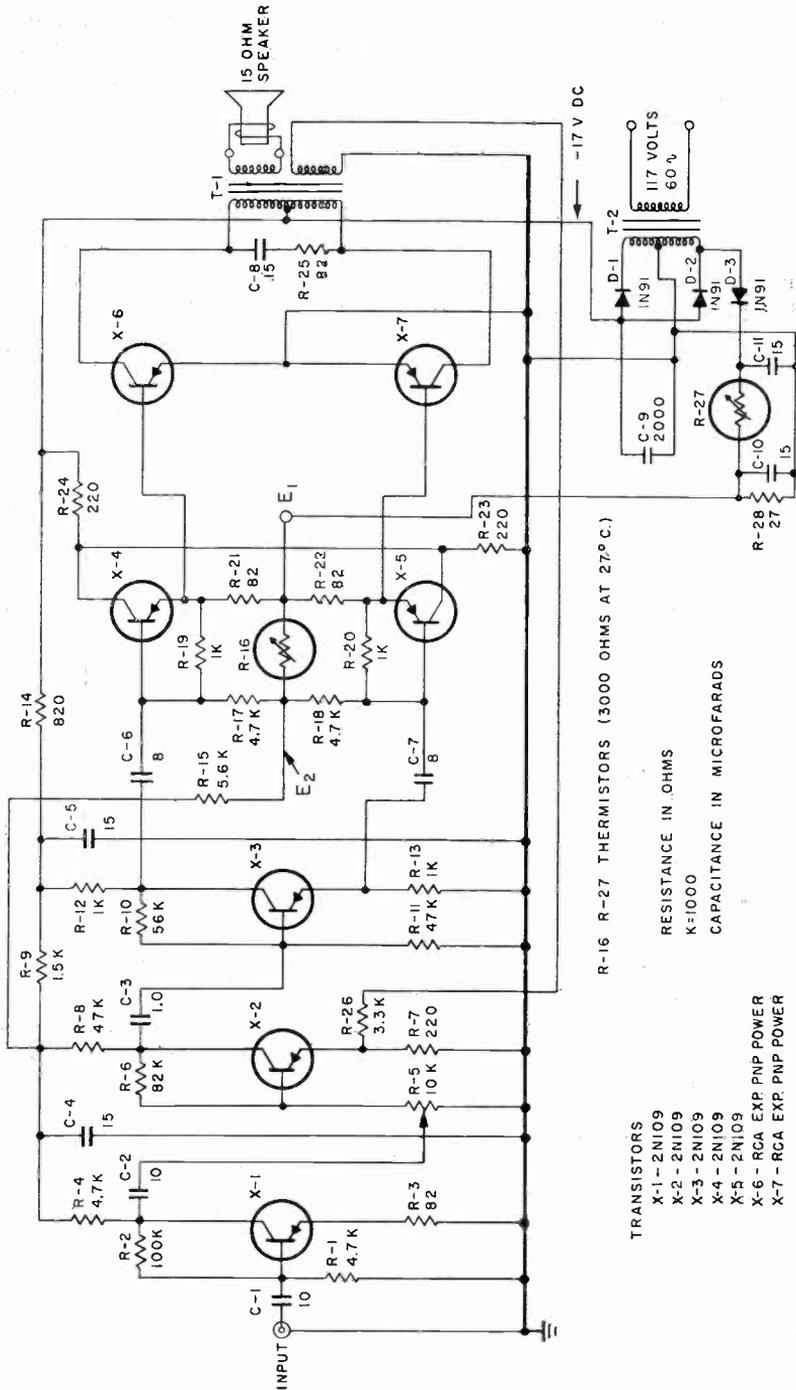


Fig. 7.—Five-watt power amplifier schematic.

plifier, at five watts output, the power transistors should have a minimum α_{cb} of 20 at a collector current of 1 ampere. The 2N109 drivers have sufficient peak current capability and gain to drive the output stage to 10 watts if a supply voltage of 28 volts is used for the power transistors. The purpose of the voltage divider (R-23, R-24) is to reduce the collector voltage on X-4 and X-5 for lower dissipation.

The collector-to-collector loading on the output stage is 96 ohms with a 15-ohm secondary load. Distortion versus loading for the amplifier is shown in Figure 9. Seventeen decibels of tertiary feedback is applied to the emitter resistor, R-7, of X-2, which reduces mid-frequency distortion to less than 1 per cent. Distortion increases at low and high frequencies because of insufficient primary inductance and

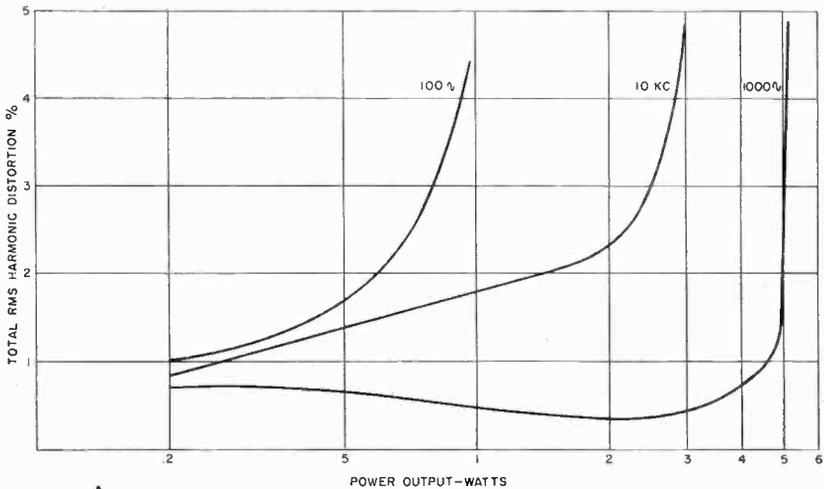


Fig. 8—Five-watt amplifier distortion versus power output.

high leakage reactance in the output transformer, and reduction in feedback. This is shown in Figure 8. The network C-8, R-25 across the primary of the output transformer maintains a high-frequency load on the amplifier to provide for feedback stability on removal of the external load.

Phase-Inverter-Driver

Transistor X-3 is used as a split-load phase inverter feeding the drivers, X-4 and X-5. Minimum distortion of the phase inverter output signal and elimination of the charge on coupling capacitors C-6 and C-7 (which would cause crossover distortion) is desired. This is accomplished by linearizing the input impedance of the class-B drivers, by means of R-19 and R-20 connected between base and emitter of X-4 and X-5.

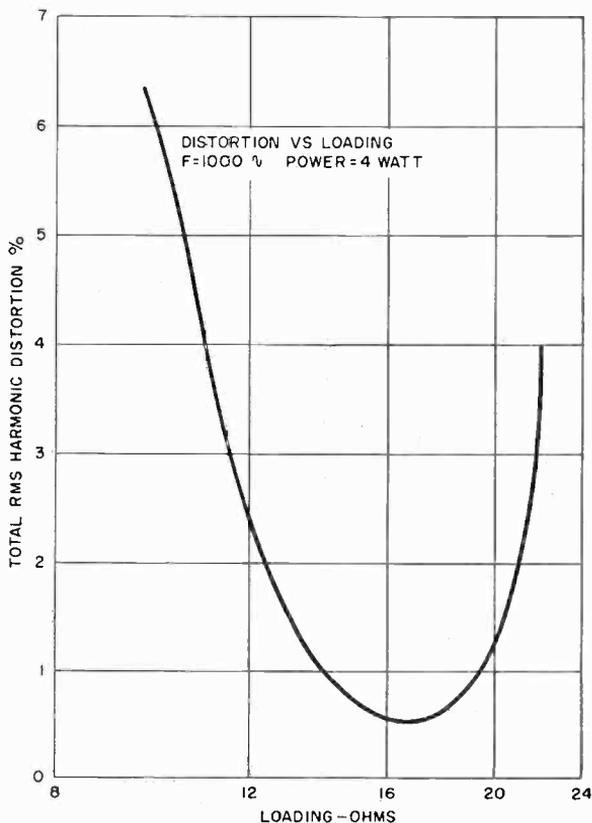


Fig. 9—Five-watt power amplifier distortion versus loading.

How this is done can be seen by analyzing the input impedance of driver stage X-4, (which will be the same for X-5).

During conduction,

$$Z_{in} \cong \frac{R_{19} R_{in\ X-4}}{R_{19} + R_{in\ X-4}} + \alpha_{cb} \frac{R_{21} R_{in\ X-6}}{R_{21} + R_{in\ X-6}},$$

where

$R_{in\ X-4}$ = input impedance of X-4,

$R_{in\ X-6}$ = input impedance of X-6 (which will vary with signal drive).

During cutoff,

$$Z_{in} \cong R_{19} + R_{21},$$

(neglecting shunting effects of $R_{in\ X-4}$ and $R_{in\ X-6}$). By equating these

relationships, values for R-19 and R-21 can be obtained to achieve a relatively constant input impedance during drive and cutoff of the output stage transistor, X-6.

With conduction maintained for the full cycle of signal flow, the tendency for a charge to develop on capacitors C-6 and C-7 producing crossover distortion will be reduced.

Heat Dissipation and Temperature Compensation

A primary consideration in the design of a reliable transistor power stage is the maintenance of low collector junction temperatures. Maintaining low junction temperatures with signal power output, can be achieved by mechanically clamping the power transistor to a heat sink (chassis). With the transistor shell electrically at collector potential, suitable insulation must be provided between the shell and chassis.

Even if great care is exercised in heat dissipation of the power transistors, there still exists the temperature-controlled variation in collector current which at high ambient temperatures tends to become regenerative with eventual thermal runaway.

Temperature compensation can be employed to maintain a constant zero-signal collector current with ambient temperature change. This compensation is accomplished by two thermistor-controlled bias supplies shown in Figure 7. With this arrangement, a positive bias voltage is applied to the bases of the power transistors and a negative voltage applied to the 2N109 bases. With signal present, the negative voltage is adjusted at room temperature until the crossover distortion just disappears. This voltage is varied as shown in Figure 10 by means of thermistor R-16 to maintain a constant zero-signal collector current with temperature change. The positive bias voltage is used to maintain a reverse bias on the power transistors which is desirable at high ambient temperatures. If the thermistors are mounted on the transistor stud, some additional temperature compensation can be achieved. With this temperature compensation, the zero-signal collector current of the power transistors can be held constant from -40°C to $+60^{\circ}\text{C}$.

Low-Level Amplifier

Transistors X-1 and X-2 are conventional common-emitter amplifiers. Stabilization against operating-point drift with temperature is obtained by d-c feedback from collector to base and in the emitter circuit.

Power Supply

Maintaining a well-regulated power supply with the supply current

varying from 50 to 500 milliamperes involves reducing the supply impedance to a few ohms. This is accomplished by using low-resistance power-transformer windings and power junction rectifiers. A 2,000 microfarad filter capacitor was required to reduce the ripple to a low value.

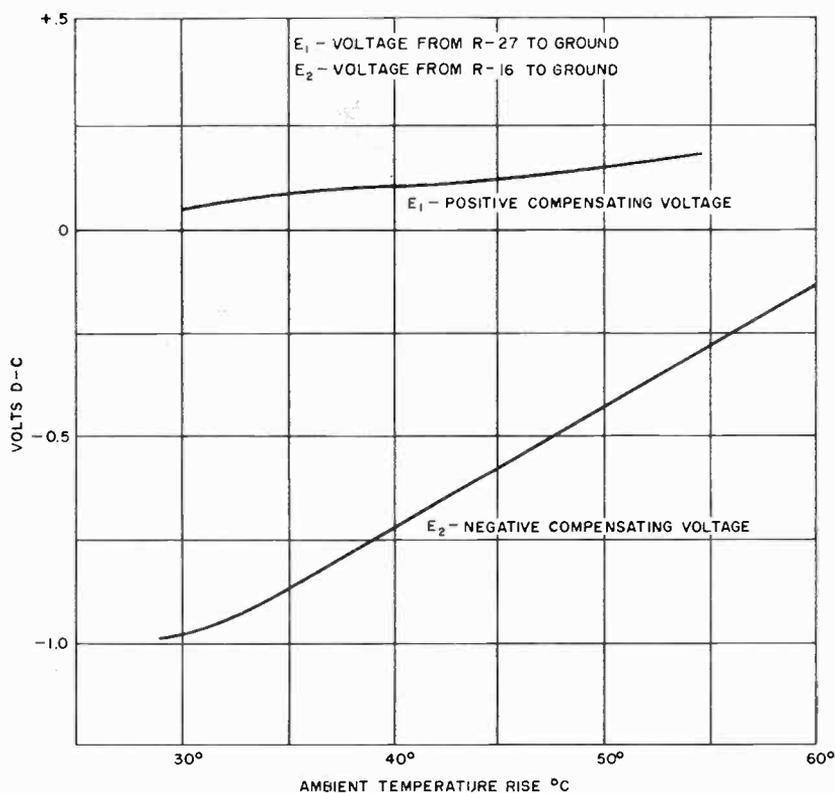


Fig. 10—Five-watt power amplifier temperature compensating voltages.

COMPLEMENTARY-SYMMETRY POWER AMPLIFIER

General Description

The five-watt complementary-symmetry amplifier is shown in Figure 11. The n-p-n and p-n-p power transistors can be seen mounted on the chassis at the right of the photograph, with the remaining portions of the circuit, exclusive of the power supply, mounted on a printed wiring board. A view of the under side of the chassis shows the wiring and power supply components.

The circuit consists of a two-stage equalized phonograph preamplifier, followed by a two-stage class-A amplifier. This last stage is

direct-coupled to two cascaded class-B stages, the first being a driver and the second the output stage.

Circuit Description

Figure 12 represents a complete schematic diagram of the amplifier. The circuit can be conveniently analyzed by considering the individual circuit groupings.

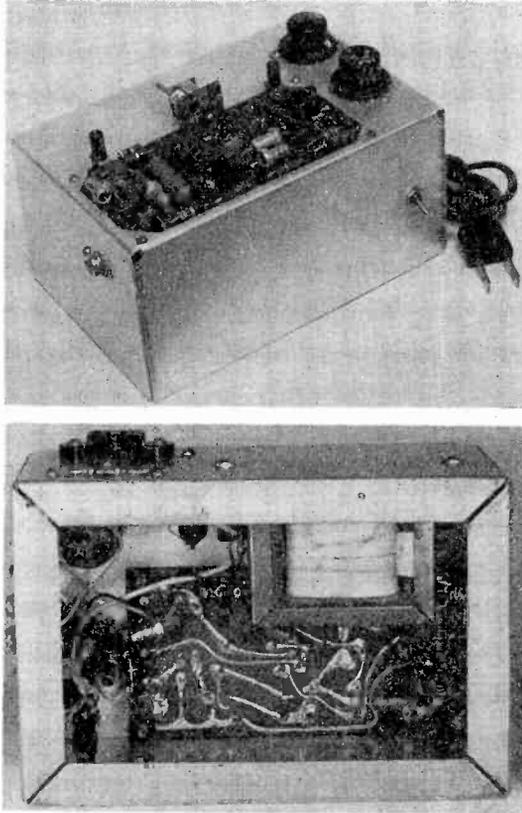


Fig. 11—Five-watt complementary symmetry power amplifier.

Power-Output Stage

Experimental p-n-p and n-p-n power transistors⁴ are used in a common-collector connection for the output stage. The common-collector connection was used in preference to common-emitter or common-base for the reasons outlined in the discussion of the 20-dbm pre-amplifier.

⁴B. N. Slade, "Recent Advances in Power Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 153-171.

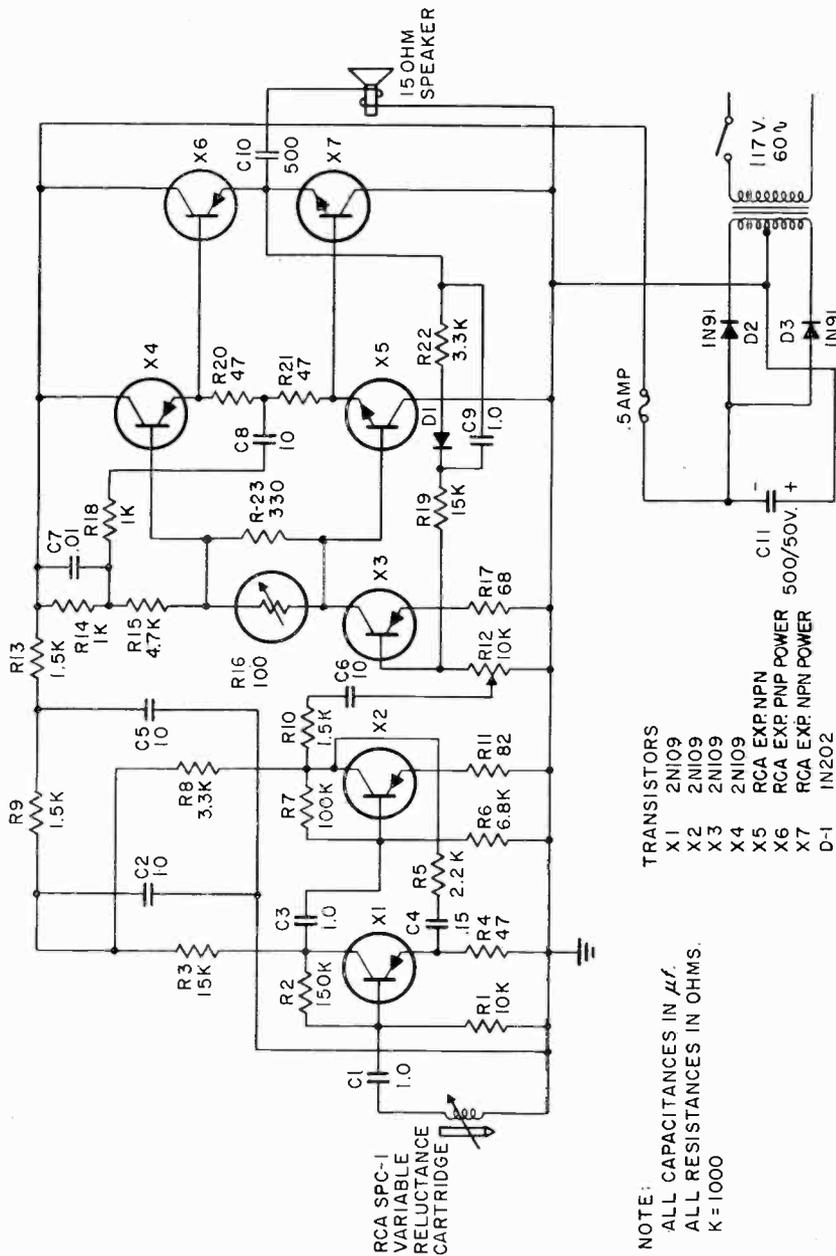


Fig. 12—Five-watt complementary symmetry power amplifier schematic.

The disadvantage of the common-collector connection is that the voltage gain is less than unity, making drive requirements difficult when a common d-c power supply is used for the last three stages of the amplifier.

For distortion characteristics as shown in Figure 13, the power transistors should have a large-signal current gain of at least 30 at one ampere of collector current. Peak collector currents at rated output will be on the order of one ampere with an output load of 15 ohms. Distortion versus loading characteristics are shown in Figure 14. The amplifier will function normally with power transistors having less than the desired gain characteristics if an increase in distortion at power levels above three or four watts can be tolerated. At levels below 4 watts, output with less than 1 per cent distortion can be obtained with lower gain power transistors.

A single-polarity power supply is used for both the n-p-n and p-n-p transistors. Proper collector-emitter bias polarities for each transistor is obtained by connecting the two units in series across the power supply without a conductive path to the load, and biasing the base circuit with approximately half the supply potential. This base voltage will appear at the power transistor emitters because of the common-collector connections.

Complementary-Symmetry Class-B Driver Stage

The driver stage consists of an experimental n-p-n* and a 2N109 p-n-p transistor connected in a common-collector configuration. The circuit is similar to that described for the power stage and it is used for the advantages outlined before. With power transistors in the output stage having a large-signal current gain of approximately 30, the drivers supply approximately 35 milliamperes peak current for full power output. This peak current requirement is within the capability of the drivers, but with lower gain power transistors the drivers will be unable to supply the peak current and will contribute to the over-all distortion. Also, the dissipation in the drivers may become excessive with low-gain power units in the output stage.

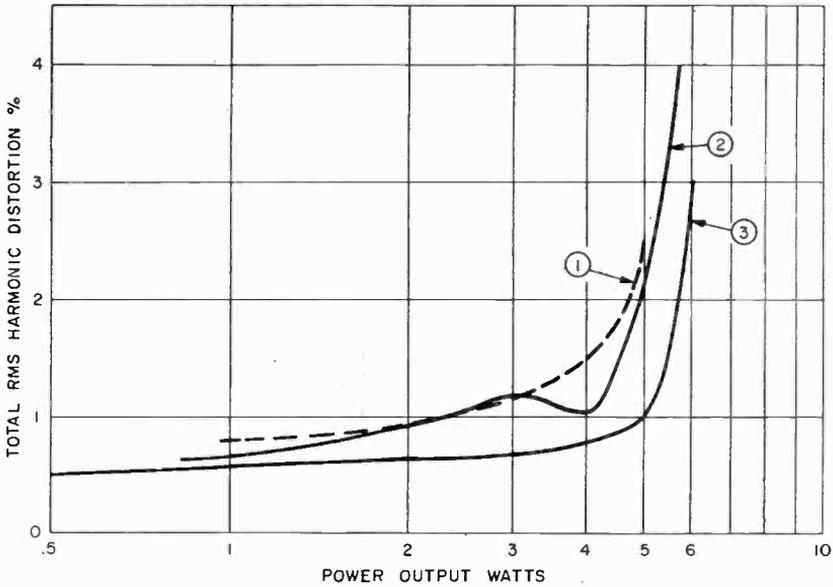
Quiescent d-c base-emitter bias voltage for the driver and output stages is employed to eliminate crossover distortion. The voltage across the thermistor R-16 in the collector circuit of the last class-A stage, X-3, is used for this purpose.

Final Class-A Stage

The final class-A stage, X-3, is direct-coupled to the class-B driver

* This transistor has characteristics approximately similar to the 2N109 except for polarity. Close match of α_{cb} with that of the 2N109 is not necessary, an unbalance of 5 to 1 being tolerable.

stages, X-4 and X-5. Because of the cascaded common-collector configuration of the driver and output stages, with voltage gain less than unity, the output voltage requirements of the class-A stage, X-3, are rather stringent. Since the X-3 stage controls the d-c operating point of the remaining stages, stabilization of this circuit is important, and is discussed further in the following section.



POWER OUTPUT VS DISTORTION
15 OHM LOAD

CURVE	FREQ.CPS
1	100
2	20,000
3	1,000

Fig. 13—Five-watt complementary symmetry power amplifier.

Temperature Stabilization and Heat Dissipation

The amplifier is stable up to 70°C with full signal. Temperature stability can generally be improved by

- A. Providing for heat removal from the collector junction.
- B. Providing circuit means for maintaining the d-c operating point constant with temperature change.
- C. Insuring that maximum transistor ratings are not exceeded.

Heat removal from the power transistor collector junction is accomplished by clamping the shell (on which the collector is mounted) to the chassis which acts as an effective heat sink. With the transistor shell electrically at collector potential, suitable means must be provided for insulating the collector from the chassis while still maintaining good thermal contact. For the p-n-p transistor, insulation is provided by .001-inch mylar film sandwiched between the transistor mounting

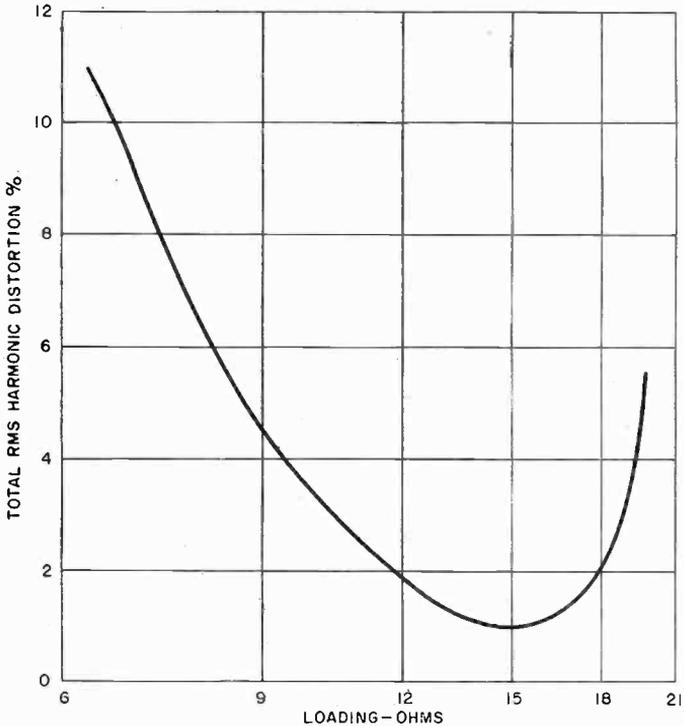


Fig. 14—Five-watt complementary symmetry power amplifier. Distortion versus loading, power output = five watts, frequency = 1,000 cycles per second.

base and chassis. Silicone oil is used between the transistor and heat sink for improved thermal contact. Fortunately, with the n-p-n power transistor collector at ground potential in the circuit, it can be mounted directly without insulation.

The power transistors, X-6 and X-7, tend to stabilize each other. With the transistors in series and capacity coupled to the load, the emitter currents for both units must be equal. The low resistance of R-18 and R-19 between X-6 and X-7 bases provides a low resistance d-c

path for I_{co} currents that tends to avoid amplification of this thermally dependent current in the collector circuit.

The class-B drivers, X-4 and X-5, have the same stability as the power stage with an additional measure of stability afforded by emitter feedback in R-20 and R-21. Signal is coupled from the drivers to the output such that no drive power is wasted in these resistors.

The zero-signal d-c operating point of both the class-B driver and output stages is controlled by the collector voltage and current of the X-3 stage. This stage requires extremely good temperature stabilization because a small change in operating point at X-3 is amplified by direct-coupling to the output stage and can cause thermal runaway.

Stabilization of X-3 is accomplished by the d-c feedback path consisting of R-19, R-22, and the "Zener diode," D-1. The Zener diode is operated in the reverse direction and has a voltage breakdown of approximately 12 volts. When connected as shown, the diode permits the stabilization feedback resistor to be 18,000 ohms while maintaining an effective bias resistance of 150,000 ohms. This provides a 10 to 1 increase in the d-c stability of X-4 over the use of a single 150,000-ohm resistor in a conventional collector-to-base voltage feedback arrangement.

The 100-ohm thermistor, R-16, is used to compensate for the slight increase in the collector current of X-3 above 65°C. The voltage across R-16 due to the collector current flowing in X-3 provides about 300 millivolts of forward bias for the driver-output stage to eliminate cross-over distortion in the output signal.

Stabilization of the remaining low-level circuits is accomplished by conventional collector-to-base voltage feedback and emitter feedback.

Signal Feedback Circuits

Fifteen decibels of negative feedback is applied from the output load to the base of X-3 by partially bypassing the d-c stabilization network with C-9. The purpose of the network R-18, C-8 is to reduce the 100 per cent negative feedback in the driver stage (X-4, X-5) by 4 decibels which, in turn, reduces the drive voltage requirement in the class-A stage, X-3. C-7 is used to bypass the effect of this network above 10 kilocycles for a greater margin of high-frequency stability.

Phonograph Preamplifier

Transistors X-1 and X-2 are used for an equalized preamplifier to operate from a variable reluctance phonograph pickup. The circuit consists of cascaded common-emitter stages with second-stage collector to first-stage emitter feedback to provide for low-frequency equaliza-

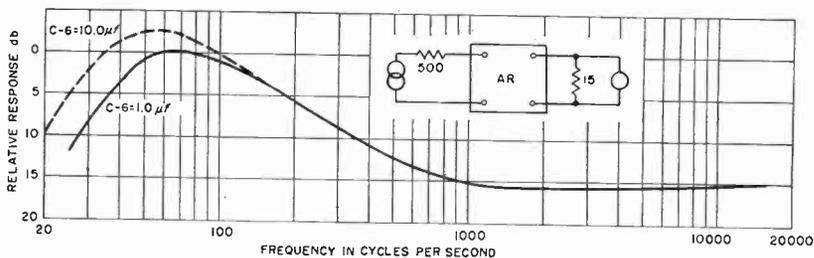


Fig. 15—Five-watt complementary symmetry power amplifier. Frequency response 500 source to 15-ohm load.

tion and a high input impedance at high frequencies for the pickup. Figure 15 shows the frequency response of the amplifier. A 10,000-ohm resistor, R-1, at the input of the amplifier shunts the pickup as shown in Figure 16 to provide the proper high frequency de-emphasis necessary for flat playback of the RIAA recording characteristic as shown in Figure 17. The voltage from the pickup at full groove modulation is about 10 millivolts at 1 kilocycle, so considerations for low-noise performance become important.* The input circuit can be modified for use as a flat amplifier by removing the bass-boost network.

Power Supply

The d-c supply voltage varies from 37 volts to 32 volts as the current varies from 10 to 250 milliamperes. To achieve reasonably good regulation, it is necessary to reduce the supply impedance to a few ohms. The large size of the power transformer resulted from regulation needs rather than power requirements. The filtering necessary to reduce ripple to low values was found to be 500 microfarads.

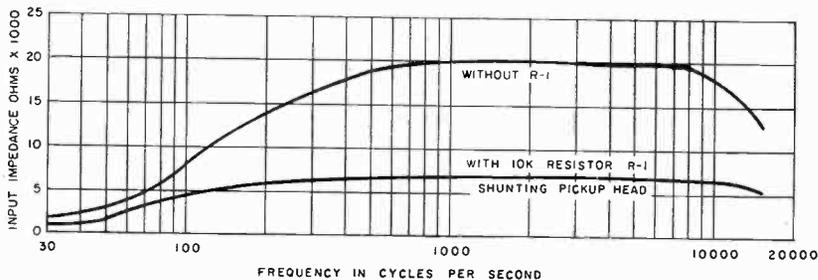


Fig. 16—Five-watt complementary symmetry power amplifier. Input impedance versus frequency.

* See discussion on 20-dbm preamplifier input stage.

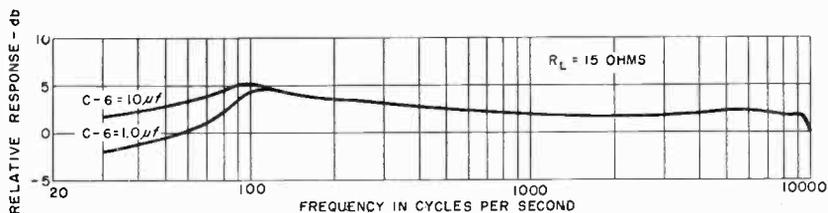


Fig. 17—Five-watt complementary symmetry power amplifier. Frequency response with SPC-1 cartridge reproducing RCA 12-5-49 test record.

The performance of the audio amplifiers discussed can be summarized as follows:

	20-dbm preamplifier	5-watt power amplifier	5-watt complementary- symmetry amplifier
Input			
source impedance	500 ohm microphone	—————	SPC-1 phonograph cartridge
input impedance	3,000 ohms	2,700 ohms	Figure 16
input voltage to obtain rated output at 1,000 cycles	0.3 mv.	45 mv.	15 mv.
Output			
load impedance	600 ohms unbalanced	15 ohms	15 ohms
output impedance at 1,000 cycles	25 ohms	3 ohms	1 ohm
Feedback	15 db	17 db	15 db
Frequency Response	Figure 4	±2 db, 20 to 20,000 cycles	Figure 15

ACKNOWLEDGMENT

Credit is due H. J. Woll and J. E. Lindsay for many technical suggestions and to A. H. Lytel for editing the text.

A 20-WATT TRANSISTOR AUDIO AMPLIFIER

By

H. C. LIN

RCA Laboratories,
Princeton, N. J.

Summary—A 20-watt audio amplifier using six p-n-p germanium junction transistors is described. Three are experimental power transistors; the others are type 2N109 transistors. The class-B output stage is connected in a single-ended push-pull circuit driving a conventional 16-ohm loudspeaker directly, without using an output transformer. The amplifier is a-c operated. It has high input impedance, suitable for crystal pickup or microphone. The operation is satisfactory for ambient temperatures from -5° to 55°C (approximately 25° to 130°F). A discussion of the power capabilities of transistors and some aspects of power transistor class-B operation precedes the description of the amplifier.

POWER CAPABILITIES OF TRANSISTORS

THE maximum power that can be dissipated in a transistor before "run away" occurs depends on a number of factors of which the most important is the ability to remove heat. This ability, when heat is removed by conduction, is measured by the thermal resistance, i.e., the internal temperature rise per unit power of dissipation. The lower the thermal resistance, the greater the power handling capability. Factors which determine the maximum power dissipation capability are collector voltage, the current amplification factor, and the reverse saturation current. An approximate expression for the maximum permissible dissipation of a germanium transistor connected in a circuit having substantially no external emitter and collector circuit resistances (such as the one to be described) can be shown to be¹

$$P_{\max} = \frac{23}{\theta} \log_{10} \left(\frac{10}{\theta V \alpha_{cB} I_S} \right) \text{ watts,} \quad (1)$$

where θ = thermal resistance in degrees centigrade per watt,

V = collector voltage in volts,

α_{cB} = a-c collector-to-base current amplification factor,

I_S = base saturation current in amperes at chassis temperature.

¹ K. E. Loofbourrow, "Class B Operation of Audio-Frequency Junction Transistors," *Electronic Design*, Vol. 3, p. 28, July, 1955.

Note that the I_S used in this formula should correspond to the temperature of the chassis on which the transistor is mounted. P_{\max} decreases at approximately the rate of $1/\theta$ watts per degree centigrade increase in chassis temperature.

Application of Equation (1) to an experimental power transistor of the type used in this amplifier serves as a pertinent example. For this unit, $\theta = 5^\circ\text{C}$ per watt, $\alpha_{cb} = 15$, $I_S = 0.45$ milliamperes at 60°C (assumed highest chassis temperature). If the collector voltage is 25 volts, from Equation (1)

$$P_{\max} = 5 \text{ watts per transistor.}$$

SOME ASPECTS OF POWER-TRANSISTOR CLASS-B OPERATION

The class-B operation of junction transistors has been discussed in detail by Loofbourrow.¹ It was pointed out that the theoretical efficiency of 78 per cent for a sine-wave output can be approached with transistor class-B amplifiers. However, for fixed power-supply voltage, the maximum dissipation occurs when the amplifier is delivering a square-wave output approximately equal to four-tenths of maximum sine-wave output power.² Thus the ratio of maximum sine-wave push-pull output to the maximum possible dissipation in each transistor is 4 to 1. If the limit of dissipation is 5 watts per transistor, the maximum output power is 20 watts. For ordinary speech or music signals, the average dissipation per transistor is far below one quarter of the maximum output power.

To operate two transistors in a push-pull class-B amplifier, either "double-ended" (Figure 1a) or a "single-ended" (Figure 1b) connection³ may be used. The single-ended connection, which requires no output transformer, eliminates both distortion and loss which would arise from the use of a transformer. Transformerless output operation also eliminates possible ill effects of transformer leakage inductance. This kind of connection is particularly suitable for power transistors working directly into a conventional loudspeaker load.

For a single-ended push-pull class-B stage, the required d-c supply voltage, V , and current, I , for a maximum sine-wave output, P , into a load resistance, R can be derived as follows: Assuming an undistorted sine-wave output voltage which swings from zero to twice the d-c

²L. J. Giacoletto, "Power Transistors for Audio Output Circuits," *Electronics*, Vol. 27, p. 144, January, 1954.

³E. T. Burton, U. S. Patent 1944216, file date May 29, 1931; also A. Peterson and D. B. Sinclair, "A Single-Ended Push-Pull Audio Amplifier," *Proc. I.R.E.*, Vol. 40, p. 7, January, 1952.

collector voltage per transistor, the load resistance should equal the ratio of V to I_m , the peak current. Since I_m equals π times the d-c collector current,

$$R = V/(\pi I). \quad (2)$$

The maximum sine-wave output is equal to $\pi/4$ times the d-c input power to the two output transistors, hence

$$P = (\pi/4) \cdot 2VI. \quad (3)$$

Combining Equations (2) and (3) gives

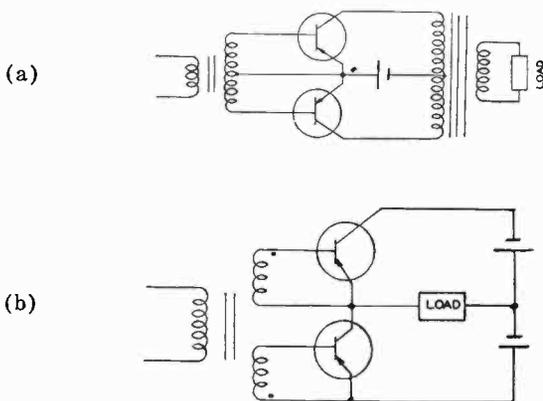


Fig. 1—Push-pull transistor amplifier.

$$V = \sqrt{2PR}, \quad (4)$$

$$I = (1/\pi) \sqrt{2P/R}. \quad (5)$$

A single-ended circuit connected as shown in Figure 1b would not operate satisfactorily at room temperature because of the absence of "threshold" bias between the base and emitter. Without the bias, the transistors operate class C instead of class B, i.e., each transistor conducts less than half the time. Such operation would give rise to "cross-over" distortion which would be particularly severe at low output levels. The required forward biasing voltage decreases with increase in temperature at a rate of approximately 2.5 millivolts per degree centigrade for germanium transistors.⁴ A suitable biasing network should there-

⁴H. C. Lin and A. A. Barco, "Temperature Effects in Circuits Using Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 369-402.

fore include a temperature-sensitive element, for example, a thermistor. The temperature-sensitive element should be thermally coupled as tightly as possible to the transistor which is to be compensated, in order that the temperature-sensitive element respond to the temperature of the transistor.

CIRCUIT DESCRIPTION

A complete schematic diagram of the 20-watt audio amplifier is shown in Figure 2. Transistors V_1 through V_3 are type 2N109 p-n-p transistors. They form a preamplifier which permits operation from a capacitive source, such as a crystal microphone or crystal phonograph pickup. Transistors V_4 through V_6 are the experimental power units.⁵ Transistor V_4 operates in a class-A driver stage, while V_5 and V_6 are employed in the class-B transformerless output stage.

Preamplifier

The input stage is a modified version of a grounded-collector amplifier, and has a high input impedance. The resistor R_2 , in series with the emitter, stabilizes the d-c operating current against temperature variations. Resistors R_3 , R_4 , and R_5 form a voltage divider for deriving proper potential for each electrode. The base resistor, R_1 , is returned to a point at emitter potential (a-c) instead of ground potential. This permits a smaller value of base return resistance to be used without causing input circuit loading. A small value of base resistance improves the operating current stability.

The second and third stages are in grounded-emitter configurations. External emitter resistances are used to stabilize the transistors against temperature variations and provide interchangeability of transistors. The base-to-emitter biases are derived from voltage dividers. The external base return resistance of every stage is made less than the product of the current amplification factor and the emitter resistance to achieve good temperature stability.⁴ Emitter resistances, with the exception of R_{18} , are adequately bypassed to avoid signal degeneration. Resistor R_{18} is not bypassed because of hum considerations, as will be explained later.

The volume control follows the first stage. Both input and output of the volume control are coupled through capacitors in order to avoid flow of direct current, which would make the control noisy. A parallel R - C network is incorporated in the input to the volume control to compensate for the high-frequency fall-off in current amplification

⁵B. N. Slade, "Recent Advances in Power Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 153-171.

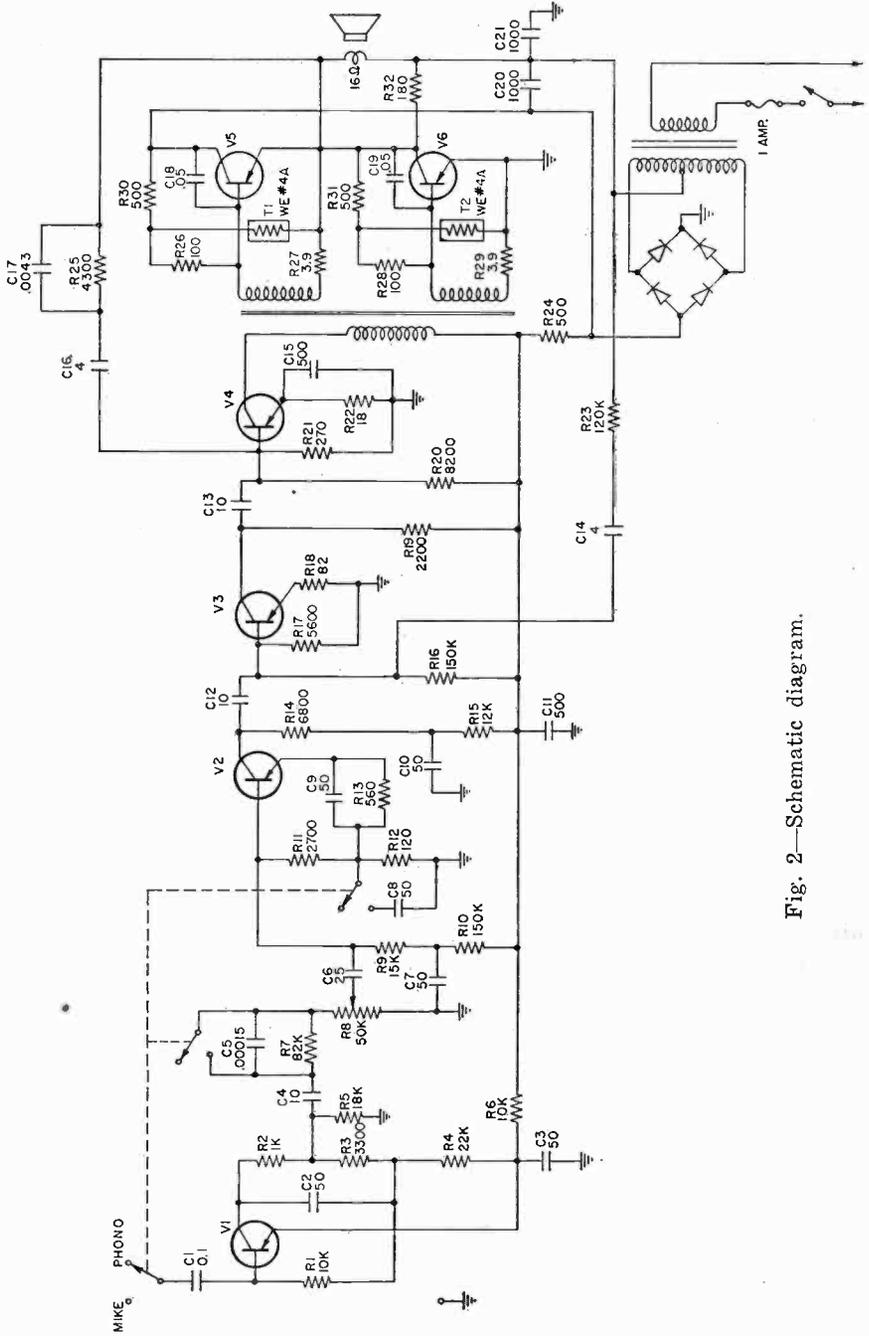


Fig. 2—Schematic diagram.

of the transistors. The value of C_5 may be changed somewhat to compensate for variations in frequency response due to the pickup and the recording. This compensating network is short-circuited in the microphone position to provide a needed increase in gain.

