p-n Junction Transistors

W. SHOCKLEY, M. SPARKS, AND G. K. TEAL
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received May 24, 1951)

The effects of diffusion of electrons through a thin *p*-type layer of germanium have been studied in specimens consisting of two *n*-type regions with the *p*-type region interposed. It is found that potentials applied to one *n*-type region are transmitted by diffusing electrons through the *p*-type layer although the latter is grounded through an ohmic contact. When one of the *p-n* junctions is biased to saturation, power gain can be obtained through the device. Used as "*n-p-n* transistors" these units will operate on currents as low as 10 microamperes and voltages as low as 0.1 volt, have power gains of 50 db, and noise figures of about 10 db at 1000 cps. Their current-voltage characteristics are in good agreement with the diffusion theory.

I. INTRODUCTION

IN this article we shall consider the phenomena which occur when voltages are applied to a semiconductor consisting of several regions of different conductivity types. Structures of this sort, consisting in particular of two regions of one conductivity type separated by another region of the opposite type, are of great practical interest in transistor electronics. Such structures can also be used to exhibit the behavior of hole and electron diffusion in rather impressive ways. In particular, the phenomenon of "internal contact potentials" can be strikingly demonstrated with such structures.

Transistors in which the nonlinear effects originate within the germanium as a result of the relationships of *p*-type and *n*-type regions are called "*p-n* junction transistors" to distinguish them from point-contact types, in which the metal semiconductor contact often plays an essential role. There are a number of possible *p-n* junction transistor structures: The *p-n-p* transistor has been discussed previously from a theoretical viewpoint.[1,2] In this article we shall consider chiefly the *n-p-n* transistor,[3] the *n*-type phototransistor with a *p-n* hook collector and a *p-n-p-n* transistor with *p*-type emitter and *p-n* hook collector.

In the following sections we shall describe first in simple terms the basic phenomena and effects with which we are concerned. We shall next describe the actual physical structure of several of the *n-p-n* transistors and their electrical characteristics. The theoretical principles will then be put in quantitative form and the current-voltage relationships derived for certain particular models. Finally a direct comparison between theory and experiment will be presented.

II. THE *n-p-n* STRUCTURE AS A TRANSISTOR AND AS A "HOOK MULTIPLIER"

In Fig. 1 we represent an *n-p-n* structure and indicate how it may be used as a transistor. Like the type-A transistor, the current paths between emitter terminal and base and between collector terminal and base have rectifying junctions. Unlike the type-A transistor, however, the rectification arises in the interior of the germanium and not at the contacts between metal leads and the germanium; which are substantially ohmic. There are other important differences between the *n-p-n* and the type-A: In the *n-p-n* the flow of injected carriers takes place chiefly by diffusion rather than by drift in an electric field; the current multiplication at the collector, which makes possible the positive feedback instability of the type-A transistor, is lacking in the *n-p-n* transistor.

In this section we shall give a brief resumé of the theory of the operation of the transistor. Additional details will be found in the cited references. In Secs. V, *et seq.*, some aspects of the theory will be treated analytically. The analysis is simplified by the use of the following assumptions:[4]

(1) The donors and acceptors are fully ionized (this is a good assumption for germanium at room temperature).

(2) The density of minority carriers is much smaller than the density of majority carriers in each region.

(3) The net rate of recombination in any region is linear in the deviation of the minority carrier density from its thermal equilibrium value. (Assumptions (2) and (3) permit us to use linear equations in dealing with the currents arising from carrier injection.)

(4) Space charge is negligible except at the space-charge regions in the *p-n* junctions themselves.

In Fig. 1 we show the energy band diagram for the structure under consideration for zero bias and for biases applied in such a way that the unit becomes an amplifying transistor. Under the latter condition the junction J_c on the right of the figure is biased in the reverse direction. This direction is such that electrons in the *n*-type collector region have low potential energy and cannot climb the potential energy hill to the base region; similarly holes are held in the base region. Electrons in the emitter region, however, may climb the small potential hill into the base region and once in this region may diffuse so that some of them will arrive at the right-hand junction. The flow over the hill depends on the height of the hill and this height may

[1] W. Shockley, Bell System Tech. J. **28**, 435–489 (1949).

[2] W. Shockley, *Electrons and Holes in Semiconductors* (D. van Nostrand Company, Inc., New York, 1950).

[3] R. L. Wallace and W. J. Pietenpol, Bell System Tech. J. July, 1951. (This article, to which we shall refer a number of times, deals with a number of practical features which we do not consider here.) It is also scheduled for the Proc. Inst. Radio Engrs. July, 1951.

[4] These assumptions are discussed further in references 1 and 2.

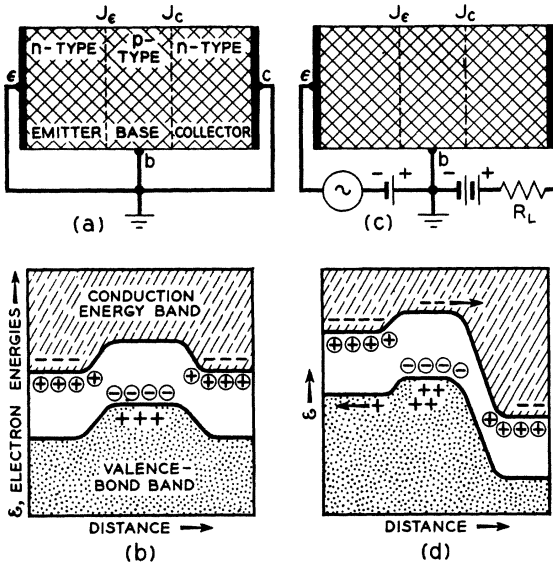

FIG. 1. The *n-p-n* structure and the energy level scheme. (a) and (b) Thermal equilibrium. (c) and (d) Biased as an amplifier.

be varied by applying a variable potential to the emitter while maintaining the base at constant potential. If the base region is very thin, few of the electrons will recombine with holes in it and, as a result, efficient transmission of electron current through the layer will occur. Furthermore, if the emitter region is more highly conducting than the base region, there will be many more electrons available to climb the hill than there are holes to climb the corresponding hill in the opposite direction. As a result, most of the current across the left junction will consist of electrons. Under these conditions the behavior of this device is closely analogous to that of a vacuum tube: the emitter region corresponds to the cathode, the base to the region around the grid wires, and the collector region corresponds to the plate. In favorably designed units the controlled electron current flowing through the base region may be very much larger than the current furnished the base region for control purposes so that the transistor has a current amplification factor that is very high. It may, therefore, be operated like a grounded cathode triode with the emitter region grounded and the signal applied to the base. In Sec. IV some current-voltage characteristics for transistors are shown and for them it may be seen that the current transmission is nearly perfect.

It is interesting to note that both for the vacuum tube and for the transistor, control is accomplished by the interaction of two forms of electron flow. In the vacuum tube, metallic flow in the grid wire controls the flow of thermionic electrons in the space between grid wires. In the transistor, hole flow in the base changes the base-emitter voltage and controls electron flow through the base layer.

The structure shown in Fig. 1 with the same operating biases may be used as a collector with high multiplication in a transistor.[5] It may also be used as a phototransistor.[6] We shall consider the application as a collector in a transistor later and shall describe here how multiplication of photocurrents can occur. For this purpose there need be no electrode connected to the base layer. If light shines on the germanium near the junctions, then the hole electron pairs generated will be separated by the field of the junctions and consequently a current of holes will flow into the base layer. These holes will accumulate in the layer and will charge it positively and thus reduce its potential energy for electrons. As a result more electrons will be able to climb the hill and flow to the collector. The effect of the added holes will die away after the light is removed due to the diffusion of holes into the left region where they combine with electrons and also due to recombination with electrons which diffuse into the base layer. If the layer is very thin, however, and the density of electrons in the left region is very high, then a very large number of electrons will be able to climb over the hill for each hole that is able to enter the emitter region and recombine. In Sec. VII we shall show that the current amplification obtained in this way is proportional to the ratio of the conductivities of the two layers and is inversely proportional to the thickness of the base layer.

An interesting consequence of the diffusion of electrons through the base layer is the occurrence of "internal contact potentials."[7] In order to illustrate these we shall suppose that the base layer is grounded and that a potential is applied to the emitter. If an additional ohmic contact is also made to the base region, it will of course show ground potential. If the contact is rectifying, however, and in particular if it carries most of its current in the form of electrons, its potential will be determined by the electron density in the base layer rather than by the potential established by the grounding contact. The *n*-region on the right represents such a contact. It is found both theoretically and experimentally that if the base layer is grounded, then potentials applied to the region on the left are transmitted through the base layer, although its electrostatic potential is practically unaltered, and exhibit themselves in the region to the right, which tends to "float" (when no current is drawn from it) at a potential approximately equal to that of the left region—at least over a certain range of voltages. Theory and experiment related to this phenomenon are given in Secs. VII and VIII.

