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

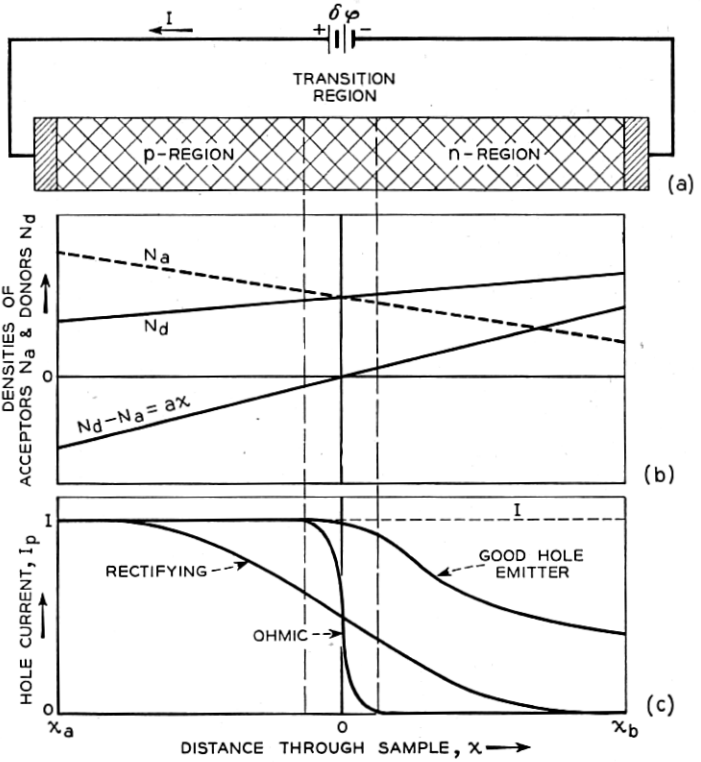

Fig. 1—The *p-n* junction.

(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.

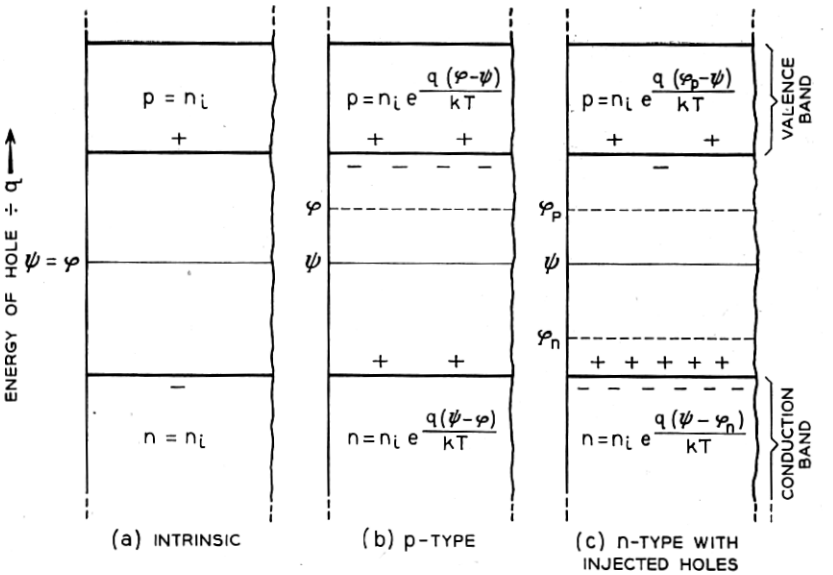

Fig. 2—Electrostatic potential ψ, Fermi level φ and quasi Fermi levels φ_p and φ_n. (In order to show electrostatic potential and energies on the same ordinates, the energies of holes, which are minus the energies of electrons, are plotted upwards in the figures in this paper.)

