The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors

By W. SHOCKLEY

In a single crystal of semiconductor the impurity concentration may vary from p-type to n-type producing a mechanically continuous rectifying junction. The theory of potential distribution and rectification for p-n junctions is developed with emphasis on germanium. The currents across the junction are carried by the diffusion of holes in n-type material and electrons in p-type material, resulting in an admittance for a simple case varying as $(1 + i\omega\tau_p)^{1/2}$ where τ_p is the lifetime of a hole in the n-region. Contact potentials across p-n junctions, carrying no current, may develop when hole or electron injection occurs. The principles and theory of a p-n-p transistor are described.

TABLE OF CONTENTS

1. INTRODUCTION

AS IS well known, silicon and germanium may be either n-type or p-type semiconductors, depending on which of the concentrations N_d of donors or N_a of acceptors, is the larger. If, in a single sample, there is a transition from one type to the other, a rectifying photosensitive p-n junction is formed.[1] The theory of such junctions is in contrast to those

[1] For a review of work on silicon and germanium during the war see H. C. Torrey and C. A. Whitmer, *Crystal Rectifiers*, McGraw-Hill Book Company, Inc., New York (1948). P-n junctions were investigated before the war at Bell Telephone Laboratories by R. S. Ohl. Work on p-n junctions in germanium has been published by the group at Purdue

of ordinary rectifying junctions because, on both sides of the junction, both electron flow and hole flow must be considered. In fact, a major portion of the hole current may persist into the n-type region and vice-versa. In later sections we show how this feature has a number of interesting consequences, which we shall describe briefly in this introduction.

A p-n junction may act as an emitter in the transistor sense, since it can inject hole current into n-type material. The a-c. impedance of a p-n junction may exhibit a frequency dependence characterized by this diffusion of holes and of electrons. For high frequencies the admittance varies approximately as $(i\omega)^{1/2}$ and has comparable real and imaginary parts. When a p-n junction makes contact to a piece of n-type material containing a high concentration of injected holes, it acts like a semipermeable membrane and tends to come to a potential which corresponds to the hole concentration.

Although some results can be derived which are valid for all p-n junctions, the diversity of possible situations is so great and the solution of the equations so involved that it is necessary to illustrate them by using a number of special cases as examples. In general we shall consider cases in which the semiconductor may be classified into three parts, as shown in Fig. 1. The meaning of the transition region will become clearer in later sections; in general it extends far enough to either side of the point at which $N_d - N_a = 0$ so that the value of $| N_d - N_a |$ at its boundaries is not much smaller than in the low resistance parts of the specimen. As stated above, appreciable hole currents may flow into the n-region beyond the transition region. For this reason, the rectification process is not restricted to the transition region alone. We shall use the word *junction* to include all the material near the transition region in which significant contributions to the rectification process occur. It has been found that various techniques may be employed to make nonrectifying metallic contacts to the germanium; when this is properly done, the resistance measured between the metal terminals in a suitably proportioned specimen is due almost entirely to the rectifying junction up to current densities of 10^{-1} amp/cm^2.

---

directed by K. Lark-Horovitz: S. Benzer, *Phys. Rev.* 72, 1267 (1947); M. Becker and H. Y. Fan, *Phys. Rev.* 75, 1631 (1949); and H. Y. Fan, *Phys. Rev.* 75. 1631 (1949). Similar junctions occur in lead sulfide according to L. Sosnowski, J. Starkiewicz and O. Simpson, *Nature* 159, 818 (1947), L. Sosnowski, *Phys. Rev.* 72, 641 (1947), and L. Sosnowski, B. W. Soole and J. Starkiewicz, *Nature* 160, 471 (1947). The theory described here has been discussed in connection with photoelectric effects in p-n junctions by F. S. Goucher, Meeting of the American Physical Society, Cleveland, March 10–12, 1949 and by W. Shockley, G. L. Pearson and M. Sparks, *Phys. Rev.* 76, 180 (1949). For a general review of conductivity in p- and n-type silicon see G. L. Pearson and J. Bardeen, *Phys. Rev.* 75, 865 (1949), and J. H. Scaff, H. C. Theuerer and E. E. Schumacher, *Jl. of Metals*, 185, 383 (1949) and W. G. Pfann and J. H. Scaff, *Jl. of Metals*, 185, 389 (1949). The latter two papers also discuss photo-voltaic barriers. The most recent and thorough theory for frequency effects in metal semiconductor rectifiers is given elsewhere in this issue (J. Bardeen, *Bell Sys. Tech. Jl.*, July 1949).

Even for distributions of impurities as simple as those shown in part (b) there are two distinctly different types of behavior of the electrostatic potential in the transition region, each of which may be either rectifying or nonrectifying. The requirement that the junction be rectifying can be stated in terms of the current distribution, certain cases of which are shown in (c). The total current, from left to right, is I, the hole and electron currents being

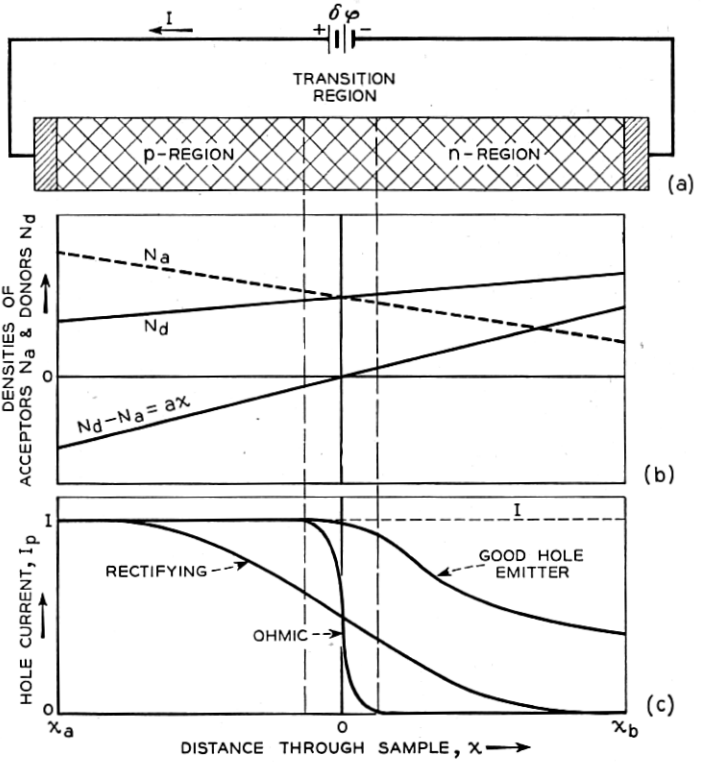

Fig. 1—The *p-n* junction.

(a) Schematic view of specimen, showing non-rectifying end contacts and convention for polarities of current and voltage.
(b) Distribution of donors and acceptors.
(c) Three possible current distributions.

I_p and I_n, with $I = I_p + I_n$. Well away from the junction in the *p*-type material, substantially all of the current is carried by holes and $I_p = I$; similarly, deep in the *n*-type material $I_n = I$ and $I_p = 0$. In general in a nonrectifying junction, the hole current does not penetrate the *n*-type material appreciably whereas in the rectifying junction it does. Under some conditions the major flow across the junction will consist of holes; such

cases are advantageous as emitters in transistor applications using n-type material for the *base*.

Where the hole current flows in relatively low resistance n-type material, it is governed by the diffusion equation and the concentration falls off as $\exp(-x/L_p)$ where L_p is the diffusion length:

$$L = \sqrt{D\tau_p}\,.$$

Here D is the diffusion constant for holes and τ_p their mean lifetime. The lifetime may be controlled either by surface recombination[2] or volume recombination. Surface recombination is important if the specimen has a narrow cross-section.

Under a-c. conditions, the diffusion current acquires a reactive component corresponding to a capacity. In addition, a capacitative current is required to produce the changing potential distribution in the transition region itself.

In the following sections we shall consider the behavior of the junction analytically, treating first the potential distribution in the transition region and the charges required change the voltage across it in a pseudo-equilibrium case. We shall then consider d-c. rectification and a-c. admittance.

2. Potential Distribution and Capacity of Transition Region

2.1 Introduction and Definitions

We shall suppose in this treatment that all donors and acceptors are ionized (a good approximation for Ge at room temperature) so that we have to deal with four densities as follows:

n = density of electrons in conduction band

p = density of holes in valence-bond band

N_d = density of donors

N_a = density of acceptors

The total charge density is

$$\rho = q\,(p - n + N_d - N_a), \tag{2.1}$$

where q is the electronic charge. We shall measure electrostatic potential ψ in the crystal, as shown in Fig. 2, from such a point, approximately[3] midway in the energy gap, that if the Fermi level φ is equal to ψ, the concentrations of holes and electrons are equal to the concentration $n_i = p_i$ char-

[2] H. Suhl and W. Shockley *Phys. Rev.* 75 1617 (1949).

[3] A difference in effective masses for holes and electrons will cause a shift of ψ from the midpoint between the bands.

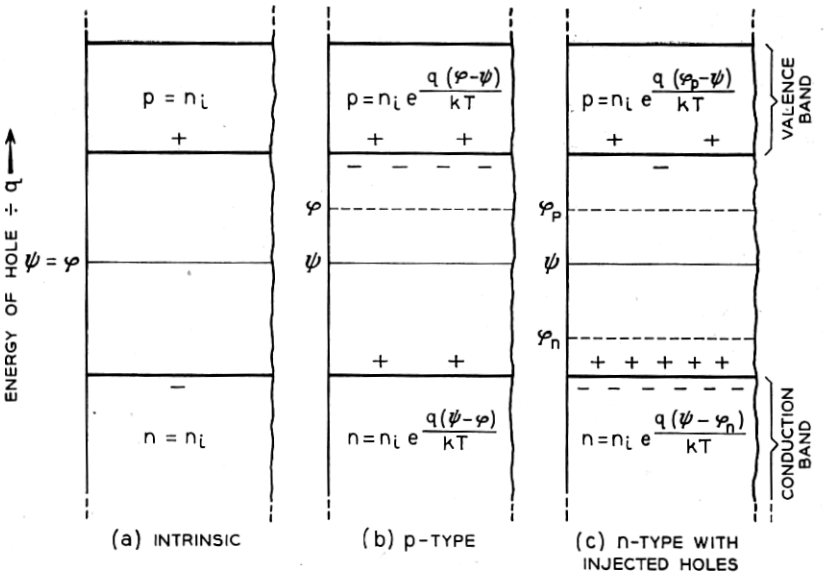

Fig. 2—Electrostatic potential ψ, Fermi level φ and quasi Fermi levels φ_p and φ_n. (In order to show electrostatic potential and energies on the same ordinates, the energies of holes, which are minus the energies of electrons, are plotted upwards in the figures in this paper.)

acteristic of a pure sample. For an impurity semi-conductor we shall have, as shown in (b),

$$\begin{aligned} p &= n_i e^{q(\varphi-\psi)/kT} && \text{(a)} \\ n &= n_i e^{q(\psi-\varphi)/kT}, && \text{(b)} \end{aligned} \tag{2.2}$$

where q is the electronic charge. Accordingly,

$$\rho = q\{N_d - N_a + 2n_i \sinh [q(\varphi - \psi)/kT]\}. \tag{2.3}$$

When the hole and electron concentrations do not have their equilibrium values, because of hole or electron injection or production of hole-electron pairs by light, etc., it is advantageous to define two non-equilibrium quasi Fermi levels φ_p and φ_n by the equations

$$\begin{aligned} p &= n_i e^{q(\varphi_p-\psi)/kT} && \text{(a)} \\ n &= n_i e^{q(\psi-\varphi_n)/kT} && \text{(b)} \end{aligned} \tag{2.4}$$

as indicated in Fig. 2 (c). In terms of φ_p and φ_n, the hole and electron currents take the simple forms:

$$I_p = -q[D\nabla p + \mu p\nabla\psi] = -q\mu p\nabla\varphi_p \tag{2.5}$$

$$I_n = bq[D\nabla n - \mu n\nabla\psi] = -qb\mu n\nabla\varphi_n \tag{2.6}$$

where the mobility μ and diffusion constant D for holes are related by Einstein's equation

$$\mu = qD/kT \tag{2.7}$$

and b is the ratio of electron mobility to hole mobility.[4]

Under equilibrium conditions $\varphi_p = \varphi_n = \varphi$ where φ is independent of position. Under those conditions, I_p and I_n are both zero according to equations (2.5) and (2.6). The electrostatic potential ψ, however, will not in general be constant and there will be unbalanced charge densities throughout the semiconductor. We shall consider the nature of the conditions which determine ψ for a general case and will later treat in detail the behavior of ψ for p-n junctions.

For equilibrium conditions, there is no loss in generality in setting φ arbitrarily equal to zero. The charge density expression (2.3) may then be rewritten as

$$\rho = \rho_d - \rho_i \sinh u \tag{2.8}$$

where

$$u \equiv q\psi/kT, \qquad \rho_i \equiv 2n_i q, \qquad \rho_d \equiv q(N_d - N_a) \tag{2.9}$$

In equation (2.8) ρ_d and u and, consequently, ρ may be functions of position. The potential ψ must satisfy Poisson's equation which leads to the equation

$$\nabla^2\psi = -4\pi\rho/\kappa \tag{2.10}$$

where κ is the dielectric constant, (2.10) can be rewritten as

$$\nabla^2 u = \frac{4\pi q\rho_i}{kT\kappa}\left(\sinh u - \frac{\rho_d}{\rho_i}\right) \tag{2.11}$$

What this equation requires in physical terms is that the electrostatic potential produces through (2.8) just such a total charge density ρ that this charge density, when used in Poisson's Equation (2.10), in turn produces ψ. It seems intuitively evident that the equation for u will always have a physically meaningful solution; no matter how the charge density ρ_d due to the impurities varies with position, the holes and electrons should be able to distribute themselves so that equilibrium is produced. For a one-dimensional case, it is not difficult to prove that a unique solution exists for $u(x)$ for any $\rho_d(x)$ (Appendix VII).

[4] We prefer b in comparison to c for this ratio since c for the speed of light also occurs in formulae involving b.

The coefficient in (2.11) has the dimensions of (length)$^{-2}$ leading us to define a quantity

$$\begin{aligned} L_D &= \sqrt{\kappa kT/4\pi q\rho_i} = \sqrt{\kappa kT/8\pi q^2 n_i} \\ &= 2.1 \times 10^{-3} \text{ cm for Si with } \kappa = 12.5,^{5}\ n_i = 2 \times 10^{10} \text{ cm}^{-3} \\ &= 6.8 \times 10^{-5} \text{ cm for Ge with } \kappa = 19,^{6}\ \ n_i = 3 \times 10^{13} \text{ cm}^{-3} \end{aligned} \tag{2.12}$$

where the subscript D for Debye emphasizes the similarity of L_D to the characteristic length in the Debye-Hückel theory of strong electrolytes. The meaning of the Debye length is apparent from the behavior of the solution in a region where ρ_d is constant, and u differs only slightly from the value u_0 which gives neutrality, with $\rho_i \sinh u_0 = \rho_d$. Under these conditions,

$$\frac{d^2u}{dx^2} = (L_D^{-2} \cosh u_0)(u - u_0) \tag{2.13}$$

so that $u - u_0$ varies as exp $(\pm x\sqrt{\cosh u_0}/L_D)$. In general, we shall be interested in cases in which the deviation of u from u_0 decays to a small value in one direction. It is evident that the distance required to reduce the deviation to $1/e$ is $L_D/\sqrt{\cosh u_0}$. If only small variations in ρ_d occur within a distance $L_D/\sqrt{\cosh u_0}$, then the semiconductor will be substantially neutral. However, if a large variation of ρ_d occurs in this distance, a region of local space charge will occur. These two cases are illustrated in connection with the potential distribution in a *p-n* junction.

2.2 Potential Distribution in the Transition Region[7]

We shall discuss the case shown in Fig. 1 for which the charge density due to donors and acceptors is given by

$$N_d - N_a = ax \tag{2.14}$$

This relationship defines a characteristic length L_a given by

$$L_a = n_i/a \tag{2.15}$$

If $L_a \gg L_D$, the condition of electrical neutrality is fulfilled (Appendix VII) and u satisfies the equation

$$\sinh u = \rho_d/\rho_i = ax/2n_i = x/2L_a$$

[5] J. F. Mullaney, *Phys. Rev.* 66, 326 (1944).
[6] H. B. Briggs and W. H. Brattain, *Phys. Rev.*, 75, 1705 (1949).
[7] Potential distributions in rectifying junctions between semiconductors and metals have been discussed by many authors, in particular N. F. Mott, *Proc. Roy. Soc.* 171A, 27 (1939) and W. Schottky *Zeits. f. Physik* 113, 367 (1939) 118, 539 (1942) and elsewhere. A summary in English of Schottky's papers is given by J. Joffe, *Electrical Communications* 22, 217 (1945). All such theories are in principle similar in involving the solution of equations like (2.11). See, for example, H. Y. Fan, *Phys. Rev.* 62, 388 (1942).

On the other hand, if $L_D \gg L_a$, a large change in impurity concentration occurs near $x = 0$ without compensating electron and hole densities occurring. Mathematically, we find that (2.11) can be expressed in the form

$$\frac{d^2u}{dy^2} = \frac{1}{K^2}(-y + \sinh u) \tag{2.16}$$

and

$$y = x/2L_a, \qquad K = L_D/2L_a \tag{2.17}$$

In Appendix VII, it is verified that the appropriate solution for $K \ll 1$ is that giving local neutrality, $u = \sinh^{-1} y$; while for $K \gg 1$, there is space charge as described below.