III. DESCRIPTION OF EXPERIMENTAL UNITS

The experimental units were made of a piece of single crystal germanium in which a thin *p*-type layer is

[5] W. Shockley, Phys. Rev. **78**, 294 (1950).

[6] J. N. Shive who has developed the phototransistor [Phys. Rev. **76**, 575 (1949)] has proposed this use of the *p-n* hook.

[7] These were first discussed in reference 1. Internal contact potentials for point contacts have been measured by G. L. Pearson and analyzed by J. Bardeen [Bell System Tech. J. **29**, 469–495 (1950)].

interposed between two n-type sections. Units of a variety of conductivity values have been prepared. Typical values for an n-p-n structure as shown in Fig. 1 are: emitter, 100 (ohm-cm)$^{-1}$; base, 1 (ohm-cm)$^{-1}$; and collector, 0.1 (ohm-cm)$^{-1}$. Typical values of lifetime of the minority carrier in the collector section are 300–400 microseconds. The experimental evidence is that lifetimes in the other sections, in which direct measurement of lifetimes is more difficult, do not differ greatly from this. The leads to the three sections are mechanically strong and ohmic in character. The structure thus operates through conditions arising within the interior of the single crystal and not because of phenomena arising at the contacts between the leads and the germanium.

IV. PERFORMANCE CHARACTERISTICS

The performance characteristics of n-p-n transistors will be considered briefly.[8] Since these devices operate uniformly over the surfaces of p-n junctions they may be greatly altered as to the size of the active area, in contrast to point-contact devices. Thus it is possible to increase power output without corresponding increases in current density. One of the larger n-p-n transistors[9] studied as an amplifier had a junction area of 0.3 sq cm, a base layer thickness of about 0.07 cm, and delivered 2.0 watts of undistorted output in class A operation. Its frequency cutoff was about 10,000 cps, this limit being in general agreement with the effect of diffusion through the p-layer as discussed in Sec. IX.

The low power potentialities of these structures have been more thoroughly investigated. A junction area of about 0.01 sq cm and a base layer thickness of about 1.5×10^{-3} cm are typical dimensions for these units. They have operated with gains of 50 db and noise figures of about 10 db to 15 db at 1000 cycles per second. Each of these quantities is an improvement of several orders of magnitude over point-contact transistors. For low signal levels they give essentially full gain with collector voltages higher than 0.1 volt and are, therefore, exceptionally good very low power amplifiers. At some sacrifice in gain they may be operated at efficiencies of 48 to 49 percent out of a theoretical maximum of 50 percent for class A. An oscillator has been constructed by R. L. Wallace, Jr., and D. E. Thomas of this laboratory which operates on 0.6-microwatt input. The same small transistors will also operate as amplifiers with maximum output powers of several hundred milliwatts. A set of operating curves is given in Fig. 2. The frequency cutoff of these units in high gain circuits is determined chiefly by the capacitance of the collector junction and is much lower than the limit set by diffusion through the base layer. A discussion of the quantitative effect will be found in the article by Wallace and Pietenpol.

V. THEORETICAL PRINCIPLES AND BOUNDARY CONDITIONS

In this and following sections we shall give analytic form to the ideas discussed in Sec. II. The principal symbols used in dealing with the theory are shown in the accompanying table of notation. In this section and the next we shall deal with the n-p-n structure in general terms and for this reason shall use subscripts "l" and "r" standing for "left" and "right" for the two n-type regions. This permits us to consider impartially cases in which either region may be biased as a collector or as an emitter.

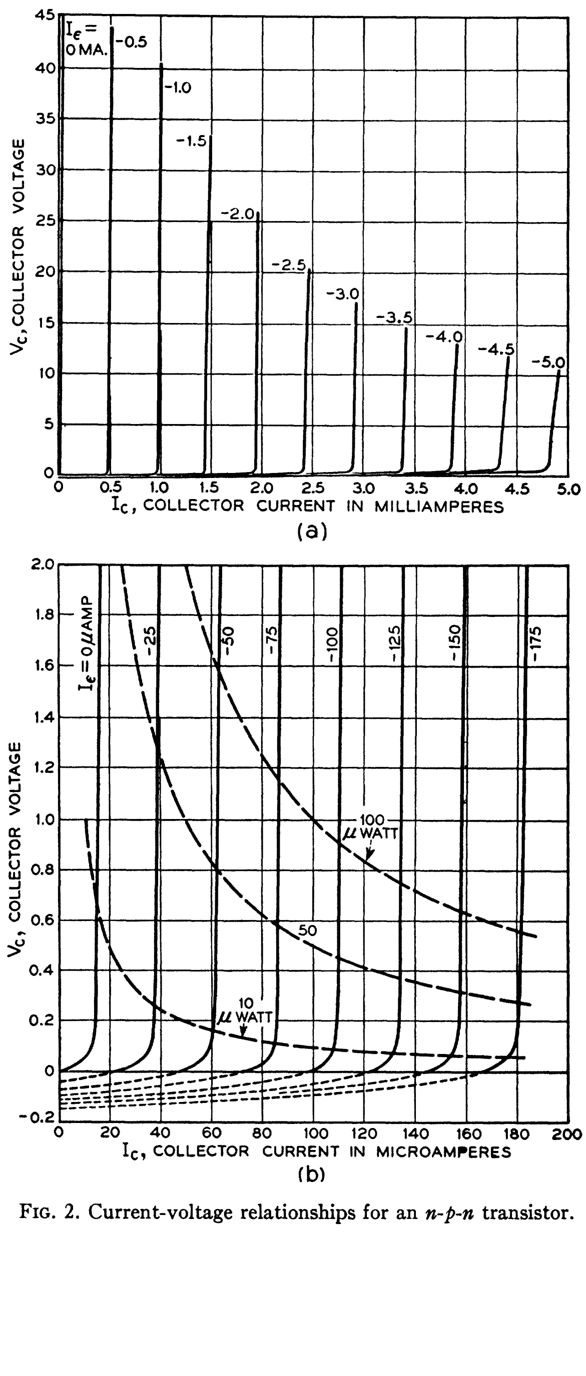

FIG. 2. Current-voltage relationships for an n-p-n transistor.

[8] An extensive presentation of circuit properties is given in reference 3.

[9] The performance of this transistor was discussed at the June, 1950, Inst. Radio Engrs. Conference on Electron Devices at the University of Michigan and also at the July, 1950, Conference on Semiconductors at Reading.

We shall first discuss the boundary conditions at the junctions when potentials are applied across them.[10] The electrostatic potential ψ in the interior of the semiconductor may have its zero chosen arbitrarily. For our purpose the zero is so chosen that $-q\psi$ is approximately the energy of an electron at a level of energy midway in the energy gap. The exact value of ψ is such that

$$n=n_i \exp q(\psi-\varphi)/kT \quad (5.1)$$

$$p=n_i \exp q(\varphi-\psi)/kT \quad (5.2)$$

under equilibrium conditions. Under nonequilibrium conditions, similar equations determine the "imrefs"[11] φ_n and φ_p in terms of ψ, n, and p as follows:

$$n=n_i \exp q(\psi-\varphi_n)/kT \quad (5.3)$$

$$p=n_i \exp q(\varphi_p-\psi)/kT. \quad (5.4)$$

NOTATION

A = cross-sectional area of unit;
$b=\mu_n/\mu_p=2.1$;
$B_l=(kT/q)[1-\exp(-q\varphi_l/kT)]$;
$B_r=(kT/q)[1-\exp(-q\varphi_r/kT)]$;
e = base of naperian logarithms;
D_n, D_p = diffusion constants for electrons and holes;
G_{ll}, G_{lr}, etc. = conductances at zero voltage;
n = density of electrons;
$n_b=n$ in the base layer;
$n_i=n$ in an intrinsic specimen;
n_1 = deviation of n from its thermal equilibrium value;
N_d, N_a = density of donors, acceptors;
p = density of holes;
q = charge of a hole = −charge of an electron;
V, v = dc and ac components of voltage;
W = thickness of base layer;
μ_n, μ_p = mobilities of electrons, holes = 3600, 1700 cm²/volt sec;
σ_i = intrinsic conductivity;
τ_{pl}, τ_{pr}, τ_{nb} = lifetimes of minority carriers;
φ = (Fermi level)/$(-q)$;
φ_n = imref for electrons;
φ_p = imref for holes;
φ_l, φ_b, φ_r = voltages of the three regions; and
ψ = electrostatic potential.

