PHYSICAL REVIEW VOLUME 99, NUMBER 2 JULY 15, 1955

Physical Theory of Semiconductor Surfaces

C. G. B. GARRETT AND W. H. BRATTAIN
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

(Received December 20, 1954)

The properties associated with the space-charge region and with surface states at a semiconductor surface are discussed. A theory of the space-charge region that takes into account charge-densities arising from immobile impurities and from both signs of mobile carrier is presented. The properties of the space-charge are discussed in terms of the surface potential and of the electrochemical potentials of holes and electrons, and related to the transport of added carriers in a homogeneous semiconductor. The change in surface conductivity arising from nonvanishing surface excesses of holes and electrons is treated. The space-charge systems at a free surface and at a *p-n* junction are compared, and the range of validity of the Mott-Schottky space-charge theory evaluated. The arrangement of surface states is discussed with reference to the Brattain-Bardeen model. Theories for the surface photoeffect and field-effect experiments are given, with and without surface states: it is concluded that the existence of surface states is without gross effect on the former, while relevant quantitative evidence from the latter is not yet available. The question of the relation between surface potential and contact potential is discussed. The properties of "channels" are discussed in terms of the theory. The paper concludes with a short section on long-time effects.

INTRODUCTION

THE purpose of this paper is to summarize the conclusions which may be drawn from physical theory as to the electrical properties of the surface of a semiconductor. The surface of a semiconductor, like most phase boundaries, is the seat of a space-charge double layer. In many one-dimensional problems, it suffices to consider the case that the field outside the semiconductor is zero: this means that the net surface charge density is also zero. The double-layer may in principle arise in several ways, such as preferential adsorption of ions of one sign, or alignment of adsorbed atoms or molecules having an electrical dipole moment. In a metal, such charge can be compensated by a high density of electrons or holes in the immediate vicinity, i.e., within a few atomic diameters, of the geometrical surface. In a semiconductor, part at least of the electron or hole distribution must be spread over a relatively wide space-charge region, owing to the combined requirements of electrostatics and statistical mechanics. At the same time, there may still be an extra population of holes and electrons bound in the near vicinity of the surface, since one must not expect bulk statistical considerations to hold over atomic distances. The concept of "surface states" was proposed some years ago by Tamm,[1] and discussed in detail by Bardeen.[2] It should, however, be noticed that the distinction between charges consisting of electrons and holes in surface states, and charges deriving from impurity atoms adsorbed onto the surface, is far from clear. Specifically, the lack of dependence on work function of the rectifying properties of metal semiconductor contacts might be due to Tamm-type surface states, or to charges associated with adsorbed gases: one does not know. One may distinguish, a little artificially perhaps, between these two kinds of surface charge as follows. That part of the surface charge which, while not forming part of the extended space-charge region, is yet in good electrical contact with the conduction or valence band, will be described as charge in surface states; that part which can only change relatively slowly, with some sort of "activation energy," is thought of as associated with adsorbed ions. How far this distinction can be maintained operationally must be left for experiment to decide. Another convenient assumption is that the surface states are associated with recombination processes at the surface; that is, the surface states are trapping centers of the type discussed by Shockley and Read.[3]

In Sec. A of this paper, we shall be concerned only with the properties of the space-charge region itself. The problem is to write down expressions for the surface excesses Γ_n and Γ_p of electrons and holes, as functions of the *surface potential* (the difference between the mean electrostatic potential just inside the surface and in the interior) and the electrochemical potentials for holes and electrons. From this one can write down (*i*) the net surface charge density due to the space-charge region, (*ii*) the shape of the space-charge region, i.e., the dependence of potential on distance, and (*iii*) (with some reservations) the excess conductivity due to mobile carriers near the surface.

In Sec. B, an attempt will be made to use the theoretical concepts to describe the results of experiments which have been or can be made on semiconductor surfaces. In the understanding of many of these measurements, it is unfortunately necessary to make some assumption about the surface states. On the basis of surface recombination measurements made on germanium surfaces prepared in a certain way, Brattain and Bardeen[4] proposed a particular model, in which the surface states are supposed to be either of donor character, lying high in the forbidden band, or of

---

[1] I. Tamm, Physik Z. Sowjetunion **1**, 733 (1932).

[2] J. Bardeen, Phys. Rev. **71**, 717 (1947).

[3] W. Shockley and W. T. Read, Phys. Rev. **87**, 835 (1952).

[4] W. H. Brattain and J. Bardeen, Bell System Tech. J. **32**, 1 (1953).

acceptor type lying low. In this paper we shall predict the results of various experiments on two extreme assumptions: (*i*) that there are negligibly few surface states, and (*ii*) that there are surface states of the type proposed by Brattain and Bardeen.

The problem of the apportionment of charge between space-charge and surface states has also been considered by Pikus,[5] who has discussed such questions as the variation of work function with temperature and the effect of applying an electric field normal to the surface. For a detailed discussion of the properties of Tamm states, as opposed to the questions of ionic charge discussed by Brattain and Bardeen, we refer the reader to this paper.

A. PROPERTIES OF THE SPACE-CHARGE REGION

1. Fundamental Considerations

The theory of the space-charge region at the surface of a semiconductor has been given by Schottky,[6] Schottky and Spenke,[7] and Mott.[8] For convenience we shall develop here the considerations applicable to rather high-resistivity germanium or silicon, at or near room temperature. It will be assumed that impurity atoms (donors and acceptors) are completely ionized at all points. The space charge then may be written down in terms of the local concentrations p and n of holes and electrons, which may be taken to obey Boltzmann statistics, so long as the impurity concentration in the semiconductor is low enough to make the Fermi-Dirac assembly nondegenerate. The considerations to be given are applicable either to a step (abrupt) junction within a single crystal between regions of differing impurity content, or to a free surface.

Following Shockley[9] we write

$$\begin{aligned} p&=n_i e^{\beta(\varphi_p-\psi)},\\ n&=n_i e^{\beta(\psi-\varphi_n)}, \end{aligned} \tag{1}$$

where φ_p and φ_n are the "quasi-Fermi levels"[9] for holes and electrons respectively; ψ is the electrostatic potential, n_i the density of holes and of electrons in intrinsic semiconductor, and $\beta=e/kT$, e being the electronic charge, k Boltzmann's constant, and T the absolute temperature. The quasi-Fermi levels have, apart from a change in sign of φ_n, the properties of electrochemical potentials. The zeros of φ_p and φ_n have been chosen in such a way that, in thermodynamic equilibrium, $\varphi_p=\varphi_n$: we shall denote this common value by φ_0. Now in equilibrium φ_0 is uniform throughout the system; in particular, it is constant right up to the surfaces of the sample. In the body of the semiconductor the electrostatic potential ψ is also uniform, but this ceases to be true in the vicinity of the surface. In order to determine the form of the electrostatic potential near the surface, one must solve Poisson's equation:

$$\nabla^2\psi=-4\pi/\epsilon\epsilon_0. \tag{2}$$

(Unrationalized quantities are used in writing this equation.) Here ϵ is the dielectric constant for the semiconductor, and ρ is the charge density at each point, which, for a completely ionized semiconductor, is given by

$$\rho=e[(p-p_0)-(n-n_0)], \tag{3}$$

where p_0 and n_0 are the hole and electron concentrations in the body of the semiconductor. With some reservations, to be discussed below, the charge distribution near the surface is completely determined (in the case of thermodynamic equilibrium) by substituting Eqs. (1) and (3) into Eq. (2), and solving with the appropriate boundary conditions.