For $L_D \gg L_a$, or $K \gg 1$, there is a space charge layer in which $N_d - N_a$ is uncompensated. To a first approximation, we can neglect the electron and hole space charge in the layer and obtain, by integrating twice,

$$\psi = -\frac{2\pi q a x^3}{3\kappa} + a_2 x, \tag{2.18}$$

where we have chosen the zero of potential as the value at $x = 0$, a condition required by the symmetry between $+x$ and $-x$ of (2.14). Although the potential rise is steep in the layer, $d\psi/dx$ should be small at the point x_m where the neutral n-type material begins. As an approximation we set $d\psi/dx = 0$ at $x = x_m$:

$$\frac{d\psi}{dx} = -\frac{2\pi q a x_m^2}{\kappa} + a_2 = 0; \tag{2.19}$$

this leads to a value for a_2 which may be inserted in (2.18) to evaluate ψ at x_m:

$$\psi_m = \frac{4\pi q a x_m^3}{3\kappa} = \frac{4\pi q}{3\kappa a^2}(a x_m)^3 = \frac{4\pi q}{3\kappa a^2} n_m^3 \tag{2.20}$$

where $n_m = a x_m$ is the density of electrons required to neutralize $N_d - N_a = a x_m$ at the edge of the space-charge layer. This value of n_m must correspond to that associated with ψ_m by (2.2)

$$n_m = n_i e^{q\psi_m/kT}. \tag{2.21}$$

We thus have two equations relating ψ_m and n_m and the parameter "a." To solve them we plot $\ln \psi_m$ versus $\ln n_m$ as shown in Fig. 3. On this figure the relationship

$$\begin{aligned} \psi_m &= \frac{4\pi q}{3\kappa}\frac{n_m^3}{a^2} \\ &= 3.18 \times 10^{-8}\frac{n_m^3}{a^2} \text{ volts for Ge} \\ &= 4.83 \times 10^{-8}\frac{n_m^3}{a^2} \text{ volts for Si} \end{aligned} \tag{2.22}$$

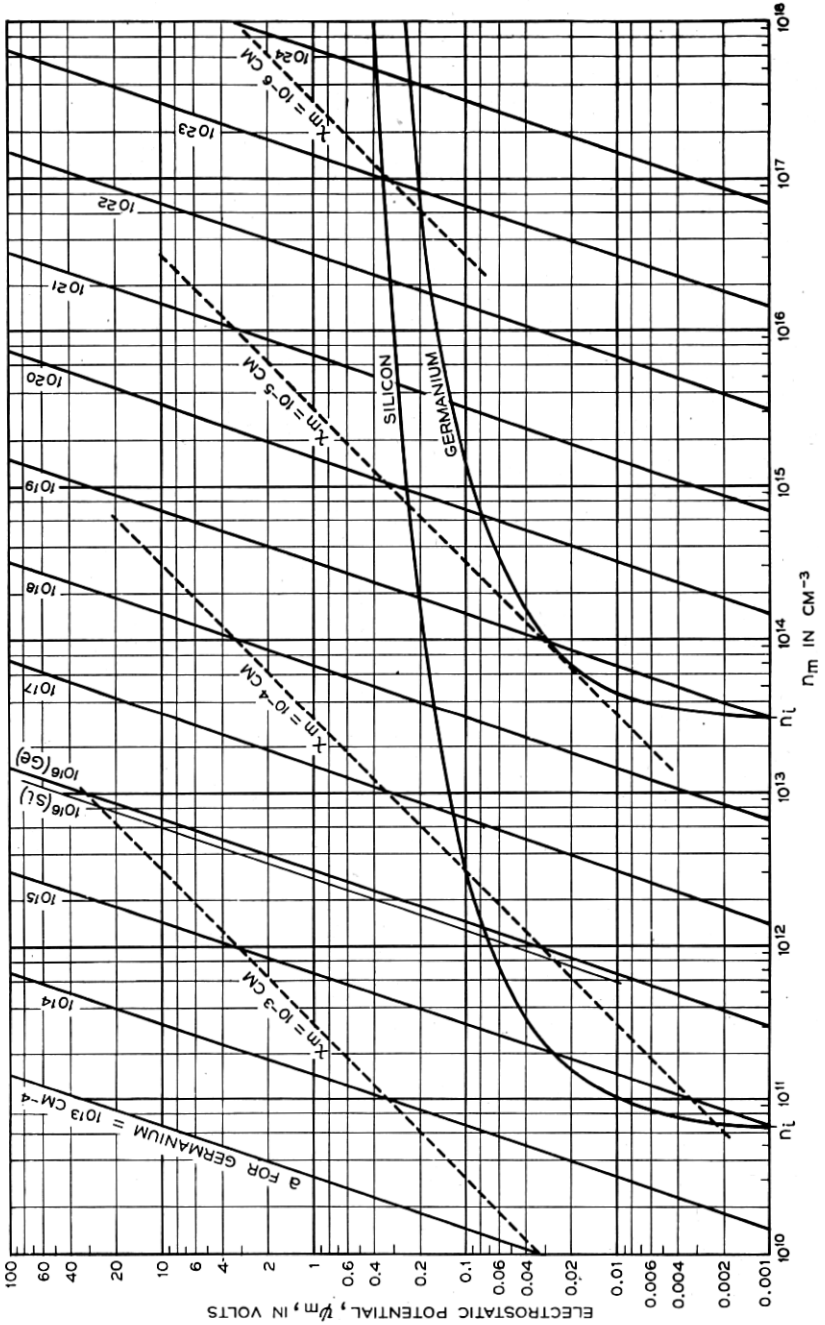

Fig. 3—Solutions for the boundaries of the space-charge region.

becomes a family of straight lines with "a" as a parameter. (Only $a = 10^{16}$ cm^{-4} is shown for Si, all the other lines being for Ge.) The half thickness x_m $(= n_m/a)$ of the space-charge region is also shown. Solutions are obtained when these lines cross the curves $n_m = n_i \exp(q\psi_m/kT)$, which are shown for room temperature. The condition that the intersection lie well to the right on the curve is equivalent to $K \gg 1$. For two Si samples cut from a melt, a was determined from measurements of conductivity[8] and was about 10^{15} to 10^{16} cm^{-4}. For these, the space change region has a half-width x_m of more than 10^{-4} cm. For other temperatures, the curves can be appropriately translated.[9]

In Fig. 4(a) we show the limiting potential shapes:

$$ax = 2n_i \sinh \frac{q\psi}{kT} \qquad \text{for } K \ll 1 \quad (2.23)$$

$$\psi = (\psi_m/2)(-(x/x_m)^3 + 3(x/x_m)) \qquad \text{for } K \gg 1 \quad (2.24)$$

In Fig. 4(b) the charge densities are shown. For the space-charge case, $|N_d - N_a|$ is greater than n or p. For a higher potential rise, i.e. larger ψ_m, the discrepancy would be greater and $N_d - N_a$ would be unneutralized except near x_m.

2.3 *The Transition-Region Capacity*

When the voltage across the junction is changing, a flow of holes and electrons is required to alter the space charge in the transition region. We shall calculate the charge distribution in the transition region with the aid of a pseudo-equilibrium model in which the following processes are imagined to be prevented: (1) hole and electron recombination, (2) electron flow across the p-region contact at x_a (Fig. 1), (3) hole flow across the n-region boundary at x_b. Under these conditions holes which flow in across x_a must remain in the specimen. If a potential $\delta\varphi$ is applied at the p end, then holes will flow into the specimen until φ_p has increased by $\delta\varphi$ so that the holes inside are in equilibrium with the contact which applies the potential. Since the specimen as a whole remains neutral, an equal electron flow will occur at x_b. When the specimen arrives at its pseudo-equilibrium steady-state, the potential distribution will be modified in the transition region and the number of holes in this region will be different from the number present under conditions of true thermal equilibrium. The added number of holes is proportional to $\delta\varphi$ for small values of $\delta\varphi$ and thus acts like the charge on a condenser. Our problem in this section is to calculate how this charge depends

[8] Unpublished data of W. H. Brattain and G. L. Pearson.

[9] The effect of unionized donors and acceptors can also be included by letting n_i include the properly weighted donor states and p_i, the acceptor states.

upon $\delta\varphi$ for various types of transition regions and to express the result as a capacity.

The justification for this pseudo-equilibrium treatment is as follows: Under actual a-c. conditions the potential drop in the *p*- and *n*-regions themselves are small because of their high conductivity so that most of the po-

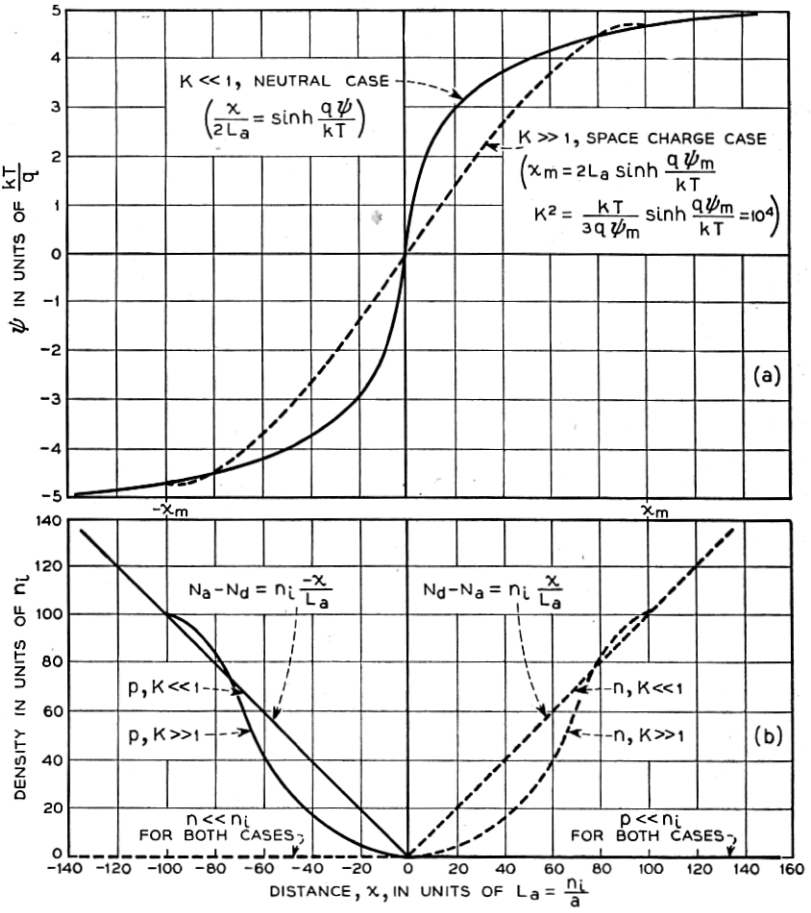

Fig. 4—Electrostatic potential and densities for *p-n* junctions.

tential drop occurs across the transition region. On the *p*-side of the transition region a large supply of holes is available to modify the potential and the fact that a current is flowing across the junction disturbs their concentration negligibly; the electrons on the *n*-side are similarly situated. Hence the distribution of holes and electrons in the transition region will be much the same as for the pseudo-equilibrium case. The question of how the hole

current required to change the potential distribution in the transition region is related to other hole currents is discussed in Section 4.1.

Under our assumptions, after the voltage $\delta\varphi$ is applied, a steady state is reached involving no current hence $\nabla\varphi_p = \nabla\varphi_n = 0$. Consequently, both φ_p and φ_n are constant and

$$\varphi_p - \varphi_n = \delta\varphi \tag{2.25}$$

since the holes are being supplied from a source at a potential $\delta\varphi$ higher than for the electrons.

We shall then have

$$p = n_i e^{q(\varphi_p - \psi)/kT} = n_1 e^{q(\varphi_1 - \psi)/kT} \tag{2.26}$$

$$n = n_i e^{q(\psi - \varphi_n)/kT} = n_1 e^{q(\psi - \varphi_1)/kT} \tag{2.27}$$

where

$$\varphi_1 = (\varphi_p + \varphi_n)/2, \quad \varphi_p = \varphi_1 + \delta\varphi/2, \quad \varphi_n = \varphi_1 - \delta\varphi/2 \tag{2.28}$$

and

$$n_1 = n_i e^{q\delta\varphi/2kT}. \tag{2.29}$$

Thus the effect of applying the potential $\delta\varphi$ in the pseudo-equilibrium case is equivalent to changing n_i to n_1 just as if the energy gap had been reduced by $q\delta\varphi$.

In the p-region, $n \ll p$ and so that $p = -ax$ is a good approximation. Similarly, in the n-region, we set $n = ax$. Hence we have in the p-region

$$\psi = \varphi_1 + (\delta\varphi/2) - (kT/q) \ln (-ax/n_i) \tag{2.30}$$

and in the n-region

$$\psi = \varphi_1 - (\delta\varphi/2) + (kT/q) \ln (ax/n_i). \tag{2.31}$$

Hence the effect of $\delta\varphi$ is to shift ψ in the p-region upwards by $\delta\varphi$ compared to ψ in the n-region. This is an example of the general result that $\psi - \varphi_p$ tends to remain constant at a given point in the p-region no matter what disturbances occur and $\psi - \varphi_n$ tends to remain constant in the n-region.

The Capacity for the Neutral Case $K \ll 1$

For the neutral case, we calculate the total number of holes, P, between x_a and x_b as a function of $\delta\varphi$. The charge of these holes is qP and the effective capacity is $q\ dP/d\ \delta\varphi$. As explained above, we are really interested in the change in number of holes in the transition region. However, the value of P is relatively insensitive to the location of the limits x_a and x_b so long as they lie in regions where the conductivity approaches the maximum values in the

p- and *n*-regions. In the following calculations, we shall consider a unit area of the junction so that values of P and of capacity are on a unit area bases.

The value of P is obtained by integrating $p\,dx$ making use of the neutrality condition to establish the functional relationship between p and x. The neutrality condition can be written as

$$ax = 2n_1 \sinh \frac{q(\psi - \varphi_1)}{kT} \equiv 2n_1 \sinh u \tag{2.32}$$

where $u \equiv q(\psi - \varphi_1)/kT$ and

$$p = n_1 e^{q(\varphi_1 - \psi)/kT} \equiv n_1 e^{-u} \tag{2.33}$$

$$n = n_1 e^{+u} \tag{2.34}$$

so that the value of P can be obtained by changing variables from x to u:

$$\begin{aligned} P &= \int_{x_a}^{x_b} p\,dx = \int_{u_a}^{u_b} p(2n_1/a) \cosh u\,du \\ &= (n_1^2/a) \int_{u_a}^{u_b} [1 + e^{-2u}]\,du = (n_1^2/a)[u_b - u_a + (e^{-2u_a} - e^{-2u_b})/2] \end{aligned} \tag{2.35}$$

For the cases of practical interest, the value of p at $x = x_a$, denoted by p_a, and the value of n at $x = x_b$, denoted by n_b, will both be large compared to n_1. Consequently, we conclude that

$$u_a = -\ln(p_a/n_1) \text{ and } u_b = \ln n_b/n_1$$

are both larger than unity in absolute value but probably less than twenty for a reasonable variation of impurity between x_a and x_b. (For example for a change in potential of 0.2 volts such as would occur between *p*- and *n*-type germanium, u_a and u_b would each be about 4 in magnitude.) Hence we obtain for P,

$$\begin{aligned} P &= (n_1^2/2a)(2(u_b - u_a) + (p_a/n_1)^2 - (n_1/n_b)^2) \\ &\cong p_a^2/2a + (n_1^2/a)(u_b - u_a) \end{aligned} \tag{2.36}$$

where we have neglected $(n_1/n_b)^2$ which is $\ll 1$ and the negligible compared to $u_b - u_a$. The term $p_a^2/2a$ is simply the integrated acceptor-minus-donor density in the *p*-region, as may be seen as follows:

$$\int_{x_a}^{0} (N_a - N_d)\,dx = \int_{x_a}^{0} (-ax)\,dx = ax_a^2/2 = p_a^2/2a. \tag{2.37}$$

The second term in (2.36) is essentially the sum of the holes of the right of $x = 0$ plus the electrons to the left of $x = 0$, whose charge is also com-

pensated by holes. The total number of holes can be expressed in terms of $\delta\varphi$ through the dependence of n_1 on $\delta\varphi$. The second term is thus

$$(n_1^2/a)[\ln(n_b/n_1) + \ln(p_a/n_1)] = (n_i^2/a)e^{q\delta\varphi/kT} \cdot [\ln(n_b p_a/n_i^2) - q\delta\varphi/kT] \quad (2.38)$$

Hence for a small change $d\delta\varphi$ in $\delta\varphi$, the change in charge $dQ = q\,dP$ and the capacity C are given by

$$C = \frac{dQ}{d\delta\varphi} = \frac{q^2}{kT}\frac{n_1^2}{a}[\ln(n_b p_a/n_i^2) - (q\delta\varphi/kT) - 1]. \quad (2.39)$$

This capacity can be reexpressed in terms of the difference in ψ between x_a and x_b : When $\delta\varphi = 0$, corresponding to the thermal equilibrium case, we have

$$p_a n_b = n_i^2\, e^{q(\psi_b - \psi_a)/kT} \quad (2.40)$$

Using this together with the definitions of L_D and L_a we obtain

$$C = \frac{\kappa[q(\psi_b - \psi_a - \delta\varphi)/kT - 1]\, e^{q\delta\varphi/kT}}{4\pi(2L_D^2/L_a)} \quad (2.41)$$

In this expression ψ_a and ψ_b are the potentials when $\delta\varphi = 0$; so that

$$\psi_b - (\psi_a + \delta\varphi)$$

is thus the increase in potential in going from x_a to x_b when $\delta\varphi$ is applied.