The imrefs or "quasi-Fermi levels" are introduced for convenience in discussing boundary conditions at the junctions and the meaning of applied voltages. In terms of the imrefs the current densities assume a particularly simple form:

$$I_n=qD_n\nabla n+q\mu_n nE=-q\mu_n n\nabla\varphi_n \quad (5.5)$$

$$I_p=-qD_p\nabla p+q\mu_p pE=-q\mu_p p\nabla\varphi_p. \quad (5.6)$$

These equations show that the current densities are those corresponding to materials with the conductivity appropriate for the electron and hole densities, respectively, and to electric fields derived from the imrefs as potentials. From these relationships it is evident that a given electron current will produce much bigger changes of the imrefs when it flows in a p-type region, where the electron density is small, than it will in an n-type region. In fact the ratio of conductivity by electrons is so great between the two regions that the imref for electrons can be regarded as substantially constant in the n-type region. In accordance with the assumption that the minority carrier density is small compared to the majority density and the assumption that the space charge is negligible, it follows that the potential ψ is substantially uniform in the interior of each region also. If the contacts on regions l and r are so far from the junctions that no injected carriers reach them and are substantially ohmic, then it also follows that the imrefs for electrons in these regions are simply the voltage applied to the two contacts. In accordance with the assumption that current to the base contact is carried by holes, it also follows that the imref for holes in this region is equal to φ_b.

We shall now apply the aforementioned conclusions to the boundary condition at J_l. For simplicity we shall assume that the base is grounded so that we shall consider in general cases in which

$$\varphi_l\neq 0, \quad \varphi_b=0, \quad \varphi_r\neq 0. \quad (5.7)$$

By the reasoning of the preceding paragraph, the imref for electrons is continuous across the junction J_l and in fact has its largest gradient only after the interior of the base region is reached. Consequently we may take φ_n in the base region near J_l as substantially equal to φ_l. The electron density in the base region near J_l is thus given by

$$n(\text{in } b \text{ near } J_l)=n_b \exp(-q\varphi_l/kT), \quad (5.8)$$

where n_b is the thermal equilibrium concentration of electrons at the corresponding point. (This equation follows directly from (5.3) together with the conclusion previously reached that ψ and φ_p in the base layer are unaffected by the applied potentials of (5.7).)

We shall be chiefly concerned with deviations of the densities from their equilibrium values and shall use the subscript 1 to indicate such densities. The deviation corresponding to (5.8) is

$$n_1=n_b[\exp(-q\varphi_l/kT)-1]\equiv -n_b qB_l/kT. \quad (5.9)$$

In this expression we have introduced the quantity B_l defined as

$$B_l\equiv(kT/q)[1-\exp(-q\varphi_l/kT)]. \quad (5.10)$$

This symbol is introduced since all the currents with which we shall be concerned depend functionally upon the voltages in the form (5.10). The coefficient kT/q is introduced so that B_l has the dimensions of a voltage and for small values of φ_l, B_l is in fact approximately equal to φ_l.

[10] The notation used here is similar to that of references 1 and 2 and the analysis is substantially an abreviation of that of reference 1.

[11] We are indebted to the most appropriate authority for suggesting this modified name for the quasi-Fermi levels.

Entirely similar reasoning leads to corresponding relationships for the hole density in the l region:

$$p(\text{in } l \text{ near } J_l)=p_l \exp(-q\varphi_l/kT) \qquad (5.11)$$

$$p_1=p_l[\exp(-q\varphi_l/kT)-1]\equiv -p_l qB_l/kT. \quad (5.12)$$

VI. THE CURRENT-VOLTAGE RELATIONSHIPS

The analysis of the last section indicates that near J_l the deviations of both hole and electron densities are proportional to B_l. These deviations lead to diffusion currents, which would vanish for the case of thermal equilibrium with $B_l=0$. As a result of the linear approximations discussed in Sec. II, these currents will be proportional to B_l. In Fig. 3 the conventions selected for signs of current are shown. We shall accordingly denote the current into region b due to hole flow across J_l as follows:

$$I_{lp}=G_{llp}B_l. \qquad (6.1)$$

For the case of a uniform cross section of area A and material of uniform conductivity and uniform lifetime τ_{pl}, the value of the coefficient may be easily derived:[12]

$$G_{llp}=q\mu_p p_{nl}A/L_{pl}=\sigma_i^2 bA/(1+b)^2\sigma_l L_{pl} \qquad (6.2)$$

in which the diffusion length is given by the equation,

$$L_{pl}=(D_p\tau_{pl})^{\frac{1}{2}}. \qquad (6.3)$$

The last form of (6.2) expresses the conductance G_{llp} in terms of the conductivity σ_i of an intrinsic sample and the actual conductivity of the n-type region. The quantity b occurring in the equation is the ratio of mobilities:

$$b=\mu_n/\mu_p. \qquad (6.4)$$

The electron current flowing across J_l can also be directly evaluated for the case of $B_r=0$. Even if no potential is applied across J_r, some of the electrons injected across J_l will arrive in the r-region. This is a consequence of the fact that the deviation n_1 is required to be zero at J_r when φ_r is zero. The electron currents across the two junctions are found to be[13]

$$I_{ln}=[(q\mu_n n_b A/L_n)\coth(W/L_n)]B_l\equiv G_{lln}B_l \qquad (6.5)$$

$$I_{rn}=-[(q\mu_n n_b A/L_n)\operatorname{csch}(W/L_n)]B_l\equiv G_{rln}B_l. \quad (6.6)$$

The conductance G_{lln} may be expressed in terms of the properties of the base layer as follows:

$$G_{lln}=[\sigma_i^2 bA/(1+b^2)\sigma_b L_n]\coth(W/L_n). \qquad (6.7)$$

A similar treatment for J_r leads to a corresponding set of equations. In terms of the G's and the B's the current-voltage relationship may be written as follows:

$$I_l=G_{ll}B_l+G_{lr}B_r \qquad (6.8)$$

$$I_r=G_{rl}B_l+G_{rr}B_r \qquad (6.9)$$

[12] Reference 1, Eqs. (4.20) and (4.21) or reference 2, p. 316. In a specimen of finite cross section, the recombination must be described in terms of a set of normal modes. For the cross sections of the small transistors, the lowest mode dominates and its lifetime may be used in the formulas derived for the one-dimensional case. See reference 1, Appendix V.

[13] Reference 1, Eq. (5.6), modified for electron diffusion.

$$G_{ll}=G_{lln}+G_{llp}, \quad G_{lr}=G_{lrn} \qquad (6.10)$$

$$G_{rl}=G_{rln}, \quad G_{rr}=G_{rrn}+G_{rrp}. \qquad (6.11)$$

For low voltages we have the approximate relationship,

$$B_l\doteq\varphi_l, \quad B_r\doteq\varphi_r \quad \text{for} \quad |\varphi_{l,r}|\ll kT/q, \qquad (6.12)$$

so that the coefficients in Eqs. (6.8) and (6.9) are simply the low voltage conductance components. If the potentials consist of a bias plus a small ac component so that we may write

$$\varphi_l=V_l+v_l, \quad \varphi_r=V_r+v_r, \qquad (6.13)$$

then the small signal equations become

$$i_l=g_{ll}v_l+g_{lr}v_r \qquad (6.14)$$

$$i_r=g_{rl}v_l+g_{rr}v_r, \qquad (6.15)$$

where the relationship between the small g's and the large G's is

$$g_{ll}/G_{ll}=g_{rl}/G_{rl}=\exp(-qV_l/kT) \qquad (6.16)$$

while a similar equation applies for the other two coefficients.

From these relationships one can also derive the result that each g is proportional to the deviation of its corresponding GB term from its saturation value corresponding to $B=(kT/q)$. This may be expressed symbolically as follows:

$$g=(\text{deviation of } GB \text{ from saturation})\cdot(q/kT). \quad (6.17)$$

For the model discussed in connection with Eqs. (6.5) and (6.6), it is evident that symmetry leads to the equation,

$$G_{lr}=G_{rl}. \qquad (6.18)$$

If the conductivity varies in an unsymmetrical way, however, in the middle layer or if the lifetime is greater at one side of the layer than the other, then we cannot reach the conclusion that the two G's are equal from symmetry arguments. It can be shown, however, that it is a consequence of the linear assumptions described in Sec. II that the symmetry relationship holds no

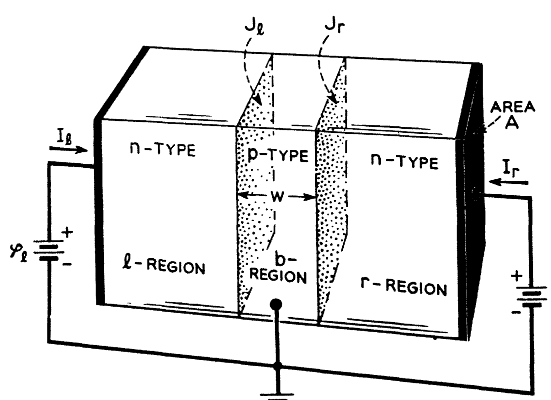

FIG. 3. Dimensions and conventions for voltage and current for an *n-p-n* structure.

matter what the geometry of the middle layer. The proof of this result is given in an appendix.[14] This leads us to the conclusion that for the idealized sort of model in which all of the currents are linear in the injected carrier density, the behavior of the transistor is described by three parameters, at least so far as low frequencies are concerned. These parameters are the four coefficients in Eqs. (6.8) and (6.9) with the relationship (6.18) between two of them.