Before following this procedure, one generalization is required. In dealing with semiconductor problems, it is often important to consider steady-state distributions which, although not corresponding to strict thermodynamic equilibrium, are sufficiently close to it to be regarded as quasi-equilibrium configurations. The lifetime of minority carriers in high-resistance germanium is so long that it is permissible to treat the holes and electrons as separate components in the thermodynamic sense, so that the sum of the electrochemical potentials (the *difference* of quasi-Fermi levels) is no longer to be held constant. Each electrochemical potential, however, is uniform throughout the system; in particular, we may consider a steady-state configuration of the semiconductor surface, such as may be produced by shining light on it, in which φ_p and φ_n are different from φ_0, but each is constant throughout the space-charge region. This approximation holds good so long as (*i*) the diffusion length of minority carriers in the semiconductor is long in comparison with the thickness of the space-charge region; (*ii*) currents (arising from either carrier) to or from the surface are not too large, nor the depletion of carriers, such as may be produced by an externally applied voltage, too extreme.[10]

In a moment we shall proceed to calculate the space-charge configuration in terms of arbitrary values for φ_p and φ_n. It is not possible, however, to choose φ_p and φ_n independently, because, in the body of the semiconductor, there must be net space-charge

---

[5] G. E. Pikus, Zhur. Eksptl. i. Teort. Fiz. **21**, 1227 (1951).

[6] W. Schottky, Z. Physik **113**, 367 (1939); **118**, 539 (1942).

[7] W. Schottky and E. Spenke, Wiss. Veröffentl. Siemens-Werken **18**, 3 (1939).

[8] N. F. Mott, Proc. Roy. Soc. (London) **A171**, 27 (1939).

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

[10] The assumption of constancy of "quasi-Fermi levels" cease to be true for a p-n junction that is subjected to a reverse bias considerably larger than kT/e. In order that the hole and electron currents I_p and I_n may be solenoidal (no recombination), the products $p\nabla\varphi_p$ and $n\nabla\varphi_n$ must be constant, and since p and n are greatly reduced in value near the junction on the n and p sides respectively, $\nabla\varphi_p$ and $\nabla\varphi_n$ must be quite large there. Calculation suggests that, in a typical case, the assumption of constancy of φ_p and φ_n is reasonably good up to about 10 times kT/e, that is, about 0.2 volt. An exact treatment of the case of still larger reverse biases has not been attempted, involving as it does the solution of the bipolar conduction problem in the presence of space charge.

neutrality, so that one electrochemical potential is determined if the other is known. The relation between them is obtained from Eqs. (1) and (3) by setting $\rho=0$, and will appear as Eq. (10) below.

There are two limitations on the validity of a treatment that starts from Eqs. (1) and (2). Equations (1) apply rigorously to the case of a statistical assembly extending indefinitely in all directions; we wish to apply them to a space-charge region in which the potential is varying very rapidly with distance. A rigorous discussion of this difficulty, which occurs also in the Debye-Hückel theory of strong electrolytes, has been given by Kirkwood.[11,12] Another limitation is set by the wave nature of the electron; one would not expect statistical considerations to apply down to distances of the order of an electron wavelength. In the absence of any alternative method, we shall assume that Eqs. (1) and (2) will be a fair approximation so long as the variation of ψ is not so great that the semiconductor becomes degenerate at the surface. The total extent of the space-charge region in high-resistivity germanium is of the order of 10^{-4} cm, but certain of the properties associated with it depend on the distribution of carriers within only 100 A of the surface. In the light of the above considerations, any conclusions concerning these latter properties must be regarded with caution.

Substituting Eqs. (1) into Eq. (3), we obtain

$$\rho=en_i[e^{\beta(\psi_0-\varphi_0)}-e^{\beta(\varphi_0-\psi_0)}+e^{\beta(\varphi_p-\psi)}-e^{\beta(\psi-\varphi_n)}], \tag{4}$$

in which we have written ψ_0 for the value of the electrostatic potential in the body of the semiconductor. Now suppose that ψ and ρ vary only in the direction perpendicular to the surface; and let x stand for the distance along this direction, with $x=0$ at the surface, and x going positive inwards. Integrating Eq. (2), we obtain

$$\left(\frac{d\psi}{dx}\right)^2=-\frac{8\pi}{\epsilon\epsilon_0}\int_{\psi_0}^{\psi}\rho d\psi. \tag{5}$$

Into this we substitute Eq. (4) and carry out the integration, so obtaining

$$\frac{d\psi}{dx}=\frac{2}{\beta\mathfrak{L}}F(y,\lambda,P,N), \tag{6}$$

where

$$\mathfrak{L}=[\epsilon\epsilon_0/2\pi en_i\beta]^{\frac{1}{2}},$$
$$\lambda=[p_0/n_0]^{\frac{1}{2}}=p_0/n_i=n_i/n_0=e^{\beta(\varphi_0-\psi_0)},$$
$$P=\beta(\varphi_p-\varphi_0),$$
$$N=\beta(\varphi_n-\varphi_0),$$
$$y=\beta(\psi-\psi_0),$$

and

$$F(y,\lambda,P,N)=\mp[\lambda e^P(e^{-y}-1)+\lambda^{-1}e^{-N}(e^y-1)+(\lambda-\lambda^{-1})y]^{\frac{1}{2}},$$

with the following sign convention: when $y<0$, the positive branch of the square-root function must be chosen; when $y>0$, the negative branch. Note that $\mathfrak{L}$ is a characteristic length for the semiconductor, having the value 1.4×10^{-4} cm for germanium at room temperature.

From Eq. (6) we can write down expressions for the surface excesses Γ_p and Γ_n of holes and electrons. The idea of "surface excess," first introduced by Gibbs,[13] implies the difference between the total amount of some component in the actual system, per unit area of surface, and that which would be found if the phases were homogeneous right up to some chosen dividing surface. We choose the dividing surface in such a way that the surface excess of the component forming the semiconductor is zero; and write Y and ψ_s for the values of y and ψ at this surface (averaged over distances of the order of atomic dimensions). Then

$$\begin{aligned}\Gamma_p&=\int_0^\infty(p-p^*)dx\\&=-\int_0^Y(p-p^*)\left(\frac{dx}{dy}\right)dy\\&=-\tfrac{1}{2}n_i\mathfrak{L}\lambda e^P\int_0^Y\frac{(e^{-y}-1)}{F(y,\lambda,P,N)}dy,\\\Gamma_n&=\int_0^\infty(n-n^*)dx\\&=-\tfrac{1}{2}n_i\mathfrak{L}\lambda^{-1}e^{-N}\int_0^Y\frac{(e^y-1)}{F(y,\lambda,P,N)}dy,\end{aligned} \tag{7}$$

where p^* and n^* are the steady-state values of hole and electron concentration at a point just inside the space-charge region, and are given by $p^*=p_0e^P$ and $n^*=n_0e^{-N}$. Notice that in the limit of $-Y$ large,

$$\Gamma_p\rightarrow n_i\mathfrak{L}e^{\frac{1}{2}\beta(\varphi_p-\psi_s)}, \tag{8a}$$

while for Y large,

$$\Gamma_n\rightarrow n_i\mathfrak{L}e^{\frac{1}{2}\beta(\psi_s-\varphi_n)}. \tag{8b}$$

The total surface charge density due to excess holes and electrons is

$$e(\Gamma_p-\Gamma_n)=en_i\mathfrak{L}F(Y,\lambda,P,N) \tag{9}$$

---

[11] J. G. Kirkwood, J. Chem. Phys. **2**, 767 (1934).

[12] The question concerns the statistical fluctuations in electrostatic potential at each point, and the resulting fluctuations in the local densities of holes and electrons. It is assumed that the mean distribution functions are obtained, to sufficient accuracy, by inserting into the Boltzmann expressions the mean electrostatic potential at each point. The error involved in so doing becomes larger, the greater the departure of the mean electrostatic potential from its value in the body of the semiconductor. It has been shown for the electrolyte case that the method usually breaks down as soon as it ceases to be a good approximation to write $e^{\beta\Delta\psi}\sim1+\beta\Delta\psi$. Fortunately, the restriction in the semiconductor system is less severe. C. Herring has pointed out to us that the approximation used is satisfactory so long as $e^2/\epsilon\epsilon_0\mathfrak{L}'\ll kT$, where $\mathfrak{L}'=(\epsilon\epsilon_0/2\pi Ne\beta)^{\frac{1}{2}}$, in which N stands for the local concentration of holes or electrons, whichever is the greater. This condition is easily satisfied in most cases of practical interest.