The volume-control setting affects the input impedance of the amplifier; the lower the setting, the higher the reflected input impedance. When a capacitive input device such as crystal pickup is used, a higher impedance input gives greater low-frequency response. Thus the volume control can be adjusted to give the desired tone compensation.

Driver

The driver stage uses a power transistor operating class A in the grounded-emitter connection. The resistor R_{22} , in series with the emitter, stabilizes the d-c operating current as temperature is varied. It is bypassed by capacitor C_{15} at audio frequencies. The base-to-emitter bias is derived from a voltage divider R_{20} and R_{21} . The output of the driver is transformer-coupled to the output stage.

Since the metal enclosure of the power transistor is electrically connected with the collector, it must be insulated from the main chassis which is at ground potential. At the same time, the metal enclosure should be in good thermal contact with the main chassis in order to remove the heat generated in the transistor. For these purposes an anodized aluminum plate, which insulates electrically but conducts thermally, is sandwiched between the power transistor and the main chassis.

Output Stage

The output stage is connected in single-ended push-pull transformerless output configuration. As explained previously, the optimum base-to-emitter bias for class-B operation decreases by approximately 2.5 millivolts with every degree centigrade of temperature rise. This bias is derived from a series-parallel combination of thermistors (T_1, T_2) and resistors (R_{26} to R_{31}).

As in the driver stage, the output transistors are mounted on anodized aluminum plates to provide low thermal resistance between the transistors and the main chassis. The thermistors are also mounted on these anodized plates to insure tight thermal coupling between the transistors and the thermistors.

The input is transformer-coupled to provide the required out-of-phase input voltage for the two transistors. The secondary windings are tightly coupled together to avoid transient voltages when current shifts from one output transistor to the other. This is accomplished

by bifilar winding. Because of the low input impedance of the power transistors (in the order of 30 ohms during conduction), the secondary is stepped down from the primary by a ratio of 5-to-1, thus providing a current gain.

Negative feedback is employed to reduce distortion. The distortion is mainly caused by fall-off in current amplification factor at high current levels. In this amplifier negative feedback is applied over two stages. This provides a greater amount of negative feedback than can be obtained within a single stage. Negative feedback also reduces the output impedance of the output stage, thereby providing greater damping for the loudspeaker.

The collector-to-base current amplification factor of the experimental power transistors falls off at higher audio frequencies. There is a certain amount of phase-shift associated with this fall-off in the amplification factor, as in a low-pass filter. This phase-shift, together with that in the driving transformer, limits the maximum amount of negative feedback that can be applied in the system without causing instability. To improve the stability, the frequency response of the final stage is extended by means of local negative feedback. This is accomplished by connecting a capacitor between the collector and the base of each output transistor (C_{18} and C_{19}). Also, a step-frequency response is introduced in the high-frequency end of the feedback loop. The step reduces the phase shift, thereby stabilizing the amplifier. The frequency step is obtained by connecting a capacitor, C_{17} , in parallel with the feedback resistor, R_{25} . The total amount of feedback is 11 decibels at an output of 20 watts and a frequency of 400 cycles. The stability margin is 9 decibels. The feedback is somewhat greater at lower levels because of higher current gains.

Power Supply and Filtering

The power supply consists of selenium rectifiers connected in a bridge circuit. Two sections of 350-milliampere selenium rectifiers are used in each arm of the bridge. This arrangement provides a center-tap for the rectified d-c output voltage. The center-tap assures equality of the voltages applied to the output transistors regardless of unbalance in characteristics of the transistors or difference in values of the filtering capacitors, C_{20} and C_{21} .

The operating collector voltages and currents of the transistors for maximum sine-wave output, are listed in Table I.

Transistor collector current is normally insensitive to collector voltage variations at collector voltages greater than a few tenths of a volt; the major cause of hum is ripple current introduced at the base

of the transistor. In a resistance-coupled amplifier, ripple may be fed to the base through the base-biasing resistance, and the resistance in series with the collector of the previous stage.

The necessary amount of filtering increases progressively toward the input end of the amplifier. As in vacuum-tube amplifiers, the filtering system can be *graded*.

Another source of hum occurs in the feedback path. The ripple voltage appearing at the center-tap of the power supply causes a ripple current to flow through the loudspeaker and the feedback network to the base of the driver. To "buck out" this source of hum, a ripple current in opposite phase is fed to the base of the preceding transistor through R_{23} and C_{14} . The required values of R_{23} and C_{14} depend upon the current amplification factor of transistor V_3 . To allow for interchangeability of transistors, degeneration is applied by means of an unbypassed emitter resistance, R_{18} . When the current amplification factor is high, the input impedance is also high. More input current is bypassed to the base return resistance, R_{17} , so that the amplified

Table I—D-C Operating Voltages and Currents

	V_1	V_2	V_3	V_4	V_5 & V_6
$-V_e$ in volts	6	10	7	27	26
$-I_e$ in milliamperes	1	1	9	35	560

output current is substantially the same as that when using a transistor with low amplification factor.

Filtering for the output stage is furnished by the capacitors, C_{20} and C_{21} . The filtering is required to reduce the hum modulation of the signal at full voltage swings. These capacitors also serve to provide a low-impedance return path for output currents. The voltage supplied to all earlier stages is filtered through R_{24} and C_{11} . Additional filtering to the first three stages is provided by R_{15} and C_{10} , R_{10} and C_7 , R_6 and C_3 .

OPERATIONAL DETAILS

In operating the amplifier, care must be taken that the output terminals are not short-circuited. If the output terminals are short-circuited, a large input signal may cause excessive transistor dissipation in the output transistors.

It is also important that the output terminals are not open circuited. Open output may cause destructively high voltage to exist across one output transistor due to unbalance in transistor characteristics, particularly if there is signal.

The output transistors should have as nearly equal current amplification factors as possible, in order that unbalanced direct current flowing in the voice coil be kept to a minimum. The permissible unbalanced current depends on the loudspeaker used and the maximum distortion which can be tolerated.

The output transistors as well as the driver transistor should satisfy the following additional requirements:

Collector breakdown voltage > 60 volts,

For 1 ampere peak collector current, peak base current should be between 25 and 65 milliamperes for frequencies up to 10 kilocycles,

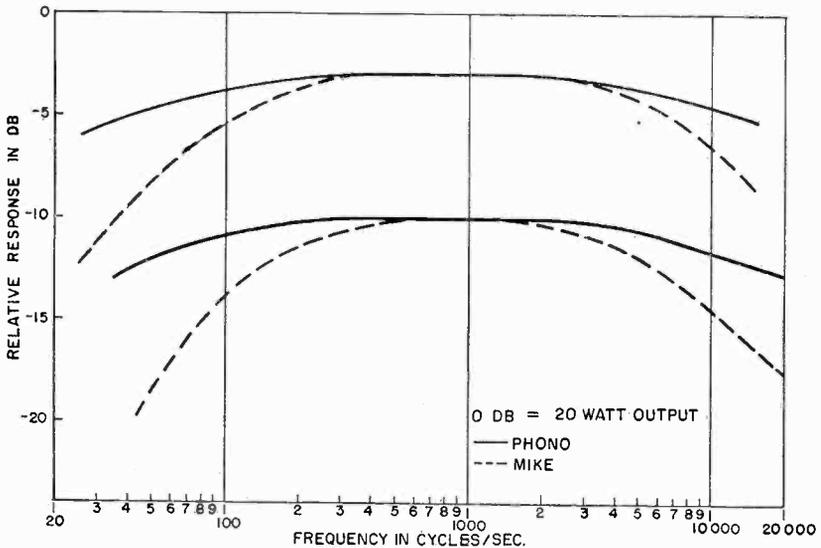


Fig. 3—Frequency response.

Peak base-to-emitter voltage for 1 ampere peak collector current*
 < 0.6 volt,

Thermal resistance $< 5^{\circ}\text{C}/\text{watt}$,

I_{c0} at V_c of -1 volt at $25^{\circ}\text{C} < 30$ microamperes.

PERFORMANCE

The over-all frequency response is shown in Figure 3. The input voltage between the base of the input transistor and ground was kept constant during the test. For phonograph input, the frequency re-

* For output pair only.

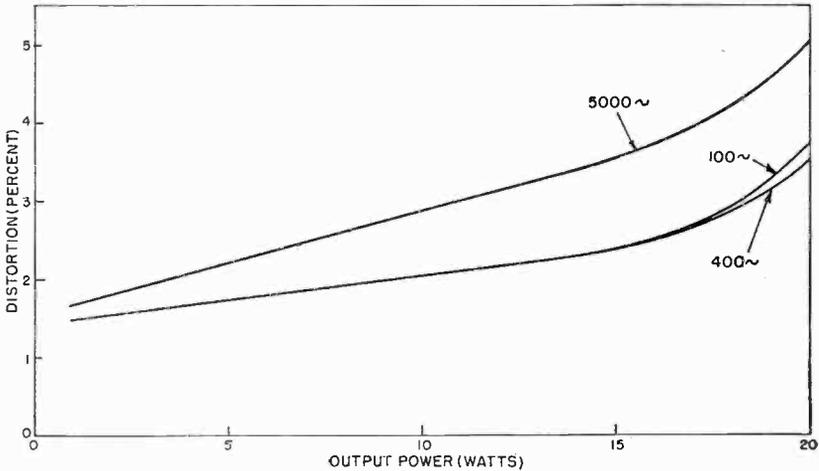


Fig. 4—Distortion versus power output.

response is within $1\frac{1}{2}$ decibels from 40 to 15,000 cycles. The high-frequency response for microphone input falls off because no compensation is used, as explained previously. The low-frequency response was purposely cut down to keep the hum within reasonable level (54 decibels below full output at maximum volume control setting). A larger bypassing capacitor than the value specified for C_8 can be used to raise the low-frequency response.

Figure 4 shows the distortion at different output levels. The dis-

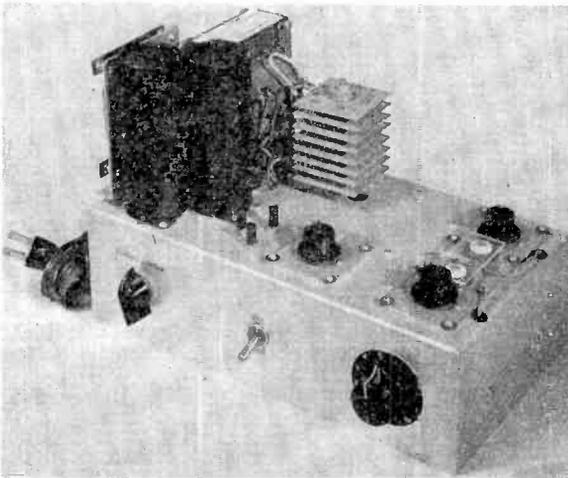


Fig. 5—Amplifier chassis.

ortion at 20 watts is less than 5 per cent, and at 10 watts, less than 3 per cent. The net distortion at 5,000 cycles is somewhat higher than that at lower frequencies for corresponding power levels, because of reduction in gain around the feedback loop.

The amplifier has been tested at ambient temperatures ranging from -5° up to $+55^{\circ}\text{C}$ with no appreciable deterioration in performance.

Table II summarizes the performance of this amplifier.[†]

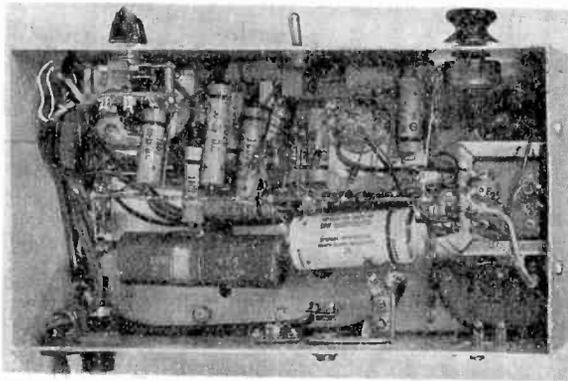


Fig. 6—Bottom view of chassis.

Table II—Amplifier Performance

Input impedance	1 megohm (phono) $\frac{1}{4}$ megohm (mike)
Output impedance	$1\frac{1}{2}$ ohms
Power gain	77 db (phono) 104 db (mike)
Harmonic distortion	< 5% at 20 watts < 3% at 10 watts
Frequency range	± 1.5 db from 40-15,000 cycles (phono)
Hum level	57 db below 20 watts (phono) 54 db below 20 watts (mike)
Maximum ambient temperature	55°C

Figure 5 shows the actual construction of the amplifier. Note the anodized plates and the thermistors. Figure 6 shows the bottom view.

[†] Measurements were made according to RMA Standard SE-101A.

MODULATED TRANSISTOR OSCILLATORS AND THEIR APPLICATIONS

By

H. C. LIN

RCA Laboratories,
Princeton, N. J.

Summary—The low power drain and small physical size of transistors make them particularly suitable for miniature transmitters intended for short-range applications. This paper deals with the working principles and modulation characteristics of both junction and point-contact transistor modulated oscillators. Both amplitude modulation and frequency modulation are considered.

Amplitude modulation of transistor oscillators can be achieved by means of collector modulation, base modulation, or emitter modulation. Factors governing the linearity of different kinds of modulation are discussed. Modulating power is high for collector modulation and low for base modulation. The input impedance to modulation at the base can be made high, of the order of 10^5 ohms.

Frequency variation due to applied amplitude modulation can be minimized by modulating two electrodes of the transistor simultaneously. A process for frequency modulating an oscillator with minimum amplitude variation is also described.

AMPLITUDE MODULATION

Junction Transistor Oscillator

FIGURE 1 shows a typical junction transistor oscillator of the reverse-feedback type. It can be modulated either at the base, emitter, or the collector. The mechanism of modulation can be understood from the following analysis.

Collector Modulation: When oscillation is sustained, the collector voltage and current can be represented approximately by Figures 2(a) and (b). At the instant of maximum collector current pulse the voltage at the collector is the difference between the collector supply voltage, V_{CC} , and the voltage drop, $I_c R_L$, across the load due to the fundamental frequency component of current flowing through the r-f load, R_L . If this minimum collector voltage, v_{cem} , lies in Region A of the collector output characteristic as shown at point "a" in Figure 3, it is approximately equal to zero. As the r-f voltage varies linearly with collector voltage, the r-f voltage also varies linearly with modulation.

If the minimum difference voltage does not lie in Region A but lies in Region B of the collector characteristic, say point "b", any

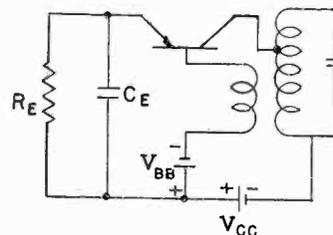


Fig. 1—Typical p-n-p junction transistor oscillator circuit.

increase of V_{CC} does not increase the collector current but only serves to increase v_{cem} . For this duration of the modulation cycle, the r-f voltage does not change. The result is that the modulation is no longer linear. Therefore, in order to have linear collector modulation, it is necessary to keep V_{CC} low or $I_c R_L$ high.

Figure 4 shows a set of collector modulation characteristics. Note that when high R_L is used, modulation is linear, whereas when low R_L is used output flattens at high amplitudes.

Base Modulation: In base modulation, the d-c base voltage, V_{BB} , is changed to cause a change in output current I_c . If the base-to-emitter impedance, r_{be} , the flow angle, 2θ , and the collector-to-base current amplification, A_{cb} , can be maintained constant as base voltage is varied, the modulation is linear. This can be seen from Figures 2(c) and (d). If the base-to-emitter impedance is constant, the base current pulse will be of the same shape as the voltage pulse lying below the line of cutoff. By Fourier analysis it can be shown that the base r-f current

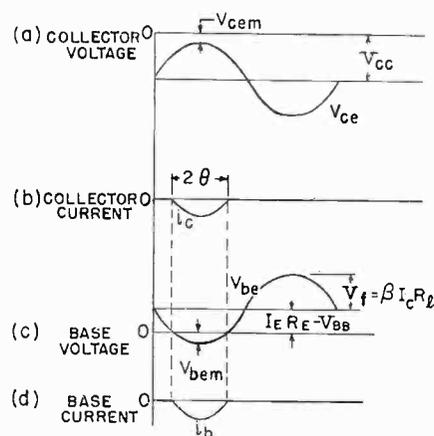


Fig. 2—Current and voltage relations at the collector and at the base of typical junction transistor oscillator.

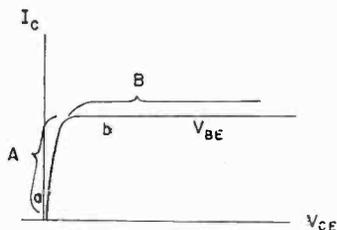


Fig. 3—Collector output characteristic of a junction transistor.

is related to this pulse as a function of the flow angle.¹ The base r-f current is amplified, producing an r-f voltage across the load R_L which is fed back to the base. Thus, if the current amplification, the load, and the r-f base-to-emitter impedance remain unchanged, the flow angle will not change for whatever values of currents, or the bias $I_B R_B - V_{BB}$ which the feedback voltage $\beta I_c R_L$ must overcome in order to cause current flow. However, if the bias is changed by changing V_{BB} , the currents I_B and I_c must increase proportionately in order to maintain the same flow angle. Thus, it is established that the output current varies linearly with the applied base voltage provided the collector-to-base current amplification and base-to-emitter impedance are constant.

In order to obtain satisfactory linear modulation, the transistor should be operated in such a way as to keep the base-to-emitter im-

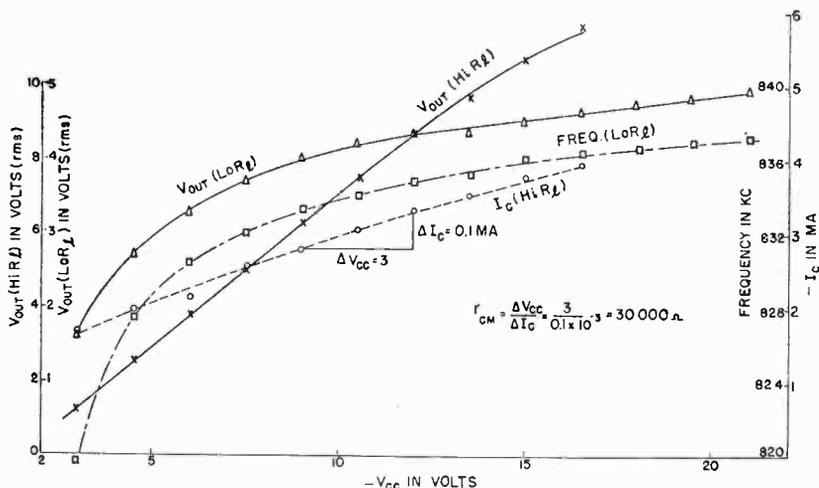


Fig. 4—Collector-modulation characteristics of a junction transistor oscillator.

¹ W. L. Everitt, *Communication Engineer*, McGraw-Hill Book Co., New York, N. Y., 1938, p. 565.

pedance, the current amplification, and the r-f load as constant as possible. As R_i is somewhat dependent on r_{be} , to which it couples, only r_{be} and A_{cb} need be considered.

A_{cb} is relatively constant when the collector current is sufficiently high or v_{cem} lies in Region B of the collector output characteristic. If $I_c R_i$ is increased or the collector supply voltage is decreased to a point where v_{cem} lies in Region A, A_{cb} can no longer be considered constant. The effect is shown in the set of base modulation characteristics in

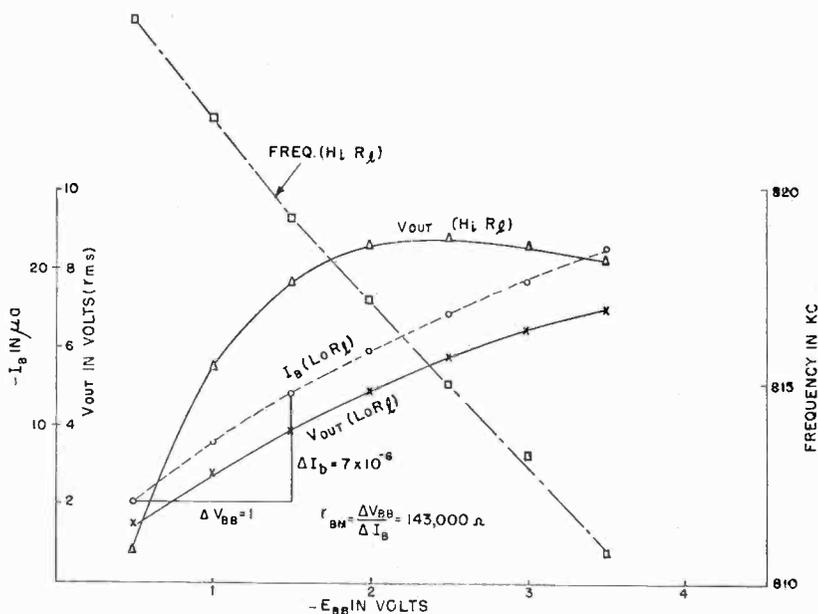


Fig. 5—Base-modulation characteristics of a junction transistor oscillator.

Figure 5. Note that when high R_i is used, the characteristics flattens in contrast to the linear characteristic obtained with low R_i .

The base-to-emitter impedance for small signal consists of the base resistance r_b and an equivalent impedance equal to the product of emitter resistance r_e and some function of current amplification.² Usually r_b is relatively constant, but r_e decreases with increase in emitter current.³ If the equivalent product impedance is made smaller than r_b , then r_{be} will not change very much. A large emitter current may be

² R. L. Wallace, Jr. and W. J. Pietenpol, "Some Circuit Properties and Applications of n-p-n Transistors," *Proc. I.R.E.*, Vol. 39, p. 753, July, 1951.

³ W. Shockley, M. Sparks, and G. K. Teal, "p-n Junction Transistors," *Phys. Rev.*, Vol. 83, p. 151, July 1, 1951.

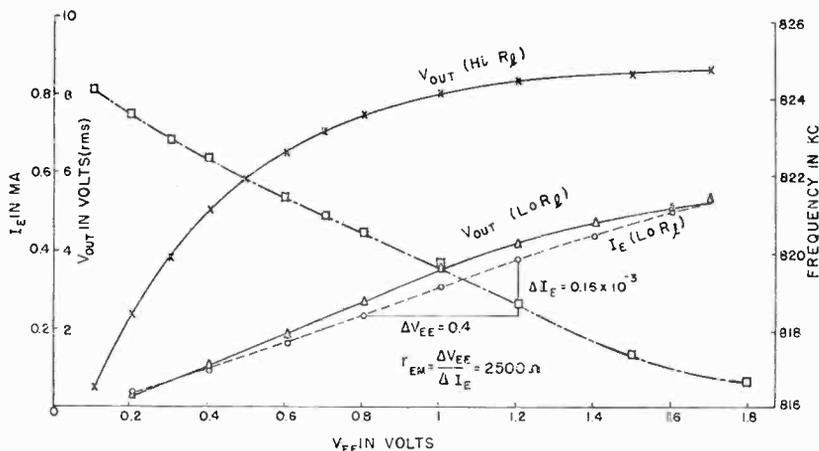


Fig. 6—Emitter-modulation characteristics of a junction transistor oscillator.

used to make r_e , hence the product, small. An impedance in series with the base will make the effective r_b high. An impedance in series with r_e will make the effective r_e less changeable with emitter current. One or more of these means may be incorporated in the design of such oscillators for achieving modulation linearity.

Emitter Modulation: The mechanism of emitter modulation is similar to that of base modulation. The factors governing the modulation linearity in the case of base modulation apply equally well to emitter modulation. A set of emitter modulation characteristics is shown in Figure 6.

Point-Contact Transistor Oscillators

Point-contact transistor oscillators are ordinarily designed to operate by virtue of their negative resistance characteristics. One type of point-contact transistor oscillator is shown in Figure 7.

The idealized emitter characteristic $I_E - V_{EE}$ of the transistor,⁴ as

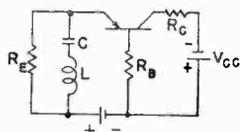


Fig. 7—Typical point-contact transistor oscillator circuit.

⁴ A. W. Lo, "Transistor Trigger Circuits," *Proc. I.R.E.*, Vol. 40, p. 1531, November, 1952.

shown in Figure 8, indicates that as the collector supply voltage, V_{CC} , varies, the negative-slope portion of the characteristic also varies.

If the collector-to-emitter current amplification factor α_{ce} and collector resistance r_c can be considered constant, the distance between P_1 and P_2 increases linearly with V_{CC} . As the negative-slope portion enlarges, the amplitude of oscillation also increases. Linear modulation is thus possible under the condition that the operating point V_E changes linearly with V_{CC} . This can be approached because I_E is proportional to V_{CC} and the bias V_E can be made proportional to I_E .

If the modulation is applied to the emitter of this oscillation, the operating point on the negative slope changes without appreciable

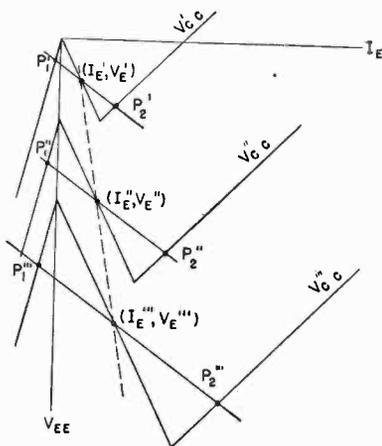


Fig. 8—Emitter input characteristics of a point-contact transistor oscillator.

change in the emitter current excursion. However, the ratio of fundamental frequency component current to that of harmonics changes and this change modulates the oscillator.

The effect of base modulation is essentially the same as emitter modulation, because a base voltage change causes changes in emitter current flowing in the resistance R_B . Figure 9 shows the modulation characteristics of this kind of oscillator.

INPUT IMPEDANCE AND MODULATING POWER

Junction Transistor Oscillator

Collector Modulation: In the case of collector modulation of junction transistor oscillators, the d-c collector current bears a certain

relationship to the r-f current depending on the flow angle and the current amplification. As was pointed out, the minimum collector voltage occurs below the knee of the collector output characteristic where the current amplification is not constant, the flow angle varies with modulation. Fortunately, the ratio of r-f current to d-c current changes very slowly with change in flow angle.¹ Hence the modulation input impedance does not have a serious change with modulation.

It can be shown analytically that the d-c current and the r-f current are of the same order of magnitude. Since the modulation input im-

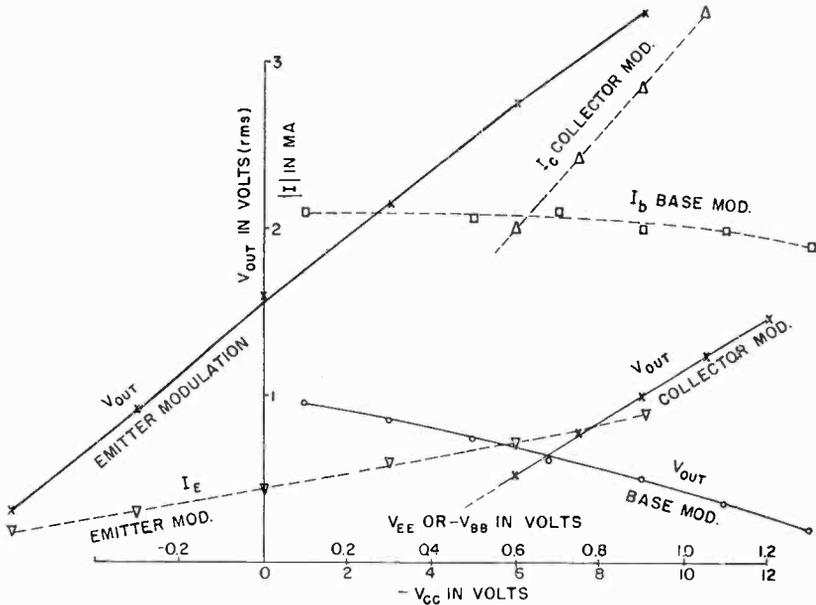


Fig. 9—Modulation characteristics of a point-contact transistor oscillator.

pedance is equal to the ratio of the change in d-c collector voltage to the change in d-c collector current and the d-c collector voltage is approximately equal to the peak r-f current times the r-f load, the modulation input impedance is also of the same order of magnitude as the r-f load resistance.

Base Modulation: In the oscillator shown in Figure 1 the external emitter resistance R_E is by-passed by C_E at radio frequency but not at audio frequency. As in the case of a base input audio amplifier, a high resistance in the emitter circuit is highly degenerative to audio-

frequency signals and makes the input impedance at the base high. The value of the input impedance is approximately equal to R_E times the low-frequency collector-to-base current amplification factor less an amount depending on the r-f feedback of the oscillator. By making R_E high and the feedback small, a high input impedance can be obtained.

Emitter Modulation: The change in emitter current is equal to the change in base current times the emitter-to-base current-amplification factor. Thus the modulation input impedance of an emitter-modulated oscillator is equal to the corresponding base-modulation input impedance divided by the emitter-to-base current-amplification factor.

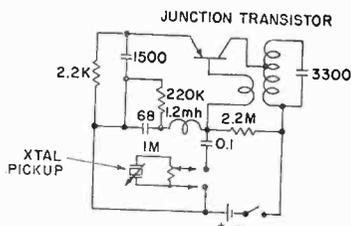


Fig. 10—"Wireless Phonograph Jack."

Measured modulation input impedance obtained by taking the slopes of V-I curves from Figures 4, 5, and 6 are

$$r_{CM} = 30,000 \text{ ohms,}$$

$$r_{BM} = 145,000 \text{ ohms,}$$

$$r_{EM} = 2,500 \text{ ohms.}$$

Note that the high base-modulation input impedance makes it possible for the oscillator to be modulated directly from a high-impedance source. Figure 10 shows the application of this property to a "Wireless Phonograph Jack." In this unit, the r-f field set up by the oscillator is picked up by a nearby radio. The radio serves as the output device of the phonograph. The basic oscillator circuit is the same as Figure 1. The bias for the base is obtained from a voltage divider consisting of the 2.2 and the .22 megohm resistors. The 68-micromicrofarad capacitor is in series resonance with 1.2-millihenry inductance to furnish the r-f bypass for the base. As the crystal pickup is approximately equivalent to a generator in series with a coupling capacitor of the order of 2,000 micromicrofarads, the frequency response depends somewhat on the setting of the volume control due to the loading of the base input impedance.

Collector-Modulated Point-Contact Transistor Oscillators

The d-c collector current corresponding to P_1 in Figure 8 is⁴

$$I_C = \frac{-V_{CC}}{R_b + R_c + r_c}.$$

At P_2

$$I_C = \frac{\alpha_{ce} V_{CC}}{\alpha_{ce} (R_b + R_c) - R_b}.$$

At the d-c operating point

$$I_C = -\frac{1}{2} \left[\frac{1}{R_b + R_c + r_c} + \frac{\alpha_{ce}}{\alpha_{ce} (R_b + R_c) - R_b} \right] V_{CC}.$$

The collector modulation resistance is, therefore

$$r_{CM} = \frac{\Delta V_{CC}}{\Delta I_C} = 2 \left/ \left[\frac{1}{R_b + R_c + r_c} + \frac{\alpha_{ce}}{\alpha_{ce} (R_b + R_c) - R_b} \right] \right.$$

Note that a higher value of R_b or α_{ce} can increase the modulating resistance.

Example: Given a point-contact transistor connected as shown in Figure 7, for which R_c (external) = 560 ohms, R_b (external) = 910 ohms, r_b (internal) = 300 ohms, r_c (internal) = 20,000 ohms, $\alpha_{ce} = 5.6$. Then effective $R_b = 910 + 300 = 1210$ ohms. From the last equation, $r_{CM} = 3120$ ohms. The measured result obtained by taking the slope of $I_C - V_C$ characteristic of Figure 8 is 3470 ohms.

Once the input impedance to modulation is known, the modulation power can be derived. Modulating power for 100 per cent modulation is equal to $(\Delta V_M)^2 / (2r_M)$ or $(1/2) (\Delta I_M)^2 r_M$, where ΔV_M is the crest modulating voltage, ΔI_M is the crest modulating current and r_M is the input resistance to modulation.

To compare base modulation with collector modulation on the same basis, assume the same R_i and same maximum r-f current in the collector at 100 per cent modulation. Since the audio-frequency collector current is of the same order of magnitude as the r-f collector current and the corresponding audio-frequency base current is reduced by a factor equal to the current amplification factor, the higher the current amplification factor, the smaller the base current. In general, the modulating power for base modulation is much less than that for collector modulation. A typical set of values of modulating power with collector supply of 22.5 volts are

P (Collector modulation) = 8 milliwatts,

P (Base modulation) = 15 microwatts,

P (Emitter modulation) = 1 milliwatt.

In the case of a collector-modulated point-contact transistor oscillator of the first type, the modulating power depends on the maximum swing in collector voltage ΔV_{CC} and is equal to

$$\frac{1}{2} \frac{(\Delta V_{CC})^2}{r_{CM}} = \left(\frac{\Delta V_{CC}}{4} \right)^2 \left[\frac{1}{R_b + R_c + r_c} + \frac{\alpha_{ce}}{\alpha_{ce}(R_b + R_c) - R_b} \right].$$

Assuming $\Delta V_{CC} = 24$ volts, the modulating power is 21 milliwatts.

FREQUENCY STABILIZATION OF AMPLITUDE-MODULATED TRANSISTOR OSCILLATOR

When modulation is applied to any electrode of a transistor oscillator, frequency modulation as well as amplitude modulation is produced. The reason for this is that a voltage change in any electrode causes a change in the effective internal capacitance.

It was found, however, that in a circuit such as that shown in Figure 1, the effect on frequency due to collector modulation is opposite to that due to base modulation. A set of constant-frequency characteristics is shown in Figure 11. Note that the base modulating voltage necessary for maintaining constant frequency is a linear function of V_{CC} . Note also that the modulation characteristic is linear over a wide range of V_{CC} .

Although the analytic relationship between V_{CC} and V_{BB} for maintaining constant frequency is not a linear one, actual experiment shows a wide range of linearity. Thus, by modulating the base and the collector simultaneously, it is possible to eliminate frequency modulation to a large degree.

This method for eliminating frequency modulation was employed in a transistor phonograph oscillator. The circuit diagram is shown in Figure 12. In this circuit, modulation is applied directly to the collector. R_2 and R_3 serve as a voltage divider to feed a portion of the modulation to the base for maintaining constant frequency.

FREQUENCY MODULATION

Frequency modulation can be produced by modulating any one of the electrodes of a transistor oscillator. However, frequency modulation is often accompanied by amplitude modulation. In order to achieve

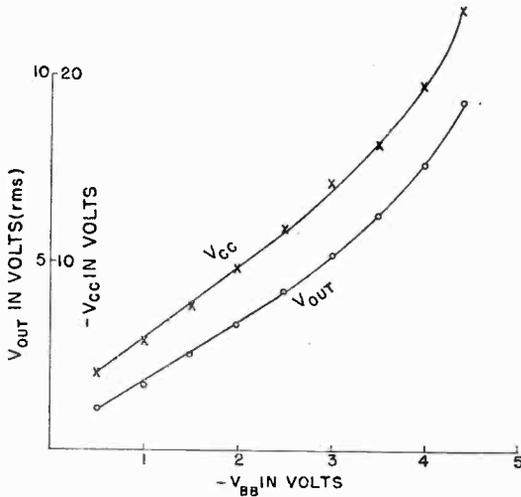


Fig. 11—Constant-frequency characteristics of an amplitude-modulated junction transistor oscillator.

satisfactory frequency modulation, attention should be directed to the elimination of amplitude modulation and the linearity of frequency deviation.

Collector Modulation

Referring again to Figure 2(a), $v_{cem} = V_{CC} - I_c R_i$. When $I_c R_i$ is made much less than V_{CC} so that v_{cem} is operating in the saturation region of the collector output characteristic, then any change in V_{CC} will not change the current and hence the amplitude of oscillation. Figure 4 shows the flattening of the modulating characteristics at higher values of V_{CC} particularly when low R_i is used. The corresponding frequency deviation is also shown in the same graph.

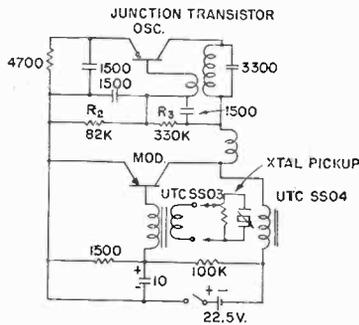


Fig. 12—Frequency-stabilized phonograph oscillator.

Base or Emitter Modulation

In base or emitter modulation, operation is dependent upon v_{cem} lying in the saturation region of collector output characteristic. Any change in base or emitter voltage then only changes the frequency. This is shown in Figures 5 and 6.

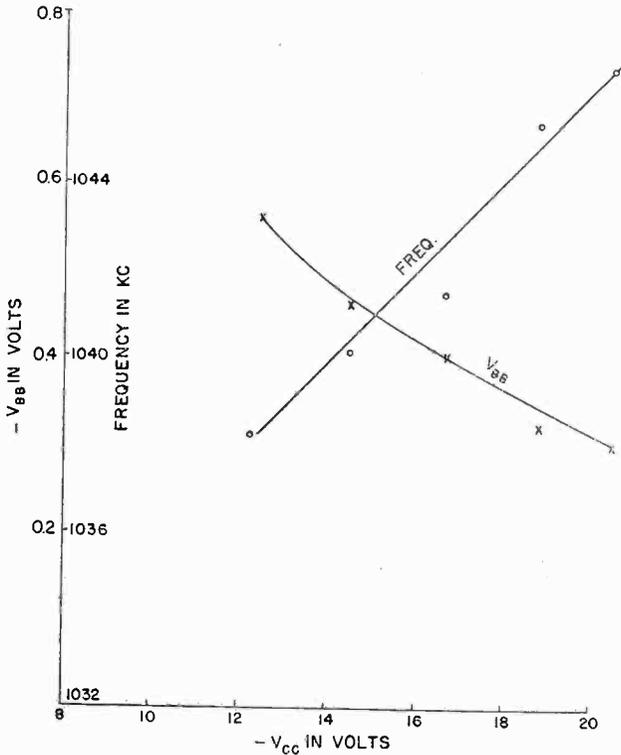


Fig. 13—Constant-amplitude characteristics of a junction transistor oscillator.

Double Modulation

An effective way to eliminate amplitude modulation is to modulate two electrodes at the same time. This is the reverse of the process used in stabilizing frequency. By proper proportioning of the modulation applied to two different electrodes, it is possible to minimize amplitude modulation and to obtain linear frequency modulation. In addition, the frequency deviation can be increased over that obtainable when modulation is applied to only one electrode. Figure 13 shows a constant-amplitude frequency-modulation characteristic. Note that there is a region where frequency deviation is quite linear.

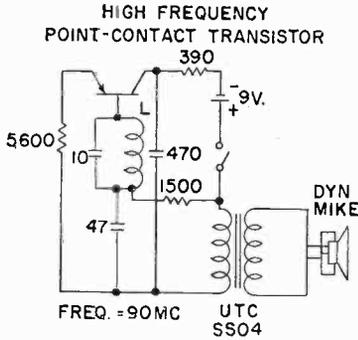


Fig. 14—Constant-amplitude FM point-contact transistor oscillator.

A practical application is the "Roving Microphone" shown in Figure 14. This is a 90-megacycle FM transmitter employing a point-contact transistor. By proper choice of R_2 and R_3 , the modulation applied to the emitter and that to the collector is in the correct proportion to produce frequency modulation with minimum amplitude modulation.

SELF MODULATION

A transistor oscillator can be made to be self-quenching simply by using a large RC time constant in the emitter circuit. The operation

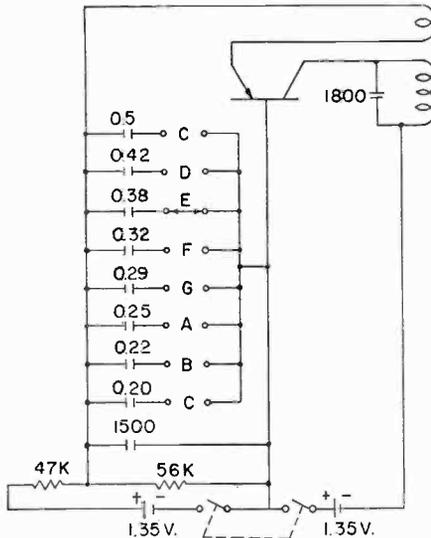


Fig. 15—Schematic diagram of a self-quenching junction transistor oscillator.

is roughly as follows.⁵ If the emitter is initially charged negatively, the transistor is cut off. The charge on the emitter capacitor gradually leaks off through the resistance in the emitter circuit reducing the bias until a point is reached where oscillation starts. As the amplitude of oscillation increases, emitter current flows and charges the capacitor. This bias builds up and damps out the oscillation, and a cycle is completed.

An application of this mode of oscillation is the 8-note "Musical Toy." The schematic of this unit is shown in Figure 15. This toy is used in conjunction with a radio. Depressing a key produces a tone in the radio corresponding to a particular time constant in the quenching circuit. When no key is depressed, the small condenser, C_o , serves to produce a continuous oscillation so as to activate the receiver automatic-gain control and to quiet the radio. The current drain of this unit is about 50 microamperes. Using two 1.35-volt, 0.3-ampere-hour mercury cells, the life of the battery should exceed 5,000 hours.

⁵ W. A. Edson, "Intermittent Behavior in Oscillators," *Bell Sys. Tech. Jour.*, Vol. XXIV, p. 1, January, 1945.

TRANSISTORIZED SYNC SEPARATOR CIRCUITS FOR TELEVISION RECEIVERS*

BY

HUNTER C. GOODRICH

RCA Victor Television Division,
Camden, N. J.

Summary—This paper presents transistorized sync separator circuits which appear to meet commercial requirements for operation under both good and adverse conditions. Tests on a limited number of transistors indicate good interchangeability and satisfactory performance up to 60°C. The impulse-noise immunity of sync separators is discussed. A method of using diode switching to control the time constant of a single-transistor separator for optimum impulse-noise performance is given.