In terms of these coefficients we may define the "alpha" of the unit using either junction as an emitter. The two α's need not be equal and the α for the left junction is given by the relationship,

$$\begin{aligned}\alpha_l &\equiv -(\partial I_r/\partial I_l) \quad [\text{for} \quad \varphi_r \text{ const}] = -G_{rl}/G_{ll} \\ &= -G_{rln}/(G_{lln}+G_{llp}) \\ &= (-G_{rln}/G_{lln})[G_{lln}/(G_{lln}+G_{llp})] = \gamma_l\beta_l. \quad (6.19)\end{aligned}$$

The definition of γ in transistor terminology is the fraction of the current at the emitter junction produced by emitter voltage that is carried by minority carriers in the base; evidently

$$\gamma_l = G_{lln}/(G_{lln}+G_{llp}). \quad (6.20)$$

The fraction of these injected carriers that reaches the collector is defined as β:

$$\beta_l = -G_{rln}/G_{lln} = \text{sech}(W/L_n). \quad (6.21)$$

The "intrinsic α" or α^* of the collector junction is defined as the ratio of change in total current per unit minority carrier current arriving at it. For a simple p-n junction collector, α^* is unity. For a "hook collector," which we treat in the next section, the arrival of injected current provokes a flow of carriers and α^* may be 100 or more.

If the middle layer is thin so that the hyberbolic functions of (6.5) and (6.6) may be approximated by their first terms, we may write

$$\beta_l \doteq 1-\tfrac{1}{2}(W/L_n)^2 \doteq 1 \quad (6.22)$$

$$\gamma_l \doteq 1/[1+(\sigma_b W/\sigma_l L_{pl})] \quad (6.23)$$

$$\alpha_l \doteq 1/[1+(\sigma_b W/\sigma_l L_{pl})] \quad (6.24)$$

$$G_{lln} \doteq \sigma_1^2 bA/(1+b)^2\sigma_b W. \quad (6.25)$$

Entirely similar expressions may be written for J_r simply by interchanging r and l.

[14] This symmetry result may be derived in a more general way from the reciprocity principle of electrical conduction provided no magnetic field is present. The necessary theorem is proved in section 5 of H. B. G. Casimir [Revs. Modern Phys. **17**, 343 1945)]. The proof there shows that in the linear range of conductivity, G_{lr} must equal G_{rl}. Since we have shown that the currents are linear functions of the B's by an independent argument, it follows that we may take $G_{lr}=G_{rl}$ in general. The method used by Casimir is based on Onsager's principle of microscopic reversibility and has an unnecessarily abstract flavor so far as the needs of this article are concerned. The desired theorem can be proved by straightforward analytical methods as is shown in the Appendix.

VII. SPECIAL OPERATING CONDITIONS

We shall next consider the consequences of the equations derived in the previous section for several limiting cases of voltages and currents.

A. Operating as an Amplifier

In order to get into the range of linear behavior it is necessary to apply a sufficiently large reverse bias to the collector so that its current saturates. This condition corresponds to the straight parts of the characteristics shown in Fig. 2. The voltage required to get into this range must exceed about $4kT/q$. After this point is reached the emitter current and collector current may be approximated by

$$I_l = G_{ll}B_l - \alpha_l G_{ll}kT/q \quad (7.1a)$$

$$\begin{aligned}I_r &= -\alpha_l G_{ll}B_l + G_{rr}kT/q \\ &= -\alpha_l I_l + (G_{rr}+\alpha_l^2 G_{ll})(kT/q). \quad (7.1b)\end{aligned}$$

Equation (7.1b) accounts for the parallel lines with currents increasing linearly in I_l shown in Fig. 2 for high collector voltages.

For applications in circuit theory, it is important to know the emitter admittance g_{ll}. To the approximation emphasized in this paper in which series ohmic resistances are neglected, this admittance is $1/r_\epsilon$ of the equivalent circuit.[3] The range of interest is usually such that $-\varphi_l > 4kT/q$ so that the application of the reasoning of (6.17) leads to

$$\begin{aligned}1/r_\epsilon = g_{ll} &= \partial I_l/\partial\varphi_l = G_{ll}\exp(-q\varphi_l/kT) \\ &\doteq I_l q/kT \doteq 40 I_l \text{mho}. \quad (7.2)\end{aligned}$$

If the unit is operated with grounded emitter and with the collector current saturated, then the input admittance is

$$\begin{aligned}dI_b/d\varphi_b &= d(-I_l-I_r)/d(-\varphi_b) \\ &= (1-\alpha_l)G_{ll}\exp(-q\varphi_l/kT) \quad (7.3)\end{aligned}$$

while the transconductance is $1/(1-\alpha_l)$ larger. Thus if $\alpha_l=0.99$, there will be a current gain of 100-fold, if the collector current is saturated. The theoretical power gain would be infinite if the collector impedance were infinite corresponding to the ideal saturation of (7.1b). Actually collector resistances of 10^7 to 10^8 ohms are observed. For currents large compared to saturation currents, (7.3) leads to an admittance of $(1-\alpha_l)I_r q/kT = 10^{-2}\times10^{-4}\times40 = 4\times10^{-5}$ mhos for $I_r=10^{-4}$ amp, a typical value for high gain performance. The power gain will be roughly the ratio of output to input impedances times the square of the current gain and will thus be about $10^7\times4\times10^{-5}\times10^4 \doteq 66$ db for a matched load. As discussed in Sec. IV, gains as high as 50 db have been obtained in practical circuits.

B. Hook Collector in p-n-p-n Transistor

Point contact transistors are frequently observed to have current multiplication in the sense that at a fixed collector bias, changes in emitter current produce changes in collector current several fold larger. One explanation is that a "p-n hook" is formed at the collector junction so that the "intrinsic α" or α^* of the collector is high. We shall illustrate this theory for the case of an n-type transistor in which holes are injected by a p-n junction emitter and the collector consists of n and p layers.

This structure is represented in Fig. 4 and is shown for bias conditions similar to those discussed for Fig. 1 except for the added emitter region and the absence of a contact to the b layer. Regarded as a transistor, the emitter, base, and collector are ϵ', b', and c'. The operating biases put reverse voltage across J_r and forward across J_l. The base layer b will float at a potential such that the net current to it is zero:

$$I_r+I_l=G_{ll}(1-\alpha_l)B_l+G_{rr}(1-\alpha_r)B_r=0. \quad (7.4)$$

For large reverse biases B_r will be positive and B_l negative and

$$\varphi_l-\varphi_r=V_{c'}, \quad (7.5)$$

where $V_{c'}$ is the voltage on the collector. For large $V_{c'}$, B_r will saturate and φ_l will be determined by

$$B_l=-G_{rr}(1-\alpha_r)(kT/q)/G_{ll}(1-\alpha_l) \quad (7.6)$$

so that the saturation current is

$$I_r(\text{sat.})=(1-\alpha_r\alpha_l)G_{rr}kT/q(1-\alpha_l). \quad (7.7)$$

If holes are injected by ϵ' so that a hole current $I_p{}^*$ arrives at J_r, then the condition,

$$I_r+I_l+I_p{}^*=0 \quad (7.8)$$

leads to a change

$$\Delta B_l=-I_p{}^*/G_{ll}(1-\alpha_l) \quad (7.9)$$

and this produces an increased electron current across J_r which leads to

$$\Delta I_{nr}=G_{rl}\Delta B_l=\alpha_l I_p{}^*/(1-\alpha_l). \quad (7.10)$$

The "intrinsic α" or α^* of the composite structure is

$$\alpha^*=(\Delta I_{nr}+I_p{}^*)/I_p{}^*=1/(1-\alpha_l). \quad (7.11)$$

For the thin layer approximation we see that

$$\alpha^*\doteq 1+(\sigma_l L_{pl}/\sigma_b W). \quad (7.12)$$

Thus for thin layers and for highly conducting l regions, α^* may be made very large.