For thermal equilibrium, $\delta\varphi = 0$ and, as discussed above, the term in $\psi_b - \psi_a$ will be about 10. Hence, using the definition $K = L_D/2L_a$, we have

$$C \cong \kappa/4\pi(4KL_D/10) \quad (2.42)$$

For $K \ll 1$, the case for which this formula is valid, C will be the capacity of a condenser whose dielectric layer is much less than L_D thick.

Capacity for Space Charge Case, $K \gg 1$

As discussed in connection with (2.30) and (2.31), the applied potential $\delta\varphi$ reduces the increase ($= 2\psi_m$) in ψ between the p-region and the n-region by $\delta\varphi/2$ on each side of $x = 0$. This is accomplished by a narrowing of the space charge layer by δx_m on each side where (according to (2.20))

$$\delta\psi_m = -\delta\varphi/2 = 4\pi q a x_m^2 \delta x_m/\kappa \quad (2.43)$$

The decrease in width δx_m brings with it an increase in number of holes $-ax\,\delta x_m$ per unit area of the junction on the p-side and an equal number of electrons on the n-side. Thus a charge of holes per unit area of $\delta Q = -qax_m\delta x_m$ must flow in from the left. The capacity per unit area is, therefore,

$$C = \delta Q/\delta\varphi = qax_m\delta x_m/\delta\varphi = \kappa/4\pi 2x_m \quad (2.44)$$

corresponding to a condenser of thickness $2x_m$. It is evident that formula (2.44) will hold for a small change $d\delta\varphi$ superimposed on a large bias $\delta\varphi$ provided that $2x_m$ is the thickness of the space charge region under the conditions when $\delta\varphi$ is applied. If ψ_{n0} is the value of ψ for $\delta\varphi = 0$, then $\psi_m = \psi_{n0} - \delta\varphi/2$; and C will vary as

$$C = \kappa[4\pi qa/3\kappa(\psi_{n0} - \delta\varphi/2)]^{1/3}/8\pi \tag{2.45}$$

so that $1/C^3$ should plot as a straight line versus $\delta\varphi$ with slope

$$-(8\pi/\kappa)^3(3\kappa/8\pi qa) = -\frac{192\pi^2}{\kappa^2 qa}. \tag{2.46}$$

In addition to the holes which flow to account for the change in ψ_m, the concentration of holes in the n-region will be increased by a factor $\exp(q\delta\varphi/kT)$. However, this increase does not lie in the transition region; we shall consider it later, in Section 4, in connection with a-c. admittance.

Comparison of the Two Capacities

It is instructive to compare the two capacities just derived. We suppose that for one value of n_i we have $K \gg 1$ so that the space charge solution is good. For this case we choose $x_a = -x_m$ and $x_b = +x_m$ so as to bound the space charge layer. We then imagine n_i to be increased, either by raising the temperature or by applying a potential difference $\delta\varphi$. The capacity then changes from

$$C_{\text{sp. chg.}} = \kappa/8\pi x_m \text{ to } C_{\text{neut.}} = 5\kappa/8\pi KL_D \tag{2.47}$$

(i.e., from (2.44) to (2.42)) so that the ratio is

$$\frac{C_{\text{neut.}}}{C_{\text{sp. chg.}}} = \frac{5x_m}{KL_D} \tag{2.48}$$

For $K < 1$, this ratio is large, both because of K in the denominator and because $x_m > L_a$ so that $x_m/L_D > L_a/L_D = 1/2\ K$.

In Section 4.4 we shall compare these capacities with that due to diffusion of holes and electrons beyond the transition region.

2.4 The Abrupt Transition

For completeness we shall consider the case in which the impurity concentration changes abruptly from p_p to n_n at $x = 0$. For this case the potential in the space-charge layer will be of the parabolic type discussed by Schottky, the potentials varying as

$$\psi = (2\pi/\kappa)q\ p_p(x - x_p)^2 + \text{constant}, \qquad x < 0 \tag{2.49}$$

$$\psi = -(2\pi/\kappa)q\ n_n(x - x_n)^2 + \text{constant}, \qquad x > 0 \tag{2.50}$$

where $x_p < 0$ and $x_n > 0$ are the ends of the space-charge layer in the p- and n-regions. The gradient of potential at $x = 0$ must be equal for the two layers leading to

$$-p_p x_p = n_n x_n \tag{2.51}$$

so that if the total width of the space charge layers is $W = x_n - x_p$, it follows that

$$x_p = -n_n W/(n_n + p_p) \text{ and } x_n = p_p W/(n_n + p_p). \tag{2.52}$$

The potential difference across the layer, which is $\psi_b - \psi_a$ is

$$\psi_b - \psi_a = (2\pi q/\kappa)(p_p x_p^2 + n_n x_n^2) = [2\pi q\ p_p n_n/\kappa(p_p + n_n)]W^2 . \tag{2.53}$$

If $p_p \gg n_n$ this reduces to

$$\psi_b - \psi_a = 2\pi q\ n_n W^2/\kappa \tag{2.54}$$

the formula given by Schottky, which should be appreciable in this case, for which all the voltage drop occurs in the n-region.

The capacity for the abrupt transition will be

$$C = \kappa/4\pi W \tag{2.55}$$

where W is obtained by solving (2.53). For this case $(1/C)^2$ should plot as a straight line versus $\psi_b - \psi_a$:

$$\frac{1}{C^2} = [8\pi(p_p + n_n)/\kappa q\ p_p\ n_n](\psi_b - \psi_a). \tag{2.56}$$

3. General Conclusions Concerning the Junction Characteristic

In this section we shall consider direct current flow through the junction and shall derive the results quoted in Fig. 1 relating the current distribution to the characteristics of the junction. We shall suppose that holes and electrons are thermally generated in pairs at a rate g and recombine at a rate rnp so that the net rate of generation per unit volume is

$$\text{(net rate of generation)} = g - rnp,$$

which vanishes at equilibrum. Obviously, $g = rn_i^2$. If relatively small concentrations δp and δn of holes and electrons are present in excess of the equilibrium values, the net rate of generation is

$$\delta\dot{p} = \delta\dot{n} = g - r(n + \delta n)(p + \delta p) = -rn\delta p - rp\delta n \tag{3.1}$$

This is equivalent to saying that excess holes in an n-type semiconductor,

and excess electrons in a p-type semiconductor, respectively, have lifetimes τ_p and τ_n given by

$$\delta\dot{p} = -\delta p/\tau_p = -rn\delta p \text{ or } \tau_p = 1/rn = p/g \tag{3.2}$$

and

$$\delta\dot{n} = -\delta n/\tau_n = -rp\delta n \text{ or } \tau_n = 1/rp = n/g. \tag{3.3}$$

We shall have occasion to use this interpretation later. (We later consider the modifications required when surface recombination occurs, Section 4.2, Appendix V, and the effect of a localized region of high recombination rate, Section 4.6, Appendix III.)

In principle, the steady-state solution can be obtained in terms of the three potentials ψ, φ_p and φ_n. These must satisfy three simultaneous ordinary differential equations, which we shall derive. As discussed in Section 2, we consider all donors and acceptors to be ionized so that Poisson's equation becomes

$$\frac{d^2\psi}{dx^2} = -\frac{4\pi q}{\kappa}\left(ax + n_i e^{q(\varphi_p-\psi)/kT} - n_i e^{q(\psi-\varphi_n)/kT}\right) \tag{3.4}$$

an equation in which the unknowns are the three functions φ_p, φ_n and ψ. The total current density, from left to right, is

$$I = I_p + I_n = -q\mu\left[p\frac{d\varphi_p}{dx} + bn\frac{d\varphi_n}{dx}\right]. \tag{3.5}$$

The elimination of p and n by equation (2.4) results in an equation involving the three unknown functions and I. The divergence of hole current, equal to the net rate of generation of holes, is

$$\begin{aligned}\frac{dI_p}{dx} &= -q\mu p\left[\frac{q}{kT}\left(\frac{d\varphi_p}{dx}\right)^2 - \frac{q}{kT}\frac{d\psi}{dx}\frac{d\varphi_p}{dx} + \frac{d^2\varphi_p}{dx^2}\right] \\ &= q(g - rnp) = qg(1 - e^{q(\varphi_p-\varphi_n)/kT}),\end{aligned} \tag{3.6}$$

with p in the second term given by (2.4) so that (3.6) is also an equation for the three unknown functions. The equation for dI_n/dx can be derived from the last two and adds nothing new. These three equations can be used to solve for $d^2\psi/dx^2$, $d^2\varphi_p/dx^2$ and $d\varphi_n/dx$ in terms of lower derivatives and I. They thus constitute a set of equations sufficient to solve the problem provided that physically meaningful boundary conditions are imposed. We shall not, however, deal directly with these equations; the main reason for deriving them was to show that the problem in question is, in principle, completely formulated. Instead of attempting to solve the equations, we shall discuss certain general features of the solutions for φ_p and φ_n, using

approximate methods, and in this way bring out the essential features of the theory of rectification.

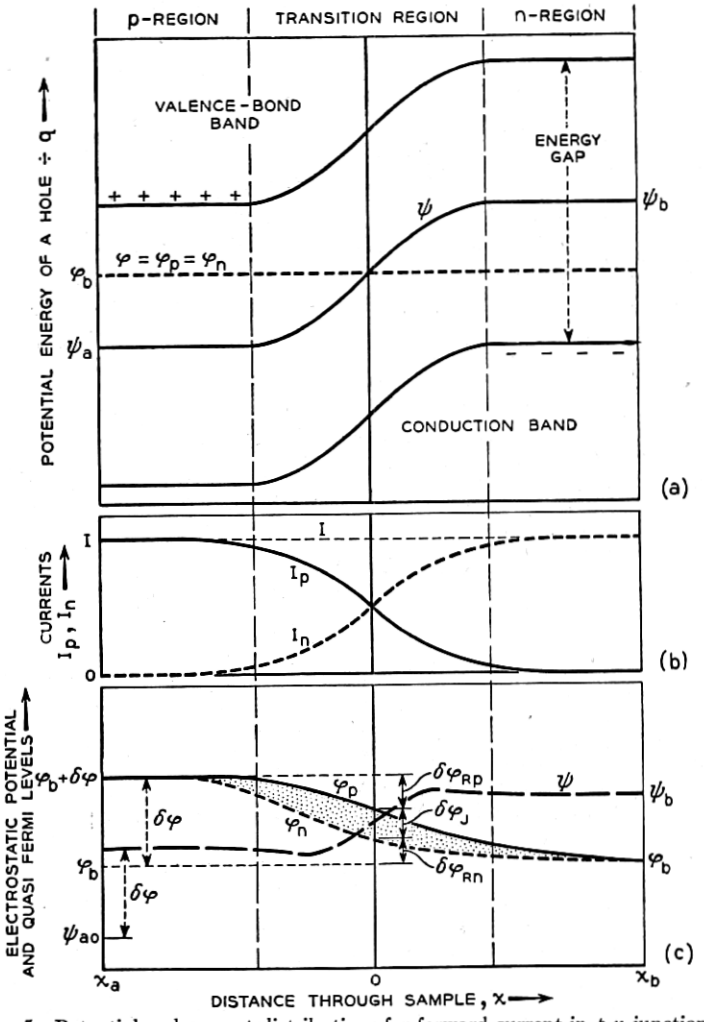

Fig. 5—Potential and current distributions for forward current in p-n junctions.
(a) p-n junction under equilibrium conditions.
(b) Division of current between holes and electrons.
(c) Distribution of potentials for forward current flow showing how the potential $\delta\varphi$ applied at x_a changes φ_p, φ_n and ψ.

In Fig. 5 we represent a general situation which may be used to illustrate the nature of the resistance of the junction. Part (a) corresponds to thermal equilibrium and shows the potential distribution and Fermi level in ac-

cordance with the scheme used in Fig. 2. Part (b) shows the current distribution for a forward current I from left to right and (c) shows the corresponding potential distribution and values of φ_n and φ_p, the total applied potential being $\delta\varphi$. Recombination prevents the hole current from penetrating far into the n-region, the depth of penetration being described by the diffusion length $L_p = \sqrt{D\tau_p} = \sqrt{Dp_n/g}$, where p_n is the hole concentration in the n-region. The electron current similarly is limited by $L_n = \sqrt{bD\tau_n} = \sqrt{bDn_p/g}$. (Diffusion lengths are evaluated for particular models of the junction in Section 4.) Far from the junction, therefore, the hole and electron concentrations have their normal values and consequently $\varphi_p = \varphi_n$ and $\varphi_p - \psi$ has its normal value. This accounts for the equal displacement $\delta\varphi$ for all three curves at $x = x_a$. The curves for φ_p and φ_n have a continuous downward trend which produces the currents

$$I_p = -q\mu p \frac{d\varphi_p}{dx} \quad \text{and} \quad I_n = -qb\mu n \frac{d\varphi_n}{dx}. \tag{3.7}$$

The area between the φ_p and φ_n curves has a special significance: This difference is related to the excess rate of recombination and the integral of this rate over the entire specimen must be sufficient to absorb the hole current $I_p = I$ entering at x_a so that the entire current at x_b is carried by electrons. In terms of $\varphi_p - \varphi_n$ and equation (3.6) we obtain

$$\begin{aligned} I &= I_p(x_a) - I_p(x_b) = \int_{x_a}^{x_b} - dI_p \\ &= gq \int_{x_a}^{x_b} (e^{q(\varphi_p - \varphi_n)/kT} - 1)\, dx. \end{aligned} \tag{3.8}$$

From (3.8) we conclude that if g is increased indefinitely for a specified current I, then $\varphi_p - \varphi_n$ must approach zero. For this case, in which the rate of recombination and generation is very high, $\varphi_p = \varphi_n$ and

$$I = I_p + I_n = -q\mu(p + bn)\, d\varphi_p/dx \tag{3.9}$$

and

$$\delta\varphi = -\int_{x_a}^{x_b} d\varphi_p = I \int_{x_a}^{x_b} dx/q\mu(p + bn) \equiv IR_0, \tag{3.10}$$

where R_0 is simply the integral of the local resistivity corresponding to densities p and n. For smaller values of g, I does not divide in the ratio $p:bn$ and $\varphi_p \neq \varphi_n$ and $\delta\varphi > IR_0$.[10]

We shall next give an approximate treatment for the case in which $\delta\varphi_J$ (J for junction), the value of $\varphi_p - \varphi_n$ at $x = 0$, is an appreciable fraction of

[10] A general proof that $\delta\varphi > IR_0$ is given in Appendix I.

the total voltage drop. For this purpose we treat $\varphi_p - \varphi_n$ as constant over a range of integration from $x = -L_n$ to $x = +L_p$ obtaining

$$\begin{aligned} I &= gq(L_n + L_p)[e^{(q\delta\varphi_J/kT)} - 1] \\ &= I_s[e^{(q\delta\varphi_J/kT)} - 1] \end{aligned} \tag{3.11}$$

where

$$I_s = gq(L_n + L_p) \tag{3.12}$$

is the current density corresponding to the total rate of generation of hole-electron pairs in a volume $L_n + L_p$ thick. We next consider $\delta\varphi_{Rp} + \delta\varphi_{Rn}$ shown in Fig. 5c, where, as the subscript R implies, these are thought of as resistive terms and are given by the integrals

$$\delta\varphi_{Rp} + \delta\varphi_{Rn} = -\int_{x_a}^{0} d\varphi_p - \int_{0}^{x_b} d\varphi_n = \int_{x_a}^{0} I_p\, dx/q\mu p + \int_{0}^{x_b} I_n\, dx/q\mu bn.$$

The denominators are both approximately $q\mu(p + bn)$ which occurs in the integral for R_0. Furthermore, for most of the first range $I_p = I$ and for most of the second $I_n = I$. Near $x = 0$, I_p or I_n must be at least $I/2$. Hence it is evident that $\delta\varphi_{Rp} + \delta\varphi_{Rn}$ cannot be much less than IR_0. We shall represent it by IR_1 where $R_0 < 2R_1 < 2R_0$.

In terms of R_1 and I_s, the relationship between current and voltage becomes

$$\delta\varphi = \delta\varphi_{Rp} + \delta\varphi_{Rn} + \delta\varphi_J = R_1 I + \frac{kT}{q} \ln\left(1 + \frac{I}{I_s}\right). \tag{3.13}$$

This corresponds to an ideal rectifier in series with a resistance R_1. The junction will, therefore, be a good rectifier if the second term represents a much higher resistance.

We shall compare the two resistances for the case corresponding to $K \ll 1$. For this case, we have $p = -ax$ and $n = +ax$ except in the narrow range $|x| < L_a = n_i/a$. The integral R_0 can be approximated by integrating dx/σ for x outside of the range $\pm L_a$ using the approximation $\pm ax$ for p and n and approximating the integral from $-L_a$ to $+L_a$ by $2L_a/\sigma$ (intrinsic). This procedure gives

$$\begin{aligned} R_1 &= \int_{L_a}^{-x_a} dx/q\mu ax + \frac{2L_a}{q\mu n_i(1 + b)} + \int_{L_a}^{x_b} dx/q\mu bax \\ &= \frac{L_a}{q\mu n_i}\left(1 + \frac{1}{b}\right) \ln\,(x_b/L_a) \end{aligned} \tag{3.14}$$

where it is supposed that $-x_a \doteq x_b$ and that $ln\,(x_b/L_a)$ is large compared to $2/(b + 2 + 1/b)$. The evaluation of L_p and L_n for use in I_s is more involved

since τ_p and τ_n are both functions of x. We shall obtain an approximate self-consistent diffusion length by assuming that the holes diffuse, on the average, to just such a depth, L_p, that in uniform material of the type found at L_p, their diffusion length would also be L_p. At a depth L_p, the value of n is aL_p so that by (3.2), τ_p is $1/raL_p = n_i^2/gaL_p$. Thus we write

$$L_p^2 = D\tau_p = Dn_i^2/gaL_p. \tag{3.15}$$

We can solve the equation (3.15) for L_p and a similar one for L_n and insert the results in equation (3.13). For small I this gives

$$\delta\varphi/I = R_1 + (kT/qI_s) = \frac{L_a}{q\mu n_i}\left(1 + \frac{1}{b}\right)\ln(x_b/L_a) + kT/(q^2 g^{2/3}(DL_a n_i)^{1/3}(1 + b^{1/3})). \tag{3.16}$$

It is seen that for g large, the second term, corresponding to the rectifying resistance, becomes small. For this case, as discussed above, $\varphi_p = \varphi_n$ and the exact integral for R_0 should be used and the junction will give poor rectification.