[13] J. Willard Gibbs, *Collected Works* (Longmans Green and Company, London, 1906), Vol. 1, p. 219.

as may be proved in a more straightforward manner by considering the value of $d\psi/dx$ at the surface. Notice that the integrals for Γ_p and Γ_n separately cannot easily be evaluated explicitly, but the difference $\Gamma_p-\Gamma_n$ can.

So far we have said nothing as to what determines the values of φ_p and φ_n. Since, however, φ_p and φ_n are constant right through the space-charge region, we may invoke the condition of space-charge neutrality in the body of the semiconductor, and write

$$(e^{-N}-1)/(e^{P}-1)=\lambda^2, \tag{10}$$

as may easily be proved from Eqs. (1) and (3), setting $\rho=0$.

2. Shape of the Space-Charge Region

The form of the space-charge region at a semiconductor surface or interface has usually been treated in the following way. If the electrostatic potential near the surface is greater than in the interior for *n*-type semiconductor, or less for *p*-type semiconductor, there will be an enrichment of the majority carrier in the surface region, which may be described by a simple Boltzmann expression. In the contrary case, there is a depletion of the majority carrier, which may be treated by thinking of an exhaustion region of appropriate depth containing no carriers of either sign. At the same time it has been realized that, on base material of high resistivity, one may have the state of affairs in which there is an appreciable concentration of minority carriers at the surface. Such a surface region is usually called an inversion layer. The purpose of this section is to consider more exactly the shape of the space-charge region, with particular reference to material that is not too far from intrinsic, in order to show clearly the limits within which the usual simple treatment is allowable.

To find ψ as a function of x, one integrates (6):

$$x=\tfrac{1}{2}\mathfrak{L}\int_0^Y \frac{1}{F(y,\lambda,P,N)}dy. \tag{11}$$

This integral cannot easily be evaluated explicitly. In order to illustrate the shape of the space-charge region, we shall consider the behavior of the integral for the case of a change in potential near the surface tending towards the production of an inversion layer. Three parts of the space-charge region may be distinguished:

(*i*) $0<|y|\ll 1$

Expanding the exponentials, and using (10), we have

$$F(y,\lambda,P,N)\rightarrow -y[\tfrac{1}{2}(\lambda^{-1}-\lambda)+\lambda e^{P}]^{\frac{1}{2}}. \tag{12}$$

The integral therefore diverges logarithmically: the space-charge region extends indefinitely far into the semiconductor. In this region, however, the properties are not greatly different from those of the bulk semiconductor.

(*ii*) *The "Parabolic" Region*

In the usual, simplified form of the space-charge theories of Mott and Schottky, it is assumed that all of the space-charge arises from absence of both holes and electrons in a certain well-defined region. Using this approximation, it is easy to show that the potential should vary as the square of the distance. This approximation is equivalent to ignoring all terms under the square root sign except $(\lambda^{-1}-\lambda)y$. It may be checked that this is a poor approximation except under the following conditions:

$$\begin{aligned} &\text{(a)}\quad y\ll -1,\quad e^{-y}/(-y)\ll \lambda^{-2}e^{-P},\\ \text{or}\quad &\text{(b)}\quad y\gg 1,\quad e^{y}/y\ll \lambda^{2}e^{N}. \end{aligned} \tag{13}$$

These conditions specify that the potential must vary in the direction tending to produce an inversion layer, by an amount that is large in comparison with kT/e, but still small in comparison with that which would actually produce an inversion layer. The conditions are most clearly satisfied at a reverse-biassed *p-n* junction, but at the free surface of a block of semiconductor that is not too far from intrinsic the range of potentials in which they are satisfied is small.[14]

In the range of x for which y satisfies condition (a), one has

$$x_1-x\rightarrow \mathfrak{L}\lambda^{\frac{1}{2}}(1+\tfrac{1}{2}\lambda^2)[(-y)^{\frac{1}{2}}-(-y_1)^{\frac{1}{2}}], \tag{14}$$

where $x=x_1$ is the limit of this region on the inside, and y_1, the value of y there, is of the order of -1.

(*iii*) *Inversion Region*

When $\lambda<1$ and y is sufficiently large and negative, or when $\lambda>1$ and y is sufficiently large and positive, the greatest contribution to the space-charge comes from carriers of the opposite kind to those predominating in the bulk. Here it is sufficient to neglect all except the first or the second term in the expression under the square-root sign in Eq. (6), so that we have: For

$$\begin{gathered} \lambda\ll 1,\quad e^{-y}/(-y)\gg \lambda^{-2}e^{-P},\\ x_2-x=\mathfrak{L}\lambda^{-\frac{1}{2}}e^{-\frac{1}{2}P}(e^{\frac{1}{2}y_2}-e^{\frac{1}{2}y}), \end{gathered} \tag{15}$$

where $x=x_2$ represents the inside limit of this third region, and y_2 is the value of y there. This third region corresponds to an inversion layer: in the case considered, a hole-rich region over an *n*-type semiconductor. The inversion layer is, as we shall see below, very thin in comparison with $\mathfrak{L}$ in a practical case, and the drop in potential across it is usually insignificant in compari-

---

[14] One physical system where such conditions do exist is the surface of the base region of an *npn* junction transistor having a channel across it. Here a reverse bias can be applied between the *p*-type body and the *n*-type surface. So long as the applied voltage is rather greater than kT/e, the parabolic approximation will hold for most of the space-charge region.

son with the drop across the parabolic region. It is important to remember however, that most of the positive space-charge may be concentrated in the inversion layer.

A similar analysis may be carried out for the case that the potential near the surface is varying in such a sense as to enhance the concentration of the majority carrier. Here there is of course no exhaustion region, and the distribution of carriers near the surface is similar to that worked out for region *(iii)* as may easily be checked from Eq. (11).

The complete space-charge region consists of regions where *(i)* and *(ii)* are applicable, with or without region *(iii)* depending on the boundary conditions. The exact profile may of course be determined by numerical integration of Eq. (11).

3. Space-Charge at a *p-n* Junction

For the sake of completeness, we shall see how the foregoing considerations apply to the case of an "abrupt" *p-n* junction,[9] in which the impurity concentration changes in a distance small in comparison with $\mathcal{L}$. The following notation will be used:

$$\begin{array}{c} x=0 \\ \hline \begin{array}{c|c} p & n \end{array} \\ \hline \rightarrow x \end{array}$$

On the *p*-side, $\lambda_p=[p_0/n_0]^{\frac{1}{2}}>1$, $Y_p=\beta(\psi_s-\psi_{0p})$. On the *n*-side, $\lambda_n=[p_0/n_0]^{\frac{1}{2}}<1$, $Y_n=\beta(\psi_s-\psi_{0n})$. For other quantities, suffixes *p* and *n* denote values referring to the *p*- and *n*-side, respectively.

It is required that, at the junction, the quantity $(d\psi/dx)$ shall be continuous. Using Eq. (6), and paying attention to the sign-convention for the square-root function (note that $Y_p/Y_n<0$), we have

$$\frac{e^{P_p}\lambda_p(e^{-Y_p}-1)+e^{-N_p}\lambda_p^{-1}(e^{Y_p}-1)+(\lambda_p-\lambda_p^{-1})Y_p}{e^{P_n}\lambda_n(e^{-Y_n}-1)+e^{-N_n}\lambda_n^{-1}(e^{Y_n}-1)+(\lambda_n-\lambda_n^{-1})Y_n}=1. \tag{16}$$

In addition, we have two relations of the form of Eq. (10) to be satisfied; and, as pointed out by Shockley,[9] the quasi-Fermi levels must be related to the applied voltage V in the following way:

$$\begin{aligned} \beta V&=P_n-P_p=N_n-N_p, \\ Y_p-Y_n&=-\beta V-\ln(\lambda_n\lambda_p^{-1}). \end{aligned} \tag{17}$$

