have

$$W^M{}_{H_1H_2}=\int_{-\frac{1}{2}}^{\frac{1}{2}} D_{H_1H_2}(z)\cos 2\pi Mz dz.$$

Accordingly the probability functions W_M which characterize the stacking disorder can be synthesized. The unknown numerical scale factor can be determined from the requirement that W_M is everywhere positive and $\int\int W_M dy_1 dy_2=1$.

In most instances it is found that the functions W_M approach a limit W_∞ as M increases. It is therefore convenient to set

$$W_M=W_\infty+\Gamma_M$$

where accordingly Γ_M represents fluctuations at small values of M so that the Fourier coefficients $\Gamma^M{}_{H_1H_2}$ go towards zero with increasing M.

One sees readily that

$$\sum_M W^M{}_{H_1H_2}\cos 2\pi Mz=W^\infty{}_{H_1H_2}\sum_M\cos 2\pi Mz+\sum_M\Gamma^M{}_{H_1H_2}\cos 2\pi Mz.$$

Thus the functions W_∞ is generally to be associated with the sharp maxima, the functions Γ_M with the diffuse maxima of the functions $P_{H_1H_2}$.

If the function W_∞ is a constant, implying zero correlation between layers far apart, $W_{H_1H_2}=0$ unless $H_1=H_2=0$. Apart from the reflections $(00H_3)$ all maxima of $P_{H_1H_2}$ will then be diffuse.

The writer has used the method outlined above to determine the stacking disorder in several crystals of the layer structure type.

PHYSICAL REVIEW VOLUME 71, NUMBER 10 MAY 15, 1947

Surface States and Rectification at a Metal Semi-Conductor Contact

JOHN BARDEEN
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received February 13, 1947)

Localized states (Tamm levels), having energies distributed in the "forbidden" range between the filled band and the conduction band, may exist at the surface of a semi-conductor. A condition of no net charge on the surface atoms may correspond to a partial filling of these states. If the density of surface levels is sufficiently high, there will be an appreciable double layer at the free surface of a semi-conductor formed from a net charge from electrons in surface states and a space charge of opposite sign, similar to that at a rectifying junction, extending into the semi-conductor. This double layer tends to make the work function independent of the height of the Fermi level in the interior (which in turn depends on impurity content). If contact is made with a metal, the difference in work function between metal and semi-conductor is compensated by surface states charge, rather than by a space charge as is ordinarily assumed, so that the space charge layer is independent of the metal. Rectification characteristics are then independent of the metal. These ideas are used to explain results of Meyerhof and others on the relation between contact potential differences and rectification.

INTRODUCTION

THE generally accepted view[1] of the nature of the rectifying contact between a metal and a semi-conductor is illustrated in Fig. 1 which applies specifically to an excess semi-conductor. Figure 1a shows an energy level diagram of the metal and semi-conductor in equilibrium, but with the contact separated. The Fermi level is the same in both the metal and the semi-conductor. As shown, the work function of the metal, χ_1, is greater than the work function of the semi-conductor, $\chi_2{}^0$, so that there is a contact difference in potential, $\chi_1-\chi_2{}^0$. It is assumed that when the metal and semi-conductor are nearly joined, the potential distribution is as shown in Fig. 1b. A double layer is formed such as to give a potential drop, φ_0, from the metal to the interior of the semi-conductor equal to the contact potential difference.

This double layer is assumed to consist of a space charge region in the semi-conductor, extending to a depth of the order of 10^{-6} to 10^{-4} cm,

[1] See, for example, N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, London, 1940), Chap. V. The theory of rectification is due in large part to W. Schottky. The most important of his papers is Zeits. f. Physik **118**, 539 (1942).

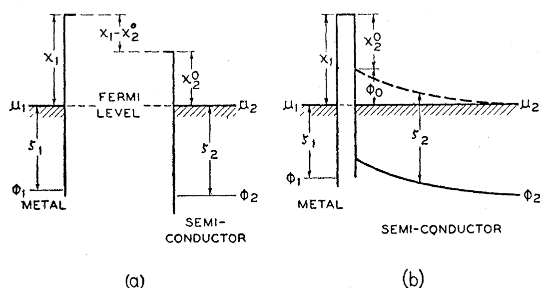

FIG. 1. Energy level diagram for metal and semi-conductor in electrical and thermal equilibrium. μ_1 and μ_2 represent the Fermi levels in metal and semi-conductor, respectively, and ϕ_1 and ϕ_2 the inner potentials. ζ_1 and ζ_2 are the chemical potentials and χ_1 and χ_2^0 the work functions. Figure 1a shows a wide gap with a contact potential difference $x_1-x_2^0$. In 1b, the gap is very small. There is a space charge region in the semi-conductor which gives an electrostatic potential energy rise at the surface $\varphi_0=\chi_1-\chi_2^0$. This is the usual picture in which no surface states on the semi-conductor are assumed.

and an induced charge on the metal surface. The space charge gives a rise in electrostatic potential energy at the surface of the semi-conductor. Electrons are depleted from the space charge region, giving a layer of high resistivity. If a potential is applied to the junction, most of the drop occurs across this barrier layer. If the potential of the semi-conductor is negative with respect to the metal, the electron energy levels in the semi-conductor are raised, and electrons may flow more easily over the potential hill into the metal. This is the direction of easy flow. On the other hand, if the semi-conductor is positive, the levels are lowered, increasing the height of the hill, and making it more difficult for electrons to travel from the semi-conductor to the metal. This is the direction of high resistance.