It is also instructive to consider L_a as a variable. Increasing L_a corresponds to making the transition from p to n more gradual. It is evident that varying L_a changes the two terms of (3.16) in opposite directions so that there will be an intermediate value of L_a for which the resistance of the junction is a minimum. As L_a approaches zero, however, the second term should be modified: If we imagine that in the transition region the concentration $(N_d - N_a)$ varies only over a finite range, bounded by fixed values n_n and p_p in the n- and p-regions, then it is clear that the limiting values of L_p and L_n should be given not by (3.15) but by $\sqrt{D\tau_p}$ and $\sqrt{bD\tau_n}$ where τ_p and τ_n are evaluated in the n-region and p-region. This leads to a limiting value for I_s, which is given in equation (4.11) of the following section. In the range for which (3.16) applies, however, the interesting result holds that widening the transition region initially decreases the resistance by furnishing a larger volume in which holes and electrons may combine or be generated.

The condition that $\delta\varphi_J$ dominate the resistance is that the second term of (3.16) be much larger than the first. This leads to the inequality

$$1 \ll \frac{kT}{q^2 g^{2/3}(DL_a n_i)^{1/3}} \cdot \frac{q\mu n_i}{L_a} = (Dn_i/gL_a^2)^{2/3} = (L_{pi}/L_a)^{4/3} \tag{3.17}$$

where we have neglected various factors involving b, which are nearly unity, and $\ln(x_b/L_a)$ (which must be about 4 for Ge since the conductivity at x_b is about exp(4) times the intrinsic conductivity). The quantity

$$L_{pi} = (Dn_i/g)^{1/2} \tag{3.18}$$

is the diffusion length for holes in the intrinsic region. The inequality states that the diffusion length must be much larger than L_a. This is equivalent to the previous statement that the hole current must penetrate the n-region for the rectifier to have a good characteristic. (If a local region of high recombination is present in the transition region, this result just quoted need not apply. See Section 4.6.)

If the hole current penetrates deeply into the n-region and R_1 is negligible, then we can conclude that the current-voltage characteristic will fit the ideal formula. For these assumptions $\delta\varphi_{Rp}$ on Fig. 5 will be small and the principal change in φ_p will occur relatively deep in the n-region, at least beyond the transition region. So long as the hole concentration introduced in the n-region is much smaller than n_n, the hole current into the n-region will be a linear function of the value of p at the right edge of the transition region, being zero when p equals p_n, the equilibrium value of p. This leads at once to a hole current proportional to $p - p_n$ and since the shift of φ_p in respect to ψ at the edge of transition region is $\delta\varphi$, $p - p_n$ is equal to $p_n(\exp(a\delta\varphi/kT)-1)$. (These ideas are discussed in detail in Section 4.) A similar relationship will hold for electrons entering the p-region; hence the total current will vary as $\exp(q\delta\varphi/kT)-1$. This is a theoretical rectification formula[11] giving the maximum rectification for carriers of charge q.

4. Treatment of Particular Models

4.1 Introduction and Assumptions

In this section we shall deal chiefly with good rectifiers so that the IR drop, discussed in connection with R_1 in Section 3, is negligible. We shall deal chiefly with the case for which the transition region is narrow compared to the diffusion length; consequently, there is little change in I_p in traversing the transition region. In Fig. 6(a) we consider a hypothetical junction in which the properties are uniform outside the transition region. The division of the specimen into three parts as shown is seen to be reasonable for germanium: In n-type germanium, the diffusion constant for holes is about 40 cm^2/sec and the lifetime is greater than 10^{-6} sec so that the diffusion distance is $L_p = \sqrt{D\tau_p} > 6 \times 10^{-3}$ cm. This is much greater than most transition regions.

The major drop in φ_p must occur to the right of the transition region. This follows from our assumptions: First, we may neglect the IR drop in the p-region; hence φ_p is substantially constant from $x = x_a$ to $x = x_{Tp}$. Second, the decrease in φ_p is much less in the transition region than in the n-region; this follows from two considerations: the resistance for hole flow is lower in

[11] C. Wagner, *Phys. Zeits.* 32, 641–645 (1931).

the transition region than in the *n*-region; the effective length of flow in the *n*-region, being L_p, is greater than the width of the transition region. Consequently, the variation of φ_p shown in Fig. 6(c) is seen to be reasonable. Similar considerations apply to φ_n. As is shown in Fig. 6(c), the application of $\delta\varphi$ does not alter $\varphi_p - \psi$ in the *p*-region nor $\varphi_n - \psi$ in the *n*-region. The

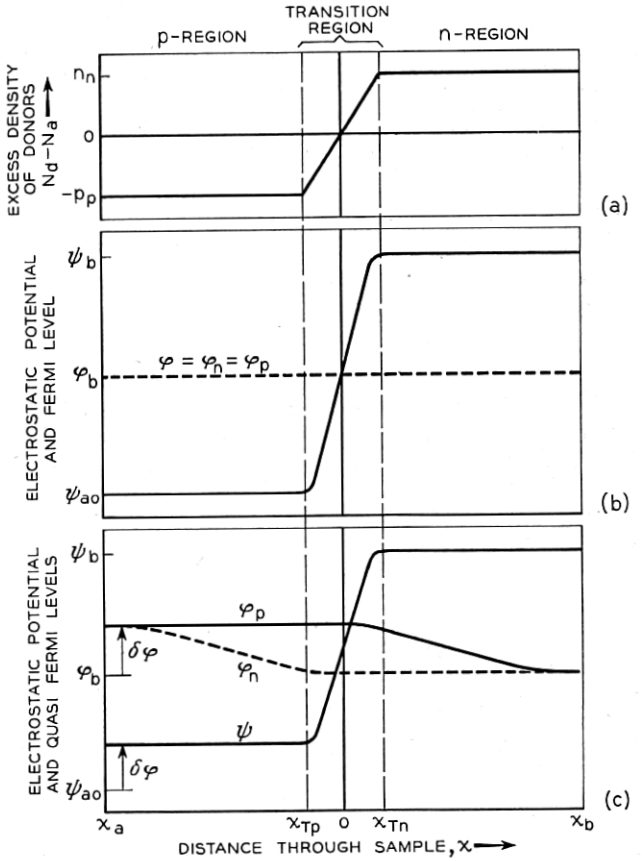

Fig. 6—Simplified model of a *p-n* junction.
(a) Distribution of donors and acceptors.
(b) Potentials for thermal equilibrium.
(c) Effect of $\delta\varphi$ applied potential in forward direction.

reason, as discussed in connection with (2.31), is that in these regions electrical neutrality requires an essentially constant value for the more abundant carrier. Hence the relationships between the φ's and ψ follow from (2.4).

The nature of the potential distribution in the transition region has no effect in the considerations just discussed. However, as shown in Section 2,

the capacity of the transition region, which we shall denote by C_T in this section, does depend on the nature of the transition region and, consequently, on the value of K.

If the sizes of the p-region and n-region are large compared to the diffusion lengths, we may assume the current at x_a to be substantially I_p only and that at x_b, I_n only. The total current entering at x_a can be accounted for as doing three things: (1) neutralizing the electron current flowing into the p-region across x_{Tp}, (2) contributing to the charge in the transition region (this corresponds to the capacity discussed in Section 2) and (3) contributing a current flow to the right across x_{Tn}.

We have selected the hole current for analysis because the hole has a positive charge and the connection between the algebra and the physical picture is simplified. For the same reason, the text emphasizes forward current, although the equations are equally applicable to reverse currents. Nothing essential is left out by this process; since the sample as a whole remains uncharged, the current I is the same for all values of x and if I_p is known, then $I_n = I - I_p$ is also determined.

4.2 Solution for Hole Flow into the n-region

We shall calculate first the hole current $I_p(x_{Tn})$ flowing across x_{Tn}. It is readily evaluated as follows: The value of $p(x_{Tn})$ is given by

$$\begin{aligned} p(x_{Tn}) &= n_i e^{q(\varphi_b+\delta\varphi-\psi_b)/kT} \\ &= p_n e^{q\delta\varphi/kT} \end{aligned} \tag{4.1}$$

where p_n is the hole concentration in the n-region for thermal equilibrium. If we apply a small a-c. signal superimposed on a d-c. bias so that

$$\delta\varphi = v_0 + v_1 e^{i\omega t} \tag{4.2}$$

where v_1 is an a-c. signal, assumed so small that linear theory may be employed (i.e. $v_1 \ll kT/q$), then

$$p(x_{Tn}) = (p_n e^{qv_0/kt})(1 + (qv_1/kT)e^{i\omega t}).$$

We resolve this density into a d-c. component p_0 and an a-c. component $p_1 e^{i\omega t}$:

$$p(x_{Tn}) = p_n + p_0 + p_1 e^{i\omega t} \tag{4.3}$$

where

$$p_0 = p_n(e^{qv_0/kT} - 1) \tag{4.4}$$

$$p_1 = (qp_n v_1/kT)e^{qv_0/kT}. \tag{4.5}$$

So long as $p(x_{Tn}) \ll n_n$, the normal concentration of electrons in the

n-region, the lifetime τ_p and diffusion constant D for a hole will be substantially unaltered by $\delta\varphi$. Application of the hole-current equation to the hole density $p(x, t)$ gives

$$I_p = -qD\frac{\partial p}{\partial x}. \tag{4.6}$$

Combining this with the recombination equation

$$\frac{\partial p}{\partial t} = \frac{p_n - p}{\tau_p} - \frac{1}{q}\frac{\partial I_p}{\partial x} = \frac{p_n - p}{\tau_p} + D\frac{\partial^2 p}{\partial x^2} \tag{4.7}$$

leads to the solution

$$p = p_n + p_0 e^{(x_{Tn}-x)/\sqrt{D\tau_p}} + p_1 e^{i\omega t+(x_{Tn}-x)(1+i\omega\tau_p)^{1/2}/(D\tau_p)^{1/2}}. \tag{4.8}$$

The quantity $\sqrt{D\tau_p}$ is the diffusion length and is denoted by L_p. (We shall use subscript p for holes in the n-region and n for electrons in the p-region for both L and τ.)

When p is large compared to p_n, but small compared to n_n, the expression for p leads to the following formula for φ_p:

$$\varphi_p = \varphi_n + v_c - (kT/q)(x - x_{Tn})/L_p + v_1 e^{i\omega t-(x-x_{Tn})[(1+i\omega\tau_p)^{1/2}-1]/L_p}. \tag{4.9}$$

This shows that the d-c. part of φ_p varies linearly in the n-region, for large forward currents, and decreases by (kT/q) in each diffusion length L_p. The transition from this linear dependence to an exponential decay for φ_p comes when $\varphi_p - \varphi_n = (kT/q)$. This behavior of the d-c. part of φ_p is useful in connection with diagrams of φ_p versus distance. (See Sections 5 and 6.)

The solution just obtained for p gives rise to a current at x_{Tn} of

$$\begin{aligned} I_p(x_{Tn}) &= -qD\frac{\partial p}{\partial x} \\ &= qp_0 D/L_p + qp_1 De^{i\omega\tau}(1 + i\omega\tau_p)^{1/2}/L_p. \end{aligned} \tag{4.10}$$

The d-c. part is calculated by substituting (4.4) for p_0:

$$\begin{aligned} I_{p0}(x_{Tn}) &= (qp_n D/L_p)(e^{qv_0/kT} - 1); \\ &\equiv I_{ps}(e^{qv_0/kT} - 1) \end{aligned} \tag{4.11}$$

and the a-c. part is similarly obtained from (4.5) for p_1:

$$\begin{aligned} I_{p1}(x_{Tn}) &= (qp_n\mu/L_p)[e^{(qv_0/kT)}](1 + i\omega\tau_p)^{1/2}v_1 e^{i\omega t} \\ &\equiv (G_p + iS_p)v_1 e^{i\omega t} \equiv A_p v_1 e^{i\omega t} \end{aligned} \tag{4.12}$$

where A_p is called the admittance (per unit area) for holes diffusing into the n-region; its real and imaginary parts are the conductance and suscept-

ance. For $\omega\tau_p$ small, the real term G_p is simply conductance per cm^2 of a layer L_p cm thick with hole conduction corresponding to the density $p_n + p_0$; it is also the differential conductance obtained by differentiating (4.11) in respect to v_0 . For the case of zero bias this establishes the result quoted in Section 1 that the voltage drop is due to hole flow in the n-region where the hole conductivity is low.

In this section we have treated τ_p as arising from body recombination. In a sample whose y and z dimensions are comparable to L_p or L_n , surface recombination may play a dominant role. However, as we show in Appendix V, the theory given here may still apply provided appropriate values for τ_p and τ_n are used.

4.3 D-C. Formulae

The total direct hole current flowing in at x_a is I_{p0} plus the current required to recombine with electrons in the p-region. This latter current is, of course, equal to the electron current flowing into the p-region. This electron current, denoted by I_{n0} or $I_{n0}(x_{Tp})$, is obtained by the same procedure as that leading to (4.11) for I_{p0} except that bD replaces D and the subscripts of L and τ are now n. Combining the two currents leads to the total direct current:

$$I_0 = I_{p0} + I_{n0} = (qL)\left(\frac{p_n}{L_p} + \frac{bn_p}{L_n}\right)(e^{qv_0/kT} - 1) \tag{4.13}$$

for the direct current per unit area for applied potential v_0 .[12] The algebraic signs are such that $I > 0$ corresponds to current from the p-region to the n-region in the specimen; $v_0 > 0$ corresponds to a plus potential applied to the p-end. The ratio of hole current to electron current across the transition region is

$$\begin{aligned}\frac{I_{p0}}{I_{n0}} &= \frac{p_n}{L_p}\cdot\frac{L_n}{bn_p} = \frac{p_p}{bn_n}\cdot\frac{\sqrt{bD\tau_n}}{\sqrt{D\tau_p}} \\ &= \frac{p_p}{bn_n}\sqrt{\frac{bn_n}{p_p}} = \sqrt{\frac{\sigma_p}{\sigma_n}}\end{aligned} \tag{4.14}$$

where we have used the relationships $n_n p_n = n_p p_p = n_i^2$ from (2.2) and $\tau_p n_n = \tau_n p_p = 1/r$ from (3.2) and (3.3). These results can be summarized by saying that the current flows principally into the material of higher re-

[12] For convenience we repeat the definitions here: $q \equiv$ magnitude of electronic charge; $D \equiv$ diffusion constant for holes; p_n and $n_n \equiv$ thermal equilibrium value of p and n, assumed constant throughout n-region ($x > x_{Tn}$); n_p and $p_p \equiv$ similar values for $x < x_{Tp}$; $L_p \equiv$ diffusion length $\equiv \sqrt{D\tau_p}$ for holes in n-region; $\tau_p \equiv$ lifetime of hole in n-region before recombination; b = electron mobility/hole mobility; L_n and τ_n similar in quantities for electrons in p-region; $\sigma_n = q\mu bn_n$ and $\sigma_p = q\mu p_p$ are the conductivities of the two regions.

sistivity. We can also say that the hole current depends only on the *n*-type material and vice versa. For a *p-n* junction emitter in a transistor with an *n*-type base, it is thus advantageous to use high conductivity *p*-type material so as to suppress an unwanted electron current.