Transistor collector voltage ratings must be at least equal to the level of sync output voltages required. For good sync rise time in the horizontal separator, transistors are required with performance which is better than that required for audio use.¹ Low I_{co} and I_{eo} are desirable, but most of the circuits described will tolerate high-temperature leakages of the order of 50 to 100 microamperes.

Because of the abrupt low-voltage "knee" of the junction transistor, one transistor will produce double-clipped sync. Since this function is usually accomplished with a two-stage vacuum-tube amplifier, the use of a transistor results in increased circuit simplicity.

GENERAL DISCUSSION

FEW circuits in a television receiver have shown as much variety as those for sync separation and automatic gain control (a-g-c).

This has resulted from the efforts of many engineers to devise economical circuits which will perform reliably under a wide variety of conditions. Some of these conditions encountered in the field are listed below:

1. A signal strength range on the order of 100,000 to 1.
2. Variations in sync percentage from the nominal 25 per cent.
3. Variation in tube (or transistor) characteristics including initial characteristics, aging, changes with temperature.
4. Variation in gain of associated circuits such as the r-f, i-f, and video stages.
5. Variations in line voltage.
6. Impulse noise of various types.
7. Airplane flutter.

* Reprinted from *RCA Review*, December, 1955.

¹ C. W. Mueller and J. I. Pankove, "A P-N-P Triode Alloy Junction Transistor for Radio-Frequency Amplification," *RCA Review*, Vol. XIV, pp. 586-598, December, 1953.

8. Reflections and multipath reception.
9. Various degrees of picture contrast desired by user.

Vacuum-tube circuits have been developed which meet these conditions in a reasonably satisfactory manner. Although such circuits may be relatively uncomplicated in structure, they are the result of much engineering and field testing, and make optimum use of the characteristics of the particular tubes employed.

Transistors differ from vacuum tubes in many important respects. As a result, most of the vacuum-tube circuits will not work effectively with transistors. The purpose of this study has been to investigate the characteristics of transistors which are relevant to sync applications and to develop circuits which will utilize these characteristics to provide optimum performance at reasonable cost. Certain transistor characteristics such as low input impedance, uncontrollable forward collector conductivity, temperature sensitive leakage, and limited voltage ratings generally represent disadvantages for sync applications. On the other hand, the high transconductance, sharp cutoff and saturation characteristics, and the absence of heaters, represent definite advantages for transistors. Also, the existence of both p-n-p and n-p-n transistor types offers increased flexibility in devising circuits. Thus the application of transistors to sync circuits presents both new problems and new opportunities to the circuit designer.

THE JUNCTION TRANSISTOR AS A PULSE AMPLIFIER

All of the three basic transistor amplifier types—common emitter, common base, and common collector—can be used as sync amplifiers. The common emitter connection gives the poorest rise and fall times for a given transistor. However, since it is the only one giving both voltage and current gain, both of which are usually required, it is the most useful. Most of the following discussion of pulse amplifier characteristics is therefore devoted to the common emitter type.

Figure 1a illustrates the common emitter circuit and Figure 1b shows one form of its equivalent circuit.¹ Equivalent-circuit parameters are given for the 2N34 and the SX-160*, the latter being an experimental alloy junction r-f transistor described by Mueller and Pankove.¹ Since the vertical straightness of the received picture depends on the timing accuracy of the sync pulses, reasonably good rise time of the horizontal sync amplifier is highly desirable. An important limitation on rise time is set by the time constant of r_{bb} and $c_{b'e}$. A low-frequency transistor such as the 2N34 may have a

* After the work described in this paper was completed, commercial units with performance comparable to that indicated here for the SX-160 have become available as RCA types 2N139 and 2N140.

rise time actually in excess of the 5-microsecond width of horizontal sync when operated with common emitter.

The 10 to 90 per cent rise and decay times of the medium-frequency experimental SX-160 with a 10,000-ohm collector load are of the order of 2 microseconds. Both rise and decay times increase with increasing collector load resistance as indicated in Figure 2. The rise time can be reduced greatly by overdriving, as will be discussed later. The equivalent circuit of Figure 1b indicates that any driving source resistance will lengthen the rise time by effectively adding to $r_{bb'}$. Tests on an SX-160 substantiated this as shown in Figure 3.

Figure 4 illustrates the collector family of curves for a junction transistor. A fortunate aspect of this characteristic for sync separators is the abrupt knee which occurs near zero collector voltage.

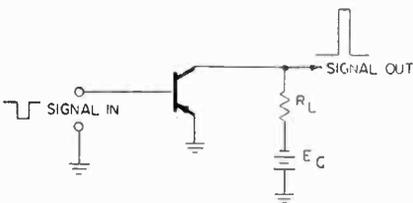


FIG. 1a

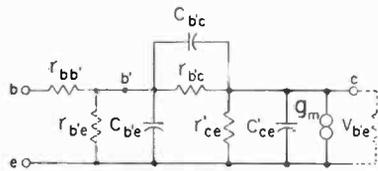


FIG. 1b

TYPICAL VALUES FOR $I_c = 1 \text{ MA.}$

		2N34	SX-160
$r_{bb'}$	OHMS	250	75
$r_{b'e}$	OHMS	1000	500
$C_{b'e}$	UUF	10,000	10000
$r_{b'c}$	MEG	1	0.1
$C_{b'c}$	UUF	35	10
g_m	MHOS	.032	.032

Fig. 1—(a) Common-emitter circuit; (b) pi-equivalent circuit.

Thus an amplifier with the proper load may be driven into increasing collector conduction until abruptly limited by collector voltage saturation. Such an amplifier, when driven from cutoff to saturation, produces a double-clipped output with an amplitude only a few tenths of a volt less than the collector supply voltage. The relatively flat collector characteristic above the knee is a result of the fact that the current carriers passing through the base layer to the collector travel by diffusion instead of being attracted by an electrostatic field as is the case in a vacuum tube. The value of the collector voltage thus has little effect on the collector current as long as it is sufficient to attract the carriers that have reached the collector junction.

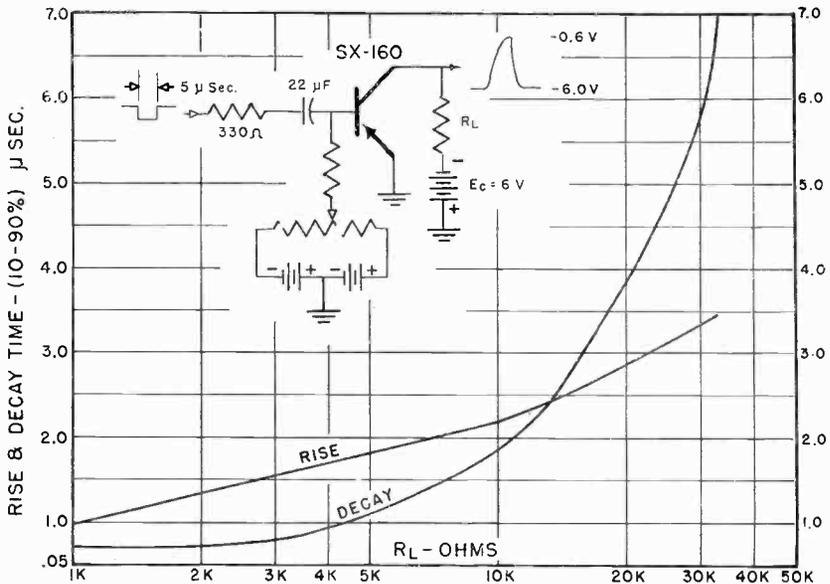


Fig. 2—Rise and decay time versus collector load.

Pulse-driving a transistor into collector saturation modifies both the rise and fall times at the output. The application of a pulse which drives the base-emitter junction farther in a forward direction than is necessary just to saturate the collector will be called *overdriving* the transistor. Overdriving produces a marked decrease in rise time.

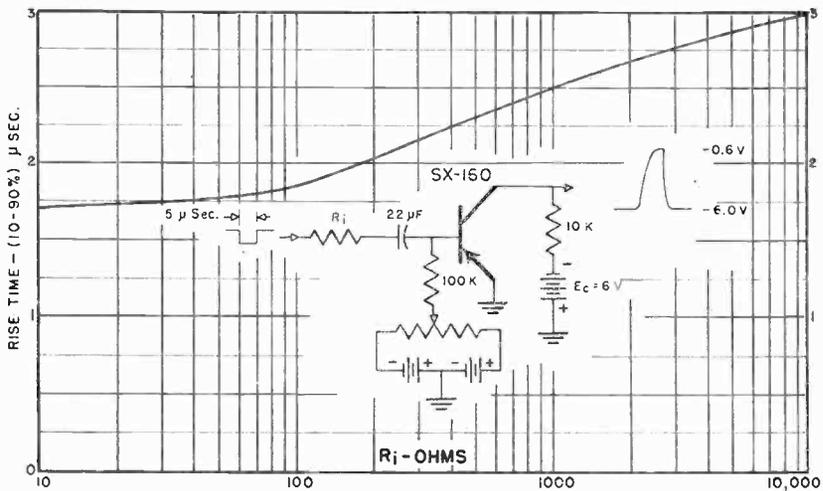


Fig. 3—Rise time versus external base resistance.

As an example, applying an input signal voltage double that required just to saturate the collector of an SX-160 reduced the rise time from 2 to 0.2 microsecond. Lesser amounts of overdriving will produce lesser degrees of pulse steepening. The variation in rise time as a function of base-driving current in excess of that required to saturate the collector on sync tips is given in Figure 5 for an SX-160 and a 2N34. Since the output transistor of a sync separator will normally be overdriven to some extent to achieve double clipping, a considerable degree of pulse steepening will result. However, the use of an inherently low-frequency transistor which requires a large degree of over-

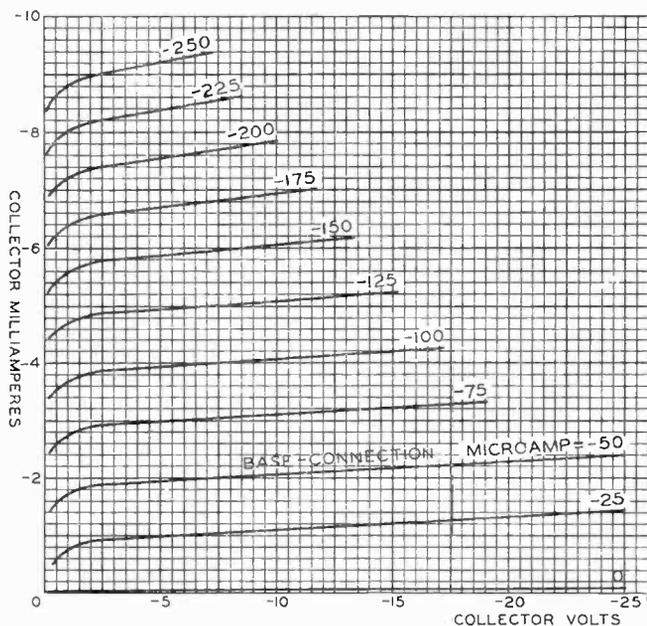


Fig. 4—Typical junction-type transistor characteristics.

driving to produce an acceptable rise time is wasteful of gain and has a deleterious effect on the trailing edge of the output.

The further effect of overdriving a junction transistor is that the trailing edge of the output pulse may be delayed, i.e., the pulse may be effectively widened. This effect has been called "back-porch" effect. Back-porch results from the fact that, while collector saturation limits the collector current to a value equal to the supply voltage divided by the load resistance, overdriving produces a minority carrier current flow into the base in excess of this limit. Some of the excess carriers are thus stored in the base and continue to flow to the collector after the input signal has returned to zero or reversed. Serious back-porch

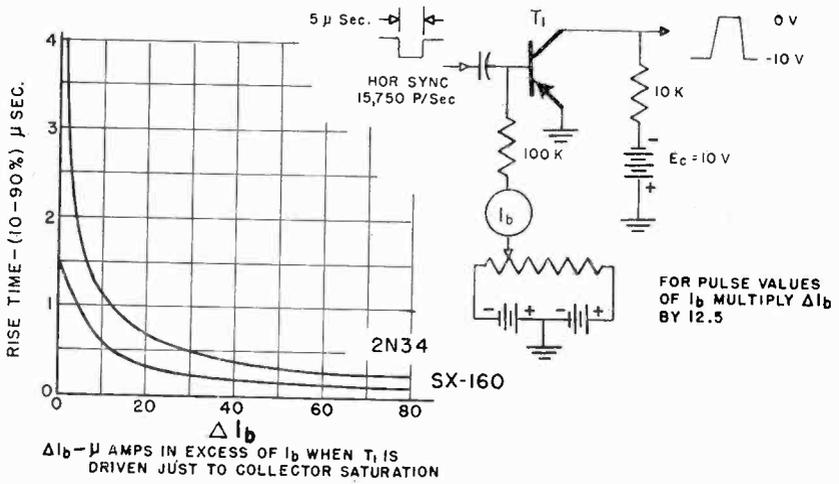


Fig. 5—Rise time with overdriving.

effect shifts the average timing of the sync output pulses and will cause a phase shift in many types of horizontal phase detectors. Other types of phase detector work primarily from the front portion of the sync pulses and will be little affected. In any case it is possible with a medium-frequency transistor such as the SX-160 to get adequate overdriving for good limiting and rise time without producing serious back-porch effect.

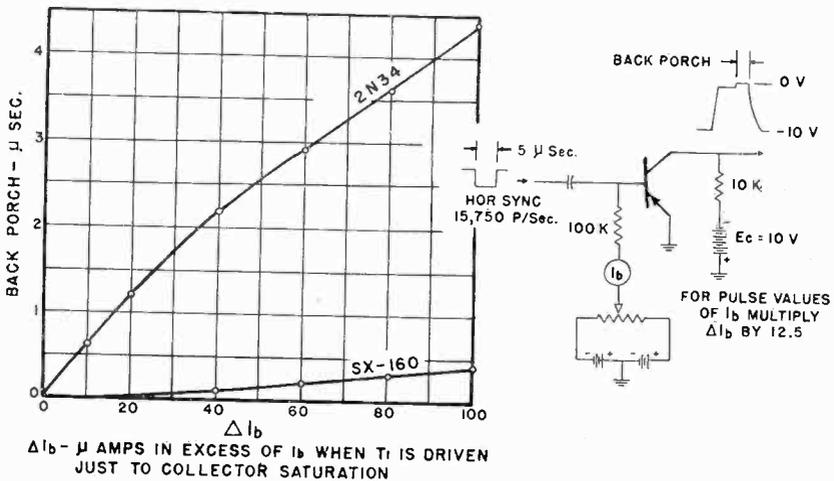


Fig. 6—Back-porch width with overdriving.

DRIVING SOURCE

Transistors operated with common emitter or base have a low input resistance. The particular value of resistance varies considerably depending on transistor characteristics, circuit parameters, and operating current. An average SX-160 as typically operated in a sync separator with common emitter and an 8200-ohm collector load has an input resistance of the order of 500 to 1000 ohms within its amplifying range. Beyond the saturation point the input resistance falls to a fraction of this value; below cutoff it may rise to over a megohm. It is evident that a low-impedance video source is desirable to drive transistorized sync and a-g-c circuits most effectively. Since at low power levels a transistor is essentially a current-driven device, a high source impedance with a signal voltage proportionally increased to supply the required signal current might also be satisfactory. However, there are two limitations to the use of such a high-impedance source: (1) the high video amplitude existing between sync pulses while the transistor is at cutoff will cause increased base-to-emitter leakage, thus disturbing the bias, and (2) the source impedance adds to the base resistance to produce, in conjunction with $C_{b'e}$, a longer rise time of the output sync pulses.

It seems basic to a-g-c circuits that the video signal must be d-c coupled to the a-g-c rectifier or amplifier. Otherwise a sudden increase in signal strength can overload the i-f amplifier, reduce the a-c video component, and block the receiver as a result of the faulty information derived from the a-c signal. Since transistors will not withstand the normal plate voltages encountered in a television receiver, this means that the video signal applied to an a-g-c or combined sync and a-g-c transistor circuit must come from a source close to ground potential.

One obvious source of such a video signal is the second detector load. Since the transistors tested had a base-to-emitter capacity on the order of 15 micromicrofarads in the cutoff condition, such a transistor must be tapped down at least to the mid-point of the detector load to prevent undue capacitive loading. Unfortunately, the detector signal level in terms of video current available to the separator is not sufficient to operate the more economical separator circuits. A further disadvantage of this signal source is that any impulse noise present has not yet been clipped by the first video amplifier.

A second possible video source is from a resistive divider from video plate to ground as shown at point A of Figure 7a. For some separator and a-g-c circuits this source is quite satisfactory. However, since the impedance of the divider must be high compared with the plate-to-B+ load to maintain the necessary video plate voltage, only

a fraction of the total video plate current is available to the transistor circuits. In cases where the signal source need not be near d-c ground, as in a capacitance-coupled separator, the full video signal current is available from a tap (B) on the normal video plate load.

A third video source, which eliminates most of the disadvantages listed above, is shown in Figure 7b. The voltage developed across R_1 , the normal cathode resistor, is not usually sufficient to supply the sync and a-g-c circuits. An additional resistor, R_4 , is added to increase the video voltage at the cathode. This does not introduce degeneration since it is not in series with the grid-to-cathode voltage applied from the second detector load. Taking the signal from point C, the source impedance is low and grounded and carries the full video current in V_1 . Optionally, a variable resistance R_5 can be inserted as a contrast control. As the resistance of R_5 is increased, the voltage across it

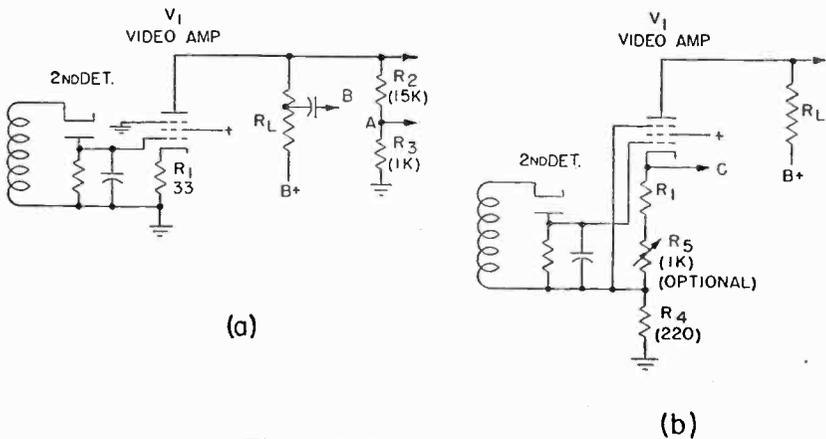


Fig. 7—Driving signal sources.

increases, but the voltage across R_4 decreases as degeneration decreases the gain of V_1 . The result is that the voltage to the separator can be kept relatively constant. If the resistance of R_5 is made too great, the current available to the separator may be insufficient, but a fair range of contrast control may be achieved. In this circuit the capacitance of the V_1 grid lead to ground is effectively increased by the added video voltage developed across R_4 . This capacitance should therefore be kept as low as possible.

TWO-STAGE SYNC SEPARATORS

A critical factor in a transistorized sync separator is maintaining the proper separation bias level. Even with a good a-g-c system the video level applied to the separator will change considerably with

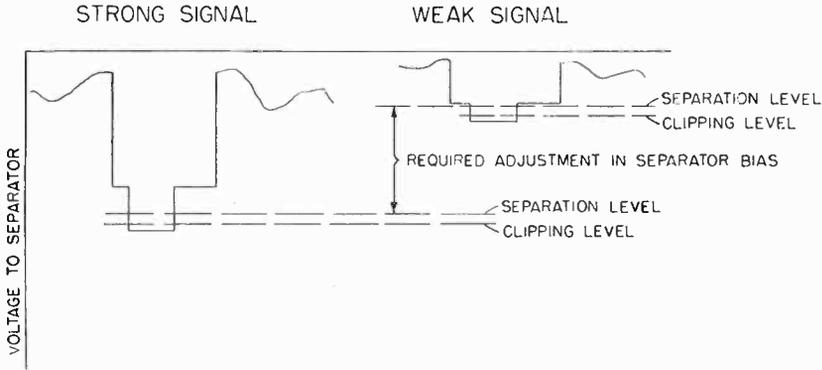


Fig. 8—Adjustment of separation level.

changes in line voltage, a-g-c control setting, and the extreme range in signal levels which may be encountered. Figure 8 illustrates how the separation level, i.e., the level where the separator begins to conduct, must change with signal level. This obviously requires the use of self-adjusting rather than fixed bias.

Figure 9 illustrates one form of two-stage sync separator. The emitter current flowing through R_1 establishes an emitter bias which biases the base-emitter junction of T_1 in the reverse direction except during sync. Collector current flows during sync to produce sync output pulses across R_2 . E_2 and R_2 are proportioned to avoid collector saturation, so that the pulse amplitude across R_2 varies somewhat with variations in input amplitude. The emitter current likewise varies with signal amplitude so that the bias is kept at the proper level for separation on all signals. This circuit is analogous to the cathode-biased vacuum-tube separator frequently used.

The first separator in Figure 9 is relatively noncritical of transistor characteristics, assuming that the transistor has adequate high-

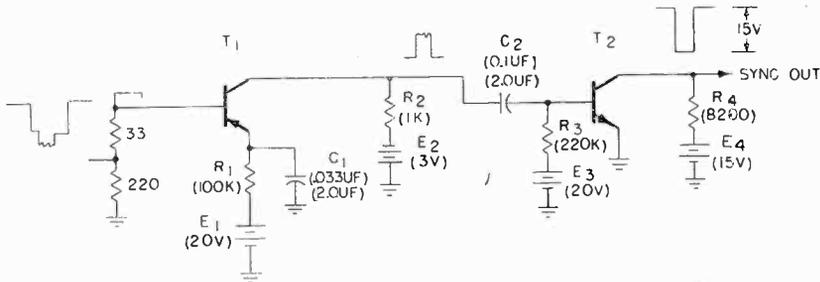


Fig. 9—Two-stage sync separator.

frequency response to give the required rise time. The collector current versus base-to-emitter voltage characteristic at low currents has been found fairly constant from unit to unit, at least for a given transistor type. The input resistance varies considerably, particularly with variations in α , but this does not affect the circuit operation if the signal has a sufficiently low impedance. An advantage of placing bias resistor R_1 in the emitter circuit is that if the d-c impedance of the base circuit is low, the collector-to-base leakage current, I_{co} , does not appreciably affect the bias. Between sync pulses the base-emitter junction is biased in the reverse direction, and any leakage at this junction, I_{co} , will flow through R_1 . In a transistor whose characteristics are satisfactory for the circuit, I_{co} should be small compared with the average emitter current, and hence have only a minor effect on bias.

In the circuit of Figure 9 a second transistor, T_2 , is used to amplify and limit the amplitude of the pulses across R_2 . T_2 is operated so that it is driven to saturation during sync by any usable signal. This provides sync of uniform amplitude and cuts off any noise pulses at sync level. Since the sync has been completely separated from the blanking pulses by T_1 , it would be possible to operate T_2 with fixed bias or with direct coupling. However, the bias would have to be set so that T_2 would saturate on the sync of weak signals; this means that on a strong signal T_2 would be heavily overdriven. The result is a considerable widening of horizontal sync or strong signals due to back-porch effect. Such heavy overdriving is avoided by using self-bias on T_2 , as is shown in Figure 9. Bias E_3 is adjusted so that T_2 will saturate on very weak sync. A transistor driven to collector saturation ceases to give normal transistor action since the collector current is fixed. The base-to-emitter characteristic under this condition is like that of a junction diode and has a low forward resistance. The effect of saturation on input resistance is shown in Figure 10. The input circuit of T_2 thus functions as a clamp circuit, producing sufficient reverse bias so that T_2 is always overdriven but never to the extent of producing excessive back-porch widening of the sync output. The average base current I_b due to conduction during sync, charges C_2 and must equal the currents discharging C_2 :

$$I_b = I_3 + I_{co} + I_{eo},$$

where

$$I_3 = \text{current through bias resistor } R_3,$$

$$I_{co} = \text{collector junction reverse leakage current,}$$

$$I_{eo} = \text{emitter junction reverse leakage current.}$$

The base current actually flowing during sync will be the average I_b divided by the duty factor of sync; this is about $12 I_b$. Variations of I_3 with signal level will be minimized by making E_3 relatively large compared with the amplitude of sync applied to T_2 . The primary effect of elevated temperature on T_2 is to increase the base current flow

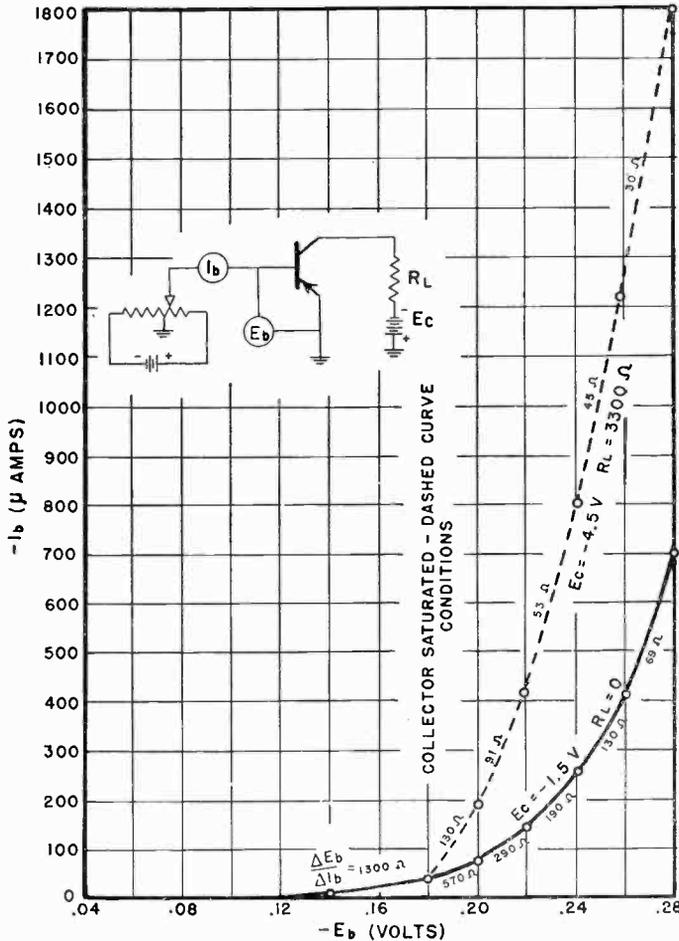


Fig. 10—Input resistance of common emitter stage (SX-160 for saturated and unsaturated collector).

during sync as a result of increased I_{co} and I_{eo} . With an experimental SX-161 transistor (the n-p-n counterpart of the experimental SX-160) an induced change in I_3 of 100 microamperes changed the output pulse width 0.5 microsecond. A similar increase in I_{co} and I_{eo} would have the same effect. The 2N35 exhibited a greater storage effect, the output pulse being widened 2 microseconds by the same change in I_3 .

The two-transistor separator of Figure 9, driven by the circuit of Figure 7b, was connected to a receiver. The sync output was applied to the balanced horizontal transistor phase detector.² An SX-160 and an SX-161 were used as T_1 and T_2 respectively. When the smaller values of C_1 and C_2 were used for separation of horizontal sync only, the performance of the circuit was quite satisfactory. Sync held over the full range of usable signal levels and produced no shift in raster when the sync in the signal was reduced to 12 per cent. Transistor interchangeability was good and simulated increases of 100 microamperes in I_{co} and I_{eo} in both stages produced only minor changes in centering. When the larger values of C_1 and C_2 are used to permit separation of both horizontal and vertical sync, performance is similar except for the decrease in horizontal impulse noise immunity resulting from the longer time constants. This effect is discussed in a later section.

The circuit of Figure 9 produces negative sync output which, as mentioned, is suitable for driving the transistor phase detector described in Reference (2). The sync polarity may be reversed by operating T_1 as a common-collector amplifier or by operating T_2 as a common-emitter amplifier. Figure 9 is presented as an example of a two-stage separator, and the discussion thereof explains some of the design factors involved. Many variations are possible. Since for each of the two stages there are three possible amplifier types (common emitter, base, or collector) two basic transistor types (p-n-p or n-p-n) and three means of biasing (self-base bias, self-emitter bias, or direct coupling), the number of possible combinations is in the hundreds. Many such combinations have obvious disadvantages, but many others would be workable circuits. Although it has not been possible to test all of these, the circuit of Figure 9 is believed to be among the better two-stage separators.

ONE-STAGE SYNC SEPARATORS

The two functions of sync separation and sync clipping are usually performed by separate vacuum tubes in commercial television receivers. A two-stage transistor separator was discussed in the previous section. However, the sharp cutoff and saturation characteristics of the junction transistor make possible its use as a one-stage separator, in which the same stage both separates and clips the sync pulses.

² B. Harris and A. Macovski, "A Symmetrical-Transistor Phase Detector for Horizontal Synchronization," *RCA Review*, Vol. XV, pp. 18-26, March, 1955.

The one-stage separator is shown in Figure 11. The bias developed by the emitter current flow through R_1 permits T_1 to conduct only during sync. At this time T_1 is driven to collector saturation, thus clipping the tops of sync as well as any noise that may be present. The output of T_1 is thus double clipped and will have a fixed output amplitude approximately equal to E_2 . A one-stage separator can also be built with the self-bias elements R_1 and C_1 in the base circuit, but this has the disadvantage that collector leakage current I_{co} passes through the bias resistor.

There are certain differences in the self-biasing action between Figure 11 and T_1 of Figure 9. In Figure 9, T_1 is not driven to saturation. The average collector current will thus vary with signal level. Since the collector current is the major component of the emitter current, it will cause adjustments in emitter bias with signal level. The change in base current required to produce a given change in

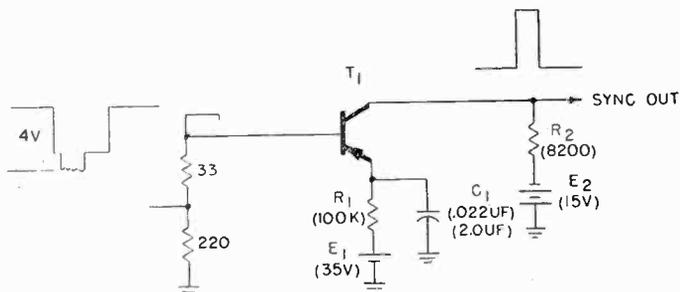


Fig. 11—One-stage sync separator.

bias current is low because of the base-to-emitter current gain of the transistor.

In Figure 11, since the collector is driven to saturation during sync at all signal levels, the collector current cannot increase with signal level. If the emitter bias is to adjust properly to varying signal levels, the adjustment must then come from changes in base-to-emitter current flow. As previously noted, after collector saturation the base-emitter junction takes on the characteristic of a forwardly biased diode. The fall in input resistance beyond saturation is shown by the change in slope of the dotted curve of Figure 10. The resistance falls well below 100 ohms. In this circuit the rectified input signal itself must provide the variations in bias, without benefit of the current gain of the transistor. The relationship between the required input signal level and source impedance and the variations encountered in signal amplitude and I_{co} will be illustrated by an example. The separator circuit of Figure 11, driven by the circuit of Figure 7a, was

installed in a receiver. The source impedance to the separator was 250 ohms and the absolute level of sync tips varied from +1 to +4 volts above ground for maximum combined excursions of a-g-c control setting and signal strength. Since the emitter bias closely follows the level of sync tips, this means a 3-volt variation across R_1 . This change in current through R_1 due to change in signal level will be called ΔI_s , in this case it is equal to 33 microamperes.

Let us allow a 50-microampere variation in I_{eo} with temperature and unit-to-unit differences. Now equating ΔI_b , the maximum resultant variation in current charging C_1 during each sync pulse, to the variations in currents discharging C_1 :

$$\Delta I_b = \Delta I_s + \Delta I_{eo} = 83 \text{ microamperes.}$$

Dividing the above average value by the duty factor of sync to get the base current flowing during each sync pulse gives about 1 milliampere. The normal video input signal is about 4 volts across the 250-ohm load. Sync normally represents 1 volt of this signal, so that 4 milliamperes of sync current flows in the source load. Since the input resistance of the transistor at saturation is very low, most of this current is available as input current to the transistor. Thus there is a safety factor to take care of weak signals or low percentage sync. The performance of the system under adverse signal conditions and normal temperature variations is good; details are given in the following section where a more refined version of the circuit is given. It may be noted here that increasing R_1 will decrease that part of ΔI_b due to signal level variations, but will not change the effect of temperature changes. If the allowance made for I_{eo} could be reduced, the signal source impedance could be raised.

It is also possible to build a common-base single-transistor separator as shown in Figure 12. This arrangement produces a shorter rise time than would be achieved with the same transistor operated with common emitter. The biasing action is very similar; however, since the video source must supply the sync output current directly, this circuit requires a greater video input current for proper separation.

SYNC SEPARATOR NOISE IMMUNITY

The synchronizing ability of a television receiver in the presence of impulse noise is a function of the noise characteristics of both the sync and the a-g-c circuits. This discussion is devoted to the former. One characteristic of a sync separator which aids noise immunity

is the ability to clip off noise so that it never exceeds the level of the sync output. The separators described in the previous sections do this. Another very important factor is that the separator not block or cut off for long periods following noise pulses. Impulse noise disturbances are normally of relatively short duration and low duty factor. The loss of sync information during the actual noise pulses usually does little harm. However, if the separator has long time constants that can charge on the noise pulses and keep the separator cut off for a long time thereafter, a serious loss of sync information results.

A long separation time constant is necessary for the separation of vertical sync. If horizontal sync is taken from the same circuit, the time constant must be on the order of 100 times that required for horizontal separation alone. This seriously affects horizontal noise

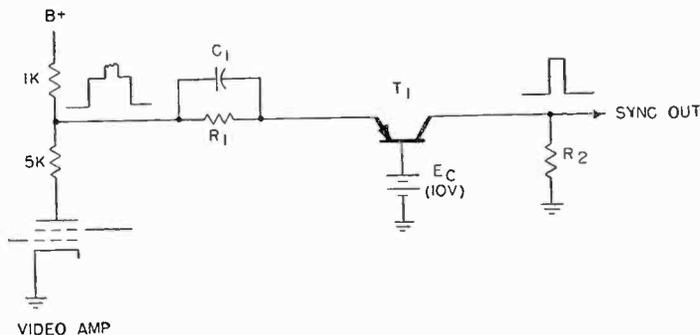


Fig. 12—Common-base one-stage sync separator.

immunity as shown by a comparison of lines 2 and 4 in Table I. One of the most effective solutions to the problem is the use of two separators, one optimized for horizontal and one for vertical. This represents an added expense, although the practicality of the one transistor separator makes it less so.

Another popular solution in vacuum-tube circuits is the use of a double time constant in the grid circuit of the separator. The application of this principle to a transistor circuit is shown in Figure 13.

In vacuum-tube circuits the equivalent in the grid circuit of R_3 , which is part of the short-time-constant combination, is made very large so that it limits the degree of charging possible in the large-time-constant circuit. The presence of this series resistance does not seriously reduce the tube gain to vertical sync because of the very high input resistance of the vacuum tube. The low input resistance of a transistor amplifier limits the usefulness of such a circuit.

Table I—Comparative Sync Separator Noise Immunity

Sync Circuit	Minimum signal level giving good horizontal synchronization in presence of impulse noise	
	2-Volt Detector Level	6-Volt Detector Level
1. Typical vacuum-tube circuit with separate horizontal and vertical sync separators and a common sync output amplifier	380 μv	200 μv
2. Figure 11, one-stage transistor separator	3000	1200
3. Figure 13, above with double time constant	800	250
4. Figure 11 with $C_1 = 0.022$ microfarads for horizontal separation only	70	70
5. Figure 14, transistor with diode-controlled time constant	300	200

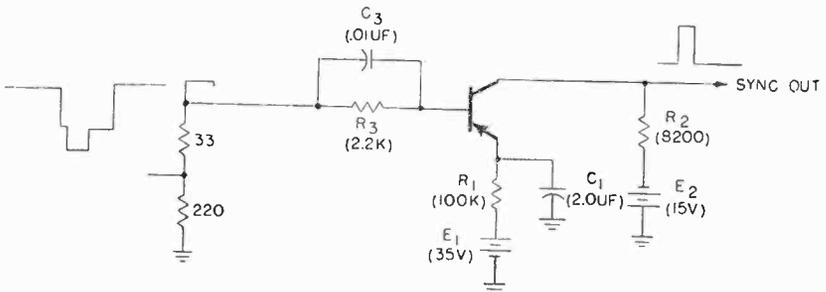


Fig. 13—One-stage separator with double time constant.

In this case, R_3 reduces the separator gain to vertical sync, and hence must be restricted in value. The values shown in Figure 13 reduced the vertical gain of the circuit by about one half. Comparison of lines 2 and 3 shows about a 3 to 1 improvement in horizontal noise immunity.

The common channel separator circuit of Figure 14 gives a marked improvement in noise immunity without appreciable loss in gain. This circuit is basically the common-emitter single-transistor separator previously described, adapted for positive power supply operation. However, it has two R - C time constant combinations in the emitter circuit which are brought into play at the proper time by diode D_1 .

During each sync pulse, the emitter current of T_1 biases D_1 in the forward direction, placing the long time constant of R_3 - R_4 - C_3 in the bias circuit. This will maintain the emitter bias for the duration of

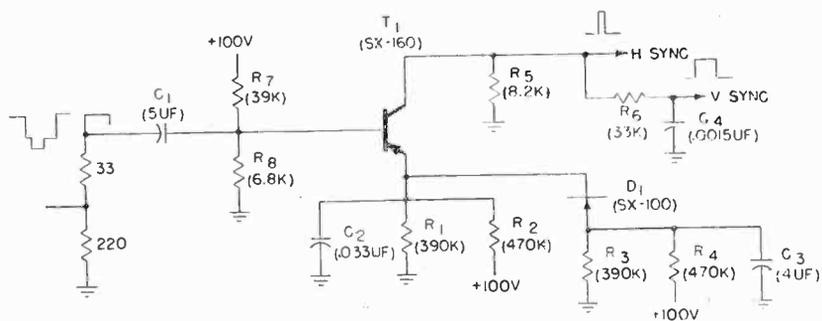
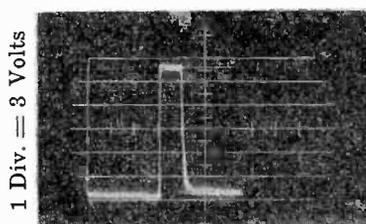
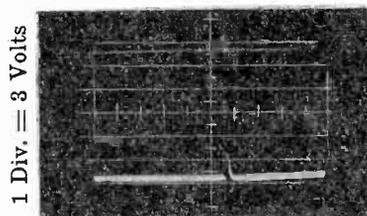


Fig. 14—One-stage sync separator with diode-controlled time constant.



1 Div. = 3 Volts
1 Div. = 5 μ Sec
(a) Horizontal rate.



1 Div. = 3 Volts
1 Div. = 1200 μ Sec
(b) Field rate.



1 Div. = 3 Volts
1 Div. = 240 μ Sec
(c) Expanded vertical sync.

Fig. 15—Sync output waveforms (one-transistor separator of Figure 14 with off-the-air signal).

the vertical sync pulse, permitting its proper separation. Any noise pulses will also cause D_1 to conduct, charging C_3 . However, after the noise pulse the excess charge on C_3 will keep D_1 open, and horizontal separation can resume as soon as the potential across the short time constant combination $R_1-R_2-C_2$ reaches the normal value. Comparison of lines 2 and 5 shows an average 8 to 1 improvement in horizontal noise immunity for the double time constant circuit.

RESULTS

The one-stage separator with diode-controlled time constant shown in Figure 14, driven by the circuit of Figure 7b, was installed in a receiver and checked against commercial requirements. It was found that the weakest intelligible signal could be held in sync and that no blanking appeared in the sync output until the sync was reduced below 15 per cent. The sync output had a 0.4-microsecond rise time and a 1-microsecond decay as shown in Figure 15. The resulting raster had no noticeable bends. In the presence of impulse noise the vertical and horizontal synchronization was comparable with that of a standard commercial receiver. The disturbance due to airplane flutter was somewhat less than in the standard receiver. Transistor interchangeability among the SX-160 transistors was good, the transistors being selected only for a V_{ce} greater than 30 volts. Heating the transistors to 60°C gave no appreciable change in performance. Of the transistors so tested, the highest I_{eo} was 80 microamperes at 60°C. The circuit showed good line-voltage tolerance, giving good separation from 85 to 130 volts. Thus it is possible to meet normal commercial performance standards with a transistor sync separator, and to do so with somewhat less circuitry than is usually employed with vacuum tubes.

TRANSISTORIZED VERTICAL DEFLECTION FOR TELEVISION RECEIVERS

By

MORRIS B. FINKELSTEIN

RCA Victor Television Division,
Camden, N. J.

Summary—Various transistorized circuits for the vertical deflection of kinescopes are examined. A circuit operating from 30 volts at 250 milliamperes is described which is capable of deflecting a standard 17-inch kinescope (50 degrees vertical deflection). Three transistors are employed: a sawtooth generator, a driver stage, and a power output stage. The circuit performance, including deflection linearity is satisfactory, but the d-c component in the yoke decenters the raster. The use of permanent-magnet devices for recentering is discussed.

INTRODUCTION

TRANSISTOR circuits have been developed for vertical deflection in a standard 17-inch television receiver. They resemble present vacuum-tube circuits in several basic respects. For example, a blocking oscillator is used to generate a sawtooth waveform which drives a power amplifier; also, a partially parabolized waveform is used together with curvature of the transfer characteristic of an amplifying device to correct for non-linearity in the system.

Differences in circuitry arise from the fact that transistors have a low input resistance and have voltage limitations which do not permit high-amplitude inductive voltage pulses. The transistor nonlinearities occur in a different manner from those of vacuum tubes.

The most promising circuits developed use three transistors: a blocking oscillator, a driver stage, and a single-ended power-amplifier stage. No output transformer is used. A permanent-magnet centering means has been developed to return the raster to normal from the deviation caused by the d-c component in the yoke (see appendix). These circuits are considered in detail.

POWER AMPLIFIER

Single-Ended Circuits

In a vertical deflection output circuit, the load is a deflection yoke or its transformed equivalent and appears as a resistor in series with an inductor. The required output current and voltage associated with

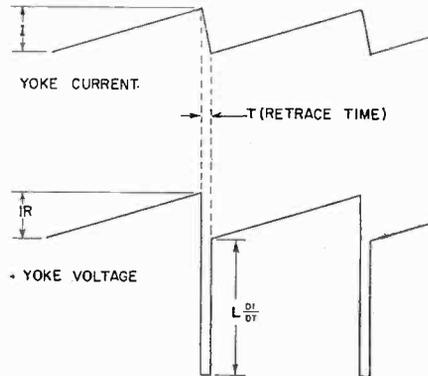


Fig. 1—Vertical yoke waveforms.

this type of load are shown in Figure 1. Since the retrace time is fixed between 200 and 400 microseconds,* the magnitudes of current and voltage across the yoke are dependent variables and choice of the best circuit connection and transformation ratio involves a compromise among these quantities and the ratings of available power transistors.