The reason for calling the structure a "p-n hook" is illustrated in Fig. 4b. The high energy potential for electrons in layer b is low potential for holes. Holes injected by ϵ' become caught in this hook and bias J_l forward so as to provoke the enhanced electron flow.

C. Phototransistor

The structure in Fig. 4 can act as a phototransistor if the hole injection by the emitter junction is simply replaced by hole electron pair generation by light. For this application, the structure has only two terminals, b' and c' of Fig. 4, the region ϵ' being absent. If the hole electron pairs are generated in the neighborhood of J_r, then they will be separated by the field in the junction with a current equivalent to the passage of one hole across J_r for each pair so separated. The hole current across the junction will under these conditions be multiplied just as for the case of the p-n-p-n transistor so that the apparent quantum efficiency for hole electron pairs generated at J_r is

$$\text{apparent quantum efficiency}=\alpha^*\doteq 1+(\sigma_l L_{pl}/\sigma_b W) \quad (7.13)$$

by the reasoning that leads to Eq. (7.12). In the following section we shall discuss some values of α^* determined by measurements of photocurrents.

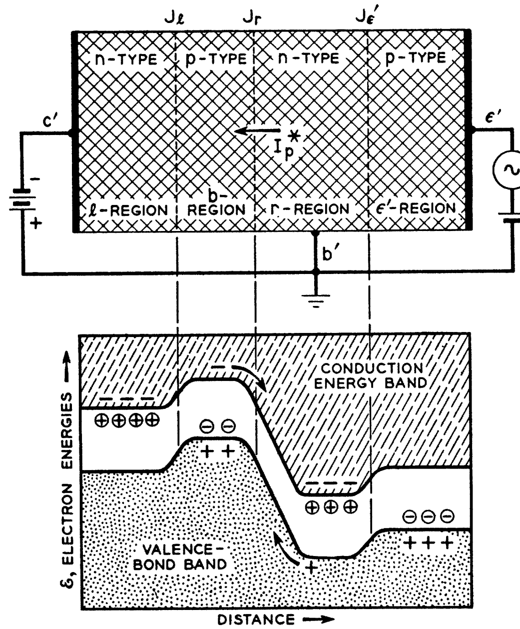

FIG. 4. A p-n-p-n transistor with a p-n hook collector.

D. Internal Contact Potentials

In order to discuss internal contact potentials we return to a consideration of the three region device of Fig. 3. If b is grounded and region r is allowed to seek its own potential, then potentials applied to region l will produce potentials on region r although the potential measured with an ohmic contact to b would everywhere be zero. The floating potential of region r is determined by setting the total current equal to zero so that we have

$$0=I_r=G_{rl}B_l+G_{rr}B_r \quad (7.14)$$

$$B_r=(-G_{rl}/G_{rr})B_l=(-G_{lr}/G_{rr})B_l=\alpha_r B_l \quad (7.15)$$

$$\exp(-q\varphi_r/kT)=1-\alpha_r+\alpha_r\exp(-q\varphi_l/kT). \quad (7.16)$$

Equation (7.16) expresses φ_r as a function of φ_l and α_r. It takes simple limiting form for extreme bias conditions:

Forward bias:

$$-q\varphi_l/kT \gg 1$$

$$\varphi_r = \varphi_l - (kT/q)\ln\alpha_r \tag{7.17a}$$

$$\alpha_r = \exp q(\varphi_l - \varphi_r)/kT. \tag{7.17b}$$

Zero bias:

$$|q\varphi_l/kT| \ll 1$$

$$\varphi_r = \alpha_r \varphi_l. \tag{7.18}$$

Reverse bias:

$$q\varphi_l/kT \gg 1$$

$$\varphi_r = -(kT/q)\ln(1-\alpha_r) \tag{7.19a}$$

$$\alpha_r = 1 - \exp(-q\varphi_r/kT). \tag{7.19b}$$

In the following section, we shall show that these expressions are approximately satisfied.

VIII. COMPARISON WITH EXPERIMENT

In this section we shall be chiefly concerned with an analysis of data on an *n-p-n* transistor and with showing that it may be interpreted on the basis of the theory discussed above. The data were taken under two sets of conditions: In the first the voltages were small compared to (kT/q) and from these the G's were determined using the approximation (6.12). In the second set, a wide range of voltages were used; for all conditions, however, one of the two B's was taken as independent and the other B was either constant or else proportional to the independent B. Consequently, each current is of the form

$$\begin{aligned} I &= c + mB = I_S + m[B - (kT/q)] \\ &= I_S - I_V \exp(-qV/kT), \end{aligned} \tag{8.1}$$

where V is the voltage upon which B depends and

$$I_S = c + (mkT/q) \tag{8.2}$$

is the "saturation" value of the current for large positive values of V and

$$I_V = mkT/q. \tag{8.3}$$

Both I_S and I_V are readily calculated in terms of the G's. In the analysis of the data the measured I_S and I_V values are compared with values computed from the G's, and the dependence of I upon V is investigated.

It should be pointed out that the values of the G's are strongly dependent on temperature. Consider, for example,

$$G_{rl} = b\sigma_i^2 A/(1+b)^2\sigma_p W. \tag{8.4}$$

TABLE 8.1. Zero bias conductances for an *n-p-n* transistor (conductances in micromhos at $T=22°$C).

	Measured	Calculated
G_{11}	8.8 ±0.5	8.8
G_{22}	33.3 ±0.5	33.3
$G_{11}(1-\alpha_1\alpha_2)$	6.9 ±0.5	6.9
$G_{22}(1-\alpha_1\alpha_2)$	26.5 ±0.5	26.4
$\alpha_2=(V_c/V_\epsilon)$ for $I_c=0$	0.86±0.02	0.89
$\alpha_1=(V_\epsilon/V_c)$ for $I_\epsilon=0$	0.22±0.02	0.23
$G_{11}+G_{12}-2G_{21}$	26.2 ±0.5	26.5

In this expression[15]

$$\sigma_i^2 \propto \exp(-\epsilon_G/kT) \tag{8.5}$$

where ϵ_G is the energy gap and

$$\sigma_p \propto T^{-\frac{3}{2}}. \tag{8.6}$$

Consequently the value of G_{rl} at $T_0+\Delta T=300°\text{K}+\Delta T$ is approximately

$$\begin{aligned} G_{rl}(T_0+\Delta T) &\doteq G_{rl}(T_0)\exp[(\epsilon_G/kT_0^2)+(3/2T_0)]\Delta T \\ &= G_{rl}(T_0)\exp(0.095+0.005)\Delta T \end{aligned} \tag{8.7}$$

so that G_{rl} increases approximately 10 percent per degree C. This increase arises chiefly from σ_i^2 and will be approximately the same for all the G's.

The fact that the G's have large temperature coefficients implies that at fixed voltages the currents will be very sensitive to temperature. This does not mean, however, that in a properly designed circuit, the behavior of the unit will be highly sensitive to temperature. The value of α_1, for example, as shown in (6.24) involves only $\sigma_b W/\sigma_\epsilon L_{p\epsilon}$ and this has only a small temperature coefficient. At a fixed emitter current, the emitter resistance is proportional to T in °K. Thus the most important quantities from a circuit point of view have small temperature coefficients.

Since in this section we are dealing with a transistor designed to have one terminal as emitter and one as collector, we shall abandon the "left" and "right" terminology and use subscript "1" for the emitter and "2" for the collector as is customary for transistors. The large signal equations are then

$$I_\epsilon = G_{11}B_1 + G_{12}B_2 \tag{8.8}$$

$$I_c = G_{21}B_1 + G_{22}B_2 \tag{8.9}$$

and the small signal equations are

$$I_\epsilon = g_{11}v_\epsilon + g_{12}v_c \tag{8.10}$$

$$i_c = g_{21}v_\epsilon + g_{22}v_c \tag{8.11}$$

where each g gepends on its corresponding voltage in the form,

$$g_{ij} = G_{ij}\exp(-qV_j/kT). \tag{8.12}$$

The low voltage conductances were measured at 2 millivolts. For this small voltage

$$B(+2\text{ mv}) = 1.95\text{ mv} \tag{8.13a}$$

$$B(-2\text{ mv}) = -2.05\text{ mv} \tag{8.13b}$$

so that the currents are nearly linear in this range, and the nonlinearity is practically eliminated by averaging the two polarities. With the collector grounded, I_ϵ was measured and G_{11} computed from I_ϵ/V_ϵ. With the collector open circuited, I_ϵ and V_c were measured; for this case $I_\epsilon/V_\epsilon = G_{11}(1-\alpha_1\alpha_2)$ and $V_c/V_\epsilon = -G_{21}/G_{22} = \alpha_2$. Similar data were taken with voltage applied to the

[15] n_i^2 varies as $T^3\exp(-\epsilon_G/kT)$, see reference 2, page 475, and $\mu_n\mu_p$ as T^{-3}, see page 287.