The above equations may now be solved for N_p, N_n, P_p, P_n, Y_p and Y_n:

$$\begin{aligned} e^{-N_p}&=e^{\beta V}\frac{(1-\lambda_n^2)+\lambda_n^2e^{\beta V}(1-\lambda_p^{-2})}{1-\lambda_n^2\lambda_p^{-2}e^{2\beta V}}, \\ e^{-N_n}&=e^{-\beta V}e^{-N_p}, \\ e^{P_n}&=e^{\beta V}\frac{(1-\lambda_p^{-2})+\lambda_p^{-2}e^{\beta V}(1-\lambda_n^2)}{1-\lambda_n^2\lambda_p^{-2}e^{2\beta V}}, \\ e^{P_p}&=e^{-\beta V}e^{-P_n}, \end{aligned} \tag{18}$$

$$\begin{aligned} Y_p&=-\frac{\beta V+\ln\lambda_n\lambda_p^{-1}}{1+(\mathcal{P}/\mathcal{N})}+\left\{\frac{1-\lambda_n\lambda_p^{-1}e^{\beta V}}{1+\lambda_n\lambda_p^{-1}e^{\beta V}}\right\} \\ &\qquad\times\left\{\frac{(\mathcal{P}/\mathcal{N})-1}{(\mathcal{P}/\mathcal{N})+1}\right\}, \\ Y_n&=\frac{\beta V+\ln\lambda_n\lambda_p^{-1}}{1+(\mathcal{N}/\mathcal{P})}-\left\{\frac{1-\lambda_n\lambda_p^{-1}e^{\beta V}}{1+\lambda_n\lambda_p^{-1}e^{\beta V}}\right\}\left\{\frac{(\mathcal{N}/\mathcal{P})-1}{(\mathcal{N}/\mathcal{P})+1}\right\}, \end{aligned} \tag{19}$$

where $\mathcal{P}=(\lambda_p-\lambda_p^{-1})$ and $\mathcal{N}=(\lambda_n^{-1}-\lambda_n)$. The foregoing expressions represent exact solutions of the problem. We now consider approximate solutions to the following topics: *(i)* the dependence of the Y's on the resistivity of the two sides, and *(ii)* the capacity of the space-charge region.

We first note that, since the maximum forward bias that can be applied across a junction is $\beta^{-1}\ln(\lambda_p/\lambda_n)$, it is permissible to drop terms of order $\lambda_n\lambda_p^{-1}e^{\beta V}$ from Eqs. (18) and (19). This excludes only the case of extreme forward bias. One then sees from Eq. (19) that, for the case $\lambda_p\lambda_n\gg1$ (implying that the *p*-type side is much more heavily doped than the *n*-type side):

$$\begin{aligned} Y_n&\sim\ln\lambda_n\lambda_p^{-1}+\beta V_a+1, \\ Y_p&\sim+1; \end{aligned} \tag{20a}$$

while in the contrary case,

$$\begin{aligned} Y_n&\sim-1, \\ Y_p&\sim-\ln(\lambda_n\lambda_p^{-1})-\beta V_a-1. \end{aligned} \tag{20b}$$

This confirms the usual conclusion that "most of the space-charge region is on the side of higher resistivity." But note that, however heavily doped one side may be, there is still a potential drop of kT/e across it when the junction is reverse-biased.

To illustrate these results, consider a junction between, say, 10 ohm-cm *n*-type and 0.1 ohm-cm *p*-type. If a reverse bias of 0.25 volt is applied, all except 0.025 volt appears across the *n*-side. There is a negative space charge (extending $\sim10^{-6}$ cm) on the *p*-side, which is largely compensated by excess holes in a narrow region (again $\sim10^{-6}$ cm) on the *n*-side of the junction; beyond this there lies the remaining positive space charge, in a region $\sim10^{-4}$ cm in width in which the "parabolic" approximation applies.

The *capacity* associated with the junction is a quantity that is rather difficult to define. The easiest approach is to consider the ac admittance of the junction $G+iS$, and take the quantity $(S/\omega)_{\omega\rightarrow0}$ as describing the low frequency shunt capacitance of the device. The term S is made up of two parts: that arising from diffusion of excess or deficit hole-electron pairs out to distances of the order of a diffusion length, and that due to the creation of a space-charge region in the immediate vicinity of the junction. The first of these, considered by Shockley[9] decreases exponentially with reverse bias, whereas the latter decreases only as

$\surd(-V)$. In consequence it is the space-charge capacity which is of principal interest for reverse biased junctions.

The analytical difficulty in discussing both sources of capacity at the same time arises from the fact that in the presentation given here we have evaded the whole question of the mechanism of carrier transport in the space-charge region, replacing it by the condition of constancy of the quasi-Fermi levels. We shall therefore consider only the space-charge capacity. First we note that $d^2\psi/dx^2$ passes through zero once and once only, to wit at the geometrical junction. The stored charge on each plate of the equivalent condenser is therefore given directly by Eq. (9), which may of course be evaluated either for the p side or for the n side. Into this one substitutes for the P's and Y's from Eqs. (18) and (19), and differentiates with respect to V. The general expression for the capacity is too complicated to be useful. One special case of interest is that in which one side (say the p side) is much more heavily doped than the other. For sufficient reverse bias, the capacity is given by

$$C\sim en_i\mathcal{L}\beta/2\lambda_n{}^{\frac{1}{2}}[-\beta V-2-\ln(\lambda_n/\lambda_p)]^{\frac{1}{2}}. \qquad (21)$$

By comparing this with the diffusion capacity discussed by Shockley, one sees that the space-charge capacity predominates as long as $2\lambda_n{}^{\frac{3}{2}}(L_p/\mathcal{L})e^{\beta V}\times\surd(-\beta V)$ is small in comparison with unity. In a practical case this condition is well satisfied for reverse voltages of 0.2 volt or more. Note that Eq. (21) is nearly the same as one gets by straight-forward application of the Mott-Schottky "exhaustion layer" theory, in which the potential drop occurs entirely on the high-resistivity side, and the potential varies parabolically with distance. One may conclude with one or two remarks as to the conditions under which the Mott-Schottky exhaustion layer theory is justified. When both sides of a p-n junction are equally heavily doped, the exhaustion layer model, and the parabolic potential, apply well to both sides, with zero or negative bias across the junction. When, however, one side of the junction is much more heavily doped than the other, the exhaustion layer model can be used only for the high-resistance side, and then only if one excludes the region in the immediate vicinity of the junction. The extent of the region over which the potential is parabolic increases, of course, with negative bias; so, at high reverse bias, and so long as one is not interested in the precise field distribution near the junction itself, it is reasonable to regard the whole of the potential drop as occurring across a Mott-Schottky exhaustion region.

4. Surface Conductivity

It is possible to evaluate the change in conductivity of the surface region due to the presence of a space-charge layer:

$$\Delta G=e\mu_p(\Gamma_p+b\Gamma_n), \qquad (22)$$

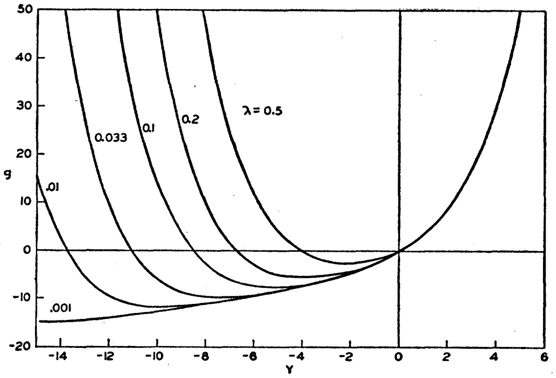

FIG. 1. The surface-conductivity integral as a function of the surface potential Y for various values of the parameter λ (for n-type).

where μ_p is the mobility for holes and b is the ratio of the electron mobility to hole mobility. This equation, however, however, assumes that it is sufficient to suppose that the mobilities of holes and electrons in a differentially thin slice of semiconductor are the same as the bulk mobilities. This will not be too bad unless the thickness of the space-charge region is comparable with the mean free path of the carriers. This latter quantity is of the order of 10^{-5} centimeter, which is not very much smaller than $\mathcal{L}$, so that this approximation is not very good. In consequence, the actual change in surface conductivity would be expected to be rather smaller than the values to be predicted below. The errors arising from this approximation have been investigated by Schrieffer.[15]