For comparison with experiment, it is advantageous to express the values of p_n and n_p in terms of the conductivities σ_n and σ_p. If the conductivity of the intrinsic material is written as

$$\sigma_i = q\mu n_i(1 + b), \tag{4.15}$$

then, if $p_n \ll n_n$ and $n_p \ll p_p$, we find

$$q\mu p_n = b\sigma_i^2/(1 + b)^2\sigma_n \tag{4.16}$$

$$q\mu b n_p = b\sigma_i^2/(1 + b)^2\sigma_p . \tag{4.17}$$

Using these equations, we may rewrite (4.11) and a corresponding equation for electron current into the *p*-region so as to express their dependence on d-c. bias v_0 and the properties of the regions:

$$\begin{aligned} I_{p0}(v_0) &= \frac{b\sigma_i^2}{(1 + b)^2\sigma_n L_p} \cdot \frac{kT}{q}(e^{qv_0/kT} - 1) \\ &\equiv G_{p0}\frac{kT}{q}(e^{qv_0/kT} - 1) \\ &\equiv I_{ps}(e^{qv_0/kT} - 1) \end{aligned} \tag{4.18}$$

$$\begin{aligned} I_{n0}(v_0) &= \frac{b\sigma_i^2}{(1 + b)^2\sigma_p L_n} \cdot \frac{kT}{q}(e^{qv_0/kT} - 1) \\ &\equiv G_{n0}\frac{kT}{q}(e^{qv_0/kT} - 1) \\ &\equiv I_{ns}(e^{qv_0/kT} - 1). \end{aligned} \tag{4.19}$$

The values of G_{p0} and G_{n0} (which are readily seen to be the values of the low-frequency, low-voltage ($v_0 < kT/q$) conductances) and the saturation reverse currents are given by

$$G_{p0} \equiv \frac{b\sigma_i^2}{(1 + b)^2\sigma_n L_p} \equiv \frac{q}{kT} I_{ps} \tag{4.20}$$

$$G_{n0} \equiv \frac{b\sigma_i^2}{(1 + b)^2\sigma_p L_n} \equiv \frac{q}{kT} I_{ns} \tag{4.21}$$

The expression for direct current then becomes

$$\begin{aligned} I_0(v_0) &= [I_{p0} + G_{n0}]\left(\frac{kT}{q}\right)[e^{qv_0/kT} - 1] \\ &= (I_{ps} + I_{ns})[e^{qv_0/kT} - 1]. \end{aligned} \tag{4.22}$$

4.4 Total Admittance

In order to calculate the alternating current, we must include the capacity of the transition region, discussed in Section 2. Denoting this by C_T, we then find for the total alternating current.

$$I_{ac} = (G_p + iS_p + G_n + iS_n + i\omega C_T)\, v_1 = Av_1 \tag{4.23}$$

where G_n and S_n are similar to G_p and S_p but apply to electron current into the p-region. The value of the hole and electron admittances can be expressed as

$$A_p = G_p + iS_p = (1 + i\omega\tau_p)^{1/2} G_{p0}\, e^{qv_0/kT} \tag{4.24}$$

$$A_n = G_n + iS_n = (1 + i\omega\tau_n)^{1/2} G_{n0}\, e^{qv_0/kT} \tag{4.25}$$

For low frequencies, such that ω is much less then $1/\tau_p$, we can expand $G_p + iS_p$ as follows:

$$G_p + iS_p = G_{p0}\, e^{qv_0/kT} + i\omega(\tau_p/2)G_{p0}\, e^{qv_0/kT} \tag{4.26}$$

Hence $(\tau_p/2)G_{p0}\, e^{qv_0/kT}$ behaves like a capacity.

It is instructive to interpret this capacity for the case of zero bias, $v_0 = 0$, for which we find:

$$C_p = \tau_p G_{p0}/2 = \tau_p q p_n \mu/2L_p = q^2 p_n L_p/2kT. \tag{4.27}$$

The last formula, obtained by noting that $\tau_p\mu = q\tau_p D/kT = qL_p^2/kT$, has a simple interpretation: qp_nL_p is the total charge of holes in a layer L_p thick. For a small change in voltage v, this density should change by a fraction qv/kT so that the change in charge divided by the change in v is $(q/kT)(qp_nL_p)$ which differs from C_p only by a factor of 2, which arises from the nature of the diffusion equation.

This capacity can be compared with $C_{T\,\text{neut.}}$, discussed in Section 2, (see equation (2.39) and text for (2.42)) for germanium at room temperature as follows:

$$\frac{C_p}{C_{T\,\text{neut.}}} = \frac{q^2 p_n L_p}{2kT} \cdot \frac{kTa}{10q^2 n_i^2} = \frac{p_n L_p a}{20n_i^2}. \tag{4.28}$$

For a structure like Fig. 6(c), the excess of donors over acceptors reaches its maximum value, equal to n_n, at x_{Tn} leading to $n_n = ax_{Tn}$. Consequently $a = n_n/x_{Tn}$. Substituting this value for a in (4.28) and noting that $p_n n_n = n_i^2$ gives

$$\frac{C_p}{C_{T\,\text{neut.}}} = \frac{L_p}{20x_{Tn}} \tag{4.29}$$

As discussed at the beginning of this section, $L_p \doteq 6 \times 10^{-3}$ cm for holes

in germanium. Hence if the transition region is 6×10^{-4} cm thick, the diffusion capacity C_p will dominate the capacitative term in the admittance.

Although A_p simulates a conductance and capacitance in parallel at low frequencies, its high-frequency behavior is quite different. In Fig. 7 the

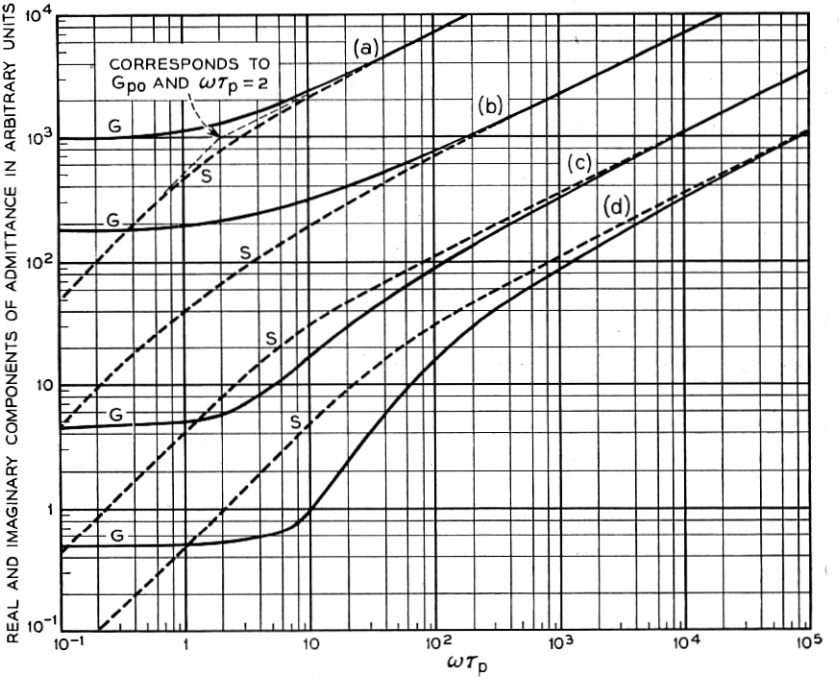

Fig. 7—Real, G, and imaginary, S, components of admittance for hole flow into n-region.
(a) $10^3 A_p/G_{po} = 10^3(1 + i\omega\tau_p)^{1/2}$ corresponding to uniform n-region.
(b) $10^2 \times$ Formula of Appendix III, corresponding to layer of high recombination rate in front of n-region. This causes G to exceed S at higher frequencies than for (*a*).
(c) $10 \times$ Equation (4.33), corresponding to a retarding field in the n-region, with $L_r = L_p/\sqrt{10}$.
(d) Equation (4.33) with $L_r = L_p/10$.

behavior of $(1 + i\omega\tau_p)^{1/2} = A_p/G_{p0}$, is shown. For high frequencies G_p and S_p are equal:

$$G_p = S_p = \sqrt{\tau_p/2}\, G_{p0} \sqrt{\omega} = \frac{b\sigma_i^2 \sqrt{\omega}}{(1 + b)^2 \sigma_n \sqrt{2D}} \qquad (4.30)$$

Thus for high frequencies the admittance is independent of τ_p and is determined by the diffusion of holes in and out of the n-region. The three straight asymptotes have a common intersection at the point G_{p0}, $\omega\tau = 2$ on Fig. 7, a fact which is useful in estimating the value of τ from such data.

For large ω, S_p varies as $\omega^{1/2}$ as shown in (4.30) whereas S_T is ωC_T. Hence

at very high frequencies C_T will dominate the admittance. At very high frequencies C_T itself will have a frequency dependence; however, for the assumptions on which the treatment of this section is based, the relaxation time for the transition region τ_T is much less than τ_p. This is a consequence of the fact that, although diffusion of holes into the transition region is required for the charging of C_T, the distance is relatively short, being in fact only that fraction of the width $x_{Tn} - x_{Tp}$ of the transition region in which ψ rises by kT/q; in germanium this will be about one-tenth of $x_{Tn} - x_{Tp}$. Since diffusion times vary as (distance)^2, the ratio of the times is

$$\frac{\tau_T}{\tau_p} = \frac{(x_{Tn} - x_{pn})^2}{100L_p^2}. \tag{4.31}$$

Hence if $L_p > x_{Tn} - x_{pn}$, τ_T will be much less than τ_p.[13]

4.5 Admittance Due to Hole Flow in a Retarding Field

In Appendix II we treat the case in which a potential gradient, due to changing concentration for example, is present in the n- and p-regions. This tends to prevent holes from diffusing deep in the n-region and for this reason the n-region acts partly like a storage tank for holes under a-c. conditions, thus enhancing S_p compared to G_p in A_p. If the electric field is $-d\psi/dx = kT/qL_r$, where L_r is the distance required for an increase of kT/q of potential (i.e. a factor of e increase in n_n), then the value of A_p is

$$A_p = [q\mu p_n/L_p] \frac{(2L_r/L_p)(1 + i\omega\tau_p)}{1 + [1 + (1 + i\omega\tau_p)(2L_r/L_p)^2]^{1/2}} \tag{4.32}$$

For $\omega\tau_p > 1$, this admittance is largely reactive provided $2L_r/L_p$ is sufficiently small.

The dependence of A_p upon ω is shown in Fig. 7 for two values of L_r/L_p. The plot shows the real and imaginary parts of

$$A_p/[2q\mu p_n L_r/L_p^2] = \frac{(1 + i\omega\tau_p)}{1 + [1 + (1 + i\omega\tau_p)(2L_r/L_p)^2]^{1/2}} \tag{4.33}$$

for $L_p/L_r = 10^{1/2}$ and $L_p/L_r = 10$, the two curves being relatively displaced vertically by one decade. The second value implies that the field keeps the holes back so that they penetrate only $\frac{1}{10}$ their possible diffusion length in no field. It is seen that for this case the storage effect is very pronounced and the susceptance S is much larger than G for high frequencies.

The function $(1 + i\omega\tau_p)^{1/2}$, discussed earlier, corresponds to the limiting case of (4.32) for $L_r = \infty$.

[13] In Appendix IV an analytic treatment of C_T is given.

4.6 The Effect of a Region of High Rate of Generation

There is evidence that imperfections, such as surfaces and cracks, add materially to the rate of generation and recombination of holes and electrons. If there is a localized region of high recombination rate in the transition region, there will be a pronounced modification of the admittance characteristics. In Fig. 8(a) such a layer is represented at $x = 0$. In Fig. 8(b) the customary plot of φ_p and φ_n versus x is shown. If we neglect the effect of the series resistance terms denoted by R_1 in Section 3, the change $\delta\varphi$ will

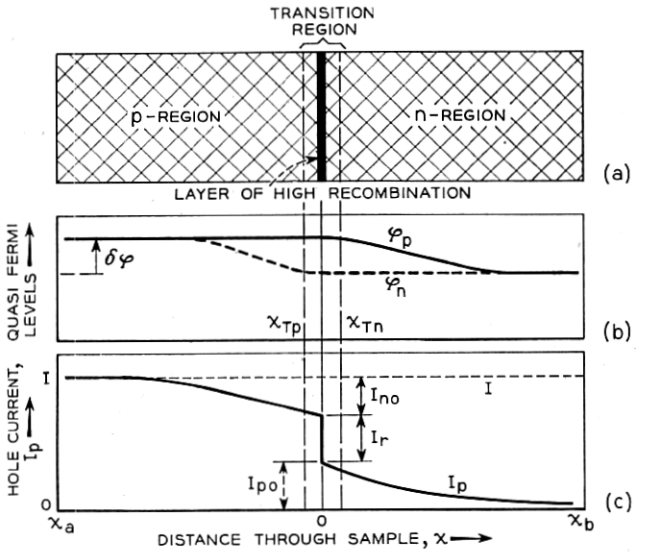

Fig. 8—The effect of a localized layer of high recombination rate on the junction characteristic.
(a) Location of layer of high recombination rate.
(b) Quasi Fermi levels.
(c) Distribution of hole current showing rapid change at layer of high recombination rate.

occur in the p-region for φ_n and in the n-region for φ_p. The hole current flowing into the n-region will thus be the same as before and will be given by equation (4.11) or (4.18) and denoted by $I_{p0}(\delta\varphi)$. Similarly, the electron current will be $I_{n0}(\delta\varphi)$. In the layer we shall suppose that there is a rate of generation of hole electron pairs equal to g_a per unit area of the layer and a rate of recombination proportional to $r_a np$ per unit area. We suppose, furthermore, that the layer is so thin that n and p are uniform throughout the layer. The net rate of generation is thus

$$g_a - r_a np = g_a[1 - e^{q(\varphi_p - \varphi_n)/kT}] \tag{4.34}$$

since for equilibrium conditions the rates balance so that $r_a n_i^2 = g_a$. The net hole current recombining in the layer per unit area is thus

$$I_r(\varphi_p - \varphi_n) = qg_a\,[e^{q(\varphi_p-\varphi_n)/kT} - 1] \tag{4.35}$$

There must, therefore, be a discontinuous decrease of hole current across the layer. The total hole current flowing in at $x = x_a$, which is also the total current I, thus does three things: for $x < x_{Tp}$, it combines with $I_{n0}(\delta\varphi)$; for $x_{Tp} < x < xT_n$, it combines with electrons at rate $I_r(\delta\varphi)$; for $x > x_{Tn}$, it flows into the n-region in amount $I_{p0}(\delta\varphi)$. This leads to

$$I = I_{n0}(\delta\varphi) + I_{p0}(\delta\varphi) + I_r(\delta\varphi). \tag{4.36}$$

In other words the layer of high recombination acts like a rectifier in parallel with $I_{n0}(\delta\varphi) + I_{p0}(\delta\varphi)$. The frequency characteristic of $I_r(\delta\varphi)$, however, will be independent of frequency and will contribute a pure conductance to the admittance of the junction.

If the layer is considered to have finite width, however, it will exhibit frequency effects just as does I_p in the n-region. In Appendix III, we treat a case in which the layer is a part of the n-region itself but has a recombination time different from the main layer. If the time is shorter, a large amount of the hole current may recombine in this layer. For high frequencies, the current may not penetrate the layer, in which case the admittance for hole current is determined by the thin layer rather than by the whole n-type region. A case of this sort is shown in Fig. 7. In this case the thickness of the layer is $\frac{1}{3}$ of its diffusion length and in it the lifetime of a hole τ_ℓ is $\frac{1}{9}$ the value τ_p in the main body of the n-region. The hole current will thus be restricted to this layer when the diffusion distance $\sqrt{D/\omega}$ is less than the layer thickness $(\frac{1}{3})\sqrt{D\tau_\ell}$; this corresponds to $\omega\tau_\ell > 9$ or $\omega\tau_p > 81$. The presence of the high rate of combination in the layer is evidenced by the tendency of G to be greater than S at high frequencies. If the layer were infinitely thin, as discussed above, it would simply add a constant conductance to the admittance.

4.7 Patch Effect in p-n Junctions

If there are localized regions of high recombination rate, a "patch effect" may be produced in an n-p junction. As an extreme example, suppose the value of g_a for the layer just considered is allowed to become very large; then the recombination resistance may become small compared to R_1 in Section 3 and the junction will become substantially ohmic. If the region of high rate of recombination is relatively small compared to the area of the rest of the junction, then the behavior of the junction as a whole may be regarded as being that due to two junctions in parallel. Over most of the area,

the currents will flow as if the patch were not present so that one component of the current will be that due to the uniform junction. In addition there will be current due to recombination and generation in the patch. The series resistance to the patch will be relatively high due to the constriction of the current paths. On the other hand, the value of $I_r(\delta\phi)$ associated with the patch may be very high. Hence the current due to the patch will be that of a low impedance ideal rectifier in series with a high resistance; and if the ratio of impedances is high enough, such a series combination amounts essentially to an ohmic leakage path. Thus patches in the *p-n* junction will tend to introduce leakage paths and destroy saturation in the reverse direction.

An extreme example of a region of high rate of recombination would be a particle of metal making a non-rectifying contact to both *p*- and *n*-type germanium. Since holes and electrons are essentially instantly combined in a metal, the boundary condition at the metal surface would be equality of φ_p and φ_n. This would mean that near the metal particle, φ_p and σ_n could not differ by $\delta\varphi$, the condition required, over some parts of the junction at least, in order for ideal rectification to occur.

A common source of imperfection in *p-n* junctions arises from dirt or fragments on the surface which overlap the junction. Even if these do not actually constitute a short circuit across the junction, they may furnish patches of the sort discussed here and modify the junction characteristic.

4.8 Final Comments

Another possible cause for frequency effects may be found in the trapping of holes or electrons.[14] When an added hole concentration is introduced into an *n*-region, a certain fraction of the holes will be captured by acceptors and later re-emitted or else recombined with electrons while trapped. Investigation of this process is given in Appendix VI. One interesting result is that the trapping of holes in a uniform *n*-region cannot produce an effective susceptance (i.e. $i\omega C$) in excess of the conductance, as can a retarding field.

Finally it should be remarked that important and significant variations of the conductivity in the *p*- and *n*-regions may be produced by hole or electron injection. Under these conditions, when the hole concentration approaches n_n, $\psi - \varphi_n$ will vary. Under these conditions R_1 may be appreciably altered. These factors favor the *p-n* junction as a rectifier since they lead to a reduction of series resistance under conditions of forward bias and thus tend to improve the rectification ratio.

[14] Frequency dependent effects in Cu_2O rectifiers have been explained in this way by J. Bardeen and W. H. Brattain, personal communication.