Figure 2 is the common-collector connection. Its operation is similar to the vacuum-tube cathode follower circuit and, as in that circuit, the output voltage is a replica of the input voltage. Since there is no voltage gain in this connection, the driving source requires the voltage capabilities presently found only in power transistors.

Figure 3 is the common-base connection. With an inductive load, the polarity of the input sawtooth waveform must be in a direction

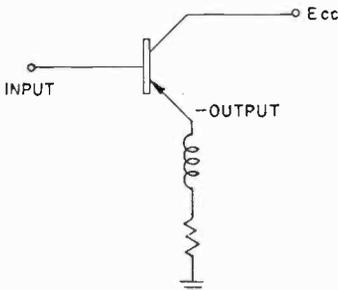


Fig. 2—Basic output circuit, common-collector connection.

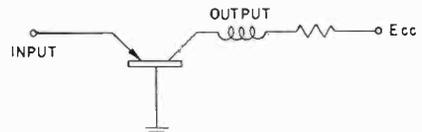


Fig. 3—Basic output circuit, common-base connection.

* Television blanking time tolerances allow up to 850 microseconds for retrace time, but vacuum-tube experience has shown that interlace is made easier by using a retrace time which falls within the double horizontal equalizing pulses following vertical sync.

such that the inductive voltage pulse $L di/dt$ will not be shorted by the collector-to-base diode, thereby lengthening retrace time. Since there is no current gain in this connection, a power transistor is again required as a driving source. Both the common-collector and common-base circuits would make inefficient use of a power-transistor driver because little driving power is needed (less than 50 milliwatts).

Figure 4, the common-emitter connection, is the one presently used. It has both voltage and current gain and can be driven by one of the presently available low power (50-milliwatt) transistors. Like the

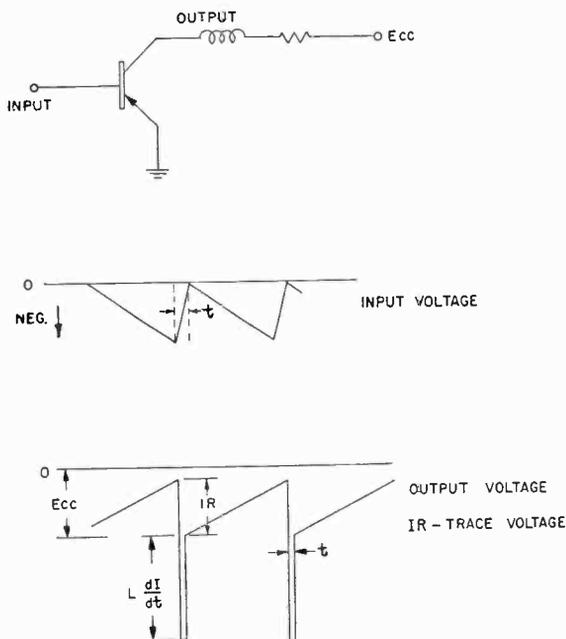


Fig. 4—Basic output circuit, common-emitter connection.

common-base circuit, it requires the proper polarity input waveform to prevent excessive damping of the output pulse. For a p-n-p transistor, the proper input is a negative-going (forward biasing) sawtooth as illustrated. Unlike vacuum-tube circuits, peaking of the input sawtooth waveform is not necessary. This is because plate current cutoff in a vacuum tube is a function of both the plate and grid voltages, whereas the collector current cutoff of a transistor is dependent only upon the emitter to base voltage. This is true up to a value of collector voltage called the breakdown point at which point there is an abrupt increase in collector current which can damage the transistor. Although it is

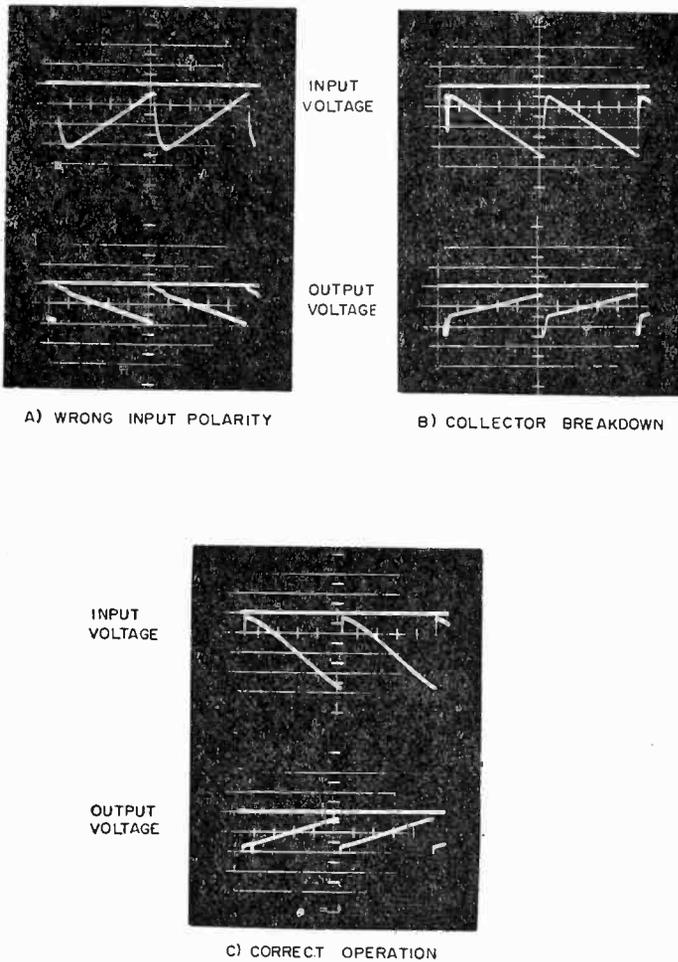


Fig. 5—Operating waveforms of vertical-output stage.

possible to exceed the breakdown voltage momentarily, such operation is detrimental to the retrace time. In the circuit of Figure 4 the maximum instantaneous voltage on the collector is the sum of the inductive retrace pulse Ldi/dt and the collector supply voltage E_{cc} . If this sum exceeds the breakdown voltage, the pulse will be damped, increasing retrace time.

The oscillograms of Figure 5 illustrate operation of the circuit in Figure 4 with: a) the wrong input polarity; b) collector breakdown; and c) correct input polarity. The lengthened retrace time of the input waveforms in Figures 5a and 5b are the result of feedback from the output circuit.

Push-Pull and Other Circuits

Thus far only single-ended circuits using one transistor have been considered. In vertical deflection using vacuum tubes it has been found economically feasible to use only one power amplifier device and this will most likely be the case with transistors. However, it is possible that more output power will be needed than can be supplied by one transistor, so it is of interest to examine multiple transistor output circuits.

Parallel operation is possible, but a significant increase in output over that of a single transistor can be attained only when the transistors are matched in gain and output resistance.

Push-pull operation of transistors (including the complementary symmetry mode) in the common-emitter connection to produce a sawtooth current waveform through an inductance faces a basic difficulty

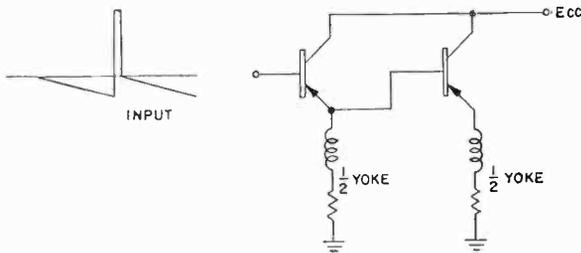


Fig. 6—Tandem operation of transistors, basic method.

not encountered in vacuum tubes — a difficulty which makes such operation impractical. In all of the circuits devised thus far, one of the output transistors damps out its collector pulse. The only way to prevent such damping is to operate the transistor at a high enough collector voltage that the pulse does not exceed it. This is very inefficient. It is possible to operate a push-pull circuit in the common-collector connection, but input voltage amplitude makes this method impractical.

Tandem operation of power transistors is feasible. Using each half of a yoke separately, a bifilar yoke, or separate primary windings of a transformer, the transistors may be connected as shown in Figure 6. Since the voltage drop from base to emitter is small, the voltage across the second half of the yoke will be almost the same as that across the first half yoke and consequently, their currents will be almost the same. A modification in Figure 7 uses a common-emitter circuit in the first stage. This simplifies input driving requirements. If this first stage is driven to collector saturation, the voltage developed across its half

yoke will have the correct d-c axis for direct coupling to the second stage. Tandem operation does not require matched transistors. Also, more stages may be added until the cumulative base-to-emitter voltage drops cause excessive unbalance in the loads.

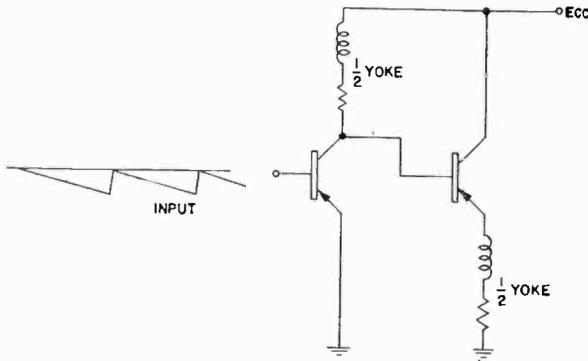


Fig. 7—Tandem operation of transistors, modified input.

Load Considerations

Using the circuit of Figure 4 and choosing a deflection angle and yoke, it is possible to calculate the magnitudes associated with the yoke waveforms in Figure 1. With a 70° kinescope (vertical deflection angle 50°) and the most sensitive 70° yoke commercially available, the following results obtain:

yoke parameters	$L = 46 \text{ mh}$	$R = 48 \text{ ohms}$
current amplitude	$I = 350 \text{ ma peak-to-peak}$	
retrace time	$t = 200 \text{ to } 400 \text{ } \mu\text{sec}$	
cutoff voltage pulse	$LdI/dt = 40 \text{ to } 80 \text{ volts}$	
trace voltage	$IR = 17 \text{ volts}$	
a-c yoke power	$I^2R/12 = 0.5 \text{ watt}$	
a-c plus d-c yoke power	$I^2R/3 = 2.0 \text{ watts}$	

A comparison of the above requirements with the ratings of available power transistors shows:

1. Power output is readily attainable;
2. Cutoff time is no limitation. Power transistors have a frequency response of at least 10 kilocycles;
3. Current ratings of the higher power transistors are generally sufficient even for direct drive to the yoke;

4. Voltage rating is a limiting factor and affects the choice of a load impedance. Present power transistors cannot withstand much more than 60 volts on the collector. This voltage is only slightly more than the minimum calculated trace plus cutoff voltage ($17 + 40 = 57$). The actual value is higher because the driving sawtooth cutoff is sharper than the gradual slope used in the calculations and also, the collector supply voltage must be somewhat higher than the trace voltage.

Since the yoke impedance itself is sufficient to produce a voltage approaching the maximum transistor collector rating, an impedance transformation from the yoke to a higher value is undesirable. Furthermore, a lower impedance is also not feasible at the present time because, although transistors are capable of large currents, the current gain decreases and linearity of the transfer characteristic changes rapidly above 500 milliamperes. If good high-current transistors are developed and if power supply problems are not encountered, a low-impedance yoke might be feasible with the advantages of lower cost and reduced pulse voltage.

Since the yoke itself is approximately the correct load for the power transistor, direct coupling of the output circuit would be advantageous. Elimination of the output transformer would not only reduce cost, but would remove a source of nonlinearity from the system. In vacuum-tube circuits, the nonlinearity caused by saturation of the iron core of the output transformer is compensated by the plate current characteristic of the output tube. However, the transistor nonlinear characteristic is such as to add to the transformer nonlinearity.

Unlike vacuum tubes, the transistor can operate efficiently into a low impedance load such as a direct-coupled yoke. Receiving type vacuum-tube triodes cannot deliver the necessary high currents without imposing severe requirements on the driving source. Pentodes or beam power tetrodes must not be driven below the knee of their plate characteristic making them inefficient for low-voltage, high-current operation.

Elimination of the output transformer has the disadvantage of causing decentering of the raster by the d-c component in the yoke. Yoke heating is caused by the added I^2R loss (four times normal). A reactor-capacitor network would remove the d-c component, but add cost and nonlinearity. This difficulty has been overcome by the development of a permanent-magnet centering method.

Input Considerations

Figure 8a shows the transfer characteristics of a typical high-gain power transistor in the grounded emitter circuit with a 50-ohm yoke

as the load. It is seen that an almost linear transfer function is obtained by applying a linear input voltage. This comes about because although the current gain decreases as current increases, the input resistance also decreases, and a low-impedance driving source will supply an increasing input current correcting for the loss in current

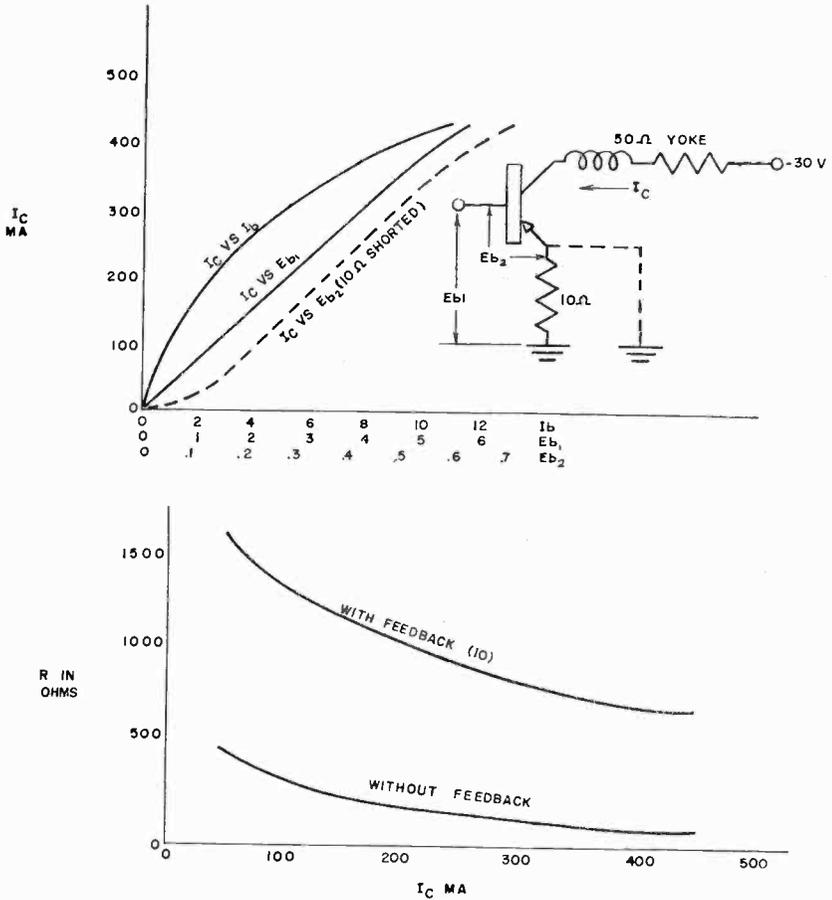


Fig. 8—Power transistor. (a) Transfer characteristics (above); (b) Input resistance (below).

gain. Unfortunately, the magnitude of the input resistance, Figure 8b, is very low and a constant voltage source for this resistance is difficult to achieve.

The addition of negative current feedback (10-ohm emitter resistor) increases the input resistance and makes the transfer curve more linear. These advantages outweigh the disadvantages of increasing the

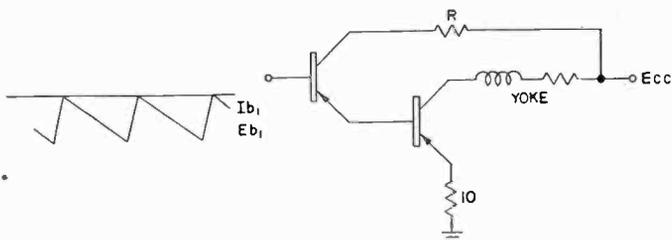


Fig. 9a—Common-collector driver circuit.

input driving power and in some cases the supply voltage, and this circuit is used in the following discussion on driving methods.

DRIVING METHODS

Among the transistor sawtooth generators developed thus far, the only one within commercial stability limits is a blocking oscillator type which has a relatively high impedance, low power output. A driver stage having power gain and providing a high-to-low impedance transformation is necessary between the sawtooth generator and power amplifier.

The common-collector circuit is inherently suitable for this type of operation. Figure 9a shows the simple driver stage requiring only the driver transistor and a resistance R which is merely a voltage-dropping resistor to prevent excessive collector dissipation.

The combined transfer curves (Figure 9b) of two p-n-p transistors

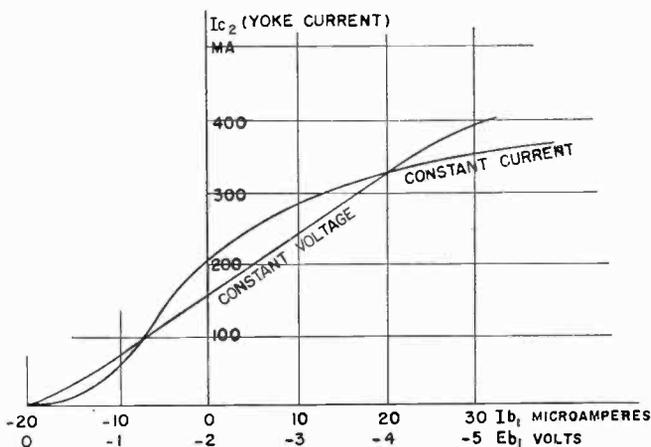


Fig. 9b—Cascaded transfer characteristic.

in cascade illustrate that good linearity can be obtained if the driver transistor is driven from a constant voltage source. The constant voltage curve is slightly S-shaped which tends to squeeze the top and bottom of the raster, but the deflection geometry of conventional kinescopes requires this sort of operation. The constant-current curve is very nonlinear and is shifted off axis by initial leakage current. It may be interesting to note that the total current gain of the circuit is extremely high, varying from 14,000 to 6,000 throughout the range. The voltage gain on the other hand is only 4.

The input resistance of this circuit is high, varying from 200,000 to 86,000 ohms, and since the output of the transistor sawtooth generator is about 10,000 ohms, reasonably constant voltage input can be achieved. It is possible to design a sawtooth generator to provide a sawtooth waveform having the correct d-c axis to permit direct coupling to the driver stage. Under this condition and with good driver and power transistors, linearity, height, and the d-c component of the yoke current are all under direct control of the sawtooth generator. The only variable controls that should be necessary are a frequency control and a height control. The latter can be obtained by varying the emitter resistor of the power stage. The driver transistor gain is not critical, but should be high enough to make the average input resistance 100,000 ohms or higher. Actual performance of the circuit is below commercial standards of linearity with the present sawtooth generator, having 2 to 1 squeezing at the bottom of the raster.

Better linearity at the cost of simplicity and stability is provided by the grounded emitter stage, one variation of which is illustrated in Figure 10a. Constant current input as provided by R_s is necessary to present a high-resistance load to the sawtooth generator. This circuit uses the curvature of the current transfer characteristic of the driver transistor to correct that of the power transistor. A graphical analysis is shown in Figure 11. The top curve is the input resistance of the power transistor plotted against input current, and the middle curve is the current transfer characteristic of a typical driver transistor. The polarity of the input sawtooth waveform to the driver is such as to drive the transistor from high conduction toward low conduction. Taking into account the load and bias conditions, a selected portion of the transfer curve will be traversed as shown. By multiplying the instantaneous values of output current to the power stage (about half of the total output current) by the instantaneous value of the input resistance, it is possible to plot a curve of input voltage to the power stage assuming a linear input current to the driver stage. The result is the bottom figure. The dashed line shows the effect of moving the

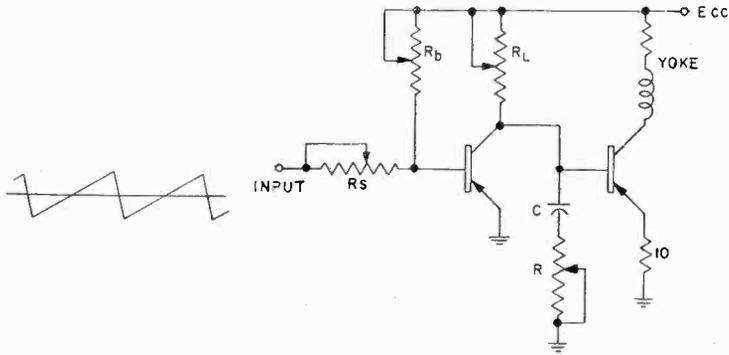


Fig. 10a—Common-emitter driver circuit.

operating point on the driver transfer curve to the left and the dotted line shows the effect of moving it to the right.

It has been shown (Figure 8a) that a linear input voltage to the power transistor produces an almost linear output current. The curve of Figure 10b taken under actual operating conditions, coincides closely with the solid curve at the bottom of Figure 11. The very beginning of the curve in Figure 10b is too nonlinear to be used and therefore the operating bias on the power transistor must be adjusted so that this amount of current rides along as an unused d-c component. This circuit is thus less efficient than the common collector driving circuit.

The beginning of the usable portion of the transfer curve has a higher slope than the rest, and correction for this part is provided by the R and C network in Figure 10a. The network acts as an integrator which draws an exponentially decreasing current, thereby subtracting from that which is available to the input of the power transistor causing more effect at the beginning of trace, which tends to straighten

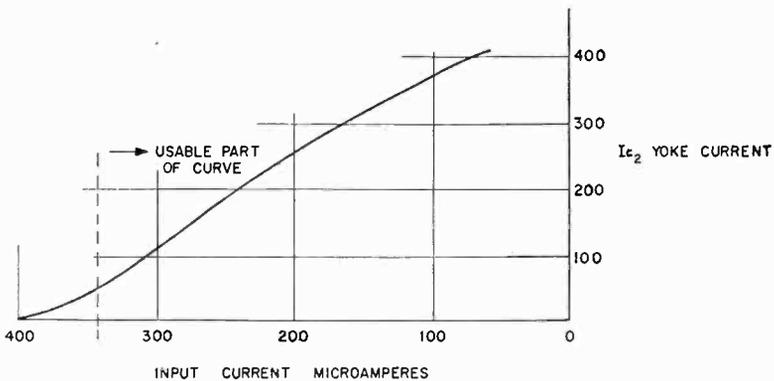


Fig. 10b—D-C cascade transfer characteristic.

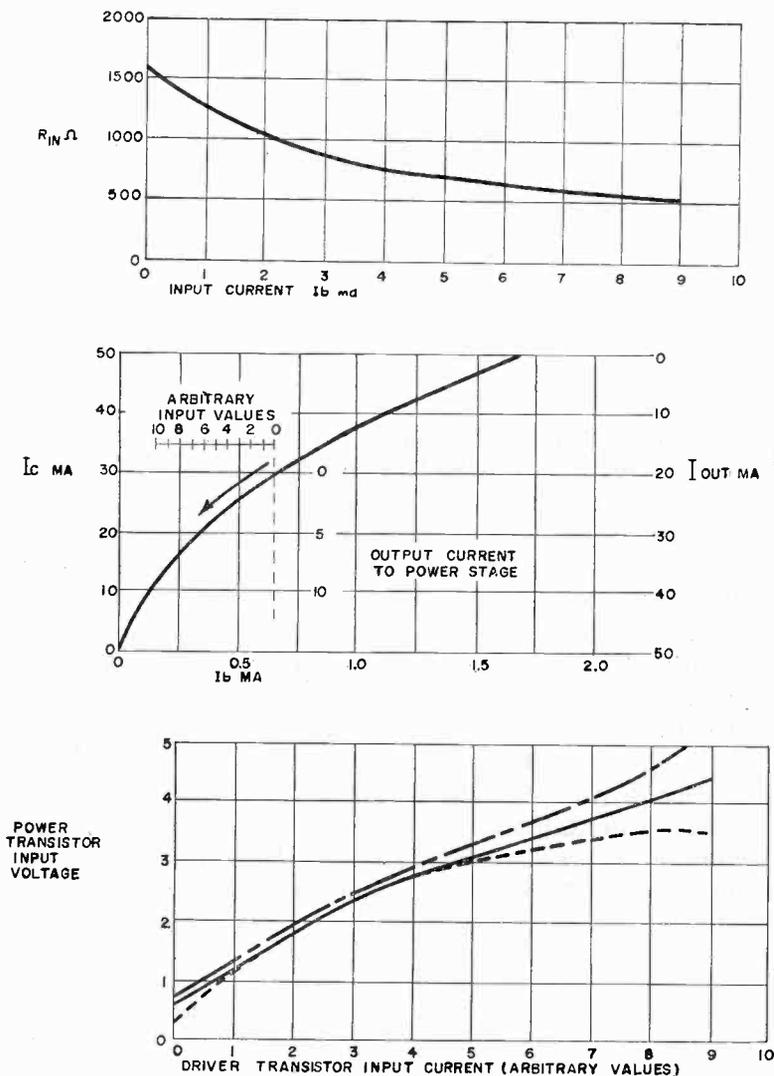


Fig. 11—Top: Power transistor input resistance versus I_b (with 10-ohm feedback). Center: Driver transistor transfer curve. Bottom: Input voltage to power transistor.

the curve. Too much integration produces a parabolic voltage causing foldover at the top of the raster.

The sawtooth generator developed for this circuit has better linearity than that developed for the common-collector circuit, and with a high-gain driver transistor linearity of 5 to 10 per cent, squeezed at the bottom, can be achieved. Replaceability of the driver transistor

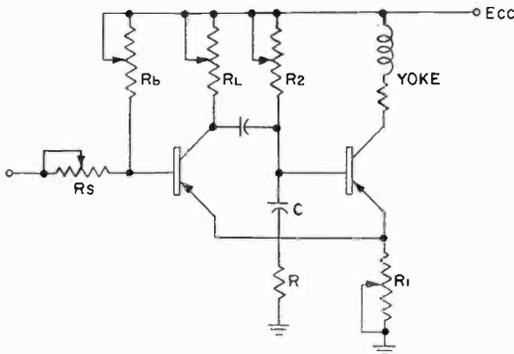


Fig. 12—Regenerative driver circuit, current feedback.

requires readjustment of R_s , R_b , R_L , and R . With present transistors there is a cumulative warm-up drift between $\frac{1}{2}$ and 1 inch. Better replaceability and stability can be obtained by capacitive coupling of the driver and output stages. However, this requires a capacitor of at least 100 microfarads and additional biasing means for the power transistor.

Lower-gain driver transistors may be used by adding positive feedback to increase the input resistance of the driver stage. Figure 12 illustrates the use of current feedback. The common emitter return to R_1 upsets bias conditions on the driver stage so that direct coupling cannot be used. Figure 13, using voltage feedback, is much less complicated. The feedback network R_A , R_B , and C includes parabolic correction and bias means for the driver transistor. This circuit provides good linearity even without an emitter resistor in the power stage, but stability is poor. Reliable operation using positive feedback has not been achieved.

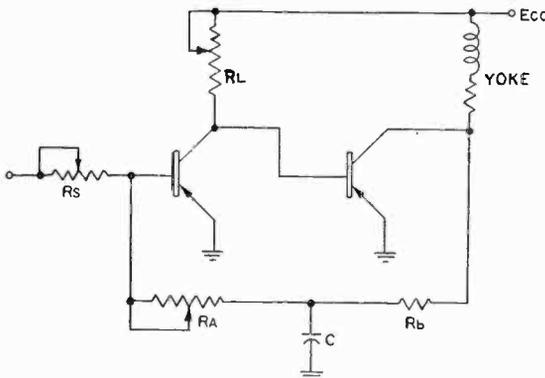


Fig. 13—Regenerative driver circuit, voltage feedback.

SAWTOOTH GENERATOR

Timing and waveshape requirements such as linearity, stability, synchronization interlace, and retrace are the same for both vacuum-tube and transistor sawtooth generators. Differences are: the transistor sawtooth normally needs no peaking, requires less amplitude, and must supply a relatively low impedance which varies with yoke current.

Some of the various relaxation oscillators which may be used to generate a sawtooth waveform are the blocking oscillator, the multivibrator, and the negative-resistance oscillator. All of these serve as switches to initiate charge or discharge of an $R-L$ or $R-C$ circuit. Since transistors are low-impedance devices, it would be desirable to use an $R-L$ circuit, but maximum ratings of transistors at present favor using the $R-C$ circuit with its high peak charging current rather than the $R-L$ circuit with its high decay voltage.

It is possible to calculate the peak charging current needed by estimating approximate values for the R and C . To obtain a linear sawtooth waveform, the $R-C$ time constant product should be considerably greater than the period of the waveform which, for 60 cycles, is 16,600 microseconds. A factor of 5 times this value will insure good linearity (2 per cent deviation), making the $R-C$ product 83,000 microseconds. To prevent loading by the driver stage, its resistance should be considerably greater than that of the $R-C$ timing combination. Since the minimum resistance of the driver stage (at maximum yoke current) is about 100,000 ohms, a factor of one-tenth this value will insure little loading, making R equal to 10,000 ohms. The value for C becomes 8.3 microfarads.

The peak current in the capacitor will be limited by the series charging resistance in the circuit, but it must be made high enough to produce sufficient stored energy in the capacitor during the charging time to reach the required voltage. Measurements indicate that a 6-volt peak-to-peak sawtooth should be sufficient. The charge on the capacitor will then be $Q = CE$, which is 49.8 microcoulombs. The charging time allowed is 400 microseconds, so the peak current will be dQ/dt which is about 125 milliamperes. Small junction transistors of the alloy type are capable of handling the peak power required. Multivibrator circuits have been tried, but good linearity and stability have not been obtained. Part of the trouble may be due to the fact that the current gain of the transistors decreases at the high currents and the charging transistor cannot maintain a good short circuit. With a blocking oscillator, however, sufficient feedback can be used to maintain the peak current even though there is a fall-off in current gain.

A satisfactory blocking oscillator circuit is shown in Figure 14. It differs from the conventional vacuum-tube blocking oscillator in that the same R - C combination is used to provide both the timing and the output waveform. Also, the capacitor is charged through the device rather than being discharged through it as in the vacuum tube circuit.

The circuit operates as follows: The base-emitter diode of the transistor is biased to conduction by bleeder current through R_2 . Collector current flows, charging C through the transformer primary which is regeneratively coupled to the base. Base current increases, causing a rise in collector current, the process being rapidly cumulative. The transistor becomes almost a short circuit and C charges through the resistance of the transformer primary and the transistor to a value of voltage determined by the fixed base bias and the induced voltage

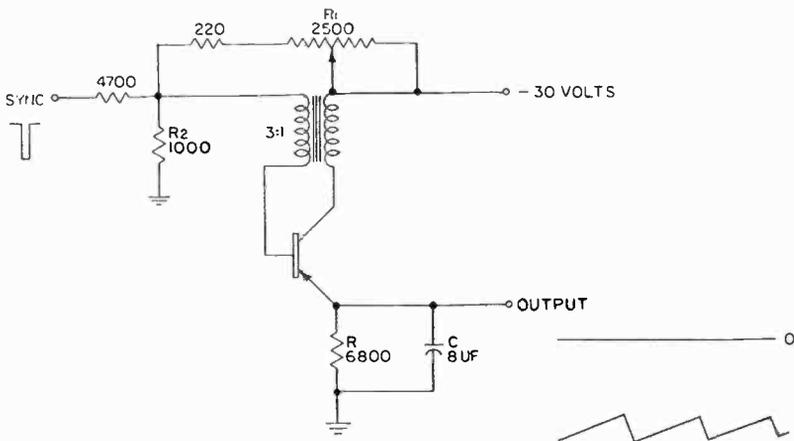


Fig. 14—Sawtooth blocking oscillator.

in the transformer secondary winding. At this point, base current ceases and the transistor is cut off and remains cut off until C discharges through R to the value of the fixed base bias. R_1 controls this bias, serving as a frequency adjustment. A negative sync polarity of 1 volt amplitude is required.

Good stability is achieved by the heavy bleeder bias which takes 10 times more current than the oscillator itself. Frequency deviation is not more than 4 cycles for a combined supply voltage change of ± 20 per cent and a temperature variation of 25°C . Replaceability is good.

The circuit of Figure 14 produces a positive-going sawtooth using a p-n-p transistor as shown. Such a sawtooth is suitable for the grounded emitter driver stages previously described. Capacitor coupling is necessary because the sawtooth is considerably displaced from the

zero axis. A grounded collector driver stage requires a negative-going sawtooth (it is assumed in all cases that p-n-p power transistors are used). An equivalent n-p-n oscillator transistor would produce such a sawtooth in the circuit of Figure 14 with proper reversal of supply voltage polarity. However, such transistors are not readily available and the p-n-p circuit (Figure 15) has been developed.

The operation of this circuit is more like that of a vacuum-tube blocking oscillator than the circuit of Figure 14 in that the timing capacitor is discharged through the device. However, the functions of timing and output are still combined in one R - C network. The transistor is held cut off by the bias developed across an emitter resistor, R_2 , controlled by the frequency adjustment, R_1 . C charges through R until the voltage across the divider network R_3 and R_4 reaches a value such that the voltage across R_4 equals the voltage R_2 .

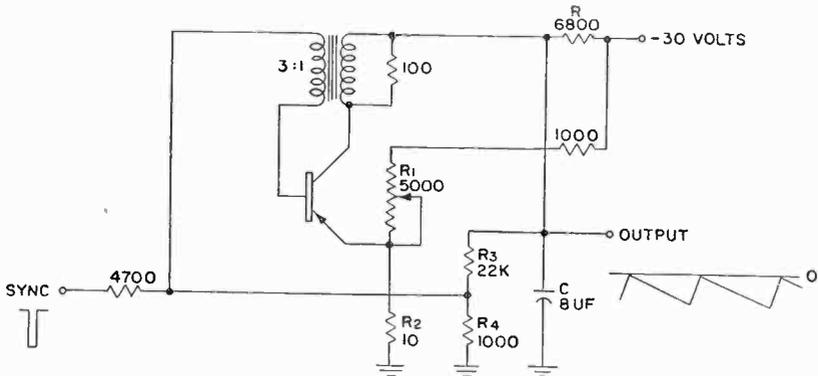


Fig. 15—Sawtooth blocking oscillator.

The transistor will start conduction, rapidly accumulating to become a short circuit, discharging C through the transformer and R_2 , which is made small to prevent limitation of the discharge current below the necessary peak value. Because R_2 is small, the amount of feedback needed to overcome its bias is small and C will discharge almost to the zero level before the transistor again reverts to cutoff. This produces a waveform suitable for direct-coupling to the following driver stage. Linearity is not as good in this circuit as in that of Figure 14, partly because of the shunting effect of the divider network, R_3 and R_4 . The resistance of the network cannot be made much higher than it is because R_4 must be kept at a low enough value to minimize transistor leakage effects and to allow sufficient base current flow to saturate the transistor. Stability, though not as good as in first circuit, is within commercial tolerance and replaceability is better.

CONCLUSIONS

The various circuits described in this report may be assembled into two basic vertical deflection systems, each using three transistors—a sawtooth generator, a driver stage, and a power output stage. These are shown in Figure 16. One system uses a common-emitter driver stage and except for a slight temperature drift, will provide acceptable 50° vertical deflection with commercially available transistors. The second system has a common-emitter driver stage, has fewer components and adjustments, and is satisfactory in all respects except linearity. Operation is from 30 volts with 250 milliamperes total current drain.

Transistors used in the driver and sawtooth generator circuits are low-power p-n-p types. Characteristics necessary for use as a sawtooth oscillator are reasonably high gain and capacity for high peak currents. The common-emitter driver stage requires a high-gain transistor with curvature in its current transfer characteristic. The common-collector driver stage requires a reasonably high-gain transistor with a linear transfer characteristic.

The power transistor used is a high-gain experimental type.* When operated in a grounded-emitter connection with 10 ohms of emitter feedback, it exhibits a fairly straight transconductance curve, drooping beyond 350 milliamperes. Collector breakdown is 60 volts. An extended linear current capacity would allow more linear operation of the common-collector driver system. Some aspects of the operation of the power output stage differ from vacuum tube practice:

1. The power transistor must be driven by a sawtooth polarity such that the inductive retrace pulse will be in the direction of reverse collector bias. Otherwise, the pulse will be excessively damped by collector diode conduction. For a p-n-p common emitter connection, the correct input is a negative-going sawtooth.
2. The output load impedance must be low enough to insure that the retrace pulse will not exceed the collector breakdown voltage.
3. Transistors work efficiently into much lower output impedances than is possible with vacuum tubes.
4. Transformer coupling is undesirable because the nonlinearity of the transformer is in a direction to add to that of the transistor output stage rather than to be compensated as in the case of a vacuum tube output stage.

* B. N. Slade, "Recent Advances in Power Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 153-171.

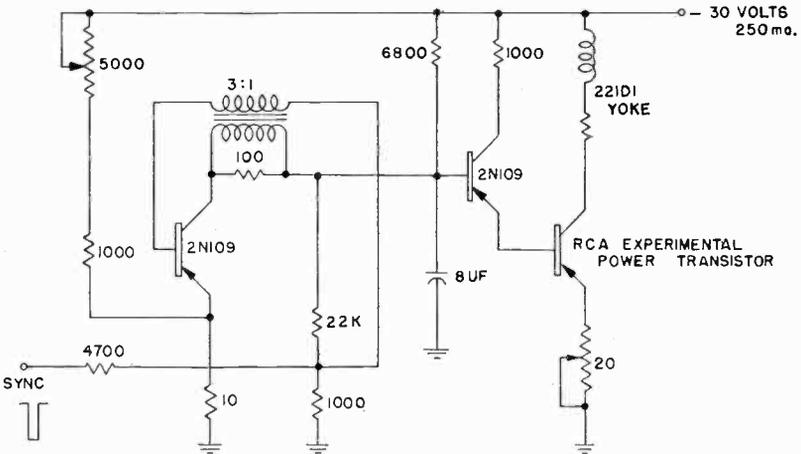
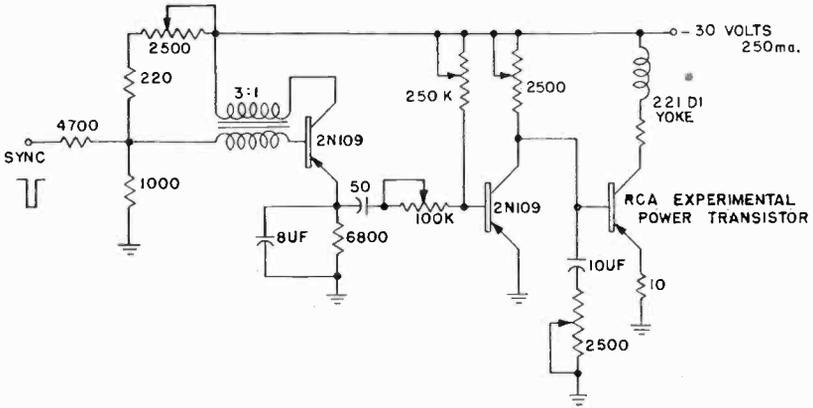


Fig. 16—Vertical-deflection systems. Common-emitter driver system (above). Common-collector driver system (below).

- 5. No peaking of the input sawtooth is needed.

APPENDIX—PERMANENT-MAGNET CENTERING METHODS

Raster shift caused by the d-c component in the yoke in a single-ended direct-drive connection is at least half the height of the raster and can be greater if part of the yoke current is not used during its sawtooth swing. Therefore, the amount of correction needed is equal to or greater than one-half the amount of peak-to-peak deflection. To produce this amount of constant flux in the yoke either by a shunt bucking voltage or by current through an auxiliary yoke winding is costly and impractical.

The magnitude of the correction required limits the correction flux to the space at or forward of the center of deflection (otherwise the beam will strike the kinescope neck). There are a number of ways of producing flux in this area:

1. Large diameter magnetic ring with transverse field, concentric with the kinescope axis and slightly ahead of the yoke. Mounting problems and the difficulty of achieving a straight flux pattern over the large area required make this method impractical.
2. Bar magnets inserted longitudinally in the yoke core. This method is capable of producing a straight field, but loss in deflection sensitivity is about 20 per cent.
3. Removal of a central cylinder of yoke core using the space for a transversely magnetized ring of permanent magnetic ferrite. Tests show that a fairly wide yoke section can be removed without affecting yoke sensitivity, but sufficient centering has not yet been achieved by this method.
4. Magnetic rings between the yoke core and kinescope neck. This method appears to be the most promising. Two rings made of permanent magnetic ferrite are used with 1.468 inches inside diameter, 1.620 inches outside diameter, and .3125 inch length. The required expansion of the yoke to provide clearance reduces sensitivity about 10 per cent (both horizontal and vertical). The raster has about $\frac{1}{2}$ inch of trapezoid caused by a slightly pincushioned field inside the rings. This corresponds to $2\frac{1}{2}$ per cent distortion or 1 per cent more than is allowed. A slight improvement in the magnetic rings or introduction of a compensating trapezoidal pattern in the yoke would bring the distortion within commercial tolerance.

AUTOMATIC-FREQUENCY CONTROL OF TELEVISION RECEIVERS USING JUNCTION DIODES

BY

WEN YUAN PAN AND OJARS RAMANIS

RCA Victor Television Division,
Camden, N. J.

Summary—The automatic-frequency-control (a-f-c) characteristics of various types of junction diodes used as frequency-correction devices in television receivers have been determined, including frequency sensitivities, oscillator loading effects and other essential properties. Practical limits of diode and oscillator-circuit parameters are proposed on the basis of receiver circuitries conforming to the present commercial practice.

A new developmental diode exhibited good a-f-c characteristics. Satisfactory results have been observed at both VHF and UHF television channels when this diode was used in a commercial television receiver in conjunction with a conventional amplifier discriminator.

INTRODUCTION

ANY practical automatic-frequency-control (a-f-c) system^{1,2} usually consists of a frequency discriminator which converts a frequency deviation into a suitable control voltage change and a frequency-correction device which couples a frequency-determining element into the oscillator circuit in accordance with the control voltage change. An investigation has been made to determine the frequency sensitivity, oscillator loading effect, and other characteristics of certain types of frequency-correction devices for the a-f-c operation in the frequency range up to 1000 megacycles.

Vacuum tubes and neon bulbs have been used as frequency-correction devices, but the relatively low frequency sensitivity, severe loading effect, and other inherent difficulties limited their usefulness for the intended purposes at very-high frequencies (VHF) and ultra-high frequencies (UHF). Devices employing mechanical movements gave rise to excessive time delay in the a-f-c action and therefore could not be used advantageously as a practical a-f-c means.

Junction diodes have been found to exhibit potentially good a-f-c characteristics at television-band frequencies. Acceptable a-f-c action could be secured with an improved junction diode³ when it was used in

¹ H. Borg, "Frequency Stabilization of V.M. Valves," *Wireless Engineer*, Vol. XXVI, p. 59, February, 1949.