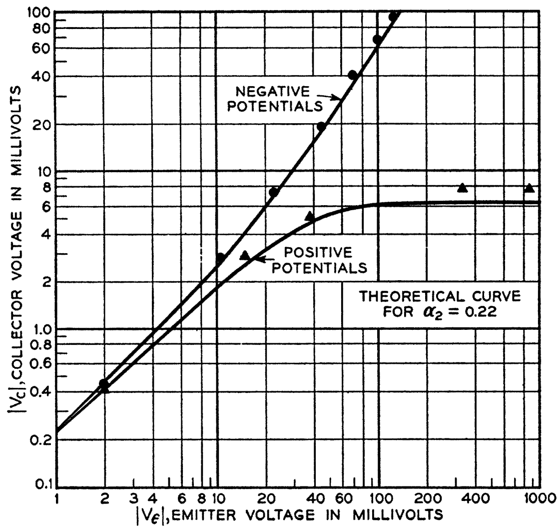

FIG. 5. Internal contact potential developed on the open circuited collector.

collector. Finally collector and emitter were connected together and the combined currents measured; this gives

$$(I_\epsilon+I_c)/V_\epsilon=G_{11}+G_{22}-G_{12}-G_{21}. \quad (8.14)$$

The values selected for the G's were: Conductances in micromhos at 22°C

$$G_{11}=8.8, \quad G_{22}=33.3, \quad G_{12}=G_{21}=-7.8. \quad (8.15)$$

These three values fit the seven measurements within the limits of experimental accuracy as shown in Table 8.1. The fact that the fit can be achieved with three constants is not a real test of the theory of Sec. VI, however, since any passive three terminal device in the absence of magnetic fields should satisfy the reciprocity condition and be described by three constants. What the table shows essentially is the consistency of the measurements.

Accurate values for the constants for the base layer were not available for the unit studied. However, the orders of magnitude of the G's are in reasonable agreement with values expected for a structure with constants lying in the ranges expected from the method of fabrication. We shall not attempt to obtain a perfect fit[16] but will choose as an example a structure with $A=0.003$ cm^2, $W=2.5\times10^{-3}$ cm, $\sigma_\epsilon=20$, $\sigma_b=10$, $\sigma_c=0.1$, $\sigma_i=(1/60)$ ohm^{-1} cm^{-1}, and lifetimes $\tau_{pc}=40$ and $\tau_{p\epsilon}=10$ microseconds. The resulting values for the G's based on Eqs. (6.2), (6.22), and (6.25) are

$$G_{11n}=7.3 \text{ micromhos} \quad (8.16a)$$

$$G_{11p}=0.5 \text{ micromhos} \quad (8.16b)$$

$$G_{22p}=45 \text{ micromhos.} \quad (8.16c)$$

[16] For the simpler case of a p-n junction, the electrical properties have been predicted with an accuracy of about 20 percent from the independently measured constants describing the junction.

The value of β is greater than 0.99 for $\tau_n=40$ microseconds and may be taken as unity so far as the G's are concerned. These values lead to

$$G_{11}=G_{11n}+G_{11p}=7.8 \quad (8.17a)$$

$$-G_{12}=-G_{12n}=\beta_1 G_{11n}=\beta_2 G_{22n}=7.3 \quad (8.17b)$$

$$G_{22}=G_{22n}+G_{22p}=7.3+45=52. \quad (8.17c)$$

According to this interpretation the failure of α_1 to be unity is due chiefly to hole flow across J_ϵ and a similar condition is true of α_2. The base current arises almost entirely from these hole flows with recombination in the base being nearly negligible.

Figures 5 and 6 show the internal contact potential effect. In Fig. 5 the potential V_ϵ is applied and V_c measured while zero current flows to the collector terminal. The data are seen to be in general agreement with the theory and with the value of α_2, corresponding to the collector junction, obtained at low voltage for Table 8.1. Figure 6 shows similar data with the voltage applied to the collector. The data are seen to differ slightly from the theoretical curves for reverse biases. The values of α obtained by applying (7.19b) to these data are

$$\alpha_2=0.226 \quad \text{and} \quad \alpha_2=0.894, \quad (8.18)$$

which shows that the fit is very sensitive to small variations in value of α. The test of Eq. (7.17b), which applies to forward bias, cannot be carried out as satisfactorily because under these conditions the currents are relatively large and the voltage drops across the series resistances of the specimen are not negligible compared to the effects studied. The values obtained are, however, approximately the same as those of Table 8.1.

In Fig. 7, the collector is biased to saturation and I_c and I_ϵ are plotted as functions of V_ϵ. For this case the

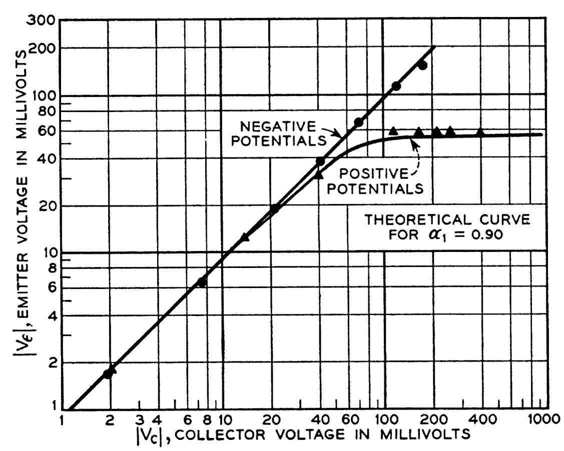

FIG. 6. Internal contact potential developed on the open circuited emitter.

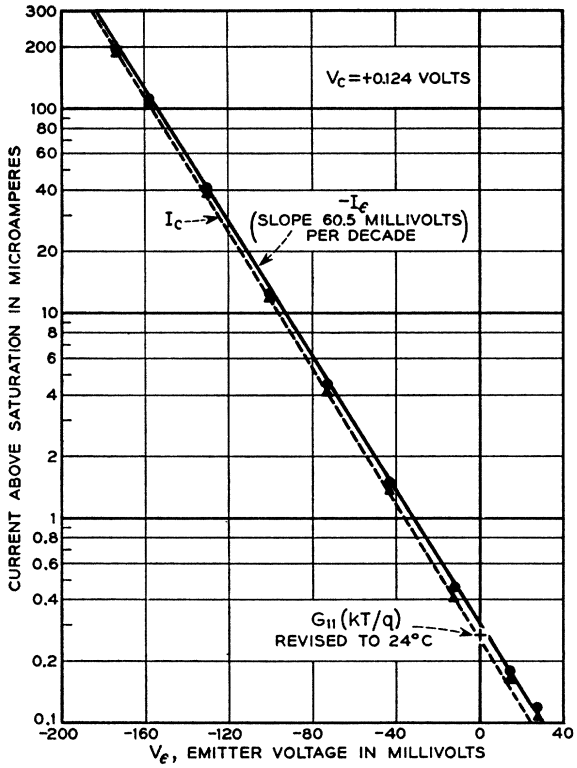

FIG. 7. Emitter and collector currents for the collector biased to saturation.

formulas should be

$$I_\epsilon = G_{11}B_1 + G_{12}(kT/q)$$
$$= G_{11}(1-\alpha_1)kT/q - (G_{11}kT/q)\exp(-qV_\epsilon/kT) \quad (8.19a)$$
$$I_c = G_{12}B_1 + G_{22}(kT/q)$$
$$= G_{22}(1-\alpha_2)(kT/q)$$
$$+ (\alpha_1 G_{11}kT/q)\exp(-qV_\epsilon/kT). \quad (8.19b)$$

It is seen that the lines agree well with the exponential forms and, furthermore, that the slope is in good agreement with theory which requires that for one decade of change in the current the voltage change should be

$$\Delta V = 2.30 \times kT/q = 2.30/39.4 = 59.0 \text{ mv} \quad (8.20)$$

for $T=297$°K, the temperature at which the data were taken. The values of I_S and I_V deduced from Table 8.1 (corrected for a ΔT of 2°C) and from the data on which Fig. 7 was based are:

	Fig. 7	Table 8.1
$G_{11}(1-\alpha_1)kT/q$	0.021 μa	0.025 μa
$G_{11}(kT/q)$	0.30 μa	0.27 μa
$G_{22}(1-\alpha_2)kT/q$	0.88 μa	0.78 μa
$\alpha_1 G_{11}kT/q$	0.27 μa	0.24 μa.