Substituting from Eqs. (6) and (7), one finds

$$\Delta G=+\tfrac{1}{2}e\mu_p n_i\mathcal{L}g\lambda^{-\frac{1}{2}}, \qquad (23)$$

where g is the integral

$$g=\lambda^{\frac{1}{2}}\int_Y^0\frac{\lambda(e^{-y}-1)+b\lambda^{-1}(e^{y}-1)}{F(y,\lambda)}dy. \qquad (24)$$

Values of this integral are shown in Fig. 1. Notice that if, for n-type semiconductor, the potential at the surface is greater than that in the interior, the change in surface conductivity arising from extra electrons at the surface is positive. If, however, the surface potential is a little bit less than the potential in the interior, the change in conductivity is negative, because there are now fewer electrons in the surface region at $x=0$, and not yet any appreciable contribution from holes. Only for very low surface potentials does the change in conductivity begin to rise again, as extra holes are created. The minimum surface conductivity is at the potential

$$Y=\ln(\lambda^2/b). \qquad (25)$$

Thus, one sees that the surface must be more p than the interior is n before there is any appreciable hole

---

[15] J. Robert Schrieffer, Phys. Rev. 94, 1420 (1954).

surface conductivity. At the extremes of surface potential the surface conductance is of the form $g\mu_p\Gamma_p$ or $gb\mu_p\Gamma_n$, where Γ_p and Γ_n are given by Eq. (8).

5. Quasi-Fermi Levels and Flow of Minority Carriers

The transport of added carriers in a homogeneous semiconductor has been discussed by van Roosbroeck.[16] We restrict the discussion to a semiconductor that is not too close to intrinsic, and suppose that any current crossing the surface is at any rate small enough for flow in the body of the semiconductor to proceed largely by diffusion rather than drift. For a semi-infinite block, the small signal solution for hole density at points inside the space-charge region is of the form

$$\Delta p = Be^{-x/L_0}, \tag{26}$$

where $L_0=\sqrt{(D_0\tau)}$ and $D_0=D_n$ for n-type, D_p for p-type.

The hole current at a point just inside the space-charge region, arising partly from flow of minority carriers into the surface traps, is

$$I_p' = I_p - \text{surface recombination flow}, \tag{27}$$

where I_p is the hole current actually crossing the surface. But, from Eq. (26), the current I_p' is given by $-eD_0 \operatorname{grad}\Delta p$, which is equal to $I_s'[(p^*/p_0)-1]$ for n-type semiconductor, where p^* is the hole concentration just inside the space-charge region, and I_s', the body saturation current, equals eD_pp_0/L_p. Now the recombination flow may be written $ev_s(p^*-p_0)$, where v_s is the surface recombination velocity, so that

$$I_p = I_s[(p^*/p_0)-1] = I_s(e^P-1), \tag{28}$$

where I_s, the total saturation current, is given by

$$I_s = ep_0[(D_p/L_p)+v_s]. \tag{29}$$

The net hole inflow I_p could arise from injection of minority carriers from a metal or electrolyte in contact with the surface. If the surface is illuminated with light that is absorbed within a distance from the surface short in comparison with a diffusion length (not necessarily short in comparison with a Debye length), it may be shown[17] that the effect is equivalent to a minority carrier flow L across the surface equal to eN, where N is the number of hole-electron pairs created per unit area per second. The effect on the distribution of carriers inside the semiconductor is then given by adding L to the current I_p in the above equations.

Equation (28) represents the linear approximation to surface and body recombination. If I_p is very large, this linear approximation is usually no longer adequate. Under many conditions one might expect a mass-action recombination law to hold,[16] so that one can write instead:

$$\begin{aligned} I_p &= (I_s/n_i^2)(p-p_0)(n_0+p), \\ &= I_s(e^P-1)(1+\lambda^2e^P). \end{aligned} \tag{30}$$

This equation is quadratic in e^P.

B. PROPERTIES OF THE SURFACE CHARGE

6. Space Charge and Charge in Surface States

Equations (7) for Γ_p and Γ_n are correct only so long as it is assumed that Eqs. (1) hold right up to the dividing surface, and that there are no free electrons or holes beyond. As mentioned in the introduction, there are reasons for believing in the existence of surface states into which electrons can be bound in such a way that they cannot take part in lateral conduction. It is reasonable to assume further that these surface states are associated through trapping with recombination of minority carriers at the surface. Brattain and Bardeen proposed that these trapping centers were either of donor type lying high in the band, or of acceptor type lying low, or both. Such a model enables one to understand how the surface recombination velocity can be insensitive to variations of surface potential produced by changes in chemical environment, as found experimentally.[4,18] The argument goes like this. Consider low-lying acceptor-like traps. The flow of current between the conduction band and the traps will be proportional both to the number of *electrons* in the conduction band at the surface, and to the number of *empty* traps; this flow will be much less than that between the traps and the valence band, which is proportional both to the number of *holes* and to the number of *occupied* traps, unless the electrochemical potential of the traps is almost equal to that of the holes: so the main resistance to recombination occurs in the former process. Since variation of surface potential affects the densities of electrons near the surface and of empty traps by compensating factors, surface recombination should be unaffected by changes in surface potential that do not affect the number of traps. Note that either donor or acceptor traps are sufficient to account for recombination in both n- and p-type.

With this model, the charge in surface traps is

$$\sigma_{\text{tr}} = e[N_be^{P-Y} - N_ae^{Y-N}], \tag{31}$$

in which N_a and N_b represent densities of donor and acceptor traps multiplied by the appropriate Boltzmann factors, and P, N, and Y have their previous meanings.

In the following sections we shall assume that the charge in surface states may be written as in Eq. (31); it is easy to proceed to the limiting case of no surface states by setting $N_a=N_b=0$.

---

[16] W. van Roosbroeck, Phys. Rev. **91**, 282 (1953).

[17] W. H. Brattain and C. G. B. Garrett, Physica **20**, 885 (1954); Bell System Tech. J. **34**, 129 (1955).

[18] R. J. Keyes and T. G. Maple, Phys. Rev. **94**, 1416 (1954).

7. Surface Photoeffect

The total charge in both space charge and traps is given by

$$\Sigma = e(\Gamma_p - \Gamma_n) + \sigma_{\mathrm{tr}}, \tag{32}$$

where $\Gamma_p - \Gamma_n$ is given by (9) and σ_{tr} by (31). We are concerned with calculating the change in ψ_s produced on illumination of the surface, the illumination being supposed to take place so abruptly that the ionic charge, and therefore the total charge Σ in the surface of the semiconductor, to which it is equal in magnitude and opposite in sign, does not have time to change. Differentiating the right-hand side with respect to L, the light current, using Eqs. (10) and (28), and setting the result equal to zero, we find (for n-type)

$$\begin{aligned}\frac{d\psi_s}{dL} = \frac{1}{\beta I_s}&[(e^Y - 1) + (e^{-Y} - 1) - Z(\lambda^2 N_a e^Y - N_b e^{-Y})]\\ &\div\{-\lambda^{-2}[(1+\lambda^2 L/I_s)e^Y - 1] + [(1+L/I_s)e^{-Y} - 1]\\ &\quad + Z[(1+\lambda^2 L/I_s)N_a e^Y + (1+L/I_s)N_b e^{-Y}]\},\end{aligned} \tag{33}$$

where

$$Z = 2F(Y,\lambda)/n_i\mathfrak{L}\lambda. \tag{34}$$

The theory of Brattain and Bardeen gave an expression equivalent to ignoring all those terms in (33) not multiplied by Z, since charge in the space-charge region was ignored. The following approximations to (33) should be noted:

$$\begin{aligned} Y \to -\infty &: \frac{d\psi_s}{dL} \sim \frac{1}{\beta I_s}\frac{1}{1+L/I_s} \to \frac{1}{\beta I_s} \text{ for } L \text{ small,}\\ Y \to +\infty &: \frac{d\psi_s}{dL} \sim \frac{-1}{\beta I_s}\frac{\lambda^2}{1+\lambda^2 L/I_s} \to -\frac{\lambda^2}{\beta I_s}\\ &\qquad\qquad \text{for } L \text{ small.}\end{aligned} \tag{35}$$

For N_a, $N_b = 0$, L small:

$$\frac{d\psi_s}{dL} = \frac{1}{\beta I_s}\frac{1 - e^Y}{1 + \lambda^{-2}e^Y}. \tag{36}$$

With the Z terms only (i.e., N_a, $N_b \to \infty$)

$$\frac{d\psi_s}{dL} = \frac{1}{\beta I_s}\frac{1 - e^{2Y'}}{1 + \lambda^{-2}e^{2Y'}}, \tag{37}$$

where $Y' = Y + \frac{1}{2}\ln(\lambda^2 N_a/N_b)$.