5. Internal Contact Potentials

The theory of p-n junctions presented above has interesting consequences when applied to the distribution of potential between two semiconductors

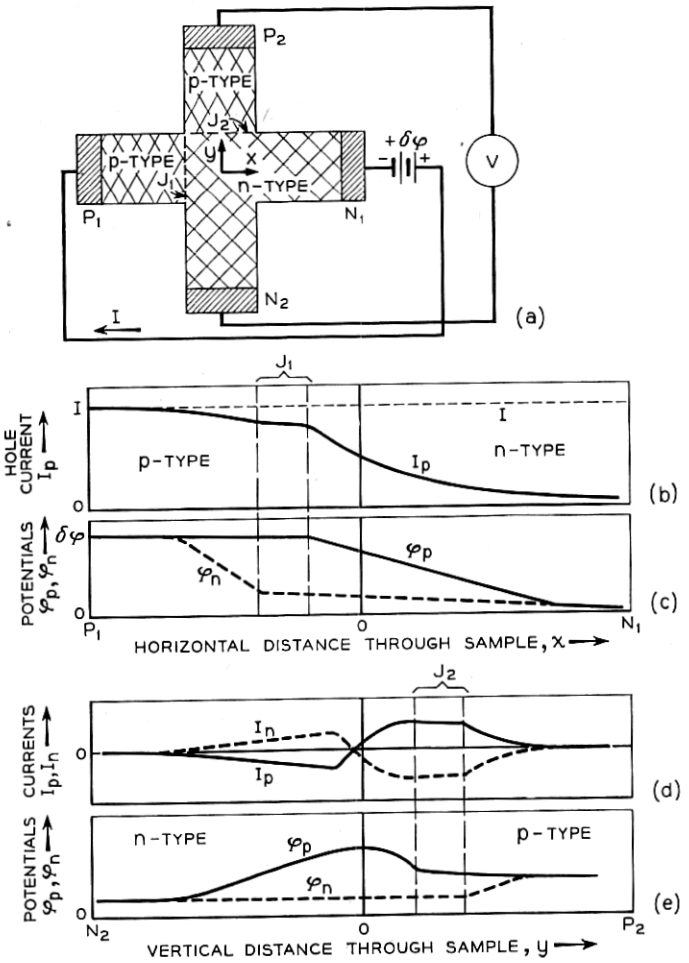

Fig. 9—Internal contact potentials showing how presence of injected holes produces a contact potential across J_2.

under conditions of hole or electron injection. In Fig. 9 we illustrate an X-shaped structure. A forward current flows across the junction P_1 and out of branch N_1. If the distance across the intersection is comparable with or small compared to the diffusion length for holes, a potential difference should be measured between P_2 and N_2. The reason for this is that holes

flow easily into P_2 since the potential distribution there favors their entrance. Since, however, P_2 is open-circuited this hole flow biases J_2 in the forward direction; since J_2 is high resistance, an appreciable bias is developed before the counter current equals the inward hole flow and a steady state is reached. No similar effect occurs in the branch N_2 ; consequently P_2 will be found to be floating (open-circuited) at a more positive potential than N_2 .

Parts (b) to (e) describe this reasoning in more complete terms. We suppose that the p-regions are more highly conducting than the n-regions so that the current across J_1 , shown in (b), is mainly holes. The potentials φ_p and φ_n along the x-axis will be similar to those of Figs. 5 and 6; (c) shows this situation and indicates that the diffusion length for electrons in the p-region is less than for holes in the n-region. Along the y axis φ_p and φ_n vary as shown in (e), the reasoning being as follows: At the origin of coordinates φ_p and φ_n have the same values as for (c). The transverse hole current (d) has a small positive component at $y = 0$ since, as mentioned above, P_2 tends to absorb holes and thus increase diffusion along the plus y-axis.Since the net transverse current is zero, $I_n = -I_p$ in (d). The φ curves of (e) have been drawn to conform to the currents in (d); φ_n is nearly constant in the n-region and φ_p is nearly constant in the p-region. As concluded in connection with Figs. 5 and 6, φ_n and φ_p are also nearly constant across the transition region. These conclusions lead to the shape of φ_n and φ_p for $y > 0$ in (e). For $y < 0$, the reasoning is the same as that used in Sections 3 and 4 and we conclude that φ_n is essentially constant. Hence, a difference in the Fermi levels at P_2 and N_2 will result.

In Fig. 10 we show a structure for which we can make quantitative calculations of the variations of φ_p and φ_n . We assume for this case that the forward current from P_1 to N does not produce an appreciable voltage drop, i.e. change in ψ and φ_n , in region N. This will be a good approximation if the dimensions are suitably proportioned. We shall next solve for the steady-state distribution of p subject to the indicated boundary conditions assuming that p is a function of x only. As we have discussed in Section 4.1, when p is small compared to n in the n-region, we can write

$$p = p_n e^{q(\varphi_p - \varphi_n)/kT} \tag{5.1}$$

In keeping with the treatment in the next section of this structure as a transistor, the terminals are designated emitter, collector and base, the potentials with respect to the base being φ_e and φ_c . The contact to N or the base is such that $\varphi_b = \varphi_n$ in this region. Hence, the boundary conditions at J_1 and J_2 are

$$p_1 = p_n e^{q\varphi_e/kT} \qquad x = -w \tag{5.2}$$

$$p_2 = p_n e^{q\varphi_c/kT} \qquad x = +w \tag{5.3}$$

The function $p(x)$ which satisfies these boundary conditions and the equation

$$D\frac{d^2p}{dx^2} - \frac{p - p_n}{\tau_p} = 0 \tag{5.4}$$

is

$$p(x) = p_n + \frac{p_1 + p_2 - 2p_n}{2\cosh(w/L_p)}\cosh(x/L_p) + \frac{p_2 - p_1}{2\sinh(w/L_p)}\sinh(x/L_p) \tag{5.5}$$

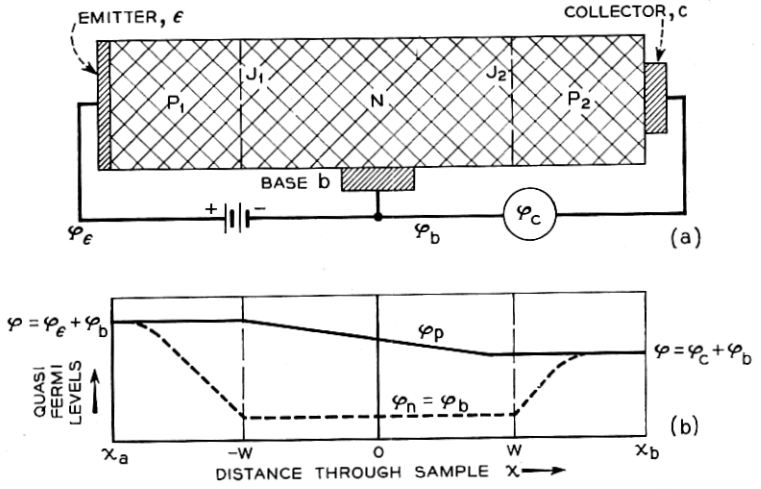

Fig. 10—Model used for calculation of internal contact potential and to illustrate *p-n-p* transistor.
(a) Semiconductor with two *p-n* junctions and ohmic metal contacts.
(b) Quasi Fermi levels showing internal contact potential between *b* and *c*.

which gives rise to a hole current across J_2 into P_2 of amount

$$\begin{aligned} I_p &= -qD\frac{dp}{dx}\bigg|_{x=w} \\ &= \frac{qD}{2L_p}\left[(p_1 - p_2)\coth\frac{w}{L_p} + (2p_n - p_1 - p_2)\tanh\frac{w}{L_p}\right] \\ &= \frac{qD}{2L_p}\left[(p_1 - p_n)\left(\coth\frac{w}{L_p} - \tanh\frac{w}{L_p}\right)\right. \\ &\qquad \left. - (p_2 - p_n)\left(\coth\frac{w}{L_p} + \tanh\frac{w}{L_p}\right)\right] \\ &= +\frac{p_n qD}{L_p}\left[\frac{[e^{q\varphi/kT} - 1]}{\sinh(2w/L_p)} - \frac{(e^{q\varphi_c/kT} - 1)}{\tanh(2w/L_p)}\right] \\ &= \operatorname{csch}(2w/L_p)I_{p0}(\varphi_\epsilon) - \coth(2wL_p)I_{p0}(\varphi_c) \end{aligned} \tag{5.6}$$

where, by $I_{p0}(\varphi)$, we mean the hole current which would flow in the forward direction across either J_1 or J_2 if uninfluenced by the other (i.e. the function of (4.11) or (4.18) and (4.20).) The equation shows that a fraction csch $(2w/L_p)$ of the current $I_{p0}(\varphi_\epsilon)$, which would be injected by φ_ϵ on P_1 in the absence of J_2, flows into P_2. The conductance of P_2 across J_2 is increased by the factor $\coth(2w/L_p)$.

The current *into* P_2 carried by electrons will be unaffected by J_1 and can be denoted by $-I_{n0}(\varphi_c)$ the minus sign resulting from the fact that currents *into* P_2 are in the reverse direction. The total current flowing *into* P_2 contains the $-I_{n0}(\varphi_c)$ and $-I_{p0}(\varphi_c)$ terms and must cancel the $+I_{p0}(\varphi_\epsilon)$ term for equilibrium. Hence:

$$I_{n0}(\varphi_c) + \coth\,(2w/L_p)\, I_{p0}(\varphi_c) = \operatorname{csch}\,(2w/L_p)\, I_{p0}(\varphi_\epsilon) \tag{5.7}$$

If $p_n \gg n_p$, the I_{n0} term can be neglected compared to $\coth\,(2w/L_p)\, I_{p0}$. Hence the value of φ_c must satisfy

$$I_{p0}(\varphi_c) = \operatorname{sech}\,(2w/L_p)\, I_{p0}(\varphi_\epsilon). \tag{5.8}$$

For $\varphi_\epsilon > kT/q$, the exponential approximation may be used for I_{p0} in both terms:

$$\varphi_c = \varphi_\epsilon - (kT/q) \ln \cosh\,(2w/L_p), \tag{5.9}$$

so that, if $(2w/L_p)$ is the order of unity, φ_c should be only about (kT/q) less than φ_ϵ. For $(2w/L_p)$ large, we get

$$\varphi_c = \varphi_\epsilon - (kT/q)\,(2w/L_p) \tag{5.10}$$

corresponding to the linear drop of φ_p, discussed in connection with equation (4.9), across the distance $2w$.

When φ_ϵ is negative, so that we have to deal with reverse current, φ_c will not decrease indefinitely but will reach a minimum value given by

$$[\exp q\varphi_c/kT] - 1 = -\operatorname{sech}\,(2w/L_p) \tag{5.11}$$

and corresponding to saturation reverse current across J_1, so that

$$\varphi_c = -(kT/q) \ln [1 + (1/2) \operatorname{csch}^2(w/L_p)]. \tag{5.12}$$

The floating potentials of p-type contacts to n-type material into which holes have been injected (or n-type contacts to p-type material with injected electrons) are reminiscent of probes in gas discharges which tend to become charged negative in respect to the space around them because they catch electrons more easily than positive ions. The situation may also be compared with that producing thermal e.m.f.'s; in fact a "concentration temperature" of the semiconductor with injected holes can be defined by finding the temperature for which $np = n_i^2(T)$. We conclude that, in the

absence of thermal equilibrium, different potentials depending on the nature of the contact are, in general, the rule rather than the exception.

The bias developed on P_2 or c will change its conductance. If we suppose that φ_e and φ_b are held constant, then the current flowing into c is obtained by the same reasoning that led to (5.7) and is

$$I_c(\varphi_c, \varphi_e) = I_{n0}(\varphi_c) + \coth \frac{2w}{L_p} I_{p0}(\varphi_c) - \operatorname{csch} \frac{2w}{L_p} I_{p0}(\varphi_e). \quad (5.13)$$

For an infinitesimal change in φ_c from the value which makes $I_c(\varphi_c, \varphi_e)$ vanish, the admittance to c is readily found from (4.18) and (4.19) to be

$$\begin{aligned} \left(\frac{\partial I_c}{\partial \varphi_c}\right)_{\varphi_e} &= I'_{n0}(\varphi_c) + \coth \frac{2w}{L_p} I'_{p0}(\varphi_c) \\ &= \left[G_{n0} + \coth \frac{2w}{L_p} G_{p0}\right] e^{q\varphi_c/kT} \end{aligned} \quad (5.14)$$

which shows that pronounced variations in admittance should be associated with variations in hole density in N in Fig. 10.[15]

6. *p-n-p* Transistors

The structure shown in Fig. 10 is a transistor with power gain provided the distance w is not too great. As a first approximation, we shall neglect the drop due to currents in the N region. If we use P_2 as the collector and call the collector current, I_c, positive when it flows into P_2 from outside, we shall have from (5.13)

$$I_c = -\operatorname{csch} \frac{2w}{L_p} I_{p0}(\varphi_e) + \coth \frac{2w}{L_p} I_{p0}(\varphi_c) + I_{n0}(\varphi_c). \quad (6.1)$$

The emitter current is similarly

$$I_e = \coth \frac{2w}{L_p} I_{p0}(\varphi_e) - \operatorname{csch} \frac{2w}{L_p} I_{p0}(\varphi_c) + I_{n0}(\varphi_e). \quad (6.2)$$

If $p_n \gg n_p$, then the I_{n0} terms can be neglected. However, the base current will not vanish but will be

$$\begin{aligned} I_b = -I_e - I_c &= \left[\operatorname{csch} \frac{2w}{L_p} - \coth \frac{2w}{L_p}\right] [I_{p0}(\varphi_e) + I_{p0}(\varphi_c)] \\ &= \frac{2 \sinh^2 w/L_p}{\sinh 2w/L_p} [I_{p0}(\varphi_e) + I_{p0}(\varphi_c)]. \end{aligned} \quad (6.3)$$

[15] The variations in admittance discussed in connection with metal point contacts in an accompanying paper in this issue (W. Shockley, G. L. Pearson and J. R. Haynes, *Bell Sys. Tech. Jl.*, July, 1949), arise from this cause; however, the nature of the contact is not as simple as here.

For w/L_p large, the junctions do not interact and the hyperbolic coefficient becomes unity and $I_b = -[I_{p0}(\varphi_\epsilon) + I_{p0}(\varphi_c)]$.

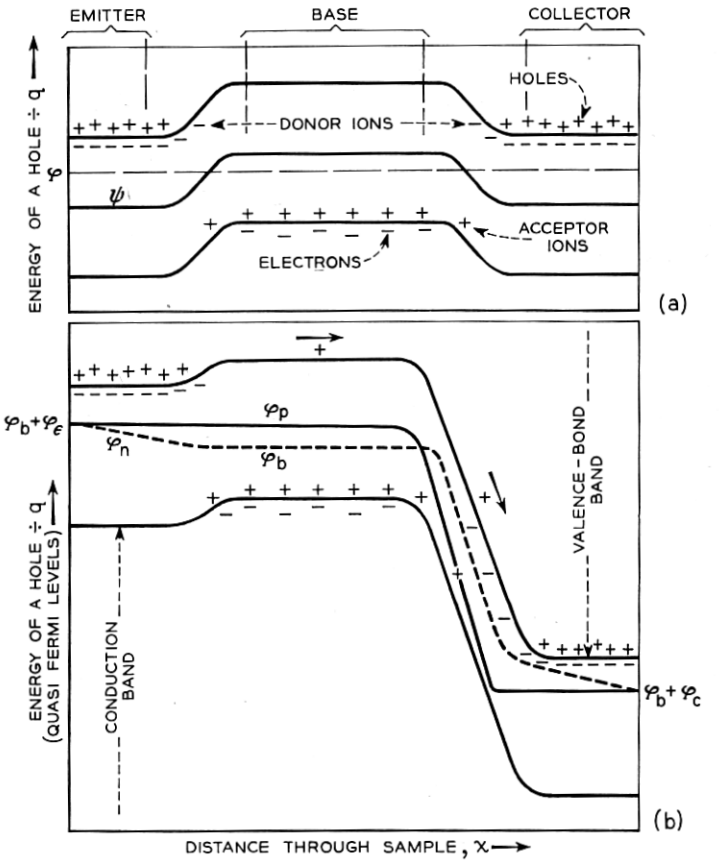

Fig. 11—*p-n-p* transistor.
(a) Thermal equilibrium.
(b) Operating condition.

If φ_c is several volts negative, so that $I_{p0}(\varphi_c)$ has its saturation value I_{ps} (see (4.11) and (4.20)), then the ratio $-\delta I_c/\delta I_\epsilon \equiv \alpha$ has the value

$$\alpha = -\frac{\delta I_c}{\delta I_\epsilon} = \frac{\operatorname{csch} \dfrac{2w}{L_p}}{\coth \dfrac{2w}{L_p}} = \operatorname{sech} \frac{2w}{L_p}. \tag{6.4}$$

For $(2w/L_p) = 0.5, 1, 2$ respectively, $\alpha = 0.89, 0.65, 0.27$. Since the output impedance R_{22} will be very high when φ_c is in the reverse direction, and the

input impedance will be low, the power gain formula[16] $\alpha^2 R_{22}/R_{11}$ will yield power gain even when α is less than unity.

In certain ways the structure of Fig. 10 resembles a vacuum tube. In Fig. 11, we show the energy band diagram, with energies of holes plotted upwards so as to be in accord with the convention for voltages. (a) shows the thermal equilibrium distribution and (b) the distribution under operating conditions. It is seen that the potential hill, which holes must climb in reaching the collector, has been reduced by φ_ϵ. The n-region represents in a sense the grid region in a vacuum tube, in which the potential and hence plate current, is varied by the charge on the grid wires. Here the potential in the n-region is varied by the voltage applied between base and emitter. In both cases one current is controlled by another. In the vacuum tube the current which charges the grid wires controls the space current. Because the grid is negative to the cathode, the electrons involved in the space current are kept away from the grid while at the same time the electrons in the grid are kept out of the space by the work function of the grid (provided that the grid does not become overheated.) In Fig. 11, the electrons flowing into the base control the hole current from emitter to collector. In this case the controlled and controlling currents flow in the same space but in different directions because of the opposite signs of their charges.

As this discussion suggests, it may be advantageous to operate the p-n-p transistor like a grounded cathode vacuum tube, with the emitter grounded and the input applied to the base.