² Charles Travis, "Automatic Frequency Control," *Proc. I.R.E.*, Vol. 23, p. 1125, October, 1935.

³ L. J. Giacoletto and J. O'Connell, "A Variable Capacitance Germanium Diode for UHF," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 221-238.

conjunction with a conventional amplifier-discriminator circuit which provided a voltage sensitivity of greater than 0.04 volt of control-voltage change per kilocycle of oscillator-frequency deviation.

FREQUENCY SENSITIVITY

A typical circuit of television local oscillators using a junction diode as the frequency-correction device is illustrated in Figure 1, where C_1 is the inter-electrode capacitance of the oscillator tube and L is the total circuit inductance. The equivalent circuit shown by Figure 2 replaces the junction diode with a diode capacitance C_d and a diode resistance r_d . In addition, the circuit capacitance C_c is equivalent to the series combination of C_1 and C_2 , and the original circuit losses are represented by the resistance r . The frequency sensitivity ϕ of such an a-f-c oscillator circuit, expressed in kilocycles oscillator-frequency correction per volt of control voltage change, can be determined readily as follows:

$$\begin{aligned}\phi &= \frac{\Delta f}{\Delta V} \\ &= \frac{\Delta f}{\Delta C} \frac{\Delta C}{\Delta V},\end{aligned}$$

where

$$\begin{aligned}C &= \frac{C_c C_d}{C_c + C_d}, \\ \frac{\Delta C}{\Delta V} &= \frac{C_c^2}{(C_c + C_d)^2} \frac{\Delta C_d}{\Delta V},\end{aligned}$$

and

$$\begin{aligned}f &= \frac{1}{2\pi\sqrt{LC}}, \\ \frac{\Delta f}{\Delta C} &= -\frac{1}{2} \frac{f}{C}.\end{aligned}$$

Therefore,

$$\phi = -\frac{1}{2} \frac{f}{C} \frac{C_c^2}{(C_c + C_d)^2} \frac{\Delta C_d}{\Delta V}. \quad (1)$$

The dependence of the diode capacitance C_d on the control voltage V , and also the change of C_d with respect to V are given by⁴

$$C_d = \frac{k_1}{\sqrt{V_o - V}}, \quad (2)$$

⁴ W. Shockley, "The Theory of p-n Junction in Semiconductor and p-n Junction Transistors," *Bell Sys. Tech. Jour.*, Vol. XXVIII, p. 435, July, 1949.

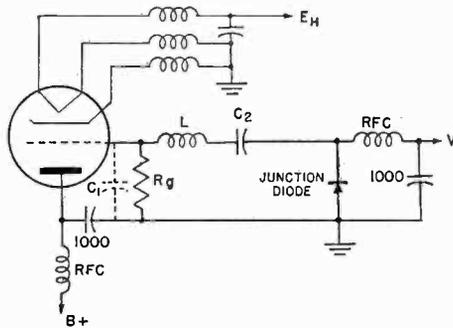


Fig. 1—Television oscillator using an automatic-frequency-control diode. where k_1 is a constant with a given junction diode, and V_0 is the diode contact potential which varies slightly with different types of diodes. When $V_0 \ll V$,

$$\frac{\Delta C_d}{\Delta V} = \frac{C_d}{2V}. \quad (3)$$

With the aid of Equation (3), the expression for the frequency sensitivity can be simplified still further,

$$\phi = -\frac{1}{4} \frac{f}{V} \frac{1}{1 + \frac{C_d}{C_c}}. \quad (4)$$

The frequency sensitivity ϕ of the a-f-c system is now expressed as a simple function of the oscillator frequency f , control voltage V , diode capacitance C_d , and the circuit capacitance C_c . On the assumption that $C_d \gg C_c$ which is usually valid in practical a-f-c oscillator circuits, the following conditions are established insofar as ϕ is concerned:

1. The frequency sensitivity is directly proportional to the local oscillator frequency, thus favoring the UHF television channels as compared to the VHF channels.

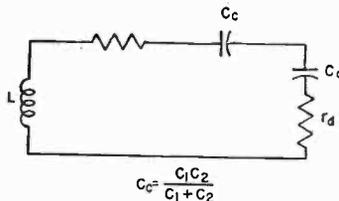


Fig. 2—Equivalent circuit.

2. A low value of circuit inductance L and a large C_c are preferred in the determination of the local oscillator frequency.

3. A small diode capacitance C_d is desired at a fixed value of the control voltage.

4. With a given junction diode, the frequency sensitivity of the a-f-c system is inversely proportional to the square root of the control voltage. This relationship is brought about by the fact that C_d also changes with V according to Equation (2).

Maximum frequency sensitivity, however, is only one of the essential requirements for a good a-f-c system. Another factor of equal importance is the loading effect of the diode on the local oscillator circuit.

LOADING EFFECT OF JUNCTION DIODE

The diode resistance r_d dissipates some available oscillator power when the diode is in the circuit. Consequently, additional input power to the oscillator tube is necessary in order to maintain the same amplitude of oscillation. If the plate voltage is kept constant, the plate current is increased and the grid current decreased, depending upon the magnitude of r_d and the equivalent coefficient of coupling between the diode and the oscillator circuit. The criterion of the diode loading effect is the change of the oscillator circuit Q by the presence of the diode as indicated in Figure 1.

Let Q = initial circuit Q of the local oscillator, assuming zero diode resistance, or $r_d = 0$,

Q' = resultant circuit Q of the local oscillator, assuming diode resistance r_d , and

$$Q_d = \frac{1}{\omega C_d r_d}$$

$$\frac{Q'}{Q} = \frac{1}{1 + \frac{r_d}{r}} \tag{5}$$

$$= \frac{1}{1 + \frac{Q}{Q_d} \frac{1}{1 + \frac{C_d}{C_c}}}$$

When $C_d \gg C_c$,

$$\frac{Q'}{Q} = \frac{1}{1 + \frac{QC_c}{Q_d C_d}} \quad (6)$$

Since the diode resistance r_d is practically independent of the control voltage over a wide operating range, the product of Q_d and C_d of a given junction diode is approximately a constant quantity. Therefore, with a particular local oscillator circuit of fixed values of Q and C_c , and using a given junction diode, the effect of the diode loading remains substantially unchanged even with a considerable variation of the operating control voltage.

The loading effect on the oscillator behavior was further investigated at 500 megacycles in a test oscillator "X" similar to the diagram shown in Figure 1, except that the junction diode was replaced with conventional circuit elements. The capacitance of the simulated diode was kept constant at C_d , and the equivalent Q_d was varied by substituting different values of r_d . Under these conditions the grid current plotted in Figure 3 gradually diminished with decreasing Q_d of the simulated diode. The manner in which the amplitude of oscillation and Q_d were related depended upon the ratio C_d/C_c at any given control voltage. It is also interesting to note that the oscillation ceased when $Q_d < 2$. The corresponding variations in plate currents are indicated in Figure 4. The rate of change of the d-c input power to the oscillator tube and the grid current is particularly pronounced when $Q_d C_d / C_c < 150$. According to Equation (6), this critical value of $Q_d C_d / C_c$ is directly proportional to the initial Q of the oscillator circuit.

Thus the factor $Q_d C_d / C_c$ must be large enough so as to neither increase the plate current excessively nor diminish the amplitude of oscillation materially. For television receivers where conventional miniature type tubes with low-loss tank circuits are generally used as all-channel oscillators, the design criterion of the a-f-c system is such that the quantity

$$\frac{C_d}{C_c} Q_d > 200$$

at any operating control voltage. Under these conditions, the d-c input power to the oscillator tube is not substantially increased and the amplitude of oscillation may still be sufficient to perform satisfactorily.

To minimize the diode loading effect, large values of Q_d and C_d must be employed. Unfortunately, the requirement for a large diode

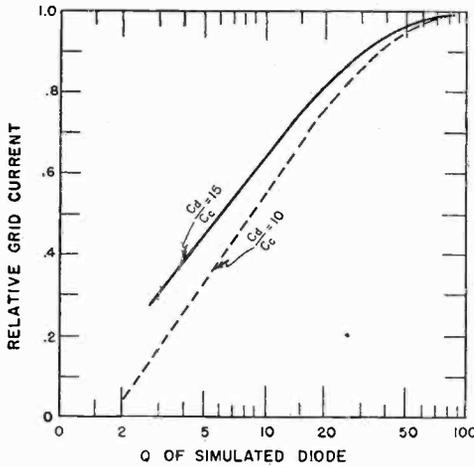


Fig. 3—Effect of added load on oscillation.

capacitance to secure minimum loading contradicts the condition for maximum frequency sensitivity. Therefore, the over-all characteristic of an a-f-c system has to be determined by taking into account both the frequency sensitivity and the diode loading.

OVER-ALL CHARACTERISTIC

The over-all characteristic μ of an a-f-c system using a junction diode is defined arbitrarily as the product of the frequency sensitivity and the resultant Q' of the local oscillator circuit, or

$$\begin{aligned} \mu &= \phi Q' \\ &= \frac{f}{4V} \frac{Q_d}{1 + \frac{r}{r_d}} \end{aligned} \tag{7}$$

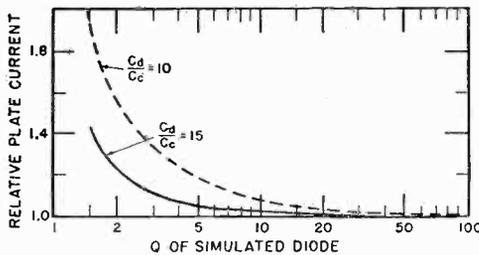


Fig. 4—Effect of added load on input power.

Case 1: $r_d \gg r$,

$$\mu = -\frac{f}{4V} Q_d. \quad (8)$$

At UHF, Equation (8) usually holds true in most practical oscillator circuits, and, therefore, discloses the important design considerations for the a-f-c diode and the control voltage source; namely, the junction diode must possess a high value for Q_d at the operating control voltage, and the control voltage must be kept as low as feasible. The over-all characteristic of the a-f-c system is again inversely proportional to the square root of V .

Case 2: $r_d \ll r$,

$$\mu = -\frac{1}{8\pi r V} \frac{1}{C_d}. \quad (9)$$

At lower frequencies, the oscillator circuit resistance r is relatively high. When such an oscillator circuit is used with a high Q_d diode, the diode loading effect is so small that μ is primarily determined by the frequency sensitivity ϕ . Therefore, a small diode capacitance C_d is desired.

In a practical a-f-c oscillator, the circuit capacitance C_c is made substantially smaller than the diode capacitance C_d in order (a) to minimize the amplitude of the oscillator voltage existing across the diode, thus preventing the rectified direct current from affecting the a-f-c action, and (b) to limit the loading effect of the diode to a predetermined level.

According to the previous discussions, the ratio C_d/C_c is limited by two boundary conditions for satisfactory operation of the frequency-correction device and the local oscillator circuit. The ratio must be at least 10 under all operating conditions; therefore, the lower limit of the diode capacitance is determined by the smallest circuit capacitance C_c that can be made practical and realizable. Conversely, the ratio C_d/C_c must not be made too large so that a sufficient frequency sensitivity of the a-f-c system can be secured at the lowest oscillator frequency.

Based on circuitries commonly employed in commercial television receivers, the arbitrary limits of $10 < C_d < 40$ micromicrofarads at the highest operating control voltage are considered to be pertinent for VHF and UHF all-channel television receivers.

TELEVISION RECEIVER PERFORMANCE

Selection of Junction Diodes

The a-f-c characteristics of various types of junction diodes have

been evaluated. The relative results at 500 megacycles with three fixed values of control voltages in volts are tabulated as follows:

Diodes	r_a (ohms)	$V = -2$		$V = -4$		$V = -10$		$C_a Q_a$
		$C_a(\mu\mu f)$	Q_a	$C_a(\mu\mu f)$	Q_a	$C_a(\mu\mu f)$	Q_a	
601	30	3.5	3.0	2.6	4.1	1.7	6.2	10.5
1N205	10	20.0	1.6	14.5	2.2	9.5	3.4	32.0
1N208	21	8.4	1.8	6.1	2.5	4.0	3.7	15.0
6002	26	8.1	1.5	6.0	2.0	3.9	3.2	12.2
6001-C	15	42.0	.5	29.5	.7	19.3	1.1	20.9
1N91	43	18.5	.4	12.3	.6	7.6	1.0	7.5
1N93	24	33.0	.4	22.0	.6	13.5	1.0	13.3
experimental	1	43.5	7.3	31.8	10.0	20.0	16.0	318.7

The experimental diode⁵ exhibited substantially better a-f-c characteristics than the commercial diodes tested. This experimental diode was used as a frequency-correction device at 500 megacycles in the test oscillator "X" which employed a circuit capacitance, C_o , of 1.5 micro-microfarads and an initial Q of 200. The frequency sensitivity and the diode loading effect are indicated in Figure 5 as a function of the control voltage. The loading effect Q'/Q was approximately 0.5 and remained practically constant, while the frequency sensitivity varied considerably with the control voltage.

In order to use the best of the commercially available diodes at 500 megacycles, it is necessary to reduce the loading by increasing the L/C_o ratio of the oscillator circuit. To bring the Q'/Q level to approximately 0.5, it is necessary to use a C_o of 0.14 micromicrofarad. Such a small capacitance is not easily achieved in practice; however, the theoretical value of performance is shown dotted in Figure 5 for the type 1N205. It will be noted that, at VHF (channels 2 to 13) some of the better commercial diodes give satisfactory performance for a-f-c, but with reduced frequency sensitivities compared with the experimental diode.

Instantaneous Oscillator Frequency

The instantaneous frequency characteristics of a typical uncontrolled television local oscillator are given by⁵

$$\Delta f = \alpha(1 - e^{-\beta t}), \quad (10)$$

where

$$\begin{aligned} \Delta f &= \text{frequency deviation of an uncontrolled oscillator,} \\ \alpha &= \text{maximum frequency deviation,} \end{aligned}$$

⁵ W. Y. Pan, "Frequency Characteristics of Local Oscillators," *RCA Review*, Vol. XVI, p. 379, September, 1955.

β = a constant determining the rate of change of oscillator frequency with respect to time,

t = time.

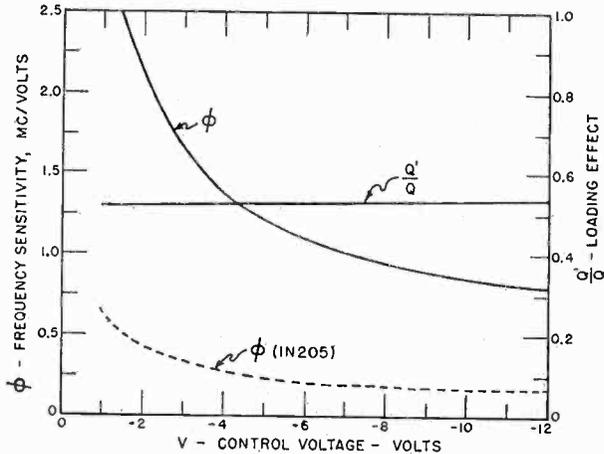


Fig. 5—Frequency sensitivity and loading effect of the experimental junction diode in a typical UHF local oscillator circuit.

This is shown dotted in Figure 6 where the constants α and β are graphically defined. If an a-f-c system is used in conjunction with this local oscillator, the resultant instantaneous frequency characteristics are modified. The δf curve for the residual oscillator frequency characteristics of a controlled oscillator differs from the Δf curve for the uncontrolled oscillator by an amount representing the a-f-c action.

$$\delta f = \alpha' (1 - e^{-\beta' t}). \quad (11)$$

The constant β' is approximately equal to β of Equation (10) but α' is substantially smaller than α unless the a-f-c system is not effective. The ratio $\Delta f/\delta f$ at any instant is determined by the voltage sensitivity v of the amplifier discriminator, and the frequency sensitivity ϕ of the local oscillator circuit including the junction diode, according to the well-known expression

$$\frac{\Delta f}{\delta f} = 1 + v\phi. \quad (12)$$

With a given a-f-c system, the instantaneous δf decreases when Δf is reduced by means other than the a-f-c system. Therefore, the conventional frequency compensations using temperature-sensitive ele-

ments are still recommended in a-f-c local oscillators in order to obtain a minimum residual frequency deviation.

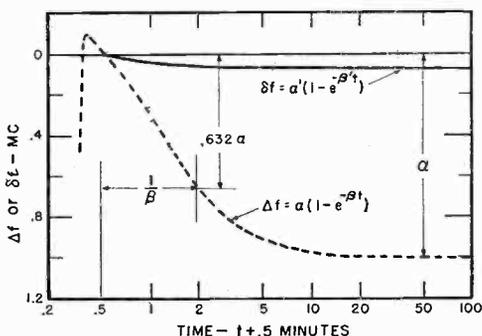


Fig. 6—Frequency characteristics of oscillators.

Observed Results

A commercial television tuner has been redesigned to incorporate the experimental junction diode as the frequency-correction device, and a control voltage range of $-8 < V < 0$ volts was secured from a conventional amplifier-discriminator circuit. The basic oscillator circuit consisted of a 6AF4 oscillator tube, an inductance, L , and a C_o of 1.5 micromicrofarads. Before the diode was added to the tuner, the oscillator tube drew an initial plate current of 12 milliamperes at 100 volts and the corresponding grid current was 0.8 milliamperes through a resistor of 8200 ohms, at a fixed frequency corresponding to channel 50.

When the diode was operative the amplitude of oscillation was reduced. By maintaining constant plate voltage, the grid current fell from 0.8 to 0.55 milliamperes and the plate current increased from 12 to 13 milliamperes. It was observed that the exact control voltage had very little influence on the oscillator strength.

At channel 50 ($f = 733$ megacycles) the dotted curve of Figure 7 indicated a frequency sensitivity at the operating point marked "p" of approximately 3,000 kilocycles per volt when the experimental diode was operative. The corresponding voltage sensitivity of the amplifier-discriminator was 0.04 volt per kilocycle. According to Equation (12), the a-f-c action $\Delta f/\delta f = 121$, which reduces a 1.2 megacycle frequency deviation of an uncontrolled oscillator to 10 kilocycles for a controlled oscillator. The experimental residual frequency deviation showed a maximum value of 17 kilocycles, probably due to the fact that the exact point of operation deviated slightly from the point "p."

At channel 13 ($f = 257$ megacycles) the corresponding frequency sensitivity using the same C_c was approximately 1000 kilocycles per volt, or $\Delta f/\delta f = 41$.

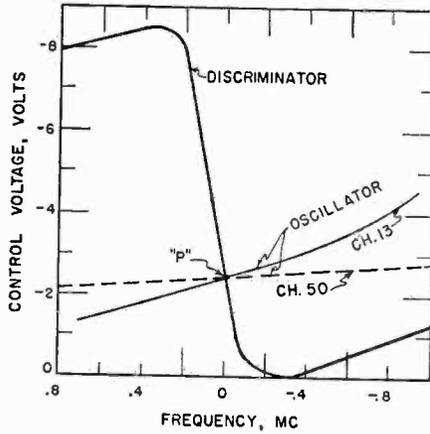


Fig. 7—Discriminator and automatic-frequency-control oscillator characteristics.

CONCLUSION

The development of the improved experimental junction diode provided a useful frequency-correction device for a-f-c application in television receivers. Satisfactory results have been observed at both VHF and UHF channels when this diode was used with a conventional amplifier-discriminator circuit.

JUNCTION TRANSISTOR SWITCHING CHARACTERISTICS

BY

D. E. DEUTCH

RCA Defense Electronic Products,
Camden, N. J.

Summary—The important switching characteristics of junction transistors are presented. Emphasis is placed upon the practical aspects of the transistor both as a switching device and in switching operation. The paper is divided into three main parts: (1) Transistor d-c properties and their voltage-current capabilities, as well as their major restrictions, are described with respect to the input and output characteristics; (2) Transistor a-c properties are then considered and expressions for transient response and switching operation are presented, including the effects of storage encountered in saturation operation; (3) Several typical switching-circuit applications are presented, which illustrate some of the useful switching properties of transistors.

INTRODUCTION

SWITCHING or pulse techniques are important in many fields of electronics, notably in radar, pulse-code communications, and digital computers. In these, and many other switching applications, transistors are becoming increasingly useful and important. This is due, not only to their basic advantages of low power consumption and small size, but also to their particularly useful electrical properties for switching operation. It is the purpose of this paper to present the important characteristics of junction transistors for switching operation and to illustrate their advantages and limitations in this service.

JUNCTION-TRANSISTOR SWITCHING OPERATION

In switching operation, the transistor is biased normally in either a high impedance cutoff (OFF) state or in a low impedance saturation (ON) state.* An input signal drives, or triggers in the case of a bistable circuit, the transistor from one of these states to the other. The d-c properties of the cutoff and saturation states are important from the standpoint of transistor dissipation and subsequent thermal conditions affecting the circuit operation. They also limit voltage—

* The operating states are henceforth defined as: Cutoff (OFF) State — The low-current operating region where the input current is zero or in a reverse direction; Saturation (ON) State — The low-voltage state where the collector junction is forward biased, and the a-c collector resistance is very small.

current capabilities and influence circuit time constants in monostable and astable circuits. The a-c properties of the transistor in the transition region between these states determines the circuit transient performance and consequently the circuit switching speed. Both the d-c and the a-c transistor properties are important to over-all switching performance and the design of switching circuits.

In this section, the properties of the transistor cutoff and saturation states and the general aspects of switching operation are described with respect to the input and output characteristic of the transistor. In the following section, the transient behavior of the transistor is presented.

The Junction Transistor as a Switching Device

Before examining the actual properties of the transistor, it is useful to consider the important features provided by the junction transistor for switching operation. These are illustrated in Figure 1. The common-base stage is essentially a current-controlled series switch, i.e., the input current, I_E , constitutes the output current, I_C , in addition to its switching-control function. The common-emitter stage, however, is a current-controlled shunt-switching device where the input current, I_B , acts only to control the output current, I_C . The OFF state for these stages is shown in Figure 1B, where the open-switch conditions represent the high-cutoff collector and emitter junction resistances. The ON state, is shown in Figure 1C, where the closed switching conditions represent the small junction impedances in the saturation state.

Although these are idealized conditions, the fact that they may actually be approached quite closely in practice with junction transistors is indicative of the value of these transistors in switching operation. The actual switching properties of the junction transistor will now be examined.

Common-Base Switching Operation

The common-base (CB) stage is important in many switching applications. It has the best transient performance for pulse amplification and is useful in gating circuits because of its low input impedance and the facility of base-bias control of input current.

Transistor collector and input characteristics for CB operation are shown in Figure 2. The cutoff state is at A on the $I_E = 0$ line of the collector characteristics. An equivalent circuit for the cutoff condition is shown in Figure 3A. The collector junction impedance consists of the junction resistance, r_{co} (essentially a leakage resistance), which is paralleled by the junction capacity, C_c , and a current generator, I_{ct} , the thermal-equilibrium current of the junction. The total collector

cutoff current, I_{c0} , includes I_{ct} , which remains essentially constant with collector voltage (above 1-2 volt), and the leakage component, I_{cl} .

The emitter junction, with no external emitter bias, is forward-biased slightly by the voltage developed by I_{c0} across r_{d0} . With reverse emitter bias, the emitter junction reverts to the high-impedance form

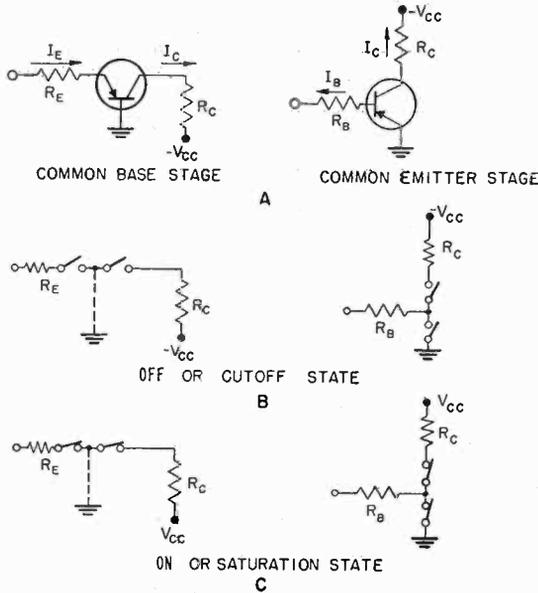


Fig. 1—The junction transistor as a switching device.

of the collector junction, including junction capacitance, C_e , and thermal and leakage currents, I_{ct} and I_{cl} .^{*} Both emitter and collector cutoff currents are then supplied by the base current, I_{b0} .^{**}

High voltage CB operation is limited generally by a surface breakdown condition, although in some cases punch-through may occur.⁴

* Forward bias is defined as the bias condition such as to establish the low impedance or high current state of a transistor junction or diode. Reverse bias is defined as the bias condition such as to establish the high impedance or low current state of a transistor junction or a diode.

** The exact cutoff current conditions are defined by the diode currents of both emitter and collector junctions together with a mutual current component.¹⁻³

¹ J. J. Ebers and J. L. Moll, "Large Signal Behavior of Junction Transistors," *Proc. I.R.E.*, Vol. 42, pp. 1761-1772, December, 1954.

² J. J. Moll, "Large Signal Transient Responses of Junction Transistors," *Proc. I.R.E.*, Vol. 42, pp. 1773-1784, December, 1954.

³ L. J. Giacoletto, "Power Transistors for Audio Output Circuits," *Electronics*, Vol. 42, pp. 144-148, January, 1954.

⁴ H. Schenkel and H. Stalz, "Voltage Punch-Through and Avalanche Breakdown and Their Effect on the Maximum Operating Voltages of Junction Transistors," *Proceedings of National Electronics Conference*, Vol. 10, 1954.

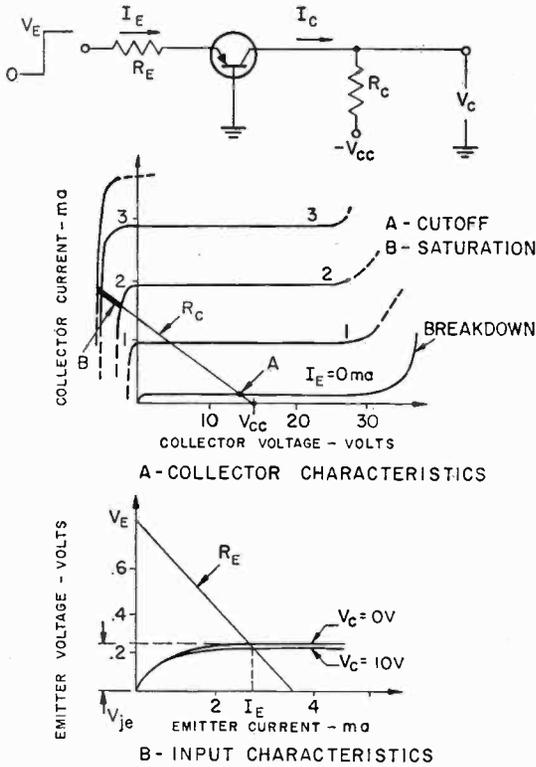


Fig. 2—Transistor characteristics, common-base operation.

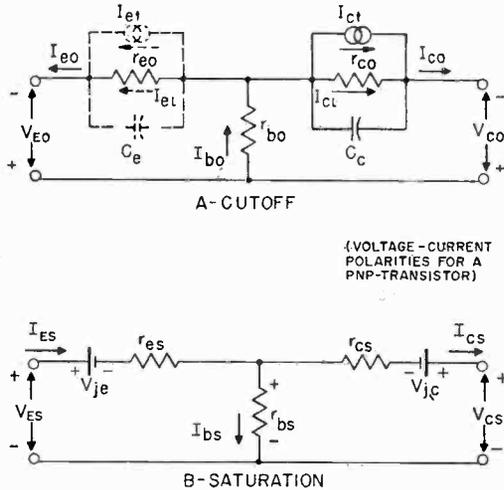


Fig. 3—Equivalent circuits for the junction transistor cutoff and saturation states.

The breakdown condition in itself is seldom destructive, but the high collector dissipation which may result is harmful.

The saturation state shown at B in Figure 2A, is the family of collector curves of low collector a-c resistance resulting from the forward-biased collector junction. An equivalent circuit for this operating state is shown in Figure 3B. The collector and emitter impedances are the forward resistances of junction diodes, which have a-c resistance, r_{es} and r_{cs} , given by

$$r_{es} = \frac{kT}{qI_{ES}}, \quad r_{cs} = \frac{kT}{qI_{CS}},$$

where T = temperature in degrees Kelvin,

k = Boltzmann's constant,

q = electronic charge,

I_{ES} and I_{CS} = emitter and collector currents, respectively, in milliamperes;

$$\frac{kT}{q} \cong 26 \text{ millivolts at } 25^\circ\text{C.}$$

Associated with each junction are the diode potentials, v_{je} and v_{jc} . The junction diodes are poled, however, such that the total emitter-collector voltage is extremely small, 10-50 millivolts.* The voltage drop produced by I_{BS} across r_{bs} raises both emitter and collector potentials above ground, thus causing the voltage reversal in the saturation region, indicated on the collector characteristics.

A typical input characteristic for CB operation is shown in Figure 2B. It is essentially the emitter-diode characteristic in series with the base resistance. Because of base-emitter current-gain however, the effective base resistance is reduced to $r_b(1-\alpha)$ and the low-frequency input impedance is $r_{in} = r_e + r_b(1-\alpha)$ **.

The voltage ordinate of the input characteristic will decrease somewhat with collector voltage. This effect, however, is usually sufficiently small to be neglected in CB operation. Except for very-low-current switching operation, the input characteristic may be approximated as an ideal diode with voltage, v_{je} , and zero a-c resistance as indicated by the dashed line.

The switching current requirements are determined by the tran-

* In grown junction transistors, the collector-emitter saturation voltage may be much greater because of extrinsic series resistance (20-200 ohms) in the collector and emitter.

** α , and β as used later, are the low-frequency ratios of a-c short circuit collector current to a-c emitter current and base current respectively.

sistor d-c current amplification factor, α_0 .* The input current required to switch the CB stage from cutoff to saturation is

$$I_B = \frac{V_{CC}}{\alpha_0 R_C - (1 - \alpha_0) r_b}.$$

The input voltage-current relationship is determined from the input characteristics such that

$$I_B = \frac{V_B - v_{je}}{R_B}.$$

$v_{je} < .25$ volt for $I_B < 10$ milliamperes in most alloy-junction transistors.

Common-Emitter Switching Operation

The common emitter (*CE*) stage, because of its current gain, is the most useful configuration for triggered switching stages such as bistable and monostable circuits. Although limited in transient response, the stage is useful for high-current switching and in gating circuits.

The cutoff and saturation states in *CE* operation are shown at A and B on the collector characteristics of Figure 4A. These operating states are defined by the corresponding equivalent circuits of Figure 3 with the base as input terminal.

In saturation, the output is virtually shorted to the emitter potential through the extremely low collector-emitter impedance (1-5 ohms). This is a valuable feature for gating and clamp circuits, and allows high-current switching at very low transistor dissipation. Actually the base resistance may become the predominant dissipative element in high-current saturation operation.

The cutoff state in *CE* operation requires special consideration. With the base open, $I_B = 0$, the cutoff collector current, I_{d0} , is the emitter current generated from the feedback of the collector-base cutoff current, I_{c0} , across the emitter junction. It includes an amplified thermal equilibrium component, I_{dt} , and leakage component, I_{dl} , $I_{d0} \approx \beta_0 I_{c0}$. It is desirable to reduce this large cutoff current for lower cutoff dissipation, larger collector voltage range, and improved thermal and breakdown characteristics.

The *CE* cutoff current is reduced considerably by using a low base-emitter resistance. This self-bias action results essentially from a division of the collector-base cutoff current I_{c0} between base and

* α_0 and β_0 are defined as current amplification factors for *CB* and *CE* operation determined by the ratios of d-c collector current to d-c emitter and base currents, respectively.

emitter paths thus reducing the component amplified at the emitter. For values of base resistance of 10,000 ohms or less, the total cutoff current is normally reduced to $5I_{c0}$ or less. The use of emitter resistance accentuates this self-bias action and permits larger base resistance for the same reduction in cutoff current. When the base is shorted to the emitter or reverse emitter-base bias is employed, cutoff conditions revert to those for the *CB* case.

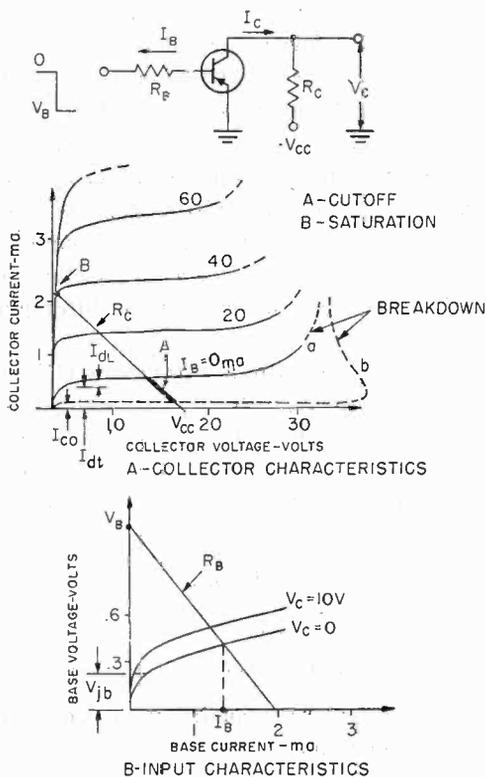


Fig. 4—Transistor characteristics, common-emitter operation.

High voltage *CE* operation is limited primarily by avalanche breakdown. With high base impedance, breakdown normally occurs as a gradual increase in cutoff current (curve "a" in the collector characteristics of Figure 4A). With low base impedance or reverse bias, breakdown may include a negative resistance region (curve b).⁵ The

⁵ M. C. Kidd, W. Hasenberg and W. M. Webster, "Delayed Collector Conduction, a New Effect in Junction Transistors," *RCA Review*, Vol. XVI, pp. 16-33, March, 1955.

breakdown voltages in *CE* operation are normally lower than those in *CB* operation because of amplification of the current initiated by the breakdown process.

High current operation is limited primarily by reduction in d-c current amplification factor, β_0 , at high collector currents. This β_0 fall-off may be as much as 50 per cent over a 20 milliamperere collector-current range. In some cases, however, notably high-frequency alloy-junction transistors, β_0 has been observed to increase with collector current up to 100 milliamperes. This is also true of some grown junction transistors over a lower current range.

Typical input characteristics for *CE* operation are shown in Figure 4B. In this case (with base current as reference) the a-c input impedance is the series base resistance plus the emitter a-c resistance increased by the degenerative emitter feedback: $r_{in} = r_b + r_e/(1 - \alpha)$.

With $V_C = 0$, the input characteristic includes the forward diode characteristics of both emitter and collector junctions. For $V_C > 0$, only the emitter diode is effective. Hence, the voltage ordinates of the input characteristics of these two cases may differ greatly as shown in Figure 4B.

For switching operation with a saturation ON state, the input current is determined by the $V_C = 0$ curve which may be approximated as a junction diode potential, V_{jb} , in series with the base resistance, r_b .

Current requirements in *CE* switching operation are determined by the d-c current amplification factor, β_0 . The input current to switch from cutoff to saturation is

$$I_B = \frac{V_{CC} - I'_{CO}R_C}{\beta_0 R_C},$$

where

$$I'_{CO} = \text{cutoff collector current.}$$

The input voltage-current conditions are determined from the input characteristics such that

$$I_B = \frac{V_B - V_{jb}}{R_B + r_b}.$$

$V_{jb} = 0.3$ to 0.4 volt for $I_B < 5$ milliamperes in most alloy-junction transistors.

Common Collector Switching Operation

The transistor cutoff and saturation characteristics in common-collector (*CC*) operation are the same as those for the *CE* case. High d-c stabilization is provided by the degenerative effect of the emitter load.

The important circuit properties of the *CC* stage are listed in Figure 5. The properties of high input and low output impedance are particularly useful in buffer or impedance-matching applications. All properties, however, are affected strongly by the magnitude and variation in d-c current amplification factor, α_0 , and the ratio of load to source resistance. This latter consideration is also important in the stage transient response, as shown later.

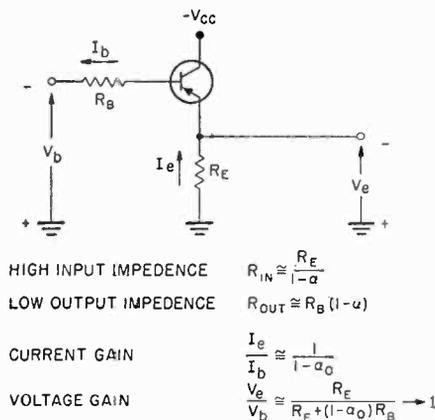


Fig. 5—Switching properties for common-collector operation.

Temperature Effects

The effects of temperature on transistor switching operating are important; three major effects occur. (1) I_{e0} and I_{c0} double (approximately) for a 10°C temperature rise. This is important from the standpoint of circuit dissipation and d-c stability in trigger circuits at high temperatures. (2) At low temperatures β_0 decreases, particularly in grown-junction transistors, dropping perhaps 50 per cent from 25°C to -50°C . This primarily influences d-c circuit stability in trigger circuits at low temperatures. (3) The emitter junction voltage at constant current varies inversely with temperature at the rate of approximately 2 millivolts per degree centigrade.⁶ This effect concerns input requirements for the switching stage, and to some degree influences circuit stability in direct-coupled-circuit operation.

Normally, the power dissipation in a switching circuit is taken as the average power as determined by the circuit duty factor and the peak power dissipation. In very-high-current switching operation,

⁶ H. C. Lin and A. A. Barco, "Temperature Effects in Circuits Using Junction Transistors," *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 369-402.

however, the dissipation during the switching transition times should be considered. The transistor thermal time constant is on the order of 1-2 milliseconds. Thus, for short-pulse, low-duty-cycle operation, average power dissipation may be very small, even for very high peak power operation. The collector-junction temperature rise is normally less than .5°C per milliwatt for low-power transistor types, and this figure may be used to determine the temperature conditions for a given average dissipation in switching operation.

JUNCTION-TRANSISTOR TRANSIENT AND SWITCHING PERFORMANCE

In the operating range between cutoff and saturation, the major limitations upon the transient response of junction transistors are imposed by the diffusion time of minority carriers through the base region and by the junction transition capacities. In this section, these factors are considered in formulating transient response and switching speed expressions for *CB*, *CE*, and *CC* operation.

Factors Determining Transistor Transient Response

A fundamental limitation upon transistor transient performance is the diffusion time of minority carriers through the base region. This action may be expressed in terms of a diffusion time constant, T_{ce} such that the complex collector-emitter current amplification factor, α_{ce} , becomes⁷

$$\alpha_{ce} = \frac{\alpha}{1 + j\omega T_{ce}} \quad (1)$$

The diffusion time constant may be related to physical properties of the transistor. Practically, however, it is determined by $T_{ce} = 1/(2\pi f_0)$ where f_0 is the frequency at which α_{ce} is 3 decibels down from α . The time delay in minority carrier diffusion also produces an effective diffusion capacity, across the emitter junction.⁸ The magnitude of this capacity is a function of emitter current and may be quite large (.001-.01 microfarad) at small emitter currents. The resulting emitter time constant, however, is independent of emitter current and is equal to the diffusion time constant, T_{ce} .

The collector-junction transition capacitance, C_{ce} , is also a major limitation in transient response. This capacitance, which stems from

⁷ R. L. Pritchard, "Frequency Variations of Junction-Transistor Parameters," *Proc. I.R.E.*, Vol. 42, pp. 786-799, May, 1954.

⁸ C. W. Mueller and J. I. Pankove, "A P-N-P Triode Alloy Junction Transistor for Radio-Frequency Amplification," *RCA Review*, Vol. XIV, pp. 586-598, December, 1953; also, *Transistors I*, RCA Laboratories, Princeton, N. J., 1956, pp. 189-201.

the charge separation across the junction, is related to the junction voltage and geometry by⁸

$$C_c \sim \frac{1}{V_{CB}^n}, \quad (2)$$

where $n = 1/2 - 1/3$, depending on the junction geometry; for alloy-junction transistors, $n = 1/2$. For small-signal operation, the value for C_c is determined from the collector-junction voltage at the d-c operating point. However, in switching operation involving a wide junction voltage range, an effective large-signal value of C_c must be considered. The average value of C_c integrated from $V_{CB} = 0$ to V_{CO} is $2C_{CO}$ where C_{CO} is the value of C_c at $V_{CB} = V_{CO}$ (for $n = 1/2$ in Equation (2)). When switching from a cutoff voltage V_{CO} to saturation, however, the effective C_c may be approximated as its cutoff value, C_{CO} . This is based on the fact that the increase in C_c is relatively small over much of the initial switching-voltage range from cutoff to saturation.

When switching from saturation to a cutoff voltage, V_{CO} , the effective C_c may be considered as $2C_{CO}$. This average value takes into account the large initial values of C_c encountered during the transition from saturation to cutoff. These large signal values represent only approximations to the actual effect of C_c but they have been correlated closely with measured results.

This effective collector capacitance together with the a-c collector resistance constitute a collector time constant, $T_c = r_c C_c$. The collector impedance is thus

$$Z_c = \frac{r_c}{1 + j\omega T_c}. \quad (3)$$

Other factors contribute to the transistor transient response, particularly the emitter junction resistance at low emitter currents, and the variation in α_0 and β_0 over the current switching range.* Where these factors are pertinent, average values over the switching range may be used to a good approximation.

A "T" equivalent circuit based on these considerations is shown in Figure 6. The $\alpha_{ce} I_e$ generator and the collector impedance are represented in terms of their respective time constants. The emitter junction is considered an ideal diode of zero a-c resistance. All parameters,

* In the following material, the d-c amplification factors α_0 and β_0 will be employed in transistor transient and switching performance.

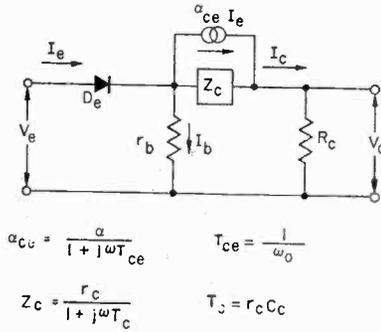


Fig. 6—Transistor T-equivalent circuit for switching operation.

except collector capacity, C_c , and therefore T_c , are approximated as linear and constant.