Let us consider the import of Eqs. (36) and (37). In each, the surface photovoltage is given as that change in surface potential which is required in order to keep the semiconductor surface charge constant when the imrefs of holes and electrons inside are altered. The reasoning leading to Eq. (37) is relatively simple; most of the surface charge is in donor and acceptor traps, and the populations of these traps would alter when the imrefs change, if it were not for the change in surface potential. The physical interpretation of Eq. (36) is harder to picture. First it is worth noting that if the space-charge region were accurately described by a Mott-Schottky exhaustion layer, and if there were no surface states, there would be no surface photovoltage at all, because the charge associated with the space-charge region would then be a function only of the surface potential, and would be unaffected by the change in imrefs produced on illuminating the sample. The surface photovoltage given by Eq. (36) arises precisely because the Mott-Schottky exhaustion layer model is not exact. In this view (with no surface states) in order to maintain the total charge in the space-charge region constant, the surface potential must change and the only way in which this can happen is by rearrangement of the distribution of that charge. One point to notice is that the limiting values of surface photovoltage are the same, whichever model one uses. As we have emphasized elsewhere,[17] the limiting values of photovoltage may be derived by a thermodynamic argument, and are therefore independent of the precise model assumed; all that the model adds is the dependence of surface photovoltage on surface potential in the intermediate region, which depends, in chemical language, on rate process considerations.

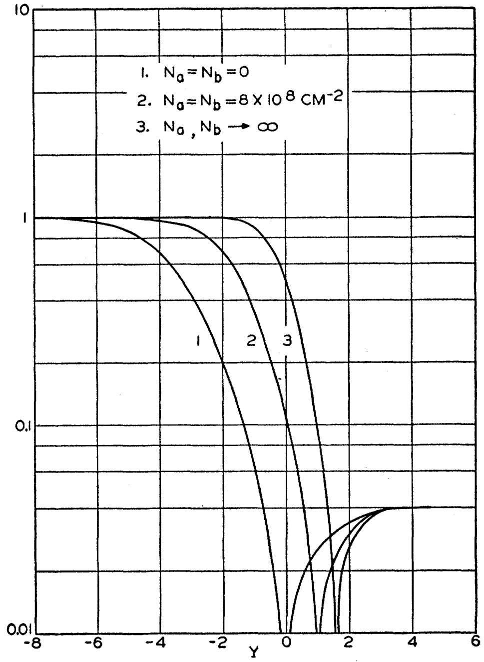

FIG. 2. Surface photoeffect as a function of Y, with and without surface states (relative values).

In Fig. 2, Eq. (33) has been evaluated for $\lambda = 0.2$ for the two extreme conditions represented by (36) and (37) and also for an intermediate case in which $N_a = N_b = 8 \times 10^8$ cm^{-2}. It will be seen that there is little difference between the shapes of the three curves.

The quantity $d\psi_s/dL$ describes the change in surface potential produced by sudden illumination of the surface. Experimentally, one can measure the change in contact potential or electrode potential. This includes changes in potential drop across the material itself, such as IR drop. Suppose that the current I, if any, is kept constant: then there still remains the change in the potential difference $[(b-1)/(b+1)]\beta^{-1}\ln(\sigma/\sigma_0)$ that arises because of the difference in mobilities of holes and electrons.[16] This gives an additional photoeffect:

$$\frac{dV_i}{dL}=\frac{1}{\beta I_s}\frac{b-1}{b}\frac{\lambda^2}{1+(\lambda^2/b)[(b+1)(L/I_s)+1]}, \tag{38}$$

so that the limiting values of the net surface photoeffect for $L\to 0$ should be

$$\begin{aligned} Y&\to-\infty, & \frac{dV}{dL}&\sim\frac{1}{\beta I_s}; \\ Y&\to\infty, & \frac{dV}{dL}&\sim-\frac{\lambda^2}{b}\frac{1}{\beta I_s}. \end{aligned} \tag{39}$$

Expressions for p-type are analogous.

8. Surface Potential and Contact Potential

Let us now turn from the question of variation of surface potential with illumination to that of the dependence of surface potential on chemical environment. One way of investigating this is to measure the contact potential difference between the semiconductor and a reference electrode, of which one has reason to believe that the work function is adequately independent of environment.

The question arises as to how much of the change in contact potential observed, for example, in the Brattain-Bardeen cycle, is change in surface potential ψ_s, and how much is to be ascribed to changes in the ionic double layer. Brattain and Bardeen[4] have concluded that only about 20 percent of the change in contact potential arises from ψ_s. Measurements by Morrison[19] of surface conductivity as a function of gaseous environment suggest that this estimate is too low. If there were no surface states, the predicted result would be close to 100 percent, since the capacity of the space-charge region is two orders of magnitude or more lower than that normally associated with an ionic layer ($\sim$30 μf/cm^2). Even with $N_a=N_b=10^9$ cm^{-2}, the capacity associated with the semiconductor surface is negligibly small except at the extremes of ψ_s. As we shall see below from the field-effect experiment, it is unlikely that the density of surface states is greater than this by more than an order of magnitude or two, so that one would not be surprised to find that changes in ψ_s account for almost all of the observed change in contact potential.

One possible test is the behaviour of contact potential in very strong light. First assume $N_a=N_b=0$. In order to derive expressions for the surface potentials in a very strong light we use the strong form for the recombination (30). On physical grounds one would expect that the difference between the surface potential and the potential in the body would become very small in strong light: now that there is a copious supply of both holes and electrons the charge in the space-charge region has become very sensitive to the exact value of surface potential. Let us then take Eq. (9) and expand the exponentials in a power series in y. In this case, the first-order terms cancel, giving

$$Y_L\sim\frac{[\lambda(e^{-Y_0}-1)+\lambda^{-1}(e^{Y_0}-1)+(\lambda-\lambda^{-1})Y_0]^{\frac{1}{2}}}{[\lambda^2L/I_s]^{\frac{1}{4}}}. \tag{40}$$

From this, one sees that $Y\to 0$ in strong light as the fourth root of the intensity of the light.

A similar analysis can be carried out for nonvanishing N_a and N_b. The point is that in all cases the effect should be to shift that portion of contact potential arising from ψ_s to a standard value ($Y=0$ for $N_a=N_b=0$; $Y'=0$ for N_a, $N_b\to\infty$) irrespective of the starting-point: any contributions from the ionic double layer, however, should be unchanged.

The experimental difficulty in this lies in the high light-intensity required. One rough experiment on an n-type sample was carried out, the results of which are shown in Fig. 3. (The measurements were made on an ac basis, with a low duty-cycle, to minimize heating.) The upper curve was obtained in an atmosphere of humid oxygen (about 70 percent relative humidity); the lower in an atmosphere of dry oxygen. For the upper curve $Y>0$, the surface is more n-type than the interior, and the surface photovoltage is small; for the lower curve $Y<0$, the surface is strongly p-type, and the small-signal photovoltage is large and of the opposite sign. The fact that, in the latter case, it was possible to shift the contact potential nearly half-way to the zero point, shows that the original estimate of Brattain and Bardeen of $\Delta\psi_s/\Delta\text{cp}\sim 0.2$ is assuredly too low, as

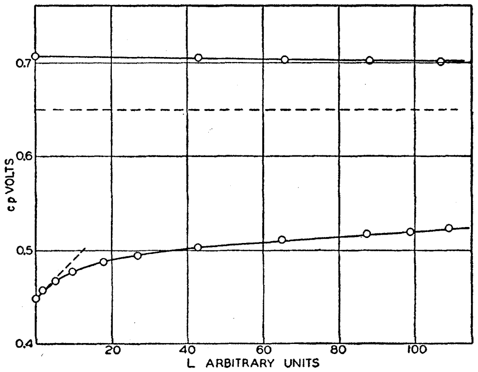

FIG. 3. Experimental behavior of contact potential in strong light.