The p-n-p transistor has the interesting feature of being calculable to a high degree. One can consider such questions as the relative ratios of width to length of the n-region and the effect of altering impurity contents and scaling the structure to operate in different frequency ranges. However, we shall not pursue these questions of possible applications further here.

Acknowledgment

The writer is indebted to a number of his colleagues for stimulating discussions and encouragement, in particular to H. R. Moore, G. L. Pearson and M. Sparks, whose experimental work, to be described in later publications, suggested development of the theory along the lines presented above. He is also indebted to J. Bardeen, P. Debye, G. Wannier and W. van Roosbroeck for theoretical comments and suggestions and especially to the last and to Mrs. G. V. Smith for valuable assistance in preparing the manuscript.

[16] Physical Principles Involved in Transistor Action, J. Bardeen and W. H. Brattain, *Phys. Rev.* 75, 1208 (1949).

APPENDIX I

A THEOREM ON JUNCTION RESISTANCE

We shall here prove that the junction resistance is never less than the value obtained by integrating the local resistivity $1/q\mu(p + bn)$. This is accomplished by analyzing the following equation which we shall discuss before giving the derivation:

$$I\delta\varphi = \frac{1}{q\mu}\int_{x_a}^{x_b}\left(\frac{I_p^2}{p} + \frac{I_n^2}{bn}\right)dx + qg\int_{x_a}^{x_b}(\varphi_p - \varphi_n)(e^{q(\varphi_p-\varphi_n)/kT} - 1)\,dx,$$

the meaning of the symbols being that shown in Fig. 5. This expression is valid even if large disturbances in p and n from their equilibrium values occur. The second integral is positive since the integrand is never negative. It may be very large if $\varphi_p - \varphi_n \gg kT/q$ in some regions. If, in the first integral, we consider that I_p and I_n may be varied subject to the restraint $I_p + I_n = I$, we may readily prove that the first integrand takes on a minimum value when

$$I_p = \frac{pI}{p + bn} \quad \text{and} \quad I_n = \frac{bnI}{p + bn}.$$

For this minimum condition, the first integral becomes

$$I^2\int_{x_a}^{x_b} dx/q\mu(p + bn) = I^2R_0$$

where R_0 is simply the integrated local resistivity. If I does divide in this way, the second integral is zero, a result which we can see as follows:

$$I_p = -q\mu p\, d\varphi_p/dx$$

$$I_n = -q\mu bn\, d\varphi_n/dx$$

$$\frac{d\varphi_p/dx}{d\varphi_n/dx} = \frac{I_p/p}{I_n/bn}$$

Hence, if the current divides in the ratio of p to bn, then $d\varphi_p = d\varphi_n$ and, since $\varphi_p = \varphi_n$ at x_a, $\varphi_p = \varphi_n$ everywhere and the second integral vanishes.

In general, of course, the conditions governing recombination prevent current division in the ratio $p{:}bn$ and then $\delta\varphi/I > R_0$.

The equation discussed above is derived as follows: We suppose that

$$\varphi_p(x_a) = \varphi_n(x_a) = \varphi_a$$

$$\varphi_p(x_b) = \varphi_n(x_b) = \varphi_b$$

Then

$$I\delta\varphi = -(I_p\varphi_p + I_n\varphi_n)\Big|_{x_a}^{x_b}$$

$$= -\int_{x_a}^{x_b} \frac{d}{dx}(I_p\varphi_p + I_n\varphi_n)\,dx$$

$$= -\int_{x_a}^{x_b}\left(\frac{dI_p}{dx}\varphi_p + \frac{dI_n}{dx}\varphi_n\right)dx - \int_{x_a}^{x_b}\left(I_p\frac{d\varphi_p}{dx} + I_n\frac{d\varphi_n}{dx}\right)dx.$$

Since

$$\frac{dI_p}{dx} = -\frac{dI_n}{dx} = qg(1 - e^{q(\varphi_p-\varphi_n)/kT})$$

and

$$\frac{d\varphi_p}{dx} = -I_p/q\mu p, \qquad \frac{d\varphi_n}{dx} = -I_n/q\mu bn$$

these two integrals are readily transformed into the ones previously discussed.

APPENDIX II

Admittance in a Retarding Field

We shall here derive the admittance equation for holes diffusing into a retarding potential $\psi = kTx/qL_r$ in which the potential increases by kT in each distance L_r. The differential equation for the a-c. component of p is

$$i\omega p = -\frac{p}{\tau_p} - \frac{\partial}{\partial x}\left[-D\frac{\partial p}{\partial x} - \mu p\frac{\partial \psi}{\partial x}\right].$$

This equation may be solved by letting $p = p_1 \exp(i\omega t - \gamma x)$ as may be seen by rewriting the equation and substituting this expression for p:

$$D\left[\frac{\partial^2 p}{\partial x^2} + \frac{1}{L_r}\frac{\partial p}{\partial x}\right] - \frac{1}{\tau_p}(1 + i\omega\tau_p)p$$

$$= -\gamma D\left[-\gamma + \frac{1}{L_r}\right]p - \frac{1}{\tau_p}(1 + i\omega\tau_p)p = 0$$

leading to

$$\gamma = \frac{1 + [1 + (2L_r/L_p)^2(1 + i\omega\tau_p)]^{1/2}}{2L_r}.$$

The corresponding current evaluated at $x = 0$ where $p = p_1 \exp(i\omega t) = (p_n q v_1/kT)\exp(i\omega t)$ is given by

$$\begin{aligned} I &= -q\left[D\frac{\partial p}{\partial x} + \mu p \frac{\partial \psi}{\partial x}\right] \\ &= -qD\left[-\gamma + \frac{1}{L_r}\right] p \\ &= \frac{q(1 + i\omega\tau_p)p}{\gamma\tau_p} \\ &= \frac{p_n q^2(1 + i\omega t)}{kT\tau_p} \cdot \frac{2L_r}{1 + [1 + (2L_r/L_p)^2(1 + i\omega\tau_p)]^{1/2}} \cdot v_1 e^{i\omega t} \\ &= \frac{q\mu p_n 2L_r}{L_p^2} \cdot \frac{(1 + i\omega\tau_p)}{1 + [1 + (2L_r/L_p)^2(1 + i\omega\tau_p)]^{1/2}} \cdot v_1 e^{i\omega t} \\ &= A_p v_1 e^{i\omega t}. \end{aligned}$$

This is equivalent to (4.32) in Section 4.

APPENDIX III

Admittance for Two Layers

We shall here treat a case in which there is a thin layer on the n-side of the transition region in which recombination occurs much more readily than deeper in the n-layer. The case of an infinitely thin plane, discussed in Section 4, is a limiting case of this model. We shall suppose that the layer extends from $x = -c$ to $x = 0$ while $x > 0$ corresponds to the n-region. We shall suppose that the potential in the layer is uniform with value ψ_1 whereas in the n-region it has value ψ_2. The lifetimes of holes will be taken τ_1 and τ_2 in the two layers. The solutions for p_1 and p_2 are evidently

$$\begin{aligned} p_1 &= p_{10} + (A\, e^{-\alpha x} + B\, e^{+\alpha x})\, e^{i\omega t} \qquad x < 0 \\ p_2 &= p_{20} + C\, e^{-\beta x + i\omega t} \qquad x > 0 \end{aligned}$$

where

$$\begin{aligned} \alpha &= (1 + i\omega\tau_1)^{1/2}/\sqrt{D\tau_1} \equiv (1 + i\omega\tau_1)^{1/2}/L_1 \\ \beta &= (1 + i\omega\tau_2)^{1/2}/\sqrt{D\tau_2} \equiv (1 + i\omega\tau_2)^{1/2}/L_2. \end{aligned}$$

The boundary condition for continuity of φ_p, required to avoid singularity in $\partial\varphi_p/\partial x$, is

$$p_2 e^{q\psi_2/kT} = p_1 e^{q\psi_1/kT}$$

and, for continuity of hole current, is $\partial p_1/\partial x = \partial p_2/\partial x$. Expressing these in terms of A, B, C, α and β for the a-c. components yields:

$$A + B = Ce^{q(\psi_1-\psi_2)/kT} \equiv CF$$

$$\alpha(A - B) = \beta C$$

so that

$$A = (F + \beta/\alpha)C/2.$$

$$B = (F - \beta/\alpha)C/2.$$

Hence the ratio $-[\partial p/\partial x]/p$ at $x = -c$ is

$$-\frac{\partial \ln p}{\partial x} = \frac{\alpha(Ae^{+\alpha c} - Be^{-\alpha c})}{(Ae^{+\alpha c} + Be^{-\alpha c})} = \frac{\alpha(F\alpha \sinh \alpha c + \beta \cosh \alpha c)}{F\alpha \cosh \alpha c + \beta \sinh \alpha c}.$$

Since at $x = -c$, the a-c. component of p_1 is $(qv_1/kT)p_{10}e^{i\omega t}$, the admittance is

$$A_p = \frac{-qD\partial p/\partial x}{v_1 e^{i\omega t}} = (q^2 Dp_{10}/kT)(-\partial \ln p/\partial x)$$

$$= (q\mu p_{10}/L_1)(1 + i\omega\tau_1)^{1/2} \frac{F\alpha \sinh \alpha c + \beta \cosh \alpha c}{F\alpha \cosh \alpha c + \beta \sinh \alpha c}.$$

For $c \to 0$, this transforms into

$$(q\mu p_{10}/L_1)\ (1 + i\omega\tau_1)^{1/2}\beta/F\alpha = (q\mu(p_{10}/F)/L_2)(1 + i\omega\tau_2)^{1/2}$$

which agrees with Section 4, since p_{10}/F then corresponds to p_n.

If c/L_1 and F are not large, an appreciable amount of recombination takes place for $x > 0$ for low frequencies. Dispersive effects will then occur corresponding to τ_2. The a-c. will not penetrate to $x = 0$, however, if $c(\omega/D)^{1/2} \gg 1$ and the dispersive effects will then be determined by τ_1.

The frequency-dependent part of the admittance,

$$(1 + i\omega\tau_1) \frac{F\alpha \sinh \alpha c + \beta \cosh \alpha c}{F\alpha \cosh \alpha c + \beta \sinh \alpha c},$$

has been computed and is shown in Fig. 7 for $\tau_p = \tau_2$, $F = 1$, $\tau_1 = \tau_2/9$ and $c/L_1 = \frac{1}{3}$. For these values about half the hole current reaches $x = 0$ for low frequencies. As the time constant for diffusion through the layer is $\tau_p/81$, as discussed in Section 4.6, the layer will act as a largely frequency-independent admittance well above the point for $\omega\tau_p = 1$. This is reflected in the behavior of the curves of Fig. 7 and, for frequencies in the $\sqrt{\omega t}$ range, it is seen that G is larger than S by about 50% of the low-frequency value of G; this split of $G + iS$ into $(\frac{1}{2})G_0$ plus approximately $(\frac{1}{2})G_0\ (1 + i\omega\tau_p)^{1/2}$ corresponds to the fact that about half the holes are absorbed in layer 1 for the assumed conditions.

APPENDIX IV

TIME CONSTANT FOR THE CAPACITY OF THE TRANSITION REGION

For this case we shall consider the case of holes in an a-c. field with potential

$$\psi = \frac{kT}{q}\left(\frac{x}{L_r} + \frac{xe^{i\omega t}}{L_1}\right)$$

where the d-c. retarding field is kT/qL_r and the a-c. field is kT/qL_1 where $1/L_1$ is considered small for the linear theory presented here. The expression for the current of holes is

$$-D\frac{\partial p}{\partial x} - \mu p\frac{\partial \psi}{\partial x} = -D\left[\frac{\partial p}{\partial x} + p\left(\frac{1}{L_r} + \frac{e^{i\omega t}}{L_1}\right)\right]$$

We shall obtain a solution for p by letting

$$p = p_0 e^{-x/L_r} + p_1[e^{-x/L_r} - e^{-\gamma x}]e^{i\omega t},$$

while neglecting recombination in this region so that $\dot{p}$ must satisfy the condition $\dot{p} = -\partial$ (hole current)$/\partial x$ leading to the differential equation

$$D\left[\frac{\partial^2 p}{\partial x^2} + \frac{\partial p}{\partial x}\left(\frac{1}{L_r} + \frac{e^{i\omega t}}{L_1}\right)\right] - \dot{p} = 0$$

There are three separate exponential dependencies of the variables leading to three equations (neglecting terms of order $(1/L_1)^2$)

e^{-x/L_r}: $$D\left[p_0\frac{1}{L_r^2} - p_0\frac{1}{L_r^2}\right] = 0$$

$e^{-x/L_r+i\omega t}$: $$D\left[p_1\frac{1}{L_r^2} - p_1\frac{1}{L_r^2} - \frac{1}{L_r L_1}p_0\right] - i\omega p_1 = 0$$

$e^{-\gamma x+i\omega t}$: $$D[\gamma^2 - \gamma/L_r]p_1 - i\omega p_1 = 0$$

The first equation is satisfied by the equilibrium distribution and the second by

$$p_1 = -p_0 D/i\omega L_1 L_r$$

and the last by

$$\gamma = \frac{1 + \sqrt{1 + 4i\omega L_r^2/D}}{2L_r}$$

It is evident that dispersive effects set in when

$$\omega = D/4L_r^2$$

This corresponds to the result used in (4.31) in which $(x_{Tn} - x_{Tp})/10$ was used for L_r. For smaller values of ω the current may be calculated and put in simple form by expanding γ up to terms including ω^2. The resulting expression for the current is

$$I = -i\omega q\, p_0\, L_r(L_r/L_1)e^{i\omega t}$$

This is interpreted as follows: The a-c. voltage across a layer L_r thick is

$$\delta\psi = (kT/q)\ (L_r/L_1)e^{i\omega t}$$

and, if we consider plus voltage as producing a field from left to right, then the a-c. voltage across L_r is $V = -\delta\psi$. Substituting this for $(L_r/L_1)\exp(i\omega t)$ gives

$$I = i\omega q p_0\, L_r(q/kT)V$$

Here qp_0L_r is the total charge in the layer L_r, (qV/kT) is an average fractional change in this charge for V so that $(qp_0L_r)\ (qV/kT) \div V$ is a capacity.

APPENDIX V

The Effect of Surface Recombination

In this appendix we shall consider the effect of surface recombination upon the characteristics of the p-n junction. As for Section 4 we shall illustrate the theory for the case of holes diffusing into n-type material. For simplicity we shall treat a square cross-section bounded by $y = \pm w$, $z = \pm w$, the current flow being along $+x$.

We shall denote the a-c. component of p as

$$p_1 \equiv p_1\,(x, y, z, t)$$

At $x = 0$, the edge of the n-region, we shall suppose that φ_p and ψ are independent of y and z so that we shall have

$$p_1(0, y, z, t) = p_{10}\, e^{i\omega t} = (p_n\, qv_1/kT)e^{i\omega t}$$

by reasoning similar to that used for equation (4.5). The boundary condition at the surface will be

$$-D\frac{\partial p_1}{\partial y} = sp_1 \qquad \text{for } y = +w$$

This states that the recombination per unit area is sp_1 and is equal to the diffusion to the surface $-D\partial p_1/\partial y$. Similar boundary conditions hold for the other surfaces. By standard procedures involving separation of variables we may verify that the solution satisfying the boundary conditions is

$$p_1 = \sum_{i,j=0}^{\infty} a_{ij}\, e^{-\alpha_{ij}x + i\omega t}\ \cos\beta_i y\ \cos\beta_j z$$

where the eigenvalues β_i are determined by the boundary condition

$$\beta_i w \tan \beta_i w = sb/D \equiv \chi.$$

We use $\theta_i = \beta_i w$ for brevity later. Because of the symmetry of the boundary conditions it is not necessary to include sine functions in the sum. The value of α_{ij} is given by

$$\alpha_{ij} = (1 + i\omega\tau_{ij})^{\frac{1}{2}}/(D\tau_{ij})^{\frac{1}{2}}$$

where τ_{ij} is the lifetime of a hole in the eigenfunction $\cos \beta_i y \cos \beta_j z$; i.e. τ_{ij} is the lifetime which makes

$$p = \exp(-t/\tau_{ij}) \cos \beta_i y \cos \beta_j z,$$

a function which satisfies the surface boundary conditions, a solution of the equation

$$\partial p/\partial t = D\nabla^2 p - p/\tau = -D(\beta_i^2 + \beta_j^2)p - p/\tau$$

where to simplify the subsequent expressions we have omitted the subscript p from τ. This equation leads to

$$\frac{1}{\tau_{ij}} = D(\beta_i^2 + \beta_j^2) + \frac{1}{\tau}.$$

The coefficients a_{ij} are readily found since the $\cos \beta_i y$ functions form an orthogonal set (as may be verified by integrating by parts and using the boundary conditions). The values are

$$a_{ij}/p_{10} = 4[\sin \theta_i \sin \theta_j]/\theta_i\theta_j[1 + (1/2\theta_i) \sin 2\theta_i]\cdot[1 + (1/2\theta_j) \sin 2\theta_j]$$

The current corresponding to this solution is

$$I_1 = -qD \iint (\partial p/\partial x)\, dy\, dz$$

integrated over the cross section at $x = 0$. This gives

$$I_1 = qDp_{10}e^{i\omega t} \Sigma\, \alpha_{ij}(a_{ij}/p_{10})(4w^2/\theta_i\theta_j) \sin \theta_i \sin \theta_j$$

Substituting for a_{ij} and inserting $p_{10} = p_n q v_1/kT$, we obtain an expression for the admittance $A_p = I_1/V_1 \exp(i\omega t)$:

$$A_p = 4w^2 q\mu p_n \Sigma_{ij}\, \alpha_{ij} \frac{4 \sin^2 \theta_i \sin^2 \theta_j}{\theta_i^2 \theta_j^2 \left[1 + \left(\frac{1}{2\theta_i}\right) \sin 2\theta_i\right]\left[1 + \left(\frac{1}{2\theta_j}\right) \sin 2\theta_j\right]}$$

where the sum plays the role formerly taken by $(1 + i\omega t)^{1/2}/\sqrt{D\tau}$ in equation (4.12); the factor $4w^2$ is the area of the junction.