The slope terms are simply the values for $V_\epsilon=0$ in Fig. 7. The saturation values were deduced directly from the data. In the voltage range used, the saturation for the collector was not perfect and the collector saturation values were corrected for a "leakage" term of about 1 megohm. (The origin of this leakage effect is not clear and it tends to saturate at higher reverse biases.)

In Fig. 8 the dependence of emitter current upon emitter voltage is again shown. The unit was at somewhat higher temperature and was measured at a higher collector voltage in one case and with the collector floating in the other. The ratio of the two terms should be

$$G_{11}/G_{11}(1-\alpha_1\alpha_2) = 1.25. \quad (8.21)$$

The observed ratio is 1.30 at $V_\epsilon=0$ which is satisfactory agreement.

In Fig. 9 the dependence of collector current upon collector voltage is shown for two cases similar to those of Fig. 8. The ratio of the two values is again 1.30 in good agreement with the prediction. It should be noted, however, that the slope requires 74 millivolts per decade, a value appropriate to an unreasonably high temperature of 103°C. This slope is established for such low currents that it seems difficult to explain it by spurious effects of series resistances.

There is an important difference in the nature of the currents of Fig. 7, which fit the theoretical slope, and those of Fig. 9, which do not. The currents of Fig. 7 consists chiefly of electrons which diffuse through the base layer and arrive at the collector; the evidence for

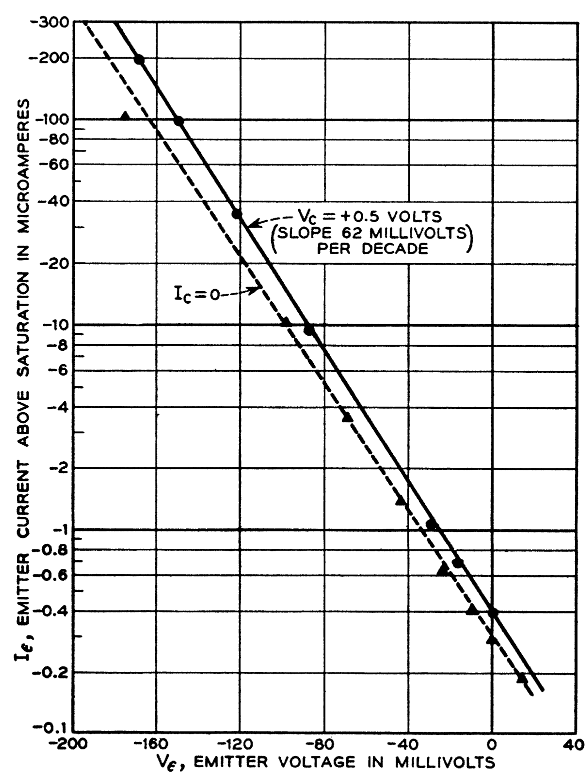

FIG. 8. Emitter current for two conditions of collectors.

this conclusion is the value 0.89 for α_1, which implies that only 11 percent of the current is carried by electrons recombining in the base layer or holes diffusing into and recombining in the emitter body (or the reverse processes, depending on the polarity). On the other hand, most of the collector current is probably carried by holes which recombine in the collector body. (This reasoning is in agreement with the attempt to interpret the G's in terms of the structure discussed in connection with (8.17a, b, c).) It is to be expected theoretically that if the recombination process involves trapping on recombination centers,[17] then the rate of recombination will increase less rapidly than linearly with injected carrier density because at high densities the traps tend to saturate. This view has received some experimental support from the work of F. S. Goucher and J. R. Haynes who find an apparent increase in lifetime with increasing carrier density. It may be that this mechanism accounts for failure of the currents to increase as rapidly as they should with increasing voltage in Fig. 9.

Further evidence for nonlinearity in the recombination of holes in the emitter is furnished by the dependence of α_1 upon emitter current. This can be seen in Fig. 7. As the emitter current increases, the ratio of collector to emitter current (above saturation) increases from 0.9 to about 0.95. This is interpreted as being due to the failure of hole current to increase as rapidly with voltage as does emitter current.

The tendency of α to increase with emitter current appears to be a general feature of n-p-n transistors. Wallace and Pietenpol[3] report values as high as 0.995 for α.

Phototransistors made of n-p-n structures are extremely responsive to light and exhibit apparent quantum efficiencies of at least several hundreds. In accordance with the interpretation of Sec. VII C, these efficiencies would lead to α_1 values deduced from $1-(1/\alpha^*)$ comparable to or larger than the highest observed in n-p-n transistors.

IX. SOME DESIGN CONSIDERATIONS

It has been the principal purpose of the preceding sections to examine the consequences of the diffusion theory and compare them with experiment. For this purpose the experimental conditions considered were the simplest: small currents and zero frequency. For practical applications high frequency and larger currents are also of interest. In this section we shall discuss briefly some of the factors of importance in design considerations.

Of great interest is the frequency cutoff. This may be determined by the external circuit or by one or another of several internal features of the transistor. The most fundamental of these latter is that set by the diffusion

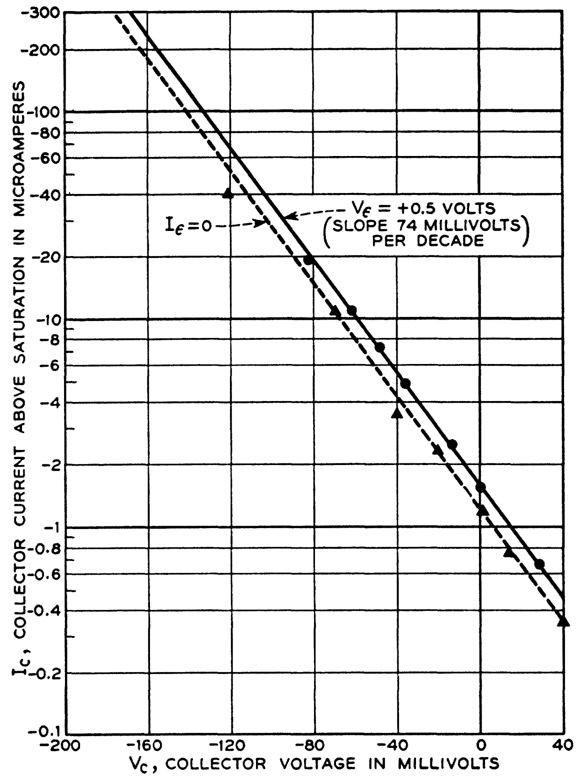

FIG. 9. Collector current for two conditions of emitter.

time through the base layer. This time is

$$\tau_D = W^2/D_n. \tag{9.1}$$

At a circular frequency ω, solutions for n_1 in the base layer are of the form,

$$\exp(i\omega t \pm x(1+i\omega\tau_n)^{\frac{1}{2}}/L_n) \tag{9.2}$$

and this leads to a β value of

$$\beta = \operatorname{sech}(1+i\omega\tau_n)^{\frac{1}{2}}W/L_n. \tag{9.3}$$

For frequencies such that $\omega\tau_n \gg 1$, this reduces to

$$B = 2/[\exp(1+i)(\omega\tau_D/2)^{\frac{1}{2}} + \exp(-1-i)(\omega\tau_D/2)^{\frac{1}{2}}]. \tag{9.4}$$

From this it is evident that for $\omega\tau_D \gg 2$ there is a phase lag of $(\omega\tau_D/2)^{\frac{1}{2}}$ radians and an equal attenuation in nepers. The power gain, which is proportional to β^2 in many cases, drops about 3 db when $\omega\tau_D = 2$ or $f = D_n/\pi W^2 \doteq 30/W^2$. For $W = 10^{-3}$ inch or 2.5×10^{-3} cm this is about 5×10^6 cps.

In addition to this fundamental limitation, there may be limitations due to capacitance and ohmic resistances. We shall illustrate this by considering the grounded emitter form of circuit, which is analogous to a grounded cathode vacuum tube circuit. For this case ac signals are applied to the base and ac voltages are developed on the collector. These voltages charge the capacitance of the base collector junction and this charging current must be furnished by hole flow in the base layer. For a layer

[17] See reference 2, page 342.