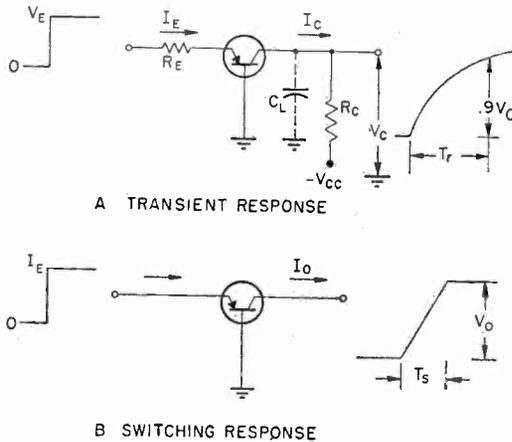


Fig. 7—Transient response, common-base operation.

Common-Base Operation

Based on the equivalent circuit of Figure 6, the collector current response of the *CB* stage in Figure 7A to a step of emitter current of magnitude I_E , as determined by Laplace analysis techniques,⁹ is given by

$$I_c = \alpha_0 I_E \left[1 + A e^{-\frac{t}{a}} - (1 + A) e^{-\frac{t}{b}} \right] \quad (4)$$

where

$$a = \frac{T_{ce} T_\sigma}{T_{ce} + T_0}$$

⁹ Gardner and Barnes, *Transients in Linear Systems*, John Wiley and Sons, Inc., New York, N. Y., 1947.

$$b = D_b (T_{ce} + T_0),$$

$$D_b = \frac{R_E + r_b}{R_E + r_b (1 - \alpha_0)} \cong \frac{R_E + r_b}{R_E},$$

$$A = \frac{a}{b - a},$$

$$T_0 = \frac{R_C}{r_c} T_C = R_C C_T,$$

$$C_T = C_c + C_L.$$

The coefficient D_b , in the time constant b , is defined here as an "input coefficient." It relates the degenerative action of emitter resistance and the regenerative effect of base resistance on the circuit transient response. With large emitter resistance, $D_b = 1$. Normally $a \ll b$ and $A \ll 1$ so that Equation (4) may be expressed as:

$$I_c = \alpha_0 I_E \left[1 - e^{-\frac{1}{b} (t-a)} \right], \quad (5)$$

where the time constant, " a ", is now a time delay in the collector current response.

Using this relation, the calculated collector current response (normalized to $\alpha_0 I_E$) is compared to the measured response in Figure 10A for a transistor with the following parameter and circuit values:

$$\alpha_0 = .978,$$

$$C_c = 45 \mu\text{f} (V_{CB} = 6\text{v}),$$

$$r_b = 280 \Omega,$$

$$f_0 = 600 \text{kc},$$

$$R_E = 15,000 \Omega,$$

$$R_C = 27,000 \Omega,$$

from which

$$a = .23 \mu\text{sec},$$

$$b = 1.18 \mu\text{sec}.$$

The collector voltage range for the measured response is from 6 volts to 1 volt. Agreement between calculated and measured responses is very close.

From Equation (5), the collector-current rise time, T_r , and fall time, T_f , are given by* $T_r, T_f = 2.3b + a$. Where low collector voltage is involved, the effective values of C_o as described previously should be used in the time constants.

In switching operation, referring to Figure 7B, the transistor is driven between cutoff and saturation or an ON clamped state. The OFF-ON switching time is then determined by the transistor transient response together with the input drive.** The ON-OFF switching time, however, includes also the storage effects encountered in saturation operation. This switching operation and the switching times involved will be treated later.

The OFF-ON CB switching time, T_s , for a step of emitter current of magnitude I_B is determined from Equation (5) as

$$T_s = b \ln \left[\frac{1}{1 - \frac{I_{CS}}{\alpha_0 I_B}} \right] + a \quad (6)$$

where I_{CS} is the ON collector current.

For the condition, $T_s \ll b$, the collector current response is given approximately by

$$I_o = \alpha_0 I_B \left[\frac{1}{b} (t - a) \right] \quad (7)$$

For a specified output voltage, V_o , and switching time, T_s , the required input current is

$$I_B = \frac{V_o}{\alpha_0 R_G} \frac{D_b}{T_s} (T_{ce} + T_0) \quad (8)$$

In terms of transistor and load capacitance, transistor cutoff frequency, and the output voltage current conditions,

$$I_B = \frac{D_b}{\alpha_0 T_s} \left[V_0 C_T + \frac{I_Q}{\omega_0} \right] \quad (9)$$

The emitter voltage for the current requirement from Equation (9) is determined from the input characteristics as

* Rise and fall times, T_r and T_f , unless otherwise noted, are the transition times from 0 to 90 per cent of final current or voltage amplitudes.

** Switching time is defined as the transition time for current or voltage to change between its minimum and maximum stable values.

$$V_E = I_E R_E + V_{je}, \quad (10)$$

where V_{je} is the idealized emitter diode voltage.

Common-Emitter Operation

The transient response of the *CE* stage, Figure 8A, using the equivalent circuit of Figure 6 with the base as input and the emitter common, is given by the same relation as for the *CB* case, Equation (4), except for the current amplification factor and the time constant, b . The collector current response to a step of base current of magnitude, I_B , expressed in time-delay form, is

$$I_c = \beta_0 I_B \left[1 - e^{-\frac{1}{b'}(t-a)} \right]$$

where $b' = D_e [T_{ce} + T_0], \quad (11)$

$$D_e = \frac{r_e + R'_B}{r_e + R'_B (1 - \alpha_0)} \cong \beta_0,$$

$$R'_B = R_B + r_b.$$

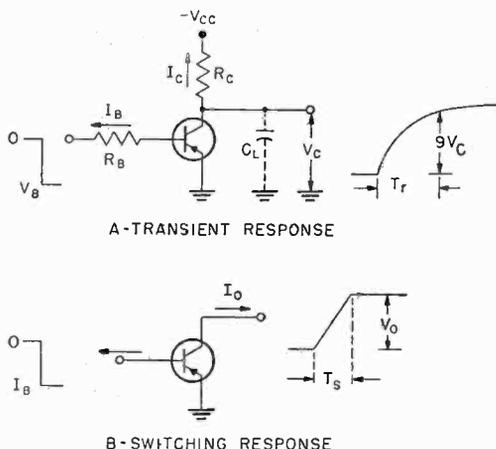


Fig. 8—Transient response, common-emitter operation.

The "input coefficient," D_e , is given here including the emitter junction resistance, r_e , (departing for the moment from the idealized emitter diode representation) to indicate the influence of this parameter at low-current operation (where r_e is large) and the effect of external emitter resistance. For the grounded emitter case, $D_e = 1/(1 - \alpha_0) \cong \beta_0$. The time constant, b' , is thus the time constant b in *CB* operation magnified by the d-c base-collector current gain. The time delay, a ,

is the same in both cases, but is of much less significance here, because of the large value of b' .

The calculated collector current response, using Equation (11), is compared with the measured response in Figure 10B for the same transistor and operating range used for the *CB* example (with $R_C = 10,000$ ohms and $R_B = 15,000$ ohms). Again close agreement is obtained between the two cases.

The 0-90 per cent collector-current rise and fall times, T_r and T_f from Equation (11), are $T_r, T_f = 2.3 b'$ (ignoring a).

In *CE* switching operation, referring to Figure 8B, the OFF-ON switching time may be reduced considerably from the 0-90 per cent video response by using large input drive. This improvement in switching time, however, may be accompanied by substantial reduction in circuit current gain and large storage delays in switching from saturation to cutoff. The *CE* switching time, T_s , for step of base current of magnitude I_B is obtained from Equation (11), and is given by

$$T_s = b' \ln \left[\frac{1}{1 - \frac{I_{CS}}{\beta_0 I_B}} \right] \quad (12)$$

where I_{CS} is the ON collector current.

For the condition, $T_s \ll b'$, the *CE* collector current response is given, approximately, by

$$I_c = \beta_0 I_B \frac{t}{b'} = \frac{I_B}{T_{ce} + T_0} t. \quad (13)$$

The OFF-ON switching response for the large input drive is thus independent of β_0 and is actually improved over that for *CB* operation for the same input current. From Equation (13), the base-current requirement for specified output voltage-current conditions and switching, T_s , is given by

$$I_B = \frac{1}{T_s} \left[V_0 C_T + \frac{I_0}{\omega_0} \right]. \quad (14)$$

The base voltage for the input currents requirement of Equation (14) is determined by

$$V_B = I_B R_B' + V_{jb}, \quad (15)$$

where V_{jb} is the idealized base input diode voltage for $V_C = 0$ as described for the input characteristics of the *CE* stage.

Common-Collector Operation

For the *CC* stage shown in Figure 9 the emitter voltage response to a step of base voltage of magnitude V_B using the equivalent circuit of Figure 6 with the base as input and the collector common, is given by

$$V_e = G_c V_B \left[1 + A e^{-\frac{t}{a'}} - (1+A) e^{-\frac{t}{b''}} \right] \quad (16)$$

where

$$a' = \frac{T_{ce} R_{eb} C_T}{T_e + T_{cc}},$$

$$b'' = D_0 [T_{ee} + T_e],$$

$$D_0 = \frac{R_E + R'_B}{R_E + R'_B (1 - \alpha_0)},$$

$$A = + \frac{T_{ce} - a'}{b'' - a'} \text{ for } T_{ce} < R_{eb} C_L,$$

$$= - \frac{T_{ce} - a'}{b'' - a'} \text{ for } T_{ce} > R_{eb} C_L,$$

$$G_c = \frac{R_B}{R_E + R'_B (1 - \alpha_0)},$$

$$T_e = R_{eb} [C_e + C_L (1 - \alpha_0)],$$

$$R_{eb} = \frac{R_E R'_B}{R_E + R'_B}.$$

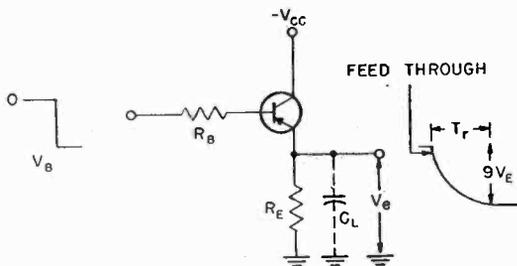


Fig. 9—Transient response, common-collector operation.

The coefficient A is generally positive for large load capacitances or high transistor frequency response and the emitter voltage response in this case includes a time delay, a' , as in the *CB* and *CE* cases. With small load capacity, or low transistor frequency response, A becomes negative and the first exponential term in Equation (16) then defines

a feedthrough effect as indicated in Figure 9. This feedthrough effect essentially results from the low base-emitter impedance prior to the establishment of the degenerative circuit action. It is accentuated for small ratios of source to load resistance, R_B/R_E . Although important from the standpoint of loading of the input signal, it does not contribute appreciably to the total output response (normally $\alpha' \ll b''$ and $A \ll 1$). Hence Equation (16) may be simplified to

$$V_e = G_o V_B \left[1 - e^{-\frac{t}{b''}} \right]. \quad (17)$$

Where A is positive in Equation (16), the time constant, α' , may be incorporated in Equation (17) as a time delay, as done in the CB and CE examples.

Using Equation (17), the calculated emitter voltage response for a step of base voltage of magnitude V_B is compared to the measured response in Figure 10C for the same transistor as in the CB case (with $R_B = 10,000$ ohms, $R_E = 1,000$ ohms, and $C_L = 40$ micromicrofarads). Within the approximations made for Equation (17), the curves agree reasonably well.

Normally, the CC stage is used in buffer or impedance-matching applications and is not over-driven as in CB and CE circuits. The response, as defined by Equation (17), therefore, determines the switching speed of the stage. From Equation (17), the 0-90 per cent emitter voltage rise and fall times $T_r = T_f = 2.3 b''$. An exception occurs for the case $R_E C_L > b''$. The emitter-base junction then becomes reverse biased during the trailing edge of an input pulse, and the output voltage decays with the load time constant, $R_E C_L$.

In CC operation, it is difficult to specify input requirements in terms of output voltage-current conditions and switching speed because of the nonlinear loading conditions at the input. Switching design however can be evaluated in terms of the load and source resistances. Setting $x = R_B/R_E$, the emitter voltage rise time and the fall times (for $R_E C_L < b''$) may be expressed approximately as

$$T_r = T_f = 2.3 U \left\{ R_E \left(C_c + \frac{C_L}{\beta_0} \right) + \frac{1}{\omega_0} \right\} \quad (18)$$

where

$$U = \frac{x + 1}{1 + \frac{x}{\beta_0}} \quad (19)$$

For specified load conditions and output response time, a value for U may be determined from Equation (18). The magnitude of x and, subsequently, R_B may then be obtained from Equation (19). In prac-

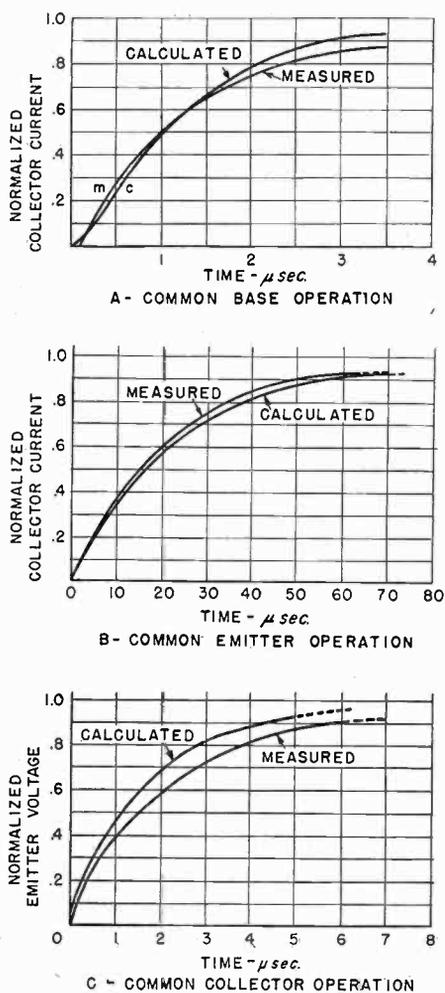


Fig. 10—Calculated versus measured transient response curves.

tice, because of the wide variation in β_0 between transistors, a plot of U versus x for different values of β_0 will facilitate the choice of x for required circuit performance.

Storage Effects in Saturation Operation

The transistor d-c properties in the saturation region were given earlier. In this section, the limitations upon switching response in this operating region will be described.^{1, 2}

When the junction transistor is driven to the saturation state, a charge storage condition is produced within the base region. The effect of this storage condition is to produce what may be appreciable delay in the transition from saturation to cutoff. This is illustrated in Figure 11. A 1.5-microsecond input pulse to the *CE* stage produces in the output signal a storage period, T_v , of 12 microseconds and a storage decay time, T_D , of over 20 microseconds.

This storage condition results from the large minority-carrier gradient established through the base region in the transistor saturation condition. The minority-carrier density is normally large at the emitter and zero at the collector, but in the saturation condition the collector junction becomes forward biased and injects into the base thereby causing a large carrier density throughout the entire base region. This excessive charge condition constitutes a stored charge.

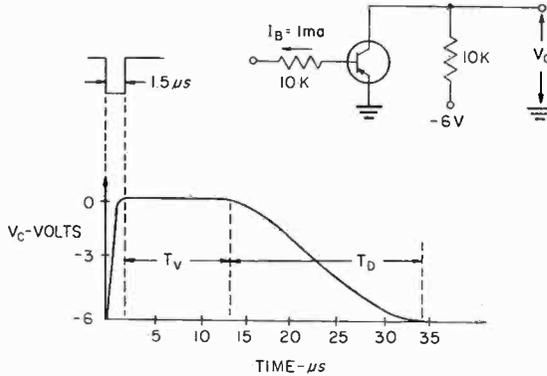


Fig. 11—Typical storage waveform.

At the termination of the input signal, the excess charge density condition will prevail until, by recombination in the base and by removal of carriers at the emitter and collector junctions, the charge density at the collector becomes zero and normal operation is restored.

The storage times may be correlated in terms of the transistor transient response, by extending the transient analysis used for the active range to the transistor conditions existing in saturation. The procedure used is outlined in Figure 12. A *CE* stage is shown here with its emitter returned to a negative voltage, V_1 . This is done to incorporate the effects of reverse input bias on storage times. The collector load current is given by the solid line in collector current diagram in Figure 12. Following the trailing edge of the input pulse, the collector load current has a storage delay time, T_v , and decay time, T_D , with total fall time, T_f .

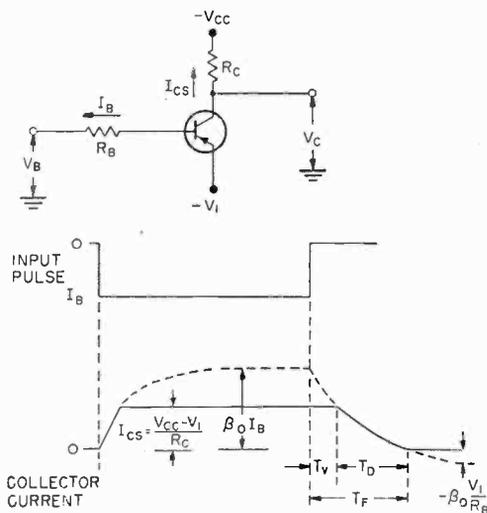


Fig. 12—Transistor current conditions in saturation.

The collector load current, I_{CS} , may be considered as resulting from a limiting action of the low collector-emitter saturation impedance upon the collector current which would result with sufficient collector-emitter voltage to prevent saturation. In this case, the $\beta_0 I_B$ curve, shown by the dashed line, would be the collector load current. The storage time in the decay of the actual load current, I_{CS} , occurs because of the transistor-limiting condition during the decay in collector current from the value $\beta_0 I_b$. Both storage delay and decay times may consequently be expressed in terms of the transient behavior of the $\beta_0 I_b$.

An equivalent circuit defining these conditions is shown in Figure 13. The emitter and collector junctions are represented as diodes J_e and J_c with the collector diode shunted by the $\beta_0 I_b$ current generator. The output voltage in saturation is clamped through the low impedance

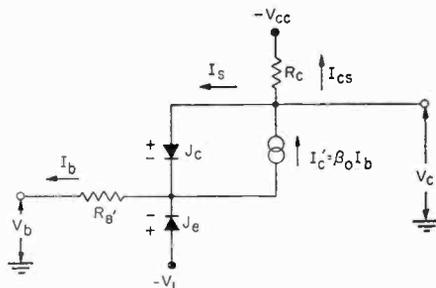


Fig. 13—Equivalent circuit for storage conditions.

of J_e and J_c , maintaining $I_{CS} = (V_{CO} - V_1)/R_C$. The generator current, I'_c , however, increases to a value, $\beta_0 I_B$, where I_B is the maximum base current. The generator current in excess of that supplied to the output, I_s , is shunted by J_c , and represents a stored current component.

At the end of the input pulse, the generator current I'_c decays with the diffusion time-constant, T_{ce} . The output voltage remains clamped by J_c and J_e , however, until I'_c becomes less than I_{CS} , whereupon J_c becomes reverse-biased, and I'_c ($= I_c$) then decays to its cutoff value. The observed storage delay, T_V , occurs during the decay of I'_c to I_{CS} , and the subsequent decay time of I_c is the storage decay time, T_D . Because of the negative emitter voltage, I_c will tend to decay to a negative value, $-\beta_0 V_1/R_B$. However, I_c is limited to its cutoff value when the emitter junction becomes reverse biased.

From these considerations, the transient behavior of the transistor may be extended to include the saturation state with the assumption that β_0 and α_0 (for the *CB* case) remain linear and constant. The only change involved is that the collector capacity is zero during the saturation condition.

The decay in the generator current, I'_c , following the trailing edge of the input pulse (assuming I'_c reaches its maximum value during the input pulse duration) is

$$I'_c = \beta_0 I_B \left[-\frac{V_1}{V_B} + e^{-\frac{t}{T_b}} \right] \quad (20)$$

where

$$\begin{aligned} T_b &= \beta_0 T_{ce} \text{ for } I'_c > I_{CS} \\ &= \beta_0 [T_{ce} + T_0] \text{ for } I'_c = 0. \end{aligned}$$

The storage time is obtained from Equation (20) by solving for $t = T_V$ at $I'_c = I_{CS}$, and is

$$T_V = \beta_0 T_{ce} \ln \left[\frac{V_B}{V_1 + \frac{R_B I_{CS}}{\beta_0}} \right] \quad (21)$$

The total fall-time $T_F = T_V + T_D$ is obtained from Equation (20) by incorporating the different T_B time constant values in the storage and storage decay periods, and is

$$T_F = \beta_0 T_{ce} \ln \frac{V_B}{V_1} + \beta_0 T_0 \ln \left[1 + \frac{R_B I_{CS}}{\beta_0 V_1} \right] \quad (22)$$

A simplified approximate expression for T_F is obtained by including collector capacity during the storage period, and is

$$T_F = \beta_0 [T_{ce} + T_0] \ln \frac{V_B}{V_1}. \quad (23)$$

This expression will yield values for T_F 10-20 per cent too large, (depending on the value of R_C). However, it is sufficiently accurate for many applications.

These storage-time equations include the effects of reverse input bias, which is employed in many switching-circuit applications. Reverse input bias is particularly effective in reducing the storage delay and particularly the storage decay time. Without the reverse bias in the CE stage of Figure 12, the storage delay is given by Equation (21), with $V_1 = 0$. The total fall time (from $V_c = 0$ to 90 per cent of V_{cc}) may be determined from Equation (20), and is approximately

$$T_F = \beta_0 [T_{ce} + T_0] \ln \left[\frac{10 \beta_0 I_B}{I_{CS}} \right] \quad (24)$$

The storage effects in CB operation may be determined in exactly the same manner. The storage time for a CB stage in Figure 7a is

$$T_v = D_b T_{ce} \ln \frac{\alpha_0 I_B}{I_{CS}}, \quad (25)$$

where I_B is the maximum input current. The total storage decay time (from $V_c = 0$ to 90 per cent of V_{cc}) is approximately

$$T_F = D_b [T_{ce} + T_0] \ln \left[\frac{10 \alpha_0 I_B}{I_{CS}} \right]. \quad (26)$$

Similar expressions may be developed for CC operation; however, the CC stage is normally not driven to saturation because of the subsequent loss in impedance-matching properties. This stage is stabilized to some degree against saturation, even for input voltages greater than that of the collector supply. This occurs because the voltage drop produced by the base current across the transistor base resistance increases the reverse voltage across the collector junction. The CB collector characteristics illustrate this action by the reversal in collector potential in the saturation region. This property is useful in a number of switching applications. The general requirement for nonsaturation operation

using this effect is that $R'_E/R_C > \beta_0$, where R_C is the collector resistance and R'_E includes both the transistor and external base resistances.

Storage times, calculated from the storage relations given above, correlate within 20 per cent of measured values. Agreement is within 5 to 10 per cent for total fall time, T_F (except Equation (23)). The greater discrepancies for storage time result primarily from the non-linear behavior of the collector junction as it becomes reverse biased with the subsequent difficulty in accurately specifying the storage time.

JUNCTION TRANSISTOR SWITCHING CIRCUITS

Several junction-transistor switching-circuits are described in this section. These circuits illustrate many of the useful properties of junction-transistors for switching applications. The circuits generally are illustrated with p-n-p transistors; however, in every case n-p-n

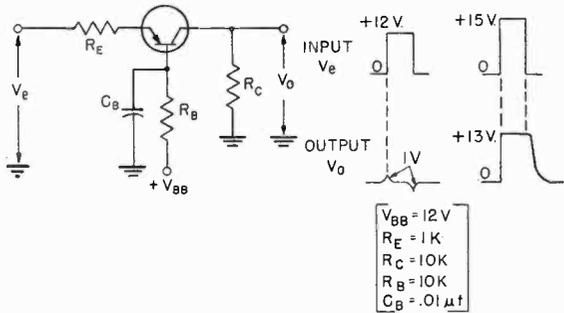


Fig. 14—Amplitude discriminator stage.

types may be used with proper bias and signal polarities.

The *CB* stage shown in Figure 14 is useful in gating and pulse-shaping applications. The positive base voltage, V_{BB} , biases both the emitter and collector junctions in a reverse direction. A positive input signal, whose amplitude is less than V_{BB} will produce no output, since the emitter junction remains reverse-biased. However, if the amplitude of the input signal exceeds V_{BB} , the transistor is forward-biased and the collector voltage rises to very nearly the input amplitude (for $R_E \ll R_C$). The purpose of the base resistance here is to maintain a high input impedance in the transistor saturation state. The circuit can therefore accommodate large input signals without limiting the output amplitude.

The circuit waveforms in Figure 14 illustrate this operation for the circuit component values shown. With a base bias of 12 volts, a 1-microsecond input pulse of less than 12 volts amplitude results in negligible output, whereas the 15-volt input pulse produces a 13-volt

output signal. The base capacitor, C_B , is employed to improve the rise time of the output signal. This capacitor provides a low base impedance for the high-frequency components of the input signal, effectively peaking the circuit response. The base time constant, $R_B C_B$, should be 2 or 3 times b where b is the time constant determining CB transient response (see Equation (5)). Because of saturation operation, the trailing edge of the output signal will include some storage delay, as indicated in the circuit waveforms. This is minimized however by the high reverse input bias produced by the base voltage. The value of the base resistance is limited by the voltage drop produced across it by the cutoff base current. This effect may become severe at elevated temperatures.

As an amplitude discriminator, this circuit will separate input signals from noise, or from other input signals of voltage amplitude less than the base bias. The circuit can also perform as a pulse shaper, in which case the base may be shorted directly to the bias voltage. In this application, a small value of R_B is used such that when the input signal amplitude exceeds the base bias, the transistor is driven quickly to its ON state with improved output rise time over the input signal.

Another useful application of the circuit is for gating. Here, the base voltage is obtained from a control source (such as a flip-flop) to perform either AND or INHIBITION gating with the input signal. The base resistance may be used to provide isolation between the gate circuit and the control source. The control signal could be applied directly to the base, with regard however, to the high base-current surges which occur at the leading and trailing edges of the input signal.

Other gate applications for the CB stage are shown in Figure 15. The OR gate, in *A*, utilizes the extremely low a-c input impedance of the CB stage for isolation between input channels. Any combination of input signals will produce an output signal. The clamp diode, D_0 , may be used to standardize the output pulse amplitude and prevent saturation.

The AND gate, in *B*, also uses the low a-c input impedance for isolation between input channels. In this operation, however, the transistor is biased ON by V_{EB} , with the output voltage clamped by diode, D_0 , to a small negative voltage, V_0 . The magnitude of emitter bias current, I_G , is designed to maintain the output clamp diode conducting until all input channels are negative, at which time D_0 is cut off and the collector voltage drops negative.

Voltage gain is obtained in all these CB circuits if the collector load resistance is greater than the input resistance. A transformer, or CC stage may be used for output coupling to obtain current gain.

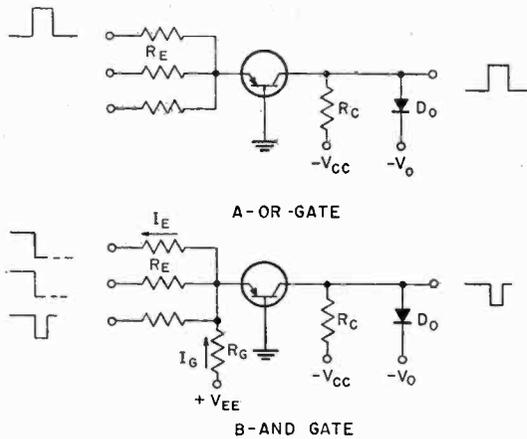


Fig. 15—Common-base gate circuits.

A very useful gating circuit employing the *CE* stage is the “shunt gate” shown in Figure 16. The transistor is normally biased to saturation by V_G . In this condition, the collector signal, V_c , is decoupled from the output by the low saturation impedance of the transistor. However, with a positive base pulse or voltage level of sufficient magnitude to reverse bias the emitter–base junction, the transistor is effectively isolated from the circuit. The collector signal then produces an output, V_o , determined by the voltage divider, R_C and R_0 . This gate action, with the given circuit component values, is illustrated in the circuit waveforms of Figure 16. The output signal amplitude is ap-

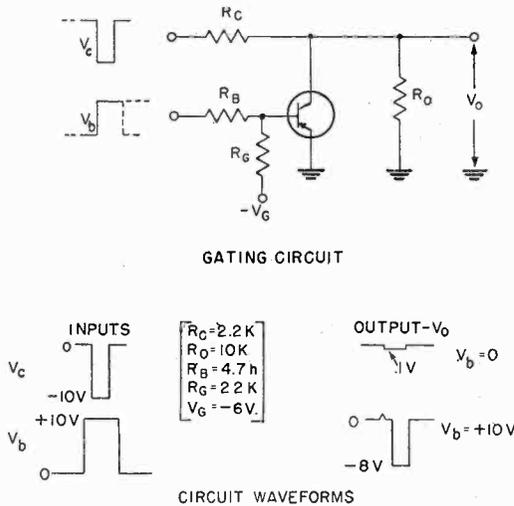


Fig. 16—Common-emitter shunt gate.

proximately 0.1 volt in the INHIBIT condition, ($V_B = 0$). In the PRIMED gate condition ($V_B = 10$ volts) the resulting output amplitude is 8 volts.

Alternately, a negative base input level may be used for saturation bias with R_G returned to a small positive voltage. Then, when the base input level is zero or positive, the transistor is cutoff by V_G . The same gating action is consequently obtained. Additional base inputs may be paralleled in this operation to obtain multiple input gate control, with the low saturation input-resistance providing high isolation between inputs. Additional transistor stages may also be paralleled, to increase the gating capacity further.

These gate circuits provide high discrimination in the output signal, between INHIBIT and PRIMED gate conditions. Very little transient distortion is introduced in the gated signal since the transistor is effectively an open circuit. Some deterioration will occur, however, with high source and load resistances, because of the cutoff junction capacities of the transistor. This type of AND gate works conveniently with the OR gate in Figure 15A, using an n-p-n transistor. The series resistance utilized in the shunt gate may in this case be the input resistance for the OR gate.

The gating speed of the shunt gate is determined in the same manner as for a CE stage, with a collector supply voltage equal to the collector signal amplitude. The switching times involved necessitate a delay between application of the base and the collector signals in the gating circuit. This delay may be considerably reduced, however, by employing a capacitor in parallel with the base resistor, R_B . The same general considerations apply for this capacity, as described in the amplitude discriminator circuit.

The circuit of Figure 17 uses the transistor storage property to obtain a short-duration memory device or delayed gate. This circuit also illustrates the use of the amplitude discriminator described previously. Quiescently, the p-n-p stage, T_2 , is cut off. Its collector voltage, V_C , is therefore highly negative and reverse biases both emitter and collector junctions of the n-p-n transistor, T_1 . In this state, a negative input signal, V_c , to T_1 , will produce no output if its maximum voltage is less than V_C . When a negative "storage" pulse, V_b , is applied to T_2 , this stage is driven to saturation. V_C then drops essentially to zero through the low saturation resistance of T_2 and remains low during the subsequent storage period. T_1 is then unbiased and the emitter input pulse will produce a virtually unattenuated output signal. At the cessation of the storage period, V_C rises to its negative cutoff value and T_1 is again inhibited.

In the circuit waveforms, a 4-volt input pulse to T_2 produces a storage period of 25 microseconds, during which time a 10-volt interrogation pulse to T_1 produces approximately a 10-volt output pulse. In the nonstorage condition, the output is less than 0.1 volt. The dashed line indicates the locus of V_0 during the storage period of T_2 .

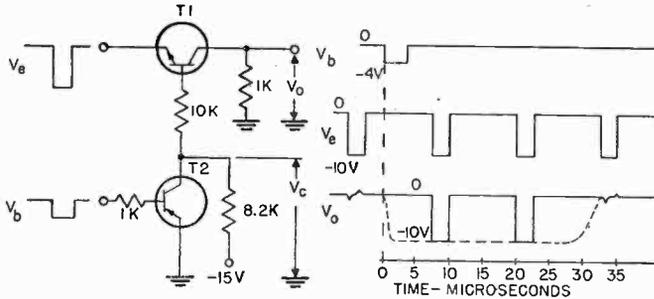


Fig. 17—Transistor memory circuit.

The circuit shown in Figure 18 utilizes minority carrier storage to achieve linear integration of short transient waveforms. In saturation, the degree of charge storage in the base region is a linear function of the total charge in the input signal for input pulse durations which are short compared to b' , where b' is the time constant determining CE transient response. Therefore, the storage charge condition is proportional to the integral, A_b , of the current waveform applied to the base. The circuit operation is based on the fact that the degree of charge storage during its decay from the initial storage condition remains linearly proportional to the integral of the base input signal.

In the circuit operation, the signal to be integrated, V_b , is applied to the base of the p-n-p transistor through R_B . An interrogation pulse, V_c , of standard amplitude is applied to the collector τ microseconds after the base signal. The amplitude of the resulting emitter pulse, V_0 , will then be a function of the remaining storage charge in the transistor at this time. The linearity of the integration process is shown in Figure 18A for the given circuit component values and a 10-volt interrogation pulse. The higher slopes of V_0 versus A_b result from the greater degrees of charge storage existing at the shorter interrogation times.

The small emitter capacitance, C_E , partially bypasses the emitter resistance, and improves the circuit response from the standpoint of base signal. The value of C_E , however, is determined primarily such as to compensate for the diffusion time constant of the transistor, which determines the transient properties of the collector current from

the interrogation pulse. This compensation may be compared to that of an R-C divider network.

Typical circuit waveforms are shown in Figure 18B. The integration process is apparent for the 40-volt-microsecond base signals. With the larger base input, the output amplitude begins to fall off due to non-linear storage effects. In the circuit operation, some feedthrough from the base signal results and the interrogation output must therefore be separated from the feed-through signal by synchronous gating.

A bistable circuit using CE stages is shown in Figure 19. The ON stage, T_1 , is biased in saturation by the base current supplied from the collector of the cutoff stage, T_2 . Note that no reverse bias is used to maintain T_2 cut off. The collector voltage of the saturation ON stage is sufficiently small to maintain a very low cutoff current in T_2 . This

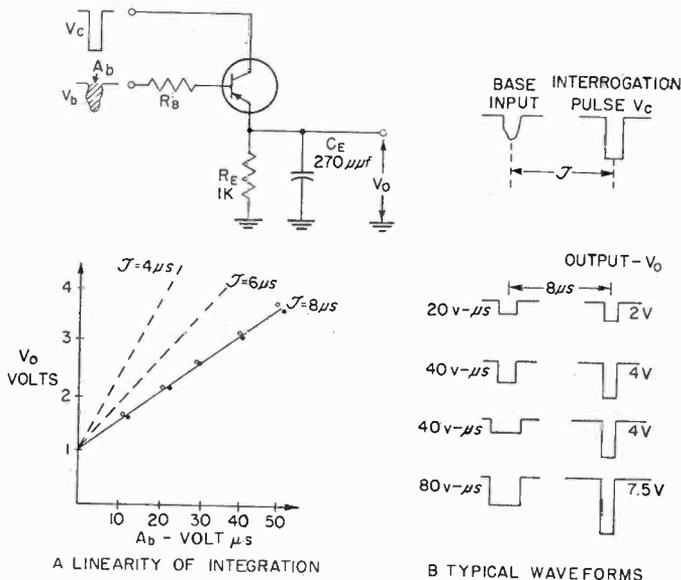


Fig. 18—Transistor integrator circuit.

type of direct-coupled biasing is extremely temperature sensitive, however, so that for high temperature operation, reverse emitter-base bias may be desirable to obtain minimum cutoff currents. Base resistors, R_G , are used here to increase the circuit cutoff stability. Where the cross-coupling resistor value is adequate for this purpose, resistors R_G may be eliminated. High d-c stability and insensitivity to noise signals is obtained from the saturation condition of the ON stage.

The trigger signals are coupled to the bases of T_1 and T_2 through resistors, R_B . This method isolates the low base impedance from the

trigger sources, at the same time providing sufficient trigger current for reliable trigger action. For small trigger-voltage amplitudes, however, the required trigger resistances may be so small as to introduce prohibitive d-c loading, due to the base-emitter voltage of the ON stage. Diode trigger coupling may then prove a better method. Cutoff clamping may be employed to standardize the OFF collector voltage. Very efficient ON clamping is provided by the low saturation impedance of the transistors.

The switching times of the circuit are determined by the transistor frequency response, and especially by the value of cross-coupling capacitors, C_s . The value of C_s is determined by the transistor frequency response and the trigger-pulse amplitude and duration.

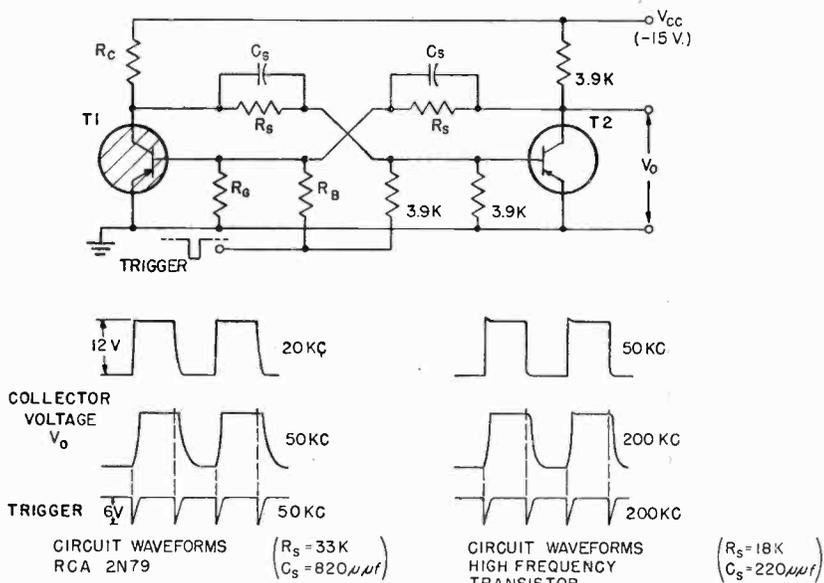


Fig. 19—Junction transistor bistable circuit.

These factors are indicated in the circuit waveforms in Figure 19. Collector voltage waveforms are shown at 20 and 50 kilocycle trigger rates using low-frequency 2N79 transistors ($f_o = 1$ megacycle), and at 50 and 200 kilocycles for high-frequency units ($f_o > 4$ megacycles). Rise and fall times are 3 and 8 microseconds in the first case, and .5 and 1 microseconds in the second case. Much faster operation is obtained with the high-frequency transistors with the reduced value of C_s .

The effect of C_s on circuit speed may be reduced considerably by employing *CC* stages in the cross-coupling networks. This technique is illustrated in the counter circuit in Figure 20. The *CC* stages de-

crease not only the a-c loading on the flip-flop stages, but also reduce greatly the d-c loading on the OFF stage. Proper bias between collectors and bases is maintained by the emitter-follower action of these stages, and the emitter terminals provide low-impedance output points. As shown in the circuit waveforms, a significant improvement in high-frequency operation is obtained by this means. With the 2N79 transistors, circuit operation at 100 kilocycles is now obtained, with rise and fall times of approximately 1.0 and 2 microseconds. With the

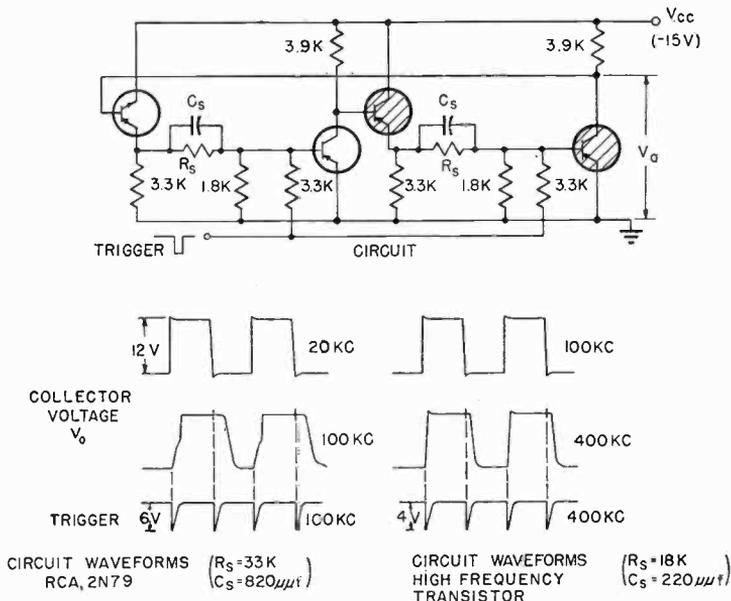


Fig. 20—Bistable circuit with common-collector coupling stages.

high-frequency units, operation at 400 kilocycles is obtained with rise and fall times of .1 and .3 microsecond. Circuit operation above 1 mega-cycle has been obtained in this circuit.

ACKNOWLEDGMENTS

The author gratefully acknowledges the assistance and ideas of R. K. Gerlach in the development of many of the circuits described and is appreciative of the suggestions of A. H. Lytel in the presentation of the subject material.

A NOVEL RING COUNTER USING JUNCTION TRANSISTORS

BY

D. E. DEUTCH

RCA Defense Electronic Products,
Camden, N. J.

Summary—A ring counter is described which employs a new method of storage propagation resulting in improved system reliability and higher counting speed capabilities. The method is to apply shift signals sequentially to alternate counter stages and thus prevent interaction between count signals to the counter stages and transfer signals produced within the counter. Several junction transistor circuits, developed for use in the ring counter, are described together with the circuit and counter performance obtained.

INTRODUCTION

RING counters have important application in many types of modern electronic equipment including storage and timing in digital computers, frequency division, counting, and time sequencing in time-multiplex communication systems.

In some of these applications, particularly frequency counting, ring counters may be replaced by binary counters using external feedback to obtain the desired decimal counts. However, for timing, information storage, and display, the direct access to ring-counter storage is often more efficient than using decoding matrices from binary counters.

Transistors greatly reduce the size and power of ring counters and thus permit ring counters to compete with the binary type. Both point-contact and junction types may be used. Point-contact transistors are useful because of their inherent bistability and high speed; only a single transistor is required for each counter stage. When junction transistors are used, two transistors are required for each stage. However, junction transistors have greater reliability and reduced power requirements, and are, at present, the most often used device for this application.

RING COUNTER OPERATION

Figure 1A shows a conventional n-stage ring counter. All stages are initially reset to their "off" or non-indicating states. The 0 stage is triggered to its "on" or indicating state by the 0 set signal. In the counting operation, the count pulses from the driver stage advance the "on" state down the successive stages of the counter, with the count denoted at any time by the particular stage which is in its "on"

state. The count signals are applied to all counter stages simultaneously, and are of such polarity as to trigger the stages "off." The "on" stage, which is triggered "off" by a count pulse, produces a transfer signal which turns "on" the following stage. This action is illustrated where stage 1, initially "on," is triggered "off" by the count pulse, V_c , producing a transfer pulse, V_t , which triggers stage 2 "on."