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

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

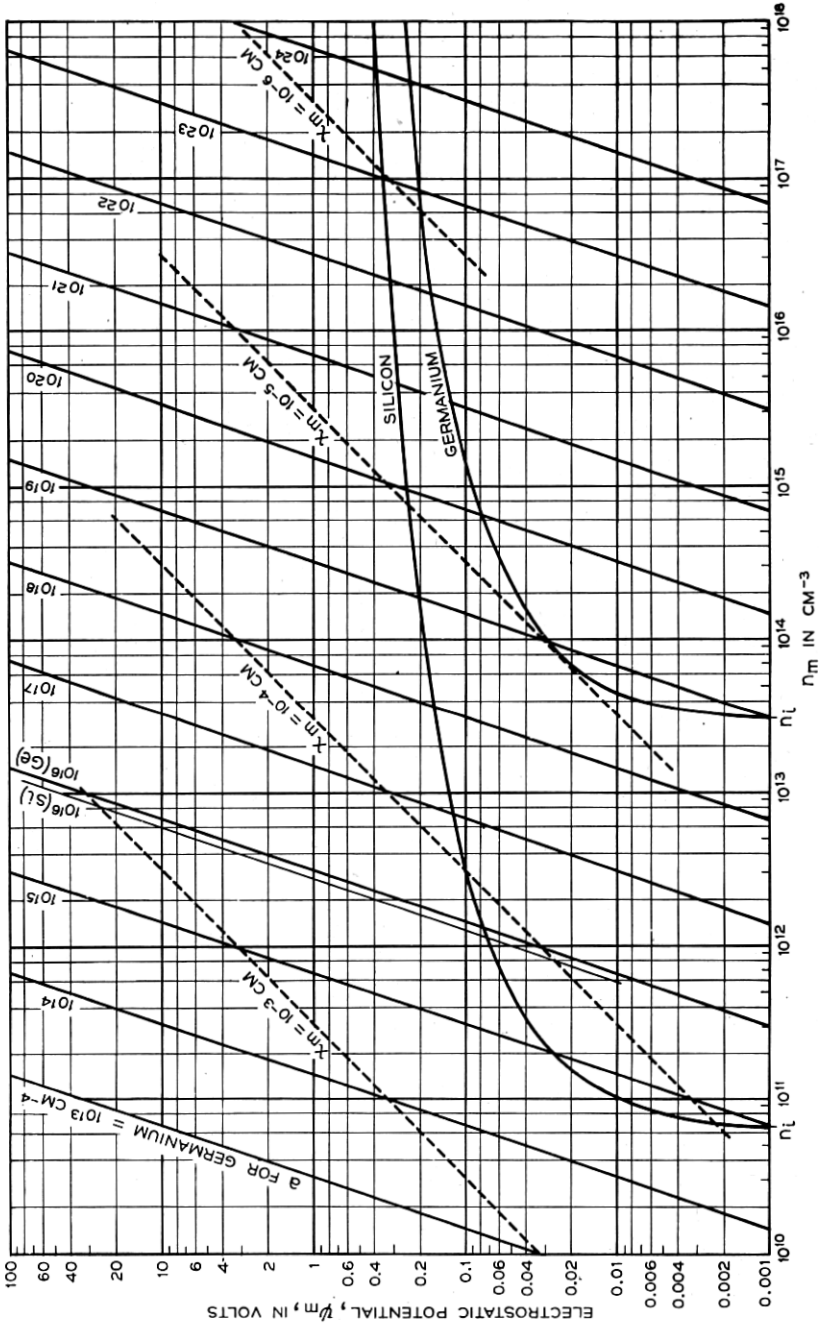

Fig. 3—Solutions for the boundaries of the space-charge region.

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-

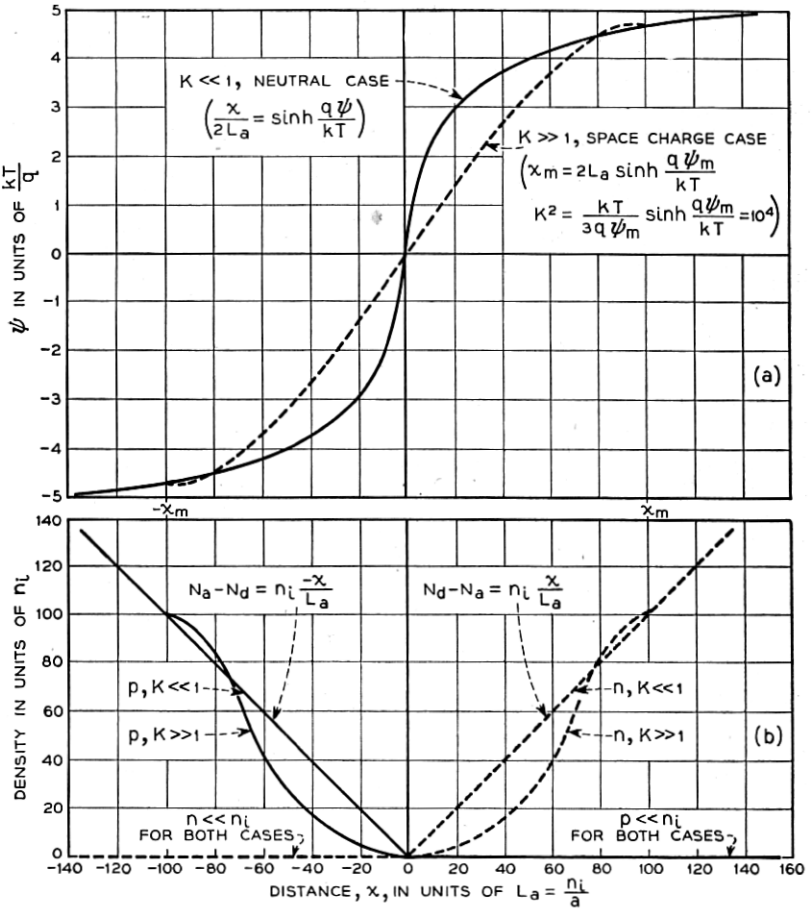

Fig. 4—Electrostatic potential and densities for *p-n* junctions.

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