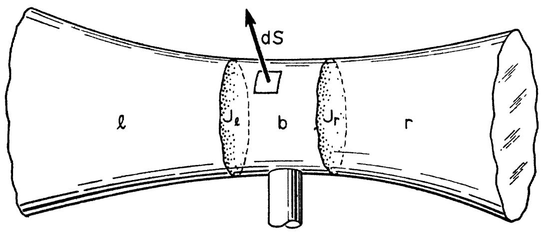

FIG. 10. The base region over whose surface the integration is made.

with $\sigma_b=10$ ohm^{-1} cm^{-1} and $W=2.5\times10^{-3}$ cm, the resistance from edge to opposite edge of a unit square is 40 ohms. If the capacitance is 2000 $\mu\mu$f/cm^2 and the voltage gain is 10, then the input signal will dominate the voltage on the base layer only if the width of the unit is less than

$$(10\times\omega\times2\times10^{-9}\times40)^{-\frac{1}{2}}=(\omega8\times10^{-7})^{-\frac{1}{2}}.$$

For a frequency of 5×10^6 cps, this leads to a width of 2 mm. The current through the base layer must also charge the emitter capacitance which may be much larger than the collector capacitance due to the effects of diffusion.

If the unit is too wide, the signal applied at the electrode on the base layer will be attenuated so that only a portion of the base layer will be operative in the desired way. The remainder of the base layer will be dominated by the capacitative voltages and these will induce currents which will lead to large "active" capacitances appearing between emitter and collector.

It is evident that the effects involved will lead to rather detailed calculations for any particular case but that the physical principles required to design *n-p-n* transistors are simply extensions of those well established for *p-n* junctions and *n-p-n* transistors for the low bias, low frequency conditions.

The low noise figures of these transistors cannot be said to be explained in the absence of an established theory of noise generation. They are, however, in rough agreement with a theory based on noise modulation of the recombination mechanism.[17] This theory predicts that each element of volume is a source of (noise current)^2 proportional to the square of the deviation of minority carrier density from its normal value. Applying this criterion to the *n-p-n* structure and comparing it to the type A indicates that the observed difference of 40 db or more between noise figures can be accounted for in terms of the change in current densities and geometries.

X. ACKNOWLEDGMENTS

We are indebted to a number of our colleagues for encouragement and assistance. In particular, we recognize the contribution of J. A. Morton of the transistor development group, whose encouragement resulted in the techniques that made good *p-n* transistors possible. We are grateful to E. Buehler and R. M. Mikulyak who processed the germanium, to W. J. Pietenpol who prepared the particular unit studied, to R. L. Wallace for help with the measurements and the manuscript, and to W. van Roosbroeck for a critical reading of the manuscript.

APPENDIX

Proof of the Equality $G_{lr}=G_{rl}$

We shall prove this equality subject to the assumption that the electron density in the base layer is small compared to the thermal equilibrium hole density. Under these conditions, the effect of injected electrons on the potential distribution may be neglected so that the variation of the electron density from its equilibrium value may be treated by linear equations. In Fig. 10 we represent the situation considered. We shall denote by $I_{nl}(B_l, B_r)$ the current across J_l into the base carried by electrons. The symmetry relation $G_{lr}=G_{rl}$ is then established by proving that

$$I_{nl}(0, B)=I_{nr}(B, 0). \tag{A1}$$

In the base layer we shall suppose that the electrostatic potential ψ and the lifetime τ are arbitrary functions of position. The boundary condition on the external surfaces and at the metal contact will be taken as

$$\mathbf{I}_n\cdot d\mathbf{S}=-qn_1s|d\mathbf{S}|, \tag{A2}$$

where $d\mathbf{S}$ is the outward normal, s the surface recombination velocity and

$$n_1=n-n_p \tag{A3}$$

is the deviation of n from the thermal equilibrium value.

We shall denote the solutions corresponding to potentials applied to the two junctions as follows:

$$B_l=B,\quad B_r=0 \qquad n_1,\quad \mathbf{I}_n \tag{A4}$$

$$B_l=0,\quad B_r=B \qquad n_1',\quad \mathbf{I}_n'. \tag{A5}$$

The currents in question are then

$$I_{nl}(0, B)=-\int_{J_l}\mathbf{I}_n'\cdot d\mathbf{S} \tag{A6}$$

$$I_{nr}(B, 0)=-\int_{J_r}\mathbf{I}_n\cdot d\mathbf{S}. \tag{A7}$$

The desired theorem is proved by considering the vector $\mathbf{A}$:

$$\mathbf{A}=(n_1/n_b)\mathbf{I}_n'-(n_1'/n_b)\mathbf{I}_n. \tag{A8}$$

Since $\mathbf{I}_n\cdot d\mathbf{S}$ is proportional to n_1 on the external surfaces, $\mathbf{A}\cdot d\mathbf{s}=0$ on these surfaces. Hence, the integral of $\mathbf{A}\cdot d\mathbf{S}$ over the surface of the base region is

$$\int\mathbf{A}\cdot d\mathbf{S}=(qB/kT)[-I_{nl}(0, B)+I_{nr}(B, 0)] \tag{A9}$$

since on J_l and J_r we have

$$\begin{aligned} n_1/n_b&=(qB/kT)\text{ on }J_l\\ &=0\text{ on }J_r \end{aligned} \tag{A10}$$

$$\begin{aligned} n_1'/n_b&=0\text{ on }J_l\\ &=(qB/kT)\text{ on }J_r. \end{aligned} \tag{A11}$$

Furthermore, it can be shown that $\nabla\cdot\mathbf{A}=0$ in the base region and hence that (A9) is zero by Gauss's theorem so that (A1) is proved. The proof that $\nabla\cdot\mathbf{A}=0$ is accomplished by showing that $\nabla\cdot(n_1/n_b)\mathbf{I}_n'$ is symmetrical in n_1 and n_1' so that the two terms in (A8) have cancelling divergences:

$$\begin{aligned} \nabla\cdot&(n_1/n_b)\mathbf{I}_n'\\ &=(1/n_b)(\nabla n_1-n_1\nabla\ln n_b)\cdot\mathbf{I}_n'+(n_1/n_b)\nabla\cdot\mathbf{I}_n'\\ &=(1/n_b)[(\nabla n_1+n_1q\mathbf{E}/kT)\cdot(q\mu_n n_1'\mathbf{E}+qD_n\nabla n_1')\\ &\qquad\qquad -q(n_1n_1'/\tau)], \end{aligned} \tag{A12}$$

which is seen to be symmetrical.

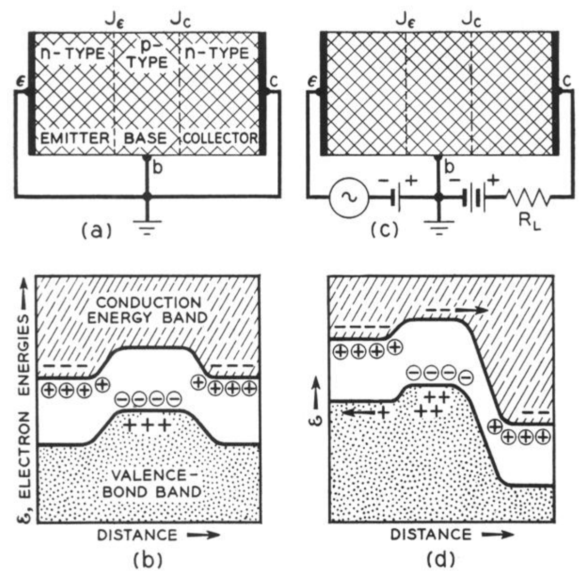

FIG. 1. The *n-p-n* structure and the energy level scheme. (a) and (b) Thermal equilibrium. (c) and (d) Biased as an amplifier.

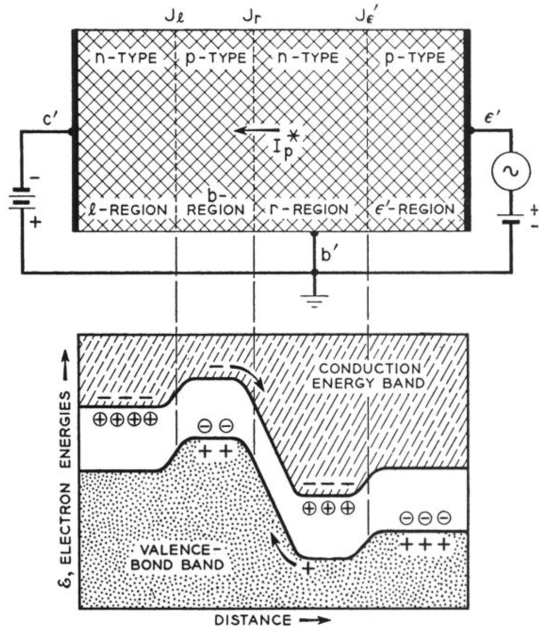

FIG. 4. A *p-n-p-n* transistor with a *p-n* hook collector.