---

[19] S. R. Morrison, J. Phys. Chem. 57, 860 (1953).

concluded by Morrison, but it was not possible to attain a light intensity sufficiently great to reach the region where Eq. (40) might apply.[20]

The conclusions of this section may be summarized thus: (1) The surface photo-effect, in combination with measurements of surface recombination velocity (in order to calculate I_s) is the most reliable tool for determining whether the surface is *p*-type or *n*-type, and how much so; (2) Such measurements do not, however, shed much light on the problem of determining the location and density of surface traps.

9. Field Effects

In the field-effect experiment of Shockley and Pearson,[21] measurements were made of the change in surface conductivity produced by applying a high normal electric field. The results of such a measurement are best described by quoting an effective mobility μ_{eff}, defined by $d\Delta G/d\Sigma$, the differentiation being carried out at constant ionic charge. Using (32), (7), (23), and (31) we find

$$\mu_{\text{eff}}=\mu_p\frac{1-\lambda^{-2}be^Y}{1+\lambda^{-2}e^Y-Z(N_ae^{2Y}+N_b)/(e^Y-1)},\tag{41}$$

where Z again has the significance of Eq. (34). Equation (41) is illustrated in Fig. 4, for the case $\lambda=0.2$, with $N_a=N_b=0$ and with $N_a=N_b=8\times10^8$ cm^{-2}. (For N_a, $N_b\rightarrow\infty$, $\mu_{\text{eff}}\rightarrow0$ for all Y).

Results of measurements on germanium generally give values of $|\mu_{\text{eff}}|$ lower than μ_p by about a factor of 2.[22] Although such experiments have not been carried out in a controlled environment such as the Brattain-Bardeen cycle, there is a suspicion that μ_{eff} is much more nearly independent of Y than one would gather from Fig. 4. If this turns out to be the case, either the Brattain-Bardeen trap model is wrong, or the traps responsible for surface recombination are different from the surface states that cause $|\mu_{\text{eff}}|$ to be less than for the case of space-charge only. It is clear, however, that N_a and N_b cannot be very much greater than 10^9 cm^{-2}, or there would be no field effect at all.[23]

Two other possible experiments may be mentioned here: (*i*) the influence of applied field on the surface photo-effect; (*ii*) the influence of light on the effective mobility as measured in the field-effect experiment.

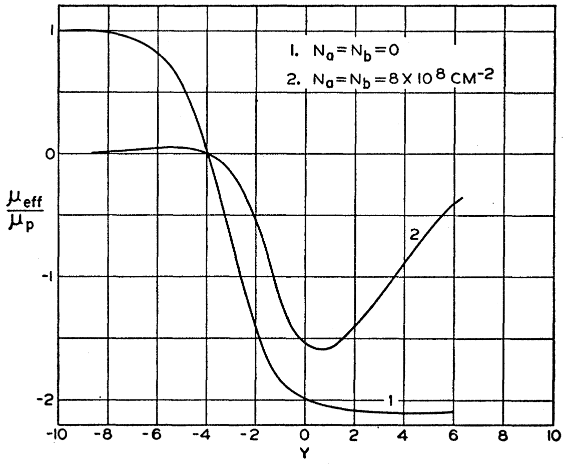

FIG. 4. The "effective mobility" in the field-effect experiment, with and without surface states.

The first of these measures the quantity

$$\left(\frac{\partial}{\partial V}\right)_L\left(\frac{\partial\psi_s}{\partial L}\right)_\Sigma=\frac{\epsilon_{\text{air}}\epsilon_0}{4\pi D}\frac{\partial^2\psi_s}{\partial\Sigma\partial L},\tag{42}$$

where D is the gap across which the voltage ΔV is applied. This may be evaluated from Eqs. (32) and (33). The resulting expressions are very cumbersome. For the case of N_a, N_b smaller than $n_i\mathfrak{L}$, and in the vicinity of the change in sign of the photo-effect ($|Y|<1$), it may be shown that

$$\Delta\left(\frac{d\psi_s}{dL}\right)=\frac{\epsilon_{\text{air}}}{\epsilon_{\text{Ge}}}\frac{\lambda^{-\frac{3}{2}}\mathfrak{L}}{\sqrt{2}D}(\beta V)\left(\frac{1}{\beta I_s}\right),\tag{43}$$

independently of the magnitude of N_a and N_b, to first order in these quantities. The effect should be easily measurable, but gives no direct information as to the surface states. The second experiment measures the quantity

$$\left(\frac{\partial}{\partial L}\right)_\Sigma\left(\frac{\partial\Delta G}{\partial\Sigma}\right)_L=\left(\frac{\partial\psi_s}{\partial L}\right)_\Sigma\left(\frac{\partial}{\partial\psi_s}\right)_L\left(\frac{\partial\Delta G}{\partial\Sigma}\right)_L+\left(\frac{\partial}{\partial L}\right)_{\psi_s}\left(\frac{\partial\Delta G}{\partial\Sigma}\right)_L.\tag{44}$$

This may be evaluated, using Eqs. (23) and (33). The general result is too complicated to be worth quoting here, but it may be shown for the case $N_a=N_b=0$ that, in the vicinity of the point at which μ_{eff} changes sign, illumination of the surface with light of intensity of the order of I_s should produce significant changes in μ_{eff}. Such an experiment ought to give an additional check on the theory.

10. "Channel" Conductance

The formulation of space-charge theory which we have presented in this paper makes possible a simple

---

[20] The lower branch of Fig. (3) fits well to an equation of the form $L=I_s(e^{\beta\Delta\psi}-1)(1+\lambda^2e^{\beta\Delta\psi})$, obtained by setting $P-Y=$ const [see Eqs. (9) and (31)] and using Eq. (31). This shows that, even at the highest light, the surface potential has not yet shifted to the point at which electrons or donor traps have become of appreciable importance.

[21] W. Shockley and G. L. Pearson, Phys. Rev. **74**, 232 (1948).

[22] J. Bardeen and S. R. Morrison, Physica **20**, 873 (1954). H. C. Montgomery and W. L. Brown, Bull. Am. Phys. Soc. **30**, No. 2, 42 (1955).

[23] Remember that N_a and N_b are the numbers of traps *multiplied by a Boltzmann factor*. The actual numbers should be much greater than this.

analysis of the question of "channel" conductance, first discussed by Brown.[24] Consider a sample of germanium containing a *p-n* junction, and suppose that the surface of the *n*-side is covered with negative ions, in such a way as to make the surface everywhere *p*-type. The properties of the structure will be affected by the fact that the holes near the surface on the *p*-type side are in electrical contact with the *n*-type side. A complete analysis demands the consideration of the three-dimensional current flow problem, allowing for the presence of space charge, and requires a knowledge of how the ionic charge will change with time as a result of the change in electrochemical potential of holes and electrons produced by applying a bias to the junction. It is easy, however, to write down the channel conductance for the one-dimensional case, in which the ionic charge and electrochemical potentials are supposed uniform in the direction parallel to the surface and constant with time. The channel conductance is then that surface conductance arising from minority carriers in the inversion region near the surface. Since the channel conductance is significant only when the inversion region is well developed, it is sufficient approximation to set the actual number of minority carriers per unit area near the surface equal to the surface excess of minority carriers. We shall suppose that the base material is *n*-type, so that holes are the minority carriers. In our agreed approximation, the channel conductance is thus equal to $e\mu_p\Gamma_p$. The magnitude of this quantity depends on the density of absorbed anions, and therefore on the chemical environment. The interesting question is then how the channel conductance depends on the voltage applied to the nearby junction in the bulk semiconductor. If the section of the channel considered is very near the bulk junction, it is sufficient to make our usual assumption that the electrochemical potentials are everywhere uniform, so that the applied voltage V and the quantity P on the *n*-type side are related approximately by $V=P/\beta$ [see Eq. (18)].[25,26] The voltage is imagined to be applied suddenly, so that the ionic charge on the surface has not time to change. The change in the space-charge region so produced still depends on the assumption made in regard to the surface states. We start first with the assumption $N_a=N_b=0$. From Eq. (9), one can write $\Gamma_p-\Gamma_n$ constant. For the case of interest (βV, $Y\ll -1$) the only terms in $F(Y)$ that need be considered are λe^{P-Y} and $\lambda^{-1}Y$, we have the following equation for Y:

$$\lambda^2 e^{\beta V-Y}-Y=\lambda^2 e^{\Delta}+\Delta, \tag{45}$$

where Δ is the value of $(-Y)$ for the case $\beta V=0$. By using Eq. (7), it may then be shown that the channel conductance G_{ch} is well described by the following approximations:

$$-\beta V<-\beta V_0,\quad -Y\sim-\beta V+\Delta,$$
$$G_{\mathrm{ch}}\sim K\{[-\beta(V+V_0)]^{\frac{1}{2}}-(-\beta V)^{\frac{1}{2}}\}. \tag{46}$$

$$-\beta V>-\beta V_0,\quad -Y\sim-\beta V_0+\Delta,$$
$$G_{\mathrm{ch}}\sim K(-\beta V_0)^{\frac{1}{2}}\{[e^{\beta(V-V_0)}+1]^{\frac{1}{2}}-1\}, \tag{47}$$

where $\beta V_0=-\lambda^2 e^{\Delta}$ and $K=e\mu_p\lambda^{-\frac{1}{2}}n_i\mathcal{L}$.

It will be seen that, in the model of no surface states, the channel conductivity should vary rather slowly with reverse bias until a voltage V_0 is reached, and should fall off very rapidly thereafter. The point is that, for zero bias, the (negative) ionic charge is compensated partly by holes, partly by ionized donors in the exhaustion region; as negative voltage is applied, the exhaustion region widens, until finally all of the (fixed) negative ionic charge can be compensated by ionized donors. There being no holes left, the channel is described as having been "pinched off."

With an appreciable density of surface states, this pinching off does not occur. Using Eq. (31), we see that there is now no linear term in Y in the expression for surface charge: therefore the approximation $-Y\sim-\beta V+\Delta$ should continue to hold indefinitely. In the limit, for $-V$ very large, we should have

$$G_{\mathrm{ch}}\sim\tfrac{1}{2}K(-\beta V_0)/(-\beta V)^{\frac{1}{2}}, \tag{48}$$

the conductivity decreasing as the inverse square root of applied voltage. For high reverse bias, however, as pointed out by Kingston,[26] the thickness t of the region in which holes can flow becomes small in comparison with the bulk mean free path l. Following Kingston, we take this thickness to mean the distance X in which the electrostatic potential falls through kT/e; then, from Eq. (11), for the limit $-\beta V\gg-\beta V_0$, one has

$$X=\tfrac{1}{2}\mathcal{L}\int_{Y+1}^{Y}\frac{dy}{(\lambda e^{\beta V-y}-\lambda^{-1}y)^{\frac{1}{2}}}\sim\frac{\lambda^{\frac{1}{2}}\mathcal{L}}{2(-\beta V)^{\frac{1}{2}}}, \tag{49}$$

so that, because of the reduction in available thickness, the hole mobility is reduced in the ratio $\lambda^{\frac{1}{2}}\mathcal{L}/2l(-\beta V)^{\frac{1}{2}}$. Thus $K=e\mu_p n_i\mathcal{L}^2/2l(-\beta V)^{\frac{1}{2}}$, and

$$G_{\mathrm{ch}}\sim\frac{e\mu_p n_i\mathcal{L}^2}{4l}\left(\frac{V_0}{V}\right) \tag{50}$$

Therefore $G_{\mathrm{ch}}\sim 3\times10^{-6}(V_0/V)$ mhos per square for germanium at room temperature. This agrees with Kingston's conclusion[26] that, at high reverse bias, the conductivity should be proportional to $1/V$. The remarkable feature of (50), however, is that the channel conductivity at $V=V_0$ is, under these conditions, *independent of the resistivity of the base material.* The reason for this is that the quantity λ enters both into the concentration of minority carriers at the surface

---

[24] W. L. Brown, Phys. Rev. **91**, 518 (1953).

[25] Since, as we shall see, the channel conductance is comparatively low, this assumption is incorrect for regions distant from the bulk junction, because, as Kingston has pointed out,[26] the current carried by the channel sets up an ohmic drop in potential along the channel, which makes the effective voltage applied between the channel and the *n*-type body smaller and smaller, the further one proceeds away from the body junction.

[26] R. H. Kingston, Phys. Rev. **98**, 1766 (1955).

and into the width of the channel, so that the channel conductance depends only on Δ, and on this only through the "pinch-off" voltage V_0.

For a channel of length c and perimeter d, the channel conductance should thus be $\approx 3\times 10^{-6}(d/c)$ at $V_0 \sim V$. In the junction transistor structure, (d/c) may be of the order of 50 or more for a channel over the base region: the channel conductance should therefore be $\sim 10^{-4}$ mhos at $V_0 \sim V$. Experiments of Kingston[26] and of Brown[24] give just this order of magnitude. In an alloyed structure, no appreciable channel can form over the base, but a channel over the emitter or collector would have $(d/c) \sim 500$, so that channel conductances of 10^{-3} mho would not be surprising.

The general conclusion of this section is to confirm Kingston's observation that, for sufficient reverse bias, $G \propto 1/V$, and to show that the coefficient in this relation depends on the bulk properties and surface treatment through the factor V_0. In comparing with the experimental results, however, one must remember that, if sufficient time is allowed after the application of bias, the ionic charge itself may change.

11. Time Effects

It has been assumed throughout that surface states are in extremely good electrochemical contact with the conduction or valence band of the semiconductor, while charges associated with ions are fixed. That the former is a reasonable assumption is shown by the observation that the "effective mobility" observed in a field-effect experiment is consistently less than the space-charge-only value even when measurements are made in a time of the order of 10^{-3} second after switching on the field.[22] The assumption about the constancy of ionic charge implies that there are no further time constants in the surface system. This is not true. When the surface of a lump of germanium is illuminated by a fairly intense light, it is found that the contact potential returns over a period of seconds or minutes towards its initial (dark) value; on switching off the light, the contact potential overshoots, and only recovers the original value some time later. Similar effects have been observed by the authors in experiments on germanium electrolyte systems.[17] These experiments can be described by saying that the ionic charge tends to change in such a sense as to restore the surface potential ψ_s to the value corresponding to chemical equilibrium. This is what one would expect from the theoretical treatments of chemisorption on a semiconductor given by Aigrain and Dugas, Engell and Hauffe and by Weisz.[27]

The question of long-time effects introduces considerable uncertainty in the application of the results of section B of this paper to actual systems. In addition, there is the complication that, in structures involving body *p-n* junctions, ions on the surface may drift under the influence of applied fields[28] thereby introducing further long-time changes in surface properties.

12. Conclusion

In this paper we have considered the theoretical interpretation of experiments on the electrical properties of semiconductor surfaces. The theory of the space-charge region may be written down fairly precisely, but considerable ignorance still prevails as to the exact structure of surface states, under given chemical conditions. Measurements of surface photoeffect are not in themselves sufficient to answer this question. The most promising line of attack appears to be a combination of the field-effect experiment with a controlled variable chemical environment.

ACKNOWLEDGMENTS

The authors have had the advantage of several discussions with R. C. Fletcher and W. L. Brown. We wish also to thank R. W. Hamming and the staff of the computation laboratory for the numerical evaluation of the surface conductivity integral (Fig. 1).

---

[27] P. Aigrain and C. Dugas, Z. Elektrochem. **56**, 363 (1952); H.-J. Engell and K. Hauffe, Z. Elektrochem. **56**, 366 (1952); **57**, 762 (1953); P. B. Weisz, J. Chem. Phys. **20**, 1483 (1952); **21**, 1531 (1953).

[28] J. T. Law, Proc. Inst. Radio Engrs. **42**, 1367 (1954).