This type of count-transfer operation places a restriction upon the operating speed of the counter and may result in marginal operation because the count pulse and the transfer pulse produced from a particular stage interact at the following stage. The transfer pulse must be

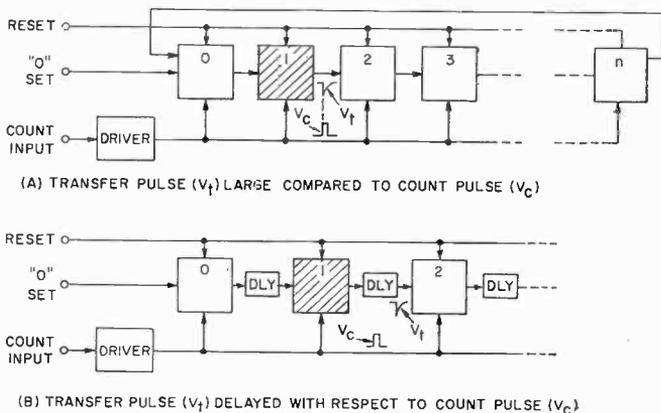


Fig. 1—Conventional counter systems.

of greater duration than the count pulse and of sufficient magnitude thereafter to reliably trigger the stage "on." The transfer pulse however, must be sufficiently short to accommodate the desired frequency of operation and to prevent undue loading of the stages (RC differentiation is normally used to obtain the transfer pulse). On the other hand, the count pulse must be of sufficient magnitude and duration to reliably trigger "off" the counter stages.

A method of eliminating interaction between count and transfer signals is to delay the transfer signal, as shown in Figure 1B. This method, however, requires a delay line between each counter stage, which is an added circuit complication, interferes with high-speed counter operation, and is an added cost factor. Proper delay line termination, to prevent reflections which could cause false triggering, may also introduce difficulties.

DUAL COUNT-TRANSFER RING COUNTER

A counter system using a dual count-transfer method to eliminate the difficulties mentioned above is shown in Figure 2 as a decade ring

configuration. The counting procedure is the same as for the previous counter types. The count is indicated by the "on" stage in the ring, and the count is transferred by triggering "off" the "on" stage with the resulting transfer signal turning "on" the following stage.

In this counter, however, the input count signals trigger a flip-flop driver stage whose outputs are connected to alternate stages of the counter, on count lines a and b. Count signals on lines a and b are derived by differentiation of the voltage steps at the flip-flop outputs. The count signals on line "a" are thus interlaced with the count signals on line "b," such that alternate stages receive a count signal from the flip-flop driver with successive input count-signals. In this manner, the stage which receives a transfer pulse from the previous stage will not simultaneously receive the count pulse which triggered off the previous stage. Further, no limitations are placed upon the counter

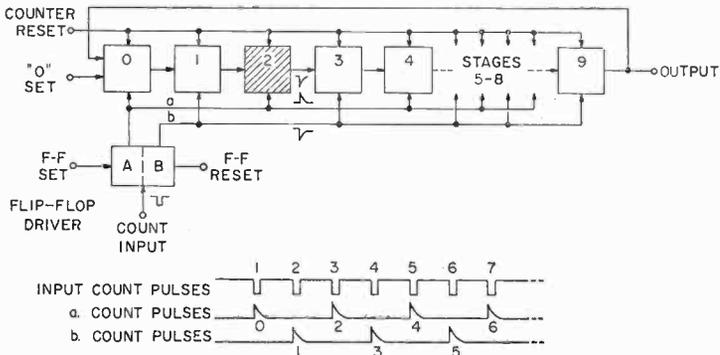


Fig. 2—Dual count-transfer counter system.

operating speed, except for that imposed by transition times of the flip-flop driver and the counter stages themselves.

This count-transfer method is not directly applicable to ring counters having an odd number of stages. In this case, there will be two adjacent stages in the ring which receive a count signal simultaneously. Figure 3 shows a revised ring-counter system which can accommodate an odd number of stages. In this system, the transfer signal from the last stage in the ring is fed back to the first stage and to the flip-flop driver through a delay line. The delay must be used for two reasons. Because stages 4 and 0 are connected to the same count pulse line, the transfer pulse from 4 must be delayed to avoid interaction at stage 0 with the count pulse initiating transfer from stage 4. The delayed transfer pulse from stage 4 must also reset

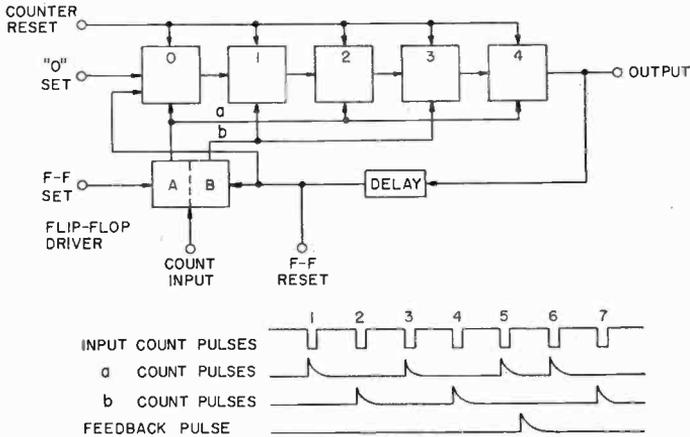


Fig. 3—Ring counter with odd number of stages.

the flip-flop driver such that the next input count pulse will properly initiate a count pulse on line a to stage 0.

TRANSISTOR DRIVER STAGE

The flip-flop driver may be any conventional type of symmetrical bistable circuit, such as the Eccles-Jordan flip-flop. A junction transistor flip-flop circuit using 2N109 transistors for low-frequency counter operation (30 kilocycles) is shown in Figure 4. The set, reset, and count trigger signals are resistance-coupled to the flip-flop stages. Note that no reverse bias is used to stabilize the "off" stage. This is permissible because the saturation collector voltage of the "on" stage (.01 to .05

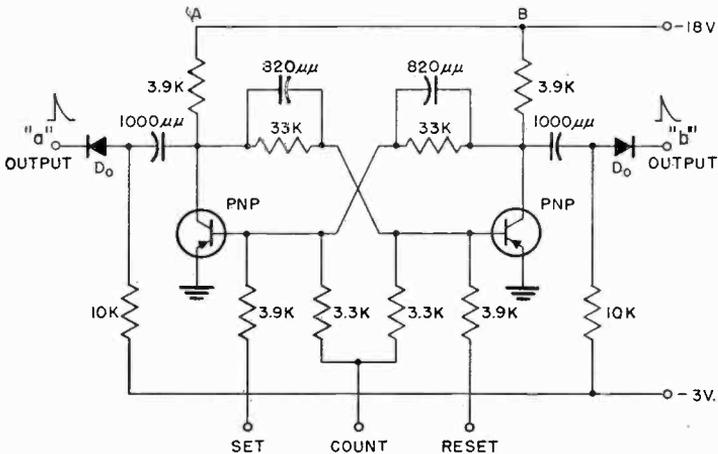


Fig. 4—Flip-flop driver stage.

volt) is much less than the conduction bias of the emitter-base junction of the "off" stage (.1 to .3 volt). Positive count pulses are derived from the flip-flop stage by RC differentiation circuits composed of the 1,000-micromicrofarad capacitors and the 10,000-ohm resistors. Diodes D_0 clip the negative portions of the differentiated waveforms. The 10,000-ohm resistors are returned to -3 volts providing a reverse bias to diodes D_0 . This prevents noise signals from the counter stages from triggering the flip-flop driver. Circuit rise time (off-on) is 1.5 microseconds, circuit fall time (on-off) is 5 microseconds, using a 2-microsecond, -6 volt trigger signal. The count signals supplied by the circuit are 6-volt differentiated pulses of 2 microseconds width at 50 per cent amplitude.

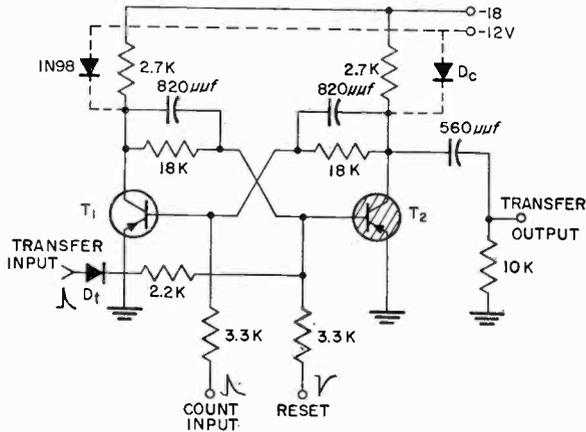


Fig. 5—Junction transistor counter stage.

JUNCTION TRANSISTOR COUNTER STAGE

A junction transistor flip-flop for use in the ring counter stages is shown in Figure 5. The circuit is basically the same as the driver stage except for reduction in the collector and cross-coupling resistors, allowing the circuit to supply one milliamperere d-c load current from either collector for gating or display functions. The collector voltage from the "off" stage may be clamped to obtain a standard output level, as indicated by diodes D_c .

In the nonstorage or reset state T_2 is "on." The positive transfer signal from the previous stage turns off T_2 . Diode D_t inhibits the negative signals in the differentiated output from the previous stage. The count signals are supplied to T_1 and turn off this side if the circuit has been set by a transfer pulse. In the reset state T_1 is "off" normally and is unaffected by the count signals. The positive voltage

step from the collector of T_2 , when turned "on" by a count signal is differentiated to provide the transfer pulse for the following stage.

This circuit has collector voltage rise and fall times of 1 microsecond and 3 microseconds when clamped to -12 volts. Transfer time between stages is approximately 4 microseconds thus permitting a maximum counting rate of 250 kilocycles.

A JUNCTION-TRANSISTOR COUNTER WITH HIGH-SPEED CARRY

BY

T. P. BOTHWELL, G. W. BOOTH, AND E. P. ENGLISH

RCA Defense Electronic Products,
Camden, New Jersey

Summary—A reliable transistor-operated counter has been developed and tested. Counting speeds of up to 140,000 counts per second have been achieved using the 2N79 transistor. Rates of several times this figure are realizable using the r-f type transistors which have recently appeared. A high-speed gating circuit for propagation of the carry bit is incorporated which allows complementation and use as a reversible counter. The design of circuits for gating, amplification, and storage functions is also discussed. Transistor characteristics required for proper circuit performance are outlined.

INTRODUCTION

BINARY counters find extensive use in a variety of digital applications; radiation measurement, time measurement, address and control circuits in digital computers to mention a few. For many, if not all of these purposes, it is desirable to count and sense the resulting output at as high repetition rates as possible. When audio-frequency transistors are used in such circuits, both carry propagation delay and impedance matching between stages present problems. A counter is described here which alleviates the carry propagation delay and impedance match problems and fixes the time at which the counter may be sensed after trigger at the approximate rise time of the flip-flop used. In addition, the configuration allows a simple conversion to reversible operation.

OPERATION OF HIGH-SPEED CARRY

This counter consists of a series connection of identical stages, as in Figure 1, each of which is a two-input adder. The function of a two-input adder is most simply explained by the truth table below. When a "one" is added to a "one," a carry is formed, but the sum is "zero." When a "one" is added to "zero," a sum of "one" is formed, but no carry. When "zero" is added, the previous sum is retained, and no carry is formed.

Addend	Augend	Sum	Carry
1	1	0	1
0	1	1	0
1	0	1	0
0	0	0	0

Thus when "one" is added to the counter, each stage in succession forms the appropriate sum, and if there is a carry, propagates this to the next stage. To understand this most readily, consider the example of a six-stage counter. When "one" is added to 000111, the result is 001000 (least significant digit on the right). Assume that the number 000111 had been stored in the flip-flops of the counter described here. To add "one" a pulse is applied to the first stage. (In the diagram of the counter, the least significant bit is on the left.) The carry is formed when the "one" state of the first flip-flop enables ("primes") the first gated pulse amplifier to pass a pulse to the second stage. Simulta-

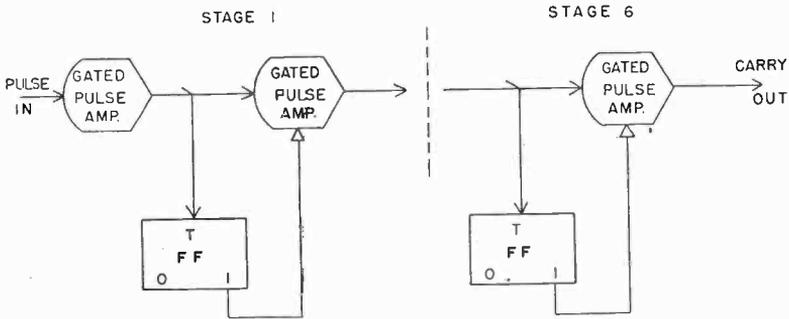


Fig. 1—Transistor counter with high-speed carry.

neously, the pulse initiates triggering of the first flip-flop to the "zero" state, thus yielding the proper sum. However, there is sufficient delay in the rise of the flip-flop to assure that the gated pulse amplifier of stage 1 acts on the basis of the *previous* state of its flip-flop. The "carry" pulse passed to the second stage next performs a function similar to that performed by the pulse applied to the first, that is, it forms a carry in the second gated pulse-amplifier and triggers the second flip-flop to the "zero" state. A similar operation occurs in the third stage. The flip-flop of the fourth stage contains "zero," so the fourth gated pulse amplifier is unprimed, and no carry pulse is propagated to the fifth stage. The carry pulse from the third stage triggers the fourth flip-flop from "zero" to "one" and the counter is ready to receive the next count. The proper sum, 001000, is stored in the flip-flops at the end of this operation.

It can be seen that since no change of state of the flip-flops is required for carry generation, the time for the carry to propagate from stage to stage is a function *only* of the delays in the pulse amplifiers. The sum of these through the six counter stages gives a total delay which corresponds to ripple time in a conventional counter, or pulse delay in a parallel high-speed counter. The total settling time of the counter in the 111111 condition will be given by six times the ripple time, plus the fall ("one" state to "zero" state) time of the last flip-flop.

When compared with the usual parallel high-speed carry scheme, such as shown in Figure 2, the scheme used here effects a considerable saving. Each flip-flop must supply current to control a single gate input, rather than between 1 and n gate inputs as is the case for an n -bit

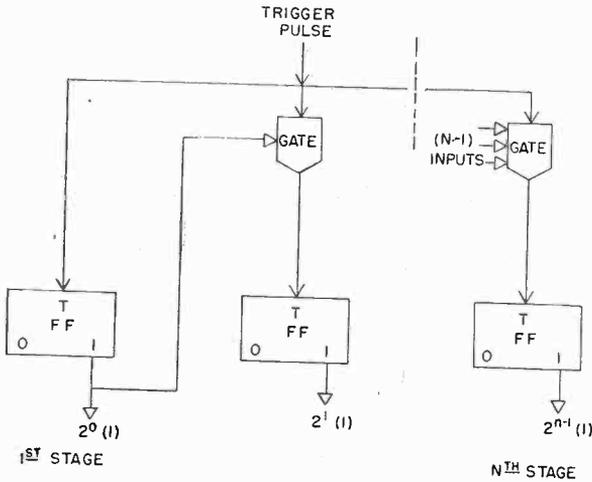


Fig. 2—Conventional counter with high-speed carry.

counter in a parallel scheme. In addition, gates used in this scheme require a uniform two inputs, those in the parallel scheme require n inputs on the n^{th} stage.

APPLICATION OF HIGH-SPEED CARRY FOR COMPLEMENTATION AND A REVERSIBLE COUNTER

The gating scheme used for the high-speed carry allows a simple means for complementation. The addition of a pair of "OR" diodes (as in Figure 10) on the gate associated with each stage will allow the "COMPLEMENT" pulse (Figure 3) to appear at the trigger input to each flip-flop regardless of the state of the preceding flip-flop. In the

scheme illustrated in Figure 3, complementation is accomplished by applying a pulse to the complement input. This sets the complement flip-flop, triggers all counter stages after a delay and finally resets the complement flip-flop. This allows the counter to be interrogated for its present state or its complement using only one set of gates for interrogation.

The scheme shown in Figure 3 also allows the use of the counter to add or subtract one on the receipt of a trigger pulse. When subtraction is desired a complement pulse precedes the trigger pulse and the counter

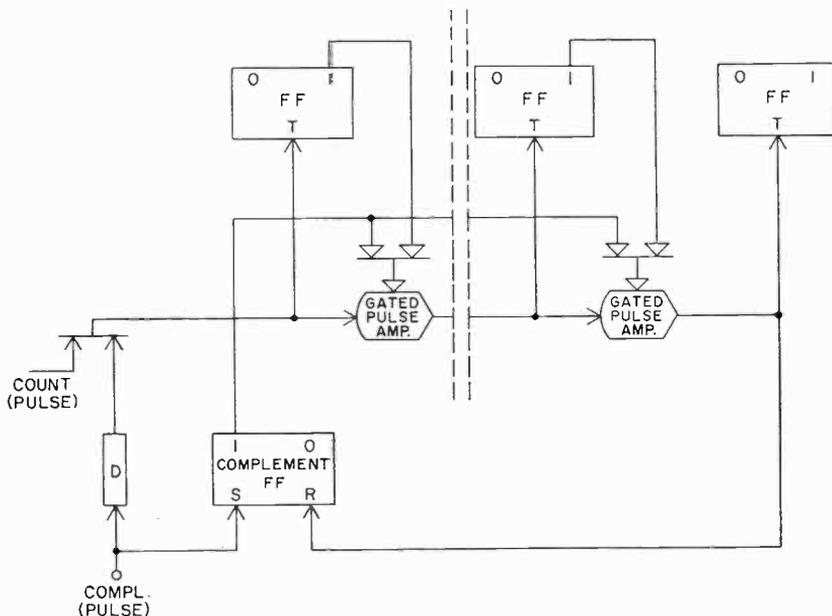


Fig. 3—Reversible transistor counter.

is recomplemented before interrogation. For addition, pulses should be fed only to the "count" input and the complement flip-flop is inactive.

FLIP-FLOP

Circuit Considerations

The flip-flop used in the counter is shown in Figure 4. Its trigger input, collector and base waveforms are in Figure 7. It employs an audio frequency p-n-p junction transistor in an Eccles-Jordan type circuit. Clamping at both ends of the collector swing is provided to assure a uniform output, to prevent saturation in the "on" transistor, and to speed turn-off time by catching the waveform on the fast portion

collector only 0.3 milliamperere load current was needed. To accommodate this, V_{CC} was lowered to 18 volts \pm 10 per cent.

Choice of a value for C_f is dictated by a compromise between trigger sensitivity and rise- and fall-time requirements. A large value of C_f increases both rise and fall times at the collector and at the same time decreases the trigger power requirement. On the other hand, too small a C_f makes triggering critical for certain transistors with very high or low cutoff frequency. It was found for example, that $C_f = 1800$ micromicrofarads required a 10 per cent tolerance on amplitude and duration of the 2-microsecond trigger pulse, with a rise time of 3 microseconds and a fall time of 6 microseconds. Although the rise and fall times were good, larger trigger amplitude tolerance was required for reliable triggering in a large system. A fairly large value of C_f (3000 micromicrofarads) was chosen at the expense of rise and fall times so that an acceptable range of trigger requirements was obtained. With this C_f , turn-on time was 3 microseconds, turn-off time 9 microseconds.

*Pulse Triggering**

Both negative and positive pulse triggering for this flip-flop were investigated. Positive triggering tends to turn "off" both transistors and utilizes the feedback capacitors principally to turn "on" the appropriate transistor when the trigger is removed. Negative triggering, on the other hand, tends to turn both transistors "on", and depends on the cross-coupling network to accomplish "turn-off" after the pulse is removed. Since the transistors turn on more rapidly than they turn off, the differentiated waveform in the feedback network is much sharper for the negative trigger pulse, and the negative trigger therefore proved the more sensitive.

The minimum and maximum trigger requirement for the worst cases of component and voltage tolerances was considered, as well as circuit performance under extreme combinations of gains of transistors. Two-microsecond pulses of between 2.2 and 3.8 volts guarantee triggering in all cases, allowing a 3-volt, 2-microsecond pulse with \pm 20 per cent tolerance as the standard trigger. In the counter, a 1.5-microsecond pulse is propagated through the amplifier chain and a 3.5-volt pulse (\pm 20 per cent) is required to trigger at this pulse width.

The flip-flop can be triggered reliably with the standard trigger up to about 125 kilocycles (8 microseconds between trigger pulses). As the pulse repetition frequency is increased beyond this point, larger trigger-pulse amplitudes are required to overcome the charge still remaining on the feedback capacitor between the conducting col-

lector and the cutoff base. The discharge time constant (RC) of this capacitor through the base and collector resistors is 7 microseconds. However, the 90 per cent discharge time is less than the usual 3 times RC , since during the switching time of the "off-going" transistor, the feedback capacitor partially discharges through the base-emitter diode of the transistor.

Transistor Selection

Following experiments with several types of p-n-p junction transistors (including the T34D, the 2N34, and the 2N49), the 2N79 was chosen for this application because of more closely controlled current-gain ratio ($\beta_{\max}/\beta_{\min}$), and cutoff current (I_{c0}).

THE PULSE AMPLIFIER

Circuit Operation

The grounded-base, transformer-coupled amplifier circuit (shown in Figure 5) was designed as the second basic circuit of the counter. Application of an appropriate level to the base of this amplifier allows its use as a two-input "AND" gate. Thus, when the base is more positive than the largest excursion of the voltage pulse on the emitter, the transistor remains cut off. However, when the base is held at a sufficiently low potential that the emitter potential rises above the base during the input pulse, the transistor rapidly turns on, producing an output pulse. In the counter application, the base of the gate is tied to the collector of the flip-flop. When the collector of the flip-flop is in its high-voltage state (-8.2 volts), the gate is "primed" and will pass a pulse from the emitter through the gate to the next stage. When the flip-flop collector is in its low voltage state (-2.8 volts), the gate is "unprimed."

The pulse transformer in the gate is a step-down type with turns ratios of 12:3:2. The 3-turn secondary is used to trigger the flip-flop; the 2-turn secondary drives the emitter of the following pulse amplifier. Primary inductance of the transformer used is 1.5 millihenries. This value, together with the 1000-ohm base impedance, will support a 2-microsecond pulse.

Power Gain

The power gain of an iterative amplifier of this type will be defined as the ratio of input pulse power to output pulse power. Since the current gain (α) of the grounded base transistor is essentially unity, the power gain of the stage is nearly equal to the voltage gain from emitter to collector. The transistor is driven from saturation to cutoff,

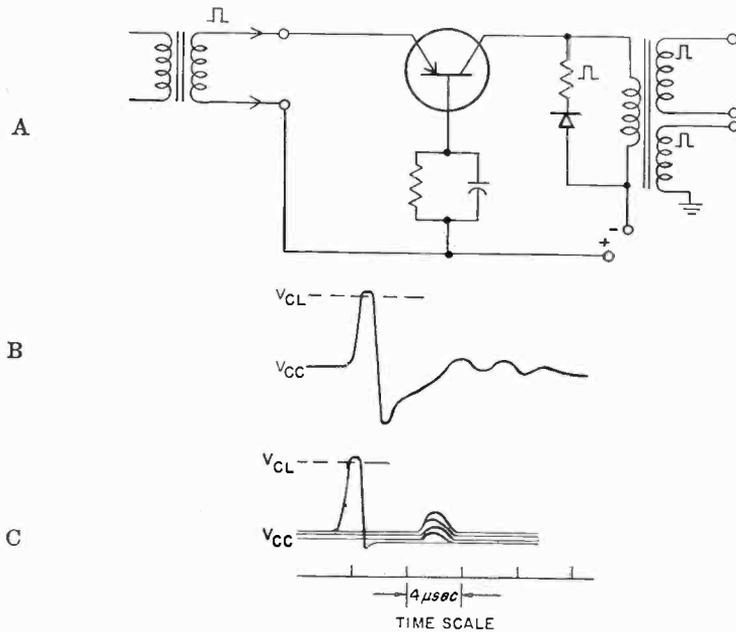


Fig. 5—(A) Basic pulse amplifier; (B) Transformer ringing; (C) Noise following pulse.

hence its collector swing is given by the sum of the input signal voltage and the collector supply potential:

$$V_c = V_e + V_{cc}$$

where V_e = input signal at emitter

V_c = output pulse amplitude at collector.

Then voltage gain (VG) is given by

$$VG = V_c/V_e = (V_e + V_{cc})/V_e = 1 + (V_{cc}/V_e)$$

and power gain, PG , is

$$PG = VG = [1 + (V_{cc}/V_e)] \alpha.$$

For an iterative amplifier the transformer turns ratio (N_1/N_2) is chosen such that the output voltage equals input voltage:

$$N_1/N_2 = (V_{cc} + V_e)/V_e = (V_{cc}/V_e) + 1.$$

Hence $PG = (N_1/N_2) \alpha$.

In this case $N_1/N_2 = 6$, $\alpha \leq 0.95$, hence

$$5.7 < PG < 6.$$

Transistor Selection

Selection of transistors for proper operation in this circuit presented an initial problem. When the gate was permanently primed (no gating function) as in Figure 5A, all 2N79's gave satisfactory operation. (Circuit operation is considered satisfactory if the pulse output is 1.5 microseconds wide with a 10-volt amplitude and 0.6-microsecond rise time on the transformer primary.) However, in the gate version, Figure 6, certain transistors showed 10-20 per cent smaller output amplitude, and rise times 0.1 to 0.4 microsecond longer than most. A partial correlation of this decreased amplitude with current gain was found, but some cases of equal current gains and unequal amplitudes showed this to be only one of the contributing factors. The effect of internal base resistance r'_b was simulated by varying a 50-ohm potentiometer in series with the base lead. It was found that degenerated performance in both amplitude and rise time could be attributed to an increased r'_b . The actual measurement of r'_b indicated however, that the variation of r'_b from unit to unit was not sufficiently large to account for the entire variation in performance. Information on frequency response of the transistors used was not available, but pulse rise times were concluded to be dependent on this parameter. Poor performance as a pulse amplifier was finally attributed to a combination of three factors: (1) high internal base resistance, (2) low current gain, both of which affect rise time and amplitude, and (3) low frequency response, which affects rise time only.

NOISE PROBLEMS

Certain problems were encountered in operation of the pulse amplifier as part of the counter which were not otherwise apparent. Spurious pulses, or "noise," originally presented a major problem. Two types of spurious pulses were encountered. The first was due to ringing of the coupling transformers, and is illustrated in Figure 5B. The ringing was eliminated by the usual technique of shunting a series diode and resistor across the primary of the pulse transformer, slightly more than critically damping the overshoot and stretching transformer recovery from 5 to 10 microseconds. This diode and resistor are included in the schematic of Figure 5A.

A second form of noise remained. This second noise pulse appears 7 to 10 microseconds after the trigger pulse, and has the form shown in Figure 5C. This phenomenon is a result of the non-zero drop across the clamp diode as the flip-flop collector is clamped at -8.2 volts. The emitter of the pulse amplifier is returned to -8.2 volts through the transformer, while the flip-flop collector (and consequently the amplifier

base) rises to about -8.4 volts, the extra 0.2 volts being the clamp-diode drop. This results in a net forward bias current of 200 microamperes in the base-emitter diode of the amplifier. This is differentiated in the pulse transformer and appears as a pulse at the input of the next amplifier. In the case where five following gates were primed (when the counter was set to all "ones") this pulse was amplified through all stages and reached sufficient size in the fifth stage to spuriously trigger the last flip-flop. The simplest solution to this problem was to reverse bias the base-emitter diode of the pulse amplifier by "bucking-out" the drop across the flip-flop clamp diode. A biased diode was introduced into the emitter circuit for this purpose. The diode, rather than a small resistor, was used in order to minimize the impedance introduced into the emitter diode loop during the pulse. This biasing diode and resistor are shown in the input transformer return in Figure 6.

PERFORMANCE

Input Current and Voltage

Figure 6 is a schematic of a single stage of the counter. With a 2 -microsecond input pulse the minimum input current requirement for the counter is approximately 6.8 milliamperes at 1.7 volts, or 12 milliwatts of power. Any current between 6.8 and 10 milliamperes gives satisfactory operation. Pulse emitter currents of 10 to 12 milliamperes were measured at various pulse amplifier stages of the counter. These vary slightly because of variations in transistor current gains. A delay of about 0.16 microsecond occurs in each stage of the amplifier chain, amounting to a total delay (ripple time) of 1 microsecond in 6 stages. The total settling time of this counter, the time after application of a trigger at which it may be sensed, is 10 microseconds. This represents the 1 -microsecond delay in pulse transmission through the amplifier chain plus the 9 -microsecond rise time of the last flip-flop.

Flip-flop waveforms are also indicated in Figure 7. The pulse input on the emitter gate and flip-flop trigger is B , the base of the gate is C , the flip-flop base is at D .

Operation of the counter with a 2 -microsecond pulse input may be seen from the waveforms in Figure 8. The difference in pulse width observed in Figures 8A through C are due to the fact that the pulse propagated through the counter chain is standardized to 1.5 microseconds width regardless of input pulse width. This is further illustrated in Figure 9 which shows all six pulse amplifier collectors for an input pulse width of 5 microseconds. All pulses beyond the 3rd stage

are found to be standardized to 1.5 microseconds. This is not the case when the same pulse amplifier is operated with a zero impedance source, such as a battery, rather than the flip-flop as a control-level source. The explanation lies in the fact that the time constant (L/R), which consists of the primary inductance (1.5 millihenries) together with the parallel value of the transistor base resistor (1000 ohms) and the reflected load (about 5000 ohms), is reduced by the non-zero impedance of the flip-flop which essentially adds to the base resistance. Hence under the load seen in the counter (flip-flop trigger input and the emitter of the next gate) the transformer flux collapses after 1.5 microseconds. This thesis is supported by the fact that no evidence of progressive pulse widening or shortening is found where the circuits are similarly loaded. The lack of a load on the last pulse amplifier

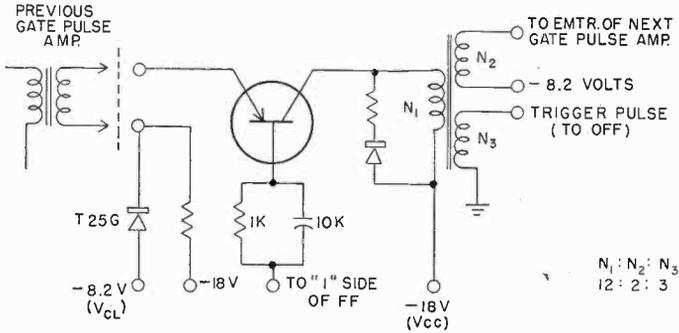


Fig. 6—Gated pulse amplifier damping and biasing diodes.

accounts for the narrower waveforms observed there in Figures 8 and 9. Loading the last stage with 150 ohms very nearly restores the pulse to the 1.5-microsecond width at that point. It is of interest to note that for identical loading the pulses in the sixth stage are of identical width in both Figures 8 and 9, that is for both 2- and 5-microsecond inputs.

For implementation of the "complement-subtract" scheme shown in Figure 3, it is necessary to OR the COMPLEMENT level with the control level from the counter flip-flop into each gate. A circuit for accomplishing this is shown in Figure 10. This is basically the same circuit as Figure 5A with the two-input diode OR added. When either the counter flip-flop or the complement flip-flop goes negative, a pulse appearing at the emitter will produce an output at the collector. The 100,000-ohm resistor is provided to discharge any stray capacity if diode back resistance is too high.

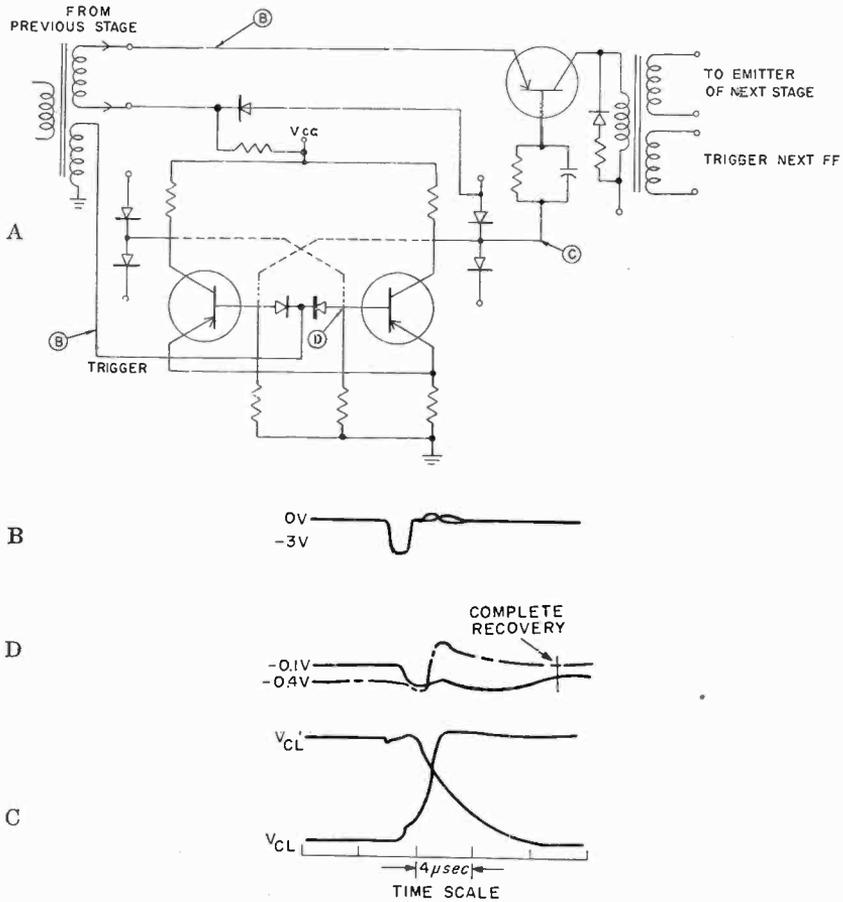
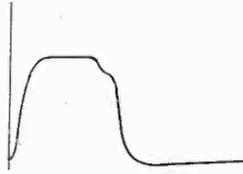


Fig. 7—(A) Flip-flop circuit; (B) Pulse on flip-flop trigger input; (C) Flip-flop collector (base of gate transistor); (D) Flip-flop base waveform showing recovery time.

Counting Rate

The maximum counting rate of the counter is, of course, dependent upon the maximum pulse repetition frequency of the individual circuits. Once any unit fails to operate, all successive units will fail because the operation of each circuit is dependent upon the successful operation of the previous circuit. In this counter, as in most, the frequency response of the first flip-flop and first gate govern maximum counter operation frequency. The first flip-flop and gate receive pulses at twice the frequency of those in the second stage, and each successive stage receives pulses at a frequency half as great as that of the immediately preceding stage. The limit of useful operation is here set by the settling time of

A. First gate collector.



B. Second gate collector.



C. Sixth gate collector.

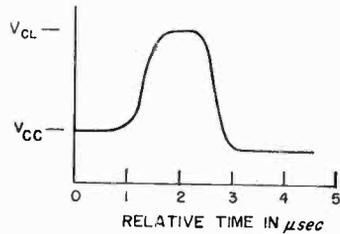


Fig. 8—Counter pulse waveforms for 2 microseconds input pulse.

the counter. Since the counter cannot be sensed earlier than 8 to 10 microseconds after the trigger due to the slow settling of the flip-flop, its maximum useful frequency for most applications is about 100 kilocycles. The 10-microsecond recovery time of the pulse amplifier requires increasing pulse power for operation above 100 kilocycles, but 140-kilocycle count rate is possible, provided no sensing of the counter is required between pulses.

Power Supply

Three power supplies are required for the entire counter and a 10 per cent regulation requirement is placed on these. Total six-bit counter power consumption is less than 1.2 watts, of which a little more than half is from the 18-volt supply.

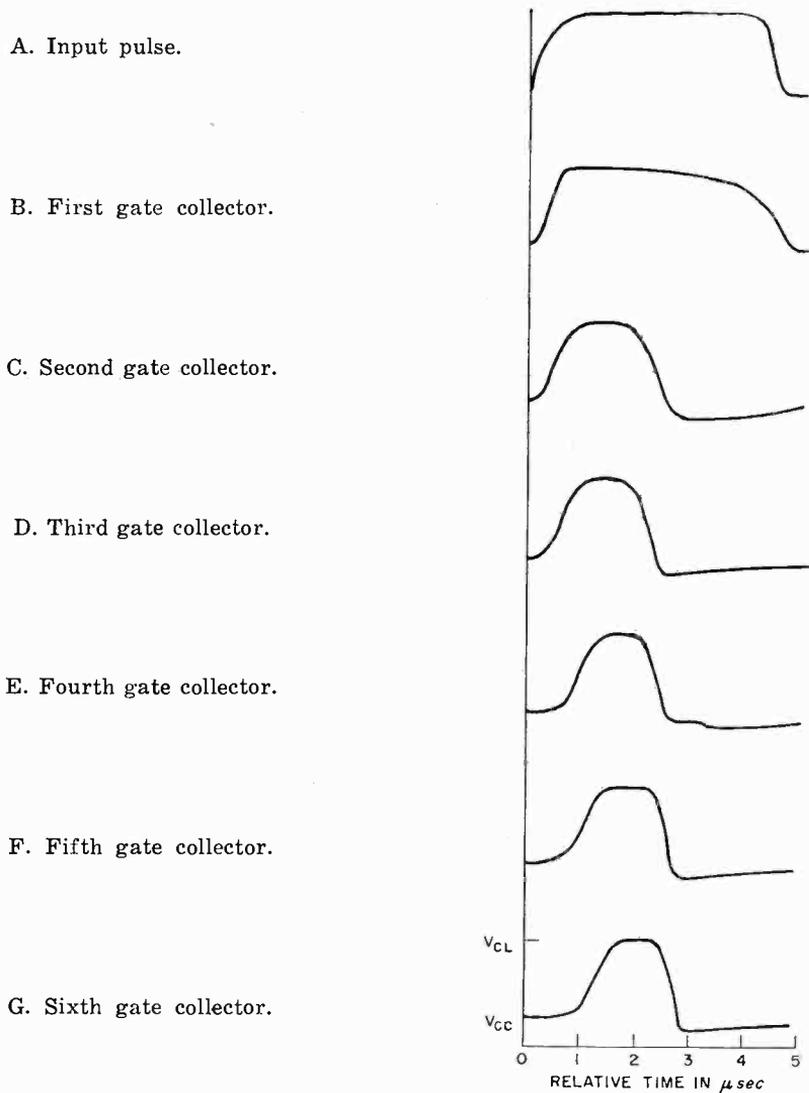


Fig. 9—Counter waveforms for 5 microseconds input pulse. (All to same scale.)

CONCLUSIONS

The counter has been built and operated using 2N79 transistors. Ripple time of 0.16 microsecond per stage has been achieved using 1.5-microsecond pulses with 0.5 microsecond rise times. Speed of the

H. First flip-flop collector.

I. Second flip-flop collector.

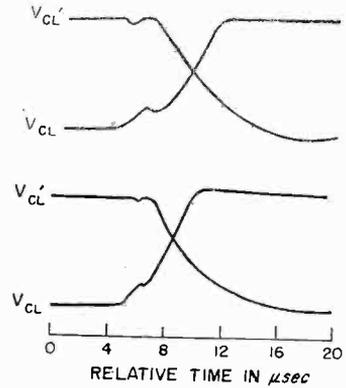


Fig. 9—Counter waveforms, pulse input 5 microseconds.

carry has been shown to be independent of the rise times of the counter flip-flops.

It has also been demonstrated that no progressive widening or narrowing of pulses need occur in a properly designed gated amplifier of the type used in this counter. Pulse width is set by the circuit constants and is essentially independent of transistor gain or storage characteristics.

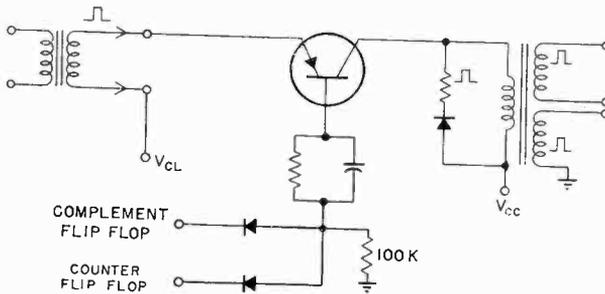


Fig. 10—Pulse amplifier modified for complementing.

ABSTRACTS

SOME NOVEL CIRCUITS FOR THE THREE-TERMINAL SEMICONDUCTOR AMPLIFIER, W. M. WEBSTER, E. EBERHARD AND L. E. BARTON

Abstract—This paper is concerned with circuits for the semiconductor amplifier which consists of two small-area probes, known as emitter and collector, placed on a block of semiconducting material to which is made a third, large-area contact called the base connection.

When such a device was announced, an amplifier circuit was suggested in which the input signal is applied to the emitter and the output taken from the collector, the base being the common electrode. The input impedance of the device when connected in this fashion is low and the output impedance is relatively high, a disadvantageous situation for most applications. In this paper two other amplifier connections are described. The base electrode is used as the input terminal in both of these circuits. In the first case, output is taken from the collector and the input impedance is of the order of the output impedance. In the second circuit, output is taken from the emitter and the input impedance is much higher than the output impedance. Equivalent circuits, both direct current and alternating current, which are adequate representations of the semiconductor amplifier are shown. Algebraic expressions and experimental values of impedance and gain for each of the basic amplifiers are presented. The empirically derived relationships which link the elements of the equivalent circuit to the applied voltages are also given.

The paper discusses two novel oscillator circuits. The first is a two-terminal sine wave generator making use of a very interesting property of negative resistance in the base lead. The second is a simple relaxation oscillator which will furnish either pulse or sawtooth wave form. It makes use of negative resistance effects in the collector circuit when a resistance is placed in the base lead.

RCA Review, Vol. X, March, 1949 (12 pages)

COUNTER CIRCUITS USING TRANSISTORS, E. EBERHARD, R. O. ENDRES AND R. P. MOORE

Abstract—Several transistor multivibrator circuits are described that are capable of performing all the functions required by a decade counter composed of bistable stages, or by a frequency divider. Their basic design features are given and their limitations are pointed out. An analysis of a bistable circuit using a single transistor is given to show how its operation may be correlated with the commonly known transistor constants. Finally, a brief description of two counters built for demonstration purposes, with appropriate wave forms, is presented. It may be concluded that as soon as transistors with stable and more nearly uniform characteristics become available they will be widely used in counter circuits.

RCA Review, Vol. X, December 1949 (18 pages)

A HIGH-PERFORMANCE TRANSISTOR WITH WIDE SPACING BETWEEN CONTACTS, B. N. SLADE

Abstract—A number of transistors having contact spacings ranging between 0.010 inch and 0.020 inch have been made. Power gains of 20 to 30 decibels and current gains as high as 25 are obtained. These values are

as good as or better than those previously reported for narrow-spaced units. An improvement in operational stability may result from the use of wide-spaced contacts through a reduction in the average value of the equivalent base resistance. Current gain falls off more rapidly with frequency in wide-spaced than in narrow-spaced transistors because of transit-time effects. These effects limit the usefulness of wide-spaced transistors to low-frequency applications. However, the technique of activating at wide spacing with an auxiliary contact, then operating with a close-spaced contact makes high current gains and reduced values of equivalent base resistance obtainable at frequencies of 1 to 5 megacycles per second.