We shall analyze the formula for the case in which recombination on the

surface is smaller than diffusion to the surface so that χ is not large. The values of θ_i, over which the sum is to be taken, may be estimated as follows: in each interval of θ_i of the form $n\pi$ to $(n + (\frac{1}{2}))\pi$, $\theta_i \tan \theta_i$ varies from 0 to ∞, giving one solution to $\theta_i \tan \theta_i = \chi$. For χ small, the solutions are approximately

$$\theta_0 \doteq \sin \theta_0 \doteq \tan \theta_0 \doteq \sqrt{\chi}$$

$$\theta_1 \doteq \pi + \chi/\pi; \; -\sin \theta_1 \doteq \tan \theta_1 \doteq \chi/\pi$$

$$\cdots \qquad \cdots$$

$$\theta_n \doteq n\pi + \chi/n\pi; \; (-1)^n \sin \theta_n \doteq \tan \theta_n \doteq \chi/n\pi$$

From this we see that the terms in the sum are as follows:

$$\alpha_{00} \cdot 4\chi^2/\chi^2 4 = \alpha_{00}$$

$$\alpha_{n0} \cdot 2(\chi/n\pi)^2/(n\pi)^2 = \alpha_{n0} 2\chi^2/n^4\pi^4$$

$$\alpha_{nm} \cdot 4\chi^4/n^4 m^4 \pi^8$$

From this it is evident that unless χ is large, the series converges very rapidly. (This conclusion is not altered when the increase in α_{nm} with $\beta_n\beta_m$ is considered.) Thus the dominant term in the admittance is

$$4w^2 q\mu p_0 \, (1 + i\omega\tau_{00})^{1/2}/\sqrt{D\tau_{00}}$$

where

$$1/\tau_{00} = 2\left(\frac{D}{w^2}\right)(\theta_0^2) + 1/\tau$$

$$\doteq 2\left(\frac{D}{w^2}\right)\frac{sw}{D} + 1/\tau$$

$$= 2\left(\frac{s}{w}\right) + 1/\tau$$

This expression is valid only for sw/D small so that $\theta_0^2 \doteq sw/D$. The term $s/(w/2)$ represents the rate of decay due to holes recombining on the surface, s having the dimensions of velocity. For $\omega \gg 1/\tau_{00}$, the admittance becomes $4w^2 q\mu p_0 (i\omega/D)^{1/2}$, the same value as given in equation (4.12) for large ω and an area $4w^2$.

The conclusion from this appendix is that for χ small, the effect of surface recombination is simply to modify the effective value of τ and otherwise leave the theory of Section 4 unaltered.

For very large values of χ, it is necessary to consider higher terms in the sum and several values of τ will be important. Under these conditions the

approximation is that, at $x = 0$, p_1 is independent of x and y may become a poor one, especially for forward currents, because the transverse currents to the edges will be important. Under these conditions the role of surface recombination will give rise to patch effects of the sort discussed in Section 4.

APPENDIX VI

The Effect of Trapping Upon the Diffusion Process

In this appendix we shall investigate the effect of the trapping of holes upon the impedance. We denote the density of mobile holes in the valence-bond band by p and the density of holes trapped in acceptors by p_a. For thermal equilibrium at room temperature there will be an equilibrium ratio, called α, for p_a/p. For germanium $\alpha \doteq 10^{-4}$ and for silicon $\alpha \doteq 0.1$ to 0.2.

We shall consider four processes which occur at rates (per particle per unit time) as follows:

ν_r direct recombination of a hole with an electron (free or bound to a donor)
ν_t trapping of a hole by an acceptor
ν_{ra} recombination of a hole trapped on an acceptor
ν_e excitation of a trapped hole into the valence-bond band.

Under equilibrium conditions as many holes are being trapped (rate $p\nu_t$) as are being excited ($p_a\nu_e$): hence $\nu_t = \alpha\nu_e$.

We shall study solutions of the customary form for the a-c. components:

$$p_1 = p_{10}\, e^{i\omega t - \gamma x}$$

$$p_{1a} = p_{1a0}\, e^{i\omega t - \gamma x}$$

These must satisfy the equations

$$\dot{p}_1 = D\nabla^2 p_1 - (\nu_t + \nu_r)p_1 + \nu_e p_{1a}$$

$$\dot{p}_{1a} = \nu_t p_1 - (\nu_e + \nu_{ra})p_{1a}$$

These lead readily to the equation for γ:

$$D\gamma^2 = i\omega + \nu_r + \nu_t - \nu_e\nu_t/(i\omega + \nu_e + \nu_{ra}) = i\omega$$

$$\cdot\left[1 + \frac{\nu_e\nu_t}{(\nu_e + \nu_{ra})^2 + \omega^2)}\right] + \nu_r + \nu_t\left[1 - \frac{\nu_e}{(\nu_e + \nu_{ra}) + \omega^2/(\nu_e + \nu_{ra})}\right]$$

From this equation we can directly reach the important conclusion that the trapping process can never lead to a capacitative term larger than the resistive term. This result is obtained by analyzing the complex phase of γ, the admittance being proportional to γ. In particular, we find that the real term in $D\gamma^2$ is always positive, as may be seen from inspection, so that the complex phase angle of γ is less than 45°.

The form reduces to a simple expression if ν_e and ν_t are very large com-

pared to ν_r, ν_{ra} and ω, a situation which insures local equilibrium between p and p_a. Under these conditions we obtain

$$D\gamma^2 = i\omega[1 + \alpha] + \nu_r + \alpha\nu_{ra}$$

Dividing by $(1 + \alpha)$ gives

$$[D/(1 + \alpha)]\gamma^2 = [Dp/(p + p_a)]\gamma^2 = i\omega + \frac{p\nu_r + p_a\nu_{ra}}{p + p_a}$$

The interpretation is that the holes diffuse as if their diffusion constant were reduced by the fraction of the time $p/(p + p_a)$ they are free to move and recombine with a properly weighted average of ν_r and ν_{ra}.

APPENDIX VII

Solutions of the Space Charge Equation

We shall first show that the space charge equation (2.11) has a unique solution for the one dimensional case. For simplicity we write (2.11) in the form

$$\frac{d^2u}{dx^2} = \sinh u - f(x) \tag{A7.1}$$

to which it can be readily reduced. We shall deal with the case for which

$$f = f_a \text{ for } x < x_a \tag{A7.2}$$

$$f = f_b \text{ for } x > x_b > x_a \tag{A7.3}$$

so that the interval (x_a, x_b) is bounded by semi-infinite blocks of uniform semiconductor. We shall require that u be finite at $x = \pm\infty$. This boundary condition requires that for large values of $|x|$

$$u = u_a + A_a e^{+\gamma_a x} \qquad x \to -\infty \tag{A7.4}$$

$$u = u_b + A_b e^{-\gamma_b x} \qquad x \to +\infty \tag{A7.5}$$

where

$$\sinh u_a = f_a, \qquad \sinh u_b = f_b$$

$$\gamma_a = |(\cosh u_a)^{1/2}|, \qquad \gamma_b = |(\cosh u_b)^{1/2}|$$

(If the opposite signs of the γ's were present, the boundary conditions would not be satisfied.) The exponential solutions are valid for $|u - u_a|$ or $|u - u_b| \ll 1$. For larger values, however, solutions exist which are obtained by integrating (A7.1) to larger or smaller values of x.

For these extended solutions the values of $u(x, A_a)$ and $u'(x, A_a)$ $(= du/dx)$

are monotonically increasing functions of A_a. This may be seen by considering $x = x_a$. For A_a sufficiently small, the value of $u(x_a, A_a)$ and $u'(x_a, A_a)$ are given simply by (A7.4). For larger values of A_a, an exact integral will be required. It is evident, however, that l solutions of the form (A7.4) are related simply by translation for $x < x_a$. Hence increasing A_a is simply equivalent to integrating (A7.1) to larger values of x and it is evident that this increases u and u' mo otonically. It may be verified that for a sufficiently large A_a the solution becc es infinite at x_a so that $u(x_a, A_a)$ $u'(x_a, A_a)$ both vary monotonically and c tinuously from $-\infty$ to $+\infty$ as A_a varies from negative to positive valu We shall refer to this property of $u(x_a, A_a)$, $u'(x_a, A_a)$ as P_1.

We next wish to show that $u(x_1, A_a)$, $u'(x_1, A_a)$ has the property P_1 for values of $x_1 > x_a$. To prove this we note that if for any x_1, $u(x_1, A_a)$ and $u'(x_1, A_a)$ are finite, the solution may be integrated somewhat further to obtain $u(x_2, A_a)$, $u'(x_2, A_a)$ for $x_2 > x_1$. From equation (A7.1) it is evident that an increase in either $u(x_1, a)$ or $u'(x_1, a)$ will result in an increase in d^2u/dx^2 in the interval $x_1 < x < x_2$ so that u and u' at x_2 are monotonically increasing functions of u and u' at x_1. Hence if u and u' at x_1 have the property P_1, so do u and u' at x_2. By extending this argument we conclude that u and u' at any value of x have the property P_1. (A rigorous proof can easily be completed along these lines provided that $|f(x)|$ is finite.)

Similarly it may be shown, starting from (A7.5), that $u(x, A_b)$ is a monotonically increasing function of A_b and $u'(x, A_b)$ is a monotonically *decreasing* function of A_b.

In order to have a solution satisfying (A7.4) and (A7.5) we must have, for any selected point x,

$$u(x, A_a) = u(x, A_b) \tag{A7.6}$$

$$u'(x, A_a) = u'(x, A_b) \tag{A7.7}$$

Now as the equation $u(x, A_a) = u(x, A_b)$ varies from $-\infty$ to $+\infty$, $u'(x, A_a)$ varies from $-\infty$ to $+\infty$ and $u'(x, A_b)$ varies from $+\infty$ to $-\infty$, monotonically and continuously. Hence there is one and only one solution of (A7.1) satisfying (A7.4) and (A7.5).

In order to verify that the solutions discussed in Section 2 are correct for large and for small K, we show schematically in Fig. A1 the solution for a representative K as a dashed line together with the curve $u = u_0(y) = \sinh^{-1} y$. In terms of u_0, equation (2.16) becomes

$$\frac{d^2u}{dy^2} = \frac{1}{K^2}(\sinh u - \sinh u_0). \tag{A7.8}$$

From the symmetry of the equation, it is evident that u must be an odd function of y and hence that the solution must pass through the origin. The boundary condition in this case will be that $u \to u_o$ for $y \to \pm\infty$ so that there will be no space charge far from the junction. We can conveniently use the origin as the point at which the solution from $y = +\infty$ joins that from $y = -\infty$; from symmetry, this requires merely that $u = 0$ when $y = 0$.

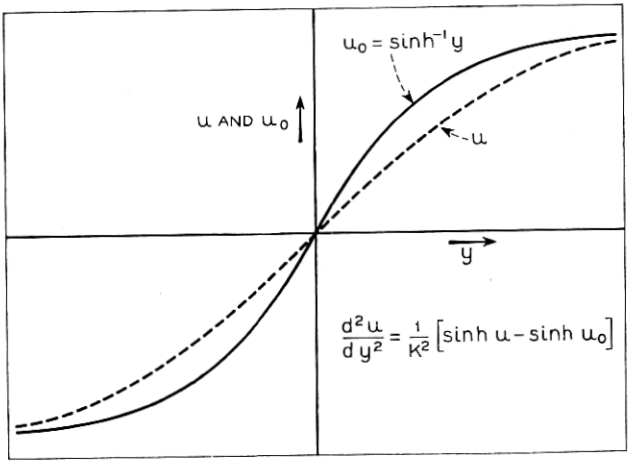

Fig. A1—Behavior of the solution of Equation (2.16) or (A7.8).

For large negative y, $u = \sinh^{-1} y$ and $du/dy = 1/\cosh u_0$ so that du/dy is small. It is at once evident that, for large values of K, u must lie above u_0 so that the integral

$$(1/K^2) \int_{-\infty}^{y} (\sinh u - \sinh u_0)\, dy = \frac{du}{dy} \qquad \text{(A7.9)}$$

will be large enough to make the solution $u(y)$ pass through the origin. If $u - u_0 > 2$ over the region of largest difference, the space charge will be largely uncompensated and the solution will correspond to that used in equation (2.18). On the other hand, as $K \to 0$, the requirement that $u(y)$ pass through the origin leads to the conclusion that $u - u_0$ must be small for all values of y. The possibility that u oscillates about u_0 need not be considered since it may readily be seen that, if for any negative value of y, say y_1, both $u(y_1)$ and $u'(y_1)$ are less than $u_0(y_1)$ and $u'(y_1)$, then $u(y)$ and $u'(y)$ are progressively less than $u_0(y)$ and $u_0'(y)$ as y increases from y_1 to 0. Hence, if for negative y the u curve goes below the u_0 curve, it cannot pass through the origin.

APPENDIX VIII

List of Symbols

(Numbers in parentheses refer to equations)

$a = (N_d - N_a)/x$ (2.14)
A = admittance per unit area of junction (4.23)
A_p = component of A due to hole flow into n-region (4.12) (4.24)
A_n = component of A due to electron flow into p-region (4.25)
A_T = component of A due to varying charge distribution in transition region
A also used as a constant coefficient in various appendices
b = ratio of electron mobility to hole mobility
b = symbol for base in Sections 5 and 6
B constant coefficient in various expansions in appendices
c = symbol for collector in Section 6; a length in Appendix III
C = capacity per unit area
C_n, C_p (4.25) (4.27) as for A_n, A_p
C_T (2.42) (2.45) (2.56) as for A_T
D = diffusion constant for holes (bD is the diffusion constant for electrons)
$e = 2.718\ldots$
f see Appendix 7
g = rate of generation of hole-electron pairs per unit volume (3.1)
G = conductance per unit area of junction
G_n, G_p as for A's
$i = \sqrt{-1}$
I = current density
I_n, I_p = current densities due to electrons and holes (2.5) (2.6) (4.10)
I_{n0}, I_{p0} I_{p1} (4.11) (4.12) (4.18) (4.19)
I_s, I_{ns}, I_{ps} saturation reverse current densities (4.11) (4.18) (4.21)
I_r see text with (4.35)
J = subscript in Section 3 for junction Fig. 5 equation (3.11)
k = Boltzmann's constant
K = space charge parameter (2.17)
L = length
$L_a = n_i/a$ (2.15)
L_D = Debye length (2.12)
L_n, L_p = diffusion lengths for electron in p-region and holes in n-region (4.8)
L_r = length required for potential increase of kT/q in region of constant field (4.32) Appendices II and IV
L_1 corresponds to a-c. field, Appendix IV
n = density of electrons

n_n, n_p = equilibrium densities of electrons in n- and p-regions
p = density of holes
p_n, p_p = equilibrium densities of holes in n- and p-regions
p_0 = d-c. component of non-equilibrium hole density (4.3)
$p_1 \exp(i\omega t)$ = a-c. component of non-equilibrium hole density (4.3)
P = total number per unit area of holes in specimen (2.35)
q = electronic charge ($q = |q|$)
$Q = qP$ = total charge per unit area (2.39)
r = recombination coefficient for holes and electrons (3.1)
R = resistance of unit area
R_0 = resistance of unit area obtained by integrating conductivity (3.10), Appendix I
R_1 = effective series resistance, discussed in connection with (3.13)
s = rate of recombination per unit area of surface per unit hole density, Appendix V
S = susceptance per unit area (imaginary part of admittance)
S_p, S_n, S_T as for A's.
t = time
T = temperature in °K
T = subscript for transition region
$u = q\psi/kT$ (2.9), $q(\psi - \varphi_1)/kT$ (2.32), Appendix VII
v_0 and $v_1 e^{i\omega t}$ = d-c. and a-c. components of voltage applied in forward direction (4.2)
W = width of space charge region in abrupt junction, Section 2.4
w = half thickness of n-region or transistor base of Sections 5 and 6.
w = half width of square rod in Appendix V.
x = coordinate perpendicular to plane of junction
y, z = transverse coordinates, Appendix V
y = reduced length (2.17), Appendix VII
α = current gain factor in transistor (6.4)
α = parameter in Appendix III and VI
α_{ij} = parameter in Appendix V
β_i = parameter in Appendix V
γ = parameter in Appendices II, IV and VII
ϵ = symbol for emitter Section 6
$\theta_i = \beta_i w$ Appendix V
κ = dielectric constant
μ = mobility of a hole ($b\mu$ = mobility of electron)
ν = rates of recombination etc., Appendix VI
ρ = charge density (2.1)
σ = conductivity

σ_i = conductivity of intrinsic material (4.15)
σ_n = conductivity of n-region $\doteq qb\mu n_n$
σ_p = conductivity of p-region $\doteq q\mu p_p$
τ = time
τ_n, τ_p = life times of electrons in p-region and holes in n-region (3.2) (3.3) (4.7)
τ_T = relaxation time of transition region, Appendix IV
$\varphi, \varphi_p, \varphi_n$ = Fermi level and quasi Fermi levels (2.2) (2.4)
$\delta\varphi$ = applied voltage across specimen in forward direction, Section 2.3, (4.2)
χ = sw/D in Appendix V
ψ = electrostatic potential (2.2)
ω = circular frequency of a-c. (4.2)