RCA Review, Vol. XI, December 1950 (10 pages)

A METHOD OF IMPROVING THE ELECTRICAL AND MECHANICAL STABILITY OF POINT-CONTACT TRANSISTORS, B. N. SLADE

Abstract—The use of thermosetting resins for embedding of point-contact transistors has resulted in a marked improvement in transistor mechanical and electrical stability. Developmental resin-embedded transistors have been subjected to severe impact and centrifuge tests with practically no change in electrical characteristics. Transistors utilizing this construction are highly resistant to the attack of water vapor and are able to withstand extended storage periods at elevated temperatures. Operation at low temperatures is satisfactory, but some changes in electrical characteristics occur at high ambient temperatures. The improvements described have extended the life of developmental transistors and indicate that their use may be feasible in applications having rigorous specifications with regard to mechanical ruggedness, high humidity, and extreme storage temperatures.

RCA Review, Vol. XII, December 1951 (9 pages)

TRANSISTOR OSCILLATORS, E. A. OSER, R. O. ENDRES AND R. P. MOORE

Abstract—The basic oscillators which utilize current-multiplication transistors are described, and their mode of operation is discussed. A fundamental mathematical criterion for oscillation in these circuits is presented, and the physical significance of current feedback is explained as it applies to their operation. The generation of sinusoidal voltages by tuned-circuit, crystal-controlled, and phase-shift-type oscillators is discussed. In connection with these circuits, means are described for improving their high-frequency operation and for obtaining frequency multiplication. Basic relaxation oscillators are presented. These oscillators may be arranged to be free running or triggered. Their operation is explained by means of the voltage and current wave forms developed at the transistor electrodes. By combining the features of the sine-wave oscillator with those of the relaxation oscillator, self-quenching oscillation, or stabilized frequency division may be obtained.

RCA Review, Vol. XIII, September 1952 (17 pages)

JUNCTION TRANSISTOR EQUIVALENT CIRCUITS AND VACUUM-TUBE ANALOGY, L. J. GIACOLETTO

Abstract—The junction transistor possesses operating characteristics that are closely comparable to a modified triode vacuum tube. A direct comparison between the two devices is particularly interesting if a π equivalent circuit is used for the transistor. The vacuum-tube analogy and transistor equivalent circuits are considered in some detail in this paper. For purpose of comparison, a tabulation of operating characteristics of a transistor and a vacuum tube has been prepared.

Proc. I.R.E., Vol. 40, November, 1952 (4 pages)

A DEVELOPMENTAL GERMANIUM P-N-P JUNCTION TRANSISTOR, R. R. LAW, C. W. MUELLER, J. I. PANKOVE AND L. D. ARMSTRONG

Abstract—A developmental germanium p-n-p junction transistor that may be readily made in the laboratory by alloying indium into opposite faces of a wafer of single-crystal n-type germanium is described. It is shown that this laboratory technique gives experimental transistors with desirable characteristics. Distribution curves of measured characteristics are given for a typical run of 118 units made and tested under similar conditions. Power gains up to 46 decibels, alpha up to 0.997, and noise factor (1 kilocycle) as low as 6 decibels were achieved.

Proc. I.R.E., Vol. 40, November, 1952 (6 pages)

P-N JUNCTIONS BY IMPURITY INTRODUCTION THROUGH AN INTERMEDIATE METAL LAYERS, L. D. ARMSTRONG

Abstract—This paper describes an experimental method for making p-n junctions by alloying and diffusing indium into n-type germanium through an intermediate thin layer of some other metal, such as gold, which has been plated on the germanium. The junction characteristics are similar to those of junctions made by other methods, but the shape may be clearly defined and controlled. New possibilities of differentiation between alloying and diffusion are other advantages of the process. Applications have been made to rectifiers and transistors.

Proc. I.R.E., Vol. 40, November, 1952 (2 pages)

THE CONTROL OF FREQUENCY RESPONSE AND STABILITY OF POINT-CONTACT TRANSISTORS, B. N. SLADE

Abstract—The frequency response and stability of point-contact transistors are determined to a large degree by control of the point-contact spacing and germanium resistivity. Stability is particularly important in amplifiers in which the impedances of the emitter and collector circuits are very small in the frequency range in which the transistor is designed to operate. Satisfactory stability has been obtained with developmental transistors having a frequency cutoff (3-decibel drop in the current amplification factor, alpha) ranging from 10 to 30 megacycles. These transistors operate under approximately the same d-c bias conditions used with lower-frequency transistors, and have an average power gain of approximately 20 decibels. By means of the methods outlined, transistors which oscillate at frequencies as high as 300 megacycles have been made.

Proc. I.R.E., Vol. 40, November, 1952 (3 pages)

TRANSISTOR TRIGGER CIRCUITS, A. W. LO

Abstract—This paper concerns trigger and pulse circuits for transistors having emitter-to-collector current gain (such as point-contact units). The circuits are designed to permit reliable operation with transistors which are not completely uniform; they also allow reasonable variation of circuit parameters or bias voltages. Quantitative analysis is made possible by use of simplified circuit theory which divides the nonlinear characteristic of a transistor and its associated feedback and external resistances into quasi-linear regions. By this analysis, one can predict both the type of operation, i.e., monostable, bistable, and astable (oscillatory), and the amplitude and waveform of the output.

A basic monostable circuit consists of a single transistor with a resistance in the base lead to provide feedback and a capacitance or, preferably, a transmission line as the emitter load. This circuit can be used to regenerate periodic or nonperiodic pulses, thereby providing a standardized output pulse shape, or to generate single pulses when initiated; it is possible to

provide very intense short pulses (up to one ampere) even with transistors of low power-handling ability. The output pulses may be arbitrarily delayed with respect to the input, and amplitude discrimination against noise or spurious signals is also possible.

Bistable circuits also use a single transistor with external resistances. It is shown that lack of reliability of previously used bistable single-transistor circuits can be overcome by proper arrangement of circuit parameters and bias supplies and, in some cases, by use of a nonlinear resistance (crystal diode) as the emitter load. Laboratory experience indicates that such single-transistor circuits can be made highly reliable and will allow interchangeability of transistors and reasonable variations in power supplies, and so on. In these respects, they have been made superior even to earlier twin-transistor bistable circuits in which reliability had been emphasized. A small, light-weight decade counter was designed with these single-transistor bistable circuits; it uses less than 2 watts of power in contrast to a well-designed electron-tube counter which needs 60 watts, uses twice as many amplifier elements, and requires considerably more weight and space.

Proc. I.R.E., Vol. 40, November, 1952 (10 pages)

JUNCTION TRANSISTOR CHARACTERISTICS AT LOW AND MEDIUM FREQUENCIES, L. J. GIACOLETTO

Abstract—The low-frequency characteristics of a junction transistor are reviewed and parameters of importance at higher frequencies are determined and examined. Some methods by which both the low- and medium-frequency junction transistor parameters can be measured are presented and discussed. On the basis of these measurements the low-frequency characteristics of the junction transistor are modified to embrace medium frequency operation. The manner in which the medium-frequency characteristics modify the circuit performance is studied.

Proc. National Electronics Conference, 1952, Vol. 8, January, 1953 (9 pages)

CATHODE-RAY TUBE PLOTS TRANSISTOR CURVES, R. D. LOHMAN, J. B. HERZOG AND J. KURSHAN

Abstract—A cathode-ray-oscilloscope curve tracer has been developed that will: 1. display any three-terminal characteristic of a transistor; 2. automatically step the bias through successive constant-current or constant voltage values; 3. accept either n- or p-type transistors; 4. provide full-scale sweep ranges of current from 0.1 milliamperes to 1 ampere and of voltage from 0.1 to 300 volts plus an additional 1,000 volts sweep range at reduced current; 5. provide steps of constant-current bias from 0.1 milliamperes to 1 ampere and of constant voltage bias from 0.1 millivolt to 10 volts; 6. protect the transistor against accidental damage.

These objectives have been met in an all-electronic system which is built into a bench-type relay rack. Either all four quadrants of a volt-ampere coordinate system may be displayed or else any one quadrant may be presented on the entire useful area of the oscilloscope tube. In doing this, the origin may be shifted electrically to the appropriate corner of a square marked on the tube face, or preferably, the deflections may be suitably reversed to convert to a first-quadrant display. A family of eleven curves is automatically displayed sequentially and for viewing purposes a long-persistence screen is used to give a continuous presentation. When a photographic record is desired, the coordinate axes can also be recorded with calibrated subdivisions superimposed.

Electronics, Vol. 26, February, 1953 (6 pages)

FACTORS IN THE DESIGN OF POINT-CONTACT TRANSISTORS, B. N. SLADE

Abstract—Electrical characteristics of point-contact transistors depend essentially on four main factors: (1) the materials used for the point contacts, (2) the spacing of the point contacts, (3) the resistivity of the germanium, and (4) the electrical forming process. Control of these four factors during transistor fabrication makes possible the control of equivalent circuit resistances, current amplification factor, static characteristic curves, and frequency response and, therefore, permits the design of different transistors each suitable for use in a specific type of circuit application. This paper discusses the design of point-contact transistors for use in radio-frequency amplifiers, oscillators, and switching or counter circuits, and the effects of electrical forming on the electrical characteristics. RCA Review, Vol. XIV, March, 1953 (11 pages)

TERMINOLOGY AND EQUATIONS FOR LINEAR ACTIVE FOUR-TERMINAL NETWORKS INCLUDING TRANSISTORS, L. J. GIA-COLETTO

Abstract—With the advent of transistors, considerably greater usage has been made of linear active four-terminal network theory. Some of this usage has been complicated because of differences in terminology. In this paper a unified system of nomenclature is developed for a linear active four-terminal network, and this system is then applied to transistors. It is hoped that this paper will serve as a step towards a standard system of nomenclature.

Section I deals with the general properties of a linear active four-terminal network. Section II is devoted to a tabulation of circuits associated with transistors. Several quantitative examples of the application of material in Sections I and II to transistor circuits are given in Section III.

RCA Review, Vol. XIV, March, 1953 (19 pages)

EQUIPMENTS FOR MEASURING JUNCTION TRANSISTOR ADMITTANCE PARAMETERS FOR A WIDE RANGE OF FREQUENCIES, L. J. GIACOLETTO

Abstract—The small-signal operation of a transistor is accurately specified by means of four complex parameters having both a resistive and a reactive component. Therefore, eight quantities must be measured, and since these quantities in a fixed environment are potentially a function of operating voltage, current, and frequency, the measurement equipment must have considerable flexibility. This paper considers the design, construction, and operation of special equipments operating on the bridge principle for measuring admittance parameters of junction transistors. These bridge equipments operate in the frequency range of approximately 1 kilocycle to 1 megacycle, although by suitable modifications, the operating frequency range can be extended.

An important feature of the operation of the equipments is the use of a multi-frequency test signal such as a square wave, pulse, or swept frequency. With this mode of operation, multi-element equivalent circuit representations can be obtained which are valid over a wide range of frequencies so that a relatively complex measurement task is considerably simplified.

RCA Review, Vol. XIV, June, 1953 (19 pages)

SYMMETRICAL PROPERTIES OF TRANSISTORS AND THEIR APPLICATIONS, G. C. SZIELAI

Abstract—Transistors have certain characteristics which are not present in vacuum tubes. Some of these characteristics may be best classified

as symmetrical properties. The first kind of symmetry may be found in the complementary characteristics of the n-p-n and p-n-p transistors. Circuits using both kinds of transistors in combination provide unique advantages in reduction of components and other circuit simplifications.

A second kind of symmetry is displayed by single units since the emitter and collector may be interchanged. This symmetry permits a current flow of either direction controlled alike by the base current. This basic property is useful in switching circuits for clamping, phase and frequency comparison, modulation, and so forth. A high-efficiency deflection-current circuit for television was developed using this principle.

Proc. I.R.E., Vol. 41, June, 1953 (8 pages)

A STUDY OF TRANSISTOR CIRCUITS FOR TELEVISION, G. C. SZIKLAI, R. D. LOHMAN AND G. B. HERZOG

Abstract—This paper describes a general study of transistors in television receivers. For this purpose the development of a completely transistorized television receiver was undertaken. An experimental model using 57 developmental transistors and five-inch kinescope housed in a cabinet $13 \times 12 \times 7$ inches was constructed. This portable receiver operates on a single channel using a self-contained loop, and has a total battery-power consumption of 13 watts, more than 25 per cent of which is consumed by the kinescope heater.

The development of a complete experimental receiver, even with a number of compromises, provided an opportunity to deal with the problems found in every stage and circuit of the receiver. Although experimental point-contact transistors have recently been developed which will provide oscillations for the entire VHF television band, considerable difficulty was found in providing wide-band r-f gain using transistors at these frequencies. This problem was much less difficult at intermediate frequencies and at the intercarrier-sound frequency. The second-detector problem of obtaining high rectification efficiency with low load impedances was solved by using a transistor detector. The video amplifier problem was complicated by the requirement for a high input impedance; however, with a combination of junction and point-contact transistors a stable, high-gain video amplifier with a relatively high input impedance was built. An audio system using complementary symmetrical junction transistors was designed to produce high output with good efficiency.

In the synchronous and deflection portion of the receiver, circuits were devised for using transistors in ways that differ from the analogues of vacuum tubes. A single transistor was used as a d-c setter, sync separator, and sync amplifier. A simple and reliable horizontal AFC system was developed by utilizing the symmetrical properties of transistors. Point-contact transistors were found to be particularly economical pulse and sawtooth oscillators. The complementary symmetry principle was used to provide vertical deflection with high linearity and efficiency. In the horizontal deflection circuits the fast high-current switching ability of transistors was used advantageously.

Proc. I.R.E., Vol. 41, June, 1953 (9 pages)

THEORETICAL RESISTIVITY AND HALL COEFFICIENT OF IMPURE GERMANIUM NEAR ROOM TEMPERATURE, P. G. HERKART AND J. KURSHAN

Abstract—The resistivity of high-quality single-crystal germanium is determined by its impurity content and, in turn, resistivity can be used as a measure of purity. The semiconductor device engineer will find it most convenient to specify germanium purity in electrical terms by its con-

ductivity type (*n* or *p*) and its resistivity at some standard temperature such as 25°C. In this paper, the temperature variation of resistivity over the range -100°C to +140°C has been calculated and plotted for both *n*-type and *p*-type germanium with different impurity content, ranging from 0.1 ohm-centimeter to 60 ohm-centimeters at 25°C.

When germanium is first purified and then intentionally doped with a single impurity, it is desirable to know the relationship between actual impurity content and resistivity or Hall coefficient. Ordinarily, with reasonably perfect crystals, resistivity can be used to specify impurity; on the other hand, for highly doped material or if crystal perfection is uncertain, the Hall coefficient is a more reliable index. These relationships have been calculated for 25°C and are plotted, again for both *n*- and *p*-types. The curves can be used either to predict electrical values from known impurity content, or to interpret measured electrical values in terms of germanium analysis.

A useful rule of thumb that applies between 20 ohm-centimeters and 0.1 ohm-centimeter at 25°C (the resistivity range useful for transistors) gives the inverse proportionality between impurity content and resistivity ρ in ohm-centimeters, as follows:

$$\text{mol-fraction of impurity} = \frac{3.8 \times 10^{-8}}{\rho} \quad (\text{for } n\text{-type germanium})$$

$$\text{mol-fraction of impurity} = \frac{8.1 \times 10^{-8}}{\rho} \quad (\text{for } p\text{-type germanium}).$$

RCA Review, Vol. XIV, September, 1953 (14 pages)

COMPLIMENTARY SYMMETRY TRANSISTOR CIRCUITS, R. D. LOHMAN

Abstract—The basic principles of complementary symmetry are discussed and the distinction is made between static and dynamic symmetry. Several prototype circuits are analyzed and practical realizations of these are illustrated. Detailed performance of a class-B audio power amplifier employing 4 transistors is included. A vertical deflection system for television is also derived from this circuit.

Electronics, Vol. 26, September, 1953 (4 pages)

MASS SPECTROMETRIC ANALYSIS OF GERMANIUM, R. E. HONIG

Abstract—A mass spectrometric method, employed in conjunction with a physical process of impurity enrichment, has yielded semiquantitative analytical results on germanium samples used in transistor work. After a two-stage treatment in a gradient furnace, the enriched sample is vaporized thermally in a mass spectrometer, at temperatures that are raised in suitable steps. As a result, most impurities are separated, in temperature and time, from the bulk germanium. In this fashion, it seems possible to extend instrumental sensitivities from a present mole fraction limit of 10^{-5} to perhaps 10^{-8} . After the sample is completely evaporated, a time integration of ion currents yields concentration estimates, whose accuracy is mainly limited by uncertainties in the ionization cross sections. In analyses made on Eagle-Picher germanium, small concentrations of the following impurities were found: arsenic, phosphorus, antimony, tin, lead, irann, and cadmium. Calibration runs that have been made on germanium, tin, and lead show that these Group IV B elements evaporate partially in molecular form, while in the case of arsenic the molecular species As_2 and As predominate.

Analytical Chemistry, Vol. 25, October, 1953 (5 pages)

VEST-POCKET TRANSISTOR ALPHA METER, B. P. KERFOOT

Abstract—To fill the need for a rapid method of testing the amplifying ability of a transistor, a simple test circuit is described. When this circuit is used in conjunction with a vacuum-tube voltmeter (such as Ballantine model 300), direct readings of the base to collector current gain property (to an accuracy of usually five per cent) can be made.

Use of a line-frequency signal is shown, although a wide range of test-signal frequencies could be employed. Bias and signal are applied as the transistor is plugged in, so that readings are instantaneous.

Electronics, December 1953 (5 pages)

A GERMANIUM N-P-N ALLOY JUNCTION TRANSISTOR, D. A. JENNY

Abstract—This paper describes a germanium n-p-n alloy junction transistor which is the counterpart to the germanium p-n-p junction transistor previously described. The importance of this new device arises from a fundamental difference between the two types. In the p-n-p transistor the active charge carriers are positive "holes"; in the n-p-n transistor the active charge carriers are negative electrons. Because these devices operate from power sources of opposite polarity, the two types may be advantageously combined in special circuits to eliminate components and fulfill unusual requirements. Because the electron mobility is more than twice that of holes, one of the factors affecting high-frequency response is more favorable for the n-p-n transistor than for its p-n-p counterpart.

This n-p-n junction transistor is made by fusing a binary lead-antimony alloy into each of the two opposite faces of a thin wafer of p-type single-crystal germanium. Since this alloy is ductile, the electrodes may be made relatively large if desired, as there is less danger of introducing differential expansion strains. The techniques and processes of assembly are similar to those employed for p-n-p junction transistors by the alloy process. However, a difference arises from the more uniform penetration afforded by the binary alloy. This leads to more planar junctions and permits better control of junction spacing.

Distribution curves on a typical lot of 100 units are given; best power gain was 45 decibels, alpha 0.997 and 1-kilocycle noise factor, 3 decibels. High alpha is maintained as the collector current is increased.

Proc. I.R.E., Vol. 41, December, 1953 (7 pages)

SOME RECENT DEVELOPMENTS IN THE CALCULATION OF CRYSTAL ENERGY BANDS — NEW RESULTS FOR THE GERMANIUM CRYSTAL, F. HERMAN

Abstract—Some current methods for determining the electronic structure of crystals are discussed briefly. Among the techniques treated are the cellular method, the orthogonalized plane wave (OPW) method, and the augmented plane wave method.

A recently completed calculation of the band structure of the germanium crystal by the OPW method is described. It is found that the band structure obtained depends quite critically upon the assumed crystal potential. The structure of the top of the valence band, which lies at the central point of the reduced zone, is strongly influenced by spin-orbit interaction and by the proximity of the lower conduction bands. The crystal eigen solutions belonging to the two lowest conduction bands converge at different rates and respond in different ways to arbitrary changes in the crystal potential. The latter fact is used as the basis for an interpretation of certain optical properties of germanium-silicon alloys.

Our solution indicates that the lowest conduction band had a well-defined minimum at the central point of the reduced zone. If our solution were carried to a higher order approximation, we would expect to find the minimum of this conduction band at positions along the [111] axes rather than at the central point of the zone. The next-to-the-lowest conduction band has six minima along the [100] axes. As we proceed from pure germanium to pure silicon through intermediate compositions of Ge-Si alloys, the states normally occupied by electrons shift from the [111] minima to the [100] minima.

Physica, Vol. XX, 1954 (12 pages)

A COMPARISON OF ANALOGOUS SEMICONDUCTOR AND GASEOUS ELECTRONICS DEVICES, W. M. WEBSTER

Abstract—From the standpoint of the fundamental electron physics involved, as well as the likely areas of application, semiconductor devices bear a strong resemblance to gas tubes. In this paper the physics of the two fields is reviewed, the analogies pointed out, and the theoretical limitations of analogous devices considered.

Advances in Electronics and Electron Physics, Vol. VI, 1954 (37 pages)

POWER TRANSISTORS OF AUDIO OUTPUT CIRCUITS, L. J. GIACOLETTO

Abstract—The operation of a power transistor is analyzed with the aid of two d-c characteristic equations suitably modified to include the base-lead resistance. The d-c characteristic equations are expressed in terms of three d-c current coefficients determined by measurements of the transistor as a passive device. It is shown that the important operating characteristics can be obtained from transistor static transfer curves. These curves can be developed from the theoretical expressions to include the effect of base-lead resistance and internal resistance of the driving source. Measured and calculated static transfer characteristics are compared. The important aspects of frequency and temperature effects are considered, the latter, in terms of a temperature correction factor applied to static transfer characteristics. The design characteristics of power transistors in class-A and push-pull class-B output stages are considered. Some attention is given to the manner in which these stages can be biased.

Electronics, Vol. 27, January, 1954 (5 pages)

BEHAVIOR OF GERMANIUM-JUNCTION TRANSISTORS AT ELEVATED TEMPERATURES AND POWER TRANSISTOR DESIGN, L. D. ARMSTRONG AND D. A. JENNY

Abstract—This paper discusses the limitations of germanium-junction transistors at elevated operating temperatures. The limiting factors are a consequence of increased thermal hole-electron pair generation at higher temperatures. This causes an increase in collector "leakage" current, which effects the base current. Thus, due to the base-lead resistance, the emitter-to-base bias conditions are changed.

The problems associated with power transistors are discussed, including cooling of the unit, current and voltage limitations, and reduction in base-lead resistance. It is pointed out that, in addition to bias changes, base lead resistance also causes a power loss which becomes important at the high currents used in power units.

Examples of liquid-convection-cooled and metallic-conduction-cooled n-p-n and p-n-p power transistors are described. Satisfactory alpha values are maintained to collector currents of over one hundred milliamperes. In class-A operation power gains exceeding 30 decibels and efficiencies close to the theoretical limit of 50 per cent are obtained at 1 watt dissipation levels.

Proc. I.R.E., Vol. 42, March, 1954 (3 pages)

HIGH-FREQUENCY OPERATION OF P-TYPE POINT-CONTACT TRANSISTORS, F. L. HUNTER AND B. N. SLADE

Abstract—This paper discusses the characteristics of developmental p-type transistors in radio-frequency-amplifier or oscillator applications. Higher frequencies can be obtained from transistors using p-type germanium rather than n-type germanium because of the greater mobility of the minority carriers (electrons) in the former as compared to the mobility of the minority carriers (holes) in the latter. Short-circuit-stable p-type amplifier transistors have been developed having a current-amplification-factor cutoff frequency of 50 to 60 megacycles. P-type oscillator transistors have been operated at frequencies up to 425 megacycles. Experimental tests have shown that both p-type and n-type oscillator transistors perform satisfactorily at temperatures up to approximately 75° C.

RCA Review, Vol. XV, March, 1954 (14 pages)

RECRYSTALLIZATION OF GERMANIUM FROM INDIUM SOLUTION, J. I. PANKOVE

Abstract—Upon cooling a germanium-in-indium solution in contact with solid germanium, germanium from the supersaturated solution recrystallized onto the solid crystal in epitaxial fashion. The growth forms a group of separate [100] needles consisting of a stack of (111) plates, many of which are hollow. Although the outer portion of the recrystallized region may be rich in defects, there is a relatively perfect continuous recrystallized layer or overgrowth region adjacent to the p-n junction. This overgrowth region is a single crystal p-type semiconductor which continues the lattice of the n-type seed. The overgrowth is thinnest in the (100) direction.

There is some indication that the alloy front, and hence the p-n junction, tend to be flattest in (111) planes. The p-n junction seems to coincide with the alloy front at least within the limits of the present experimental resolution.

RCA Review, Vol. XV, March, 1954 (11 pages)

A SYMMETRICAL-TRANSISTOR PHASE DETECTOR FOR HORIZONTAL SYNCHRONIZATION, B. HARRIS AND A. MACOVSKI

Abstract—This paper describes the application of a developmental symmetrical transistor as a balanced phase detector for controlling the horizontal oscillator of a television receiver. A high degree of balance and uniformity of performance under unfavorable conditions of transistor symmetry and temperature has been obtained. The transistor phase detector was used to control a stabilized multivibrator type of horizontal oscillator having a sensitivity of approximately 150 cycles per second per volt. Approximately ten volts of single-ended sawtooth and thirty volts of single-ended sync were required to produce a pull-in of 120 to 180 cycles per second.

Wide variations in the absolute values of transistor parameters and in the degree of symmetry resulted in less than 0.5-volt unbalance in the absence of sync. Laboratory measurements indicated that the performance was relatively unaffected by temperatures up to 55° C. These results were obtained with transistors having values of r_c , l_{co} , and α_{cb} which varied as much as 3:1 when the collector-emitter roles were interchanged.

Symmetrical transistors are constructed in the same way as conventional junction transistors with the exception that the collector and emitter junctions are made identical. The circuit was tested with two varieties of symmetrical transistors. One had both junctions 15 mils in diameter, the other had both junctions 45 mils in diameter. The two varieties of transistors performed equally well.

RCA Review, Vol. XV, March, 1954 (9 pages)

TRANSISTOR EQUATIONS USING h-PARAMETERS, C. C. CHENG

Abstract—This paper presents a simplified method for determining operating characteristics of transistor circuits. The proposed method, which is based on the use of four h -parameters of the common-emitter base-input circuit, is also applicable to the common-base emitter-input circuit and the common-collector base-input circuit. Equations are given for both high-frequency and low-frequency operation. Measurements and calculations are simplified by the use of the four fundamental parameters for all equations. Electronics, Vol. 27, April 1954 (3 pages)

RESISTIVITY STRIATIONS IN GERMANIUM CRYSTALS, P. R. CAMP

Abstract—Germanium metal, both single crystalline and polycrystalline, is frequently found to contain sharp fluctuations in impurity concentration. These are superimposed on the gradual variations expected as the result of ordinary segregation processes.

A rather simple technique for detecting fluctuations of this kind has been developed. It is more sensitive than resistivity scanning and does not involve the use of radioactive tracers. This technique has been applied to a number of samples. The origin, value, and elimination of these fluctuations are discussed briefly.

Jour. Appl. Phys., Vol. 25, April, 1954 (4 pages)

ON THE VARIATION OF JUNCTION-TRANSISTOR CURRENT AMPLIFICATION FACTOR WITH EMITTER CURRENT, W. M. WEBSTER

Abstract—Existing theories of the junction transistor fail to predict the very significant variation of current-amplification factor, α_{cb} , as the emitter current is varied. This variation has been very troublesome in power transistors, particularly at high emitter currents where the α_{cb} fall-off may be so severe as to limit usefulness. At low currents, α_{cb} also drops off, an effect of importance in very low-power applications. By taking into account modification of the base region by the injected charge carriers, an explanation is found for the observed variation. Electric fields in the base region decrease the mean transit time for minority carriers on their way to the collector. This reduces the effect of surface recombination and increases current-amplification factor as the emitter current rises. Another effect, however, is in the opposite direction; this second effect is due to an increase in conductivity of the base material which increases the rate of volume recombination and also lowers emitter efficiency. The combination of these effects yields calculated curves which show a maximum and agree well with experiment. The work is applicable to both p - n - p and n - p - n types, and it is shown that the latter is inherently less sensitive to emitter current density.

Proc. I.R.E., Vol. 42, June, 1954 (7 pages)

THE EFFECT OF JUNCTION SHAPE AND SURFACE RECOMBINATION ON TRANSISTOR CURRENT GAIN, A. R. MOORE AND J. I. PANKOVE

Abstract—An experimental and theoretical study is presented which shows that the current gain of an alloy transistor is greatly affected by the geometry of emitter and collector junctions and by surface treatment of the base germanium, but is hardly affected at all by bulk recombination (lifetime) in the base. The current gain is computed for specific three-dimensional geometries by an electric analog method which assumes that surface recombination is the major factor in minority-carrier loss. By this method, a new way of measuring surface-recombination velocity, s , from simple measurements on transistors has been devised. The value of s is

obtained directly from a suitable calibration curve, and thus may be useful as quality control on surface condition.

The transit-time path-length dispersion of minority carriers in a transistor structure with nonparallel junctions has been computed. The results show that the effect is significant only above 1 mc/second in a typical structure.

Proc. I.R.E., Vol. 42, June, 1954 (7 pages)

AN EXPERIMENTAL TRANSISTOR PERSONAL BROADCAST RECEIVER, L. E. BARTON

Abstract—This paper describes a laboratory-model AM broadcast receiver which used nine alloy-junction transistors and two compensating diodes. Six of the transistors were of the laboratory p-n-p type, designed for radio-frequency amplifiers, and three of the transistors were of the conventional p-n-p type selected for class-B audio driver and output service. The use of class-B output permitted a total battery drain below 12 milliamperes from six 1.5-volt type C cells in series. The battery life was approximately 500 hours, and the battery cost relatively small. The maximum audio-power output was 150 milliwatts into a four-inch by six-inch oval speaker. The sensitivity and signal-to-noise ratio were comparable to that of conventional battery-operated receivers and conventional a-c/d-c receivers. The receiver may be used in the place of a-c/d-c receivers without the necessity of a power outlet or a connecting power cord at a battery cost approximately the same as the cost of power for an a-c/d-c receiver.

Proc. I.R.E., Vol. 42, July, 1954 (5 pages)

PHYSICAL THEORY OF NEW CIRCUIT REPRESENTATION FOR JUNCTION TRANSISTORS, J. ZAWELS

Abstract—A circuit representation for a junction transistor is derived which places the mechanism of transistor action in evidence. The circuit is a direct interpretation of the diffusion equation and the boundary conditions which include the effects of base width modulation. It is found that the active part of the circuit is independent of frequency. A modification of the circuit for common-emitter operation is shown and experimental results are given.

Jour. of Appl. Phys., Vol. 25, August, 1954 (6 pages)

STUDY OF P-N-P ALLOY JUNCTION TRANSISTOR FROM D-C THROUGH MEDIUM FREQUENCIES, L. J. GIACOLETTO

Abstract—With the increase in commercial importance of transistors, it has become necessary that the various factors that enter into the construction of a transistor be ascertained and delineated. This study was undertaken with the object of arriving at various relationships which could be used in the design of transistors of predetermined characteristics. Although d-c characteristics were measured and studied, the main emphasis is on small-signal operation. Detailed measurements were made of a junction transistor, and these measurements are compared with theoretical calculations. Existing theories are modified and extended as it appears necessary. Various measurements are used to evaluate important constants of the materials used and the dimensions involved.

Generally, good agreement is found between theory and measurements. It should accordingly be feasible to carry out a large part of the design of a transistor without recourse to the construction of devices. In addition, the manner in which the transistor operation is dependent upon operating voltage, current, frequency, and temperature can be predicted within a workable degree of accuracy so that both device and circuit designers can carry out the calculations necessary to obtain the answers desired.

RCÁ Review, Vol. XV, December, 1954 (57 pages)

RECOMBINATION PROCESSES IN INSULATORS AND SEMICONDUCTORS, A. ROSE

Abstract—The discrete states in the forbidden zone are divided into ground states and shallow trapping states. The major recombination traffic passes through the ground states. The shallow trapping states cause the observed decay time of free carrier concentrations to exceed the lifetime of a free carrier in the conduction (or valence) band. At low rates of excitation (free carrier concentrations less than ground state concentrations) the electron lifetime and hole lifetime are independent and, in general, significantly different. At high rates of excitation the free electron and hole lifetimes are equal. For an insulator having one class of ground states (a class being defined by the capture cross sections for electrons and holes) the high-light lifetime is bracketed by the two low-light lifetimes.

The behavior of a model having one class of ground states can be described relatively simply and quantitatively. The behavior of a model having more than one class of ground states becomes sufficiently complex that only special cases can be treated easily. More than one class of ground states, however, is required to account for infrared quenching, "superlinearity" and the ability of added ground states to reduce the rate of recombination. These phenomena involve a redistribution of electrons and holes amongst the classes of ground states. Such redistributions can give some meaning to the phrases: "filling of traps" or "saturation of centers."

The recombination behavior of a semiconductor is significantly different from that of an insulator. For example, superlinearity can occur in a semiconductor having only one class of ground states. Also, the photocurrents in a semiconductor can be intrinsically more noisy than the photocurrents in an insulator.

Phys. Rev., Vol. 97, January 15, 1955 (12 pages)

SEMICONDUCTORS AND THE TRANSISTOR, E. W. HEROLD

Abstract—After a short survey of the history of semiconductors and their use in rectifiers and transistors, the periodic table is used to show the paucity of elements which are semiconducting. Only silicon and germanium are presently of great interest. On the other hand, compounds and mixtures are manifold and are being widely studied. An explanation is given of rectifying properties between junctions of differing impurity types and the techniques which produce such junctions. The characteristics of commercial point contact and junction transistors are reviewed and laboratory developments of greatly improved high-frequency and high-power units are outlined. The paper concludes with a brief description of a silicon "atomic battery" and some remarks about the future.

Jour. Frank. Inst., Vol. 259, February 1955 (20 pages)

COMPARATIVE HIGH-FREQUENCY OPERATION OF JUNCTION TRANSISTORS MADE OF DIFFERENT SEMICONDUCTOR MATERIALS, L. J. GIACOLETTO

Abstract—The high-frequency performance of junction transistors is determined in part by the time required for the injected minority carriers to traverse the base region. Large minority carrier mobility contributes to reducing this transit time. However, the high-frequency performance is also affected by factors such as base-lead resistance and collector-to-base junction capacitance, and it is shown that large majority carrier mobility will reduce the detrimental effects of these factors.

An important result of the analysis of high-frequency operation is the discovery that both minority and majority carrier mobilities are of about equal importance. It follows that the high-frequency performance of an

n-p-n transistor should be about the same as that of a geometrically identical *p-n-p* transistor. Furthermore, for evaluation of new semiconductors, a knowledge of the values of both mobilities is required, and a figure of merit is proposed which is formed by the product of the two drift mobilities divided by the square root of the dielectric constant.

RCA Review, Vol. XVI, March, 1955 (9 pages)

DELAYED COLLECTOR CONDUCTION, A NEW EFFECT IN JUNCTION TRANSISTORS, M. G. KIDD, W. HASENBERG AND W. M. WEBSTER

Abstract—A new mode of operation for junction transistors called "delayed collector conduction" (DCC) is found at higher-than-normal collector voltages. At these voltages, the collector current suddenly rises, as though a breakdown had occurred. This effect can be controlled, however, by adjustment of the base current, and is nondestructive. Because this new type of junction-transistor amplification has high input impedance and low output impedance, junction transistors can be used in simple high-speed switching circuits formerly employing only point-contact transistors. The DCC effect can be explained by an extension of the theory of junction transistors to include avalanche multiplication at the collector junction. Although the DCC mechanism is similar to breakdown in a cold-cathode gas tube, junction transistors operating in this mode differ from gas tubes in that the control electrode (base) retains control after conduction has been initiated.

The low output impedance of DCC operation has been applied in amplifiers, switches, and controllable voltage regulators. Pulse and sine-wave oscillators and astable, monostable, and bistable multivibrators have been designed utilizing the negative characteristic. Rise times of 0.01 microsecond observed for audio-frequency junction transistors compare favorably with point-contact-transistor behavior.

RCA Review, Vol. XVI, March, 1955 (18 pages)

SATURATION CURRENT IN ALLOY JUNCTIONS, W. M. WEBSTER

Abstract—According to theory, the current which flows across a *p-n* junction when it is biased in the reverse direction should be nearly independent of the applied voltage. This current is called the saturation current, I_s . In actual practice, the current often increases with voltage (due to leakage paths across the junction at the surface) and eventually large currents flow as the breakdown voltage of the junction is approached. However, for "good" junctions well below their breakdown voltage, the reverse current follows theory so that the magnitude of I_s is of considerable practical interest.

A previous equation for saturation current has been given which applies where the effect of free surfaces may be ignored and where there are quite thick (greater than a diffusion length) layers of material on either side of the junction. For alloy junctions made on thin pieces of semiconductor, these conditions are violated.

In this paper a new equation is developed which applies specifically to diodes made by alloying circular junctions on thin wafers. Over its range of applicability, it gives accurate values of I_s . It is shown that most of the saturation current comes from thermal generation at the free surfaces of the base wafer. The equation for I_s consists of a geometrical term and a coefficient which depends on the physical constants of the material. The principal dependences are these: I_s increases linearly with base wafer resistivity and exponentially with temperature. It also increases, but more slowly, with wafer thickness and surface recombination velocity.

For the collector of an alloy junction transistor, the basic equation is the same as for a diode, except for a small correction involving the emitter area. For the emitter junction, reverse saturation current is often not of great interest but can also be calculated; this requires a modification of the basic equation which is also given in this paper.

Proc. I.R.E., Vol. 43, March, 1955 (4 pages)

MEASUREMENT OF MINORITY CARRIER LIFETIME AND SURFACE EFFECTS IN JUNCTION DEVICES, S. R. LEDERHANDLER AND L. J. GIACOLETTO

Abstract—The characteristics of junction devices are influenced to a considerable degree by the lifetime of the minority carriers. Accordingly, methods for the measurement of this quantity are of considerable importance. Methods have been described for the measurement of the lifetime of minority carriers when these carriers are produced within the volume of a semiconductor. When the minority carriers are introduced near the surface of a semiconductor the resulting effect lifetime may be determined to a large extent by the nature of the surface. For most junction devices, it is the effective lifetime that is of primary importance.

This paper describes a simple method for the measurement of effective lifetimes of injected minority carriers. The measurements may be applied to practical junction structures as, for example, an alloyed junction transistor. Measurements may be made on either completed or partially completed devices. The resulting data are potentially of value as quality controls during the fabrication of transistors and similar devices.

In many cases, the effective lifetime is a good indication of the surface conditions, and immediate evaluation of these conditions may be obtained at various stages of device processing. With selected geometries, the measurement method may be applied to determine absolute values of surface recombination velocities and should therefore be in studying surface conditions and treatments.

The measurement method is described in terms of junction devices using germanium as the semiconductor. However, the method is equally applicable to junction devices made with other semiconductor materials.

Proc. I.R.E., Vol. 43, April, 1955 (6 pages)

THE EFFECTIVE SURFACE RECOMBINATION OF A GERMANIUM SURFACE WITH A FLOATING BARRIER, A. R. MOORE AND W. M. WEBSTER

Abstract—The effect of heavily doped (alloyed) p-type and n-type surface layers on n-type base, and of metallic plating on n-type base, on the surface recombination velocity, s , has been computed on the basis of one-dimensional junction theory. The results indicate that s should be of the order 1 centimeter per second for the heavily doped surfaces, and several thousand centimeters per second for the electroplated surface. The low s comes about for the same reason that the injection efficiency of alloy junctions is high; the alloy junction is a very efficient emitter of minority carriers into the base and a poor acceptor of majority carriers from the base because of the high doping level in the alloyed region. Since recombination in the surface layer of minority carriers from the base requires both majority and minority carriers, the restriction of the flow of either reduces the surface recombination.

However, measurements of s by diffusion and pulse methods on alloy junction surfaces indicate that their apparent recombination is almost the same as adjacent untreated surface, e.g., 300-500 centimeters per second. It is shown that lateral current flow, due to minority carrier gradients

parallel to the junction interface, and neglected in one-dimensional theory, gives rise to circulating currents which translate the minority carriers to the nearest high-recombination surface. This hole translation property of the floating p-layer is used to explain the erroneously high lifetimes often observed by diffusion measurements on silicon and p-type germanium, and certain discrepancies in effective life measurement on completed transistors.

Proc. I.R.E., Vol. 43, April, 1955 (9 pages)

SIMPLIFIED DESIGN PROCEDURES FOR TUNED TRANSISTOR AMPLIFIERS, C. C. CHENG

Abstract—This paper describes a systematic and simplified method for the design of tuned amplifier systems using transistors. The method consists of three steps: (1) determination of the requirements of the amplifier system, (2) selection of transistor amplifiers, and (3) selection of coupling networks. Design procedures for various types of coupling networks are presented in tabular form for easy reference.

RCA Review, Vol. XVI, September, 1955 (21 pages)

THE NATURAL EQUIVALENT CIRCUIT FOR JUNCTION TRANSISTORS, J. ZAWELS

Abstract—By considering individually and qualitatively each first-order phenomenon in junction transistor action, a circuit is synthesized in which each element represents a particular phenomenon. This approach insures an exact equivalent circuit which is fundamental to all junction transistors (not only to those having a plane parallel geometry) and which may be extended to include point contact transistors and other semiconductor devices.

It is found that the equivalent circuit reduces without loss of accuracy to a passive network in cascade with a frequency-independent amplifier. This configuration has considerable practical advantages particularly in common-base operation. Circuits which are especially useful for common-emitter and common-collector operation, and which employ the same basic passive elements, are also shown.

Finally the h-parameter expressions for the circuit are examined and a method is shown whereby the magnitude of the elements of the equivalent circuit may be simply computed from low-frequency h-parameter spot measurements.

RCA Review, Vol. XVI, September, 1955 (19 pages)

METHODS FOR REVEALING P-N JUNCTIONS AND INHOMOGENEITIES IN GERMANIUM CRYSTALS, J. I. PANKOVE

Abstract—When a p-n junction is biased in the reverse direction under an electrolyte, the junction contour appears as the n-type region etches off, while the p-type region remains relatively undisturbed. The ease of distinguishing the n and p regions can be improved by using a hot germanium-rich electrolyte. A plurality of junctions in the same crystal can be revealed by coating the crystal with a thin layer of electrolyte and passing a current through the crystal. Regions of different impurity concentrations can be shown by etching the crystal through the film of electrolyte with respect to a small negative electrode which scans the crystal.

RCA Review, Vol. XVI, September, 1955 (5 pages)

TRANSISTORS I

RCA LABORATORIES