According to this view, the equilibrium height of the potential hill, φ_0, and therefore also the degree of rectification, depend on the work function of the metal. For an excess semi-conductor, the larger the work function of the metal, the larger is the potential rise, and the larger is the reverse resistance of the barrier. If the work function of the metal is less than that of the semi-conductor there is a potential drop instead of a potential rise, and no rectification will occur. For a defect semi-conductor, in which the current is carried by holes, just the reverse is true: low metal work function gives high rectification.

A number of investigations have been carried out which have verified these conclusions in some cases, and in other cases have not. H. Schweickert, as quoted by Schottky,[2] has found a correlation between the resistance of selenium rectifiers in the blocking direction and the work function of the metal. Selenium is a defect conductor, and high reverse resistance was found for low work function metals such as K, Na, Li, Ba, and low reverse resistance was found for such high work function metals as Ag, Au, Ni. The metal electrodes were put on by evaporation.

W. H. Brattain,[3] working in this laboratory, has found a good correlation between degree of rectification and work function for metal contacts evaporated on cuprous oxide (a defect conductor) and on both N- and P-type silicon.[4] Metals used, listed in order of decreasing degree of rectification, were Al, Ag, and Pt on cuprous oxide; Pt, Be, Ag, Mg, and Al on N-type silicon; and Mg, Cd, Ag, and Pt on P-type silicon. He found that when contact is made to the semi-conductor by a metal junction in air, the rectification is practically independent of the work function of the metal used. Results somewhat similar to those of Schweickert have been obtained by J. N. Shive[3] who studied the rectification characteristics of a number of contacts made by evaporation of various metals on selenium. Metal contacts, listed in order of decreasing degree of rectification, are Be, Zn, Pb, and Au.

A. V. Joffe[5] has studied contact potential differences and the resistances of the contacts formed from a large number of different semi-conductors and metals. Most gave very poor rectification characteristics. While there was some qualitative correlation of contact potential differences with contact resistance, quantitative agreement with theories of Schottky[1] and Davydov[6] was poor.

[2] W. Schottky, Physik Zeits. 41, 570 (1940).

[3] Unpublished work done at the Bell Telephone Laboratories in 1940 under the general direction of J. A. Becker. The author is indebted to Drs. Brattain and Shive for permission to quote their results.

[4] The designations N- and P-type refer to the direction of rectification. An N-type is an excess semi-conductor, and the direction of easy flow occurs when the semi-conductor is negative relative to the metal. A P-type semi-conductor is a hole conductor, and the direction of easy flow is opposite. In both cases, the direction of easy flow is that in which the carrier moves from the semi-conductor to the metal.

[5] A. V. Joffe, J. Phys. USSR 10, 49 (1946).

[6] B. Davydov, J. Phys. USSR 4, 355 (1941).

W. E. Meyerhof[7] has recently been making an extensive study of the relation between contact potential difference and rectification for metal point contacts applied to Si and Ge. He has determined the potential rise, φ_0, from the rectification characteristics at different temperatures and has also measured the ordinary contact potential difference, or Volta potential, by a modified Kelvin bridge method. Preliminary results show little correlation between the two sets of measurements. In fact, he has found that φ_0 is practically independent of the work function of the metal for metal point—Si rectifiers. The same metal may rectify with both P- and N-type silicon, in opposite directions.

These negative results indicate that a closer analysis of the nature of the contact between a metal and semi-conductor is warranted. The main purpose of the present paper is to investigate the effect of electronic states on the surface of the semi-conductor on φ_0. There has been considerable discussion in the literature concerning the possibility of surface states,[8] but little direct evidence as to their existence.

IMPURITY LEVELS AND SURFACE STATES

According to the modern theory of semi-conductors, there is an energy gap between the highest filled band of levels and the lowest state of the conduction band. In the language appropriate to excess semi-conductors, conductivity results from electrons thermally excited to the conduction band from impurity levels. The impurity levels, which are intermediate in energy between the filled band and conduction band, represent states in which the electron is localized around a foreign atom or other defect in the crystal lattice. In normal semi-conductors there may be the order of 10^{-7} to 10^{-3} impurity levels for each atom of the crystal.

In addition to the impurity levels in the interior of the crystal, there may be localized states on the surface with energies in the "forbidden" region between the filled and conduction bands. Shockley[8] and others have investigated the conditions under which surface levels may be expected on an ideal crystal. His analysis, based primarily on a one-dimensional model, indicates that "in a plot of the energy spectrum *versus* interatomic distance the surface levels appear only at lattice constants so small that the boundary curves of the allowed energy bands have crossed." The number of such surface states is equal to the number of surface atoms, and, in a neutral crystal, the surface states are half filled. The conditions for surface states are fulfilled, for example, in a diamond-type lattice.

Surface states may also result from surface imperfections, from foreign atoms on the surface, etc. On general grounds, there is good reason to suppose that the ratio of the number of surface levels to surface atoms may be much higher than the ratio of the number of impurity levels to atoms in the interior.

The energy levels corresponding to the surface states may be discrete, or they may have a continuous distribution with all energies in the gap between the filled and conduction bands. There is little evidence of either an experimental or a theoretical nature on this point. Well-defined impurity states of the same nature, sparsely distributed over the surface, would all have about the same energy and form a discrete level. If the impurity atoms are densely distributed over the surface, so that there is considerable interaction between them, a continuous distribution in energy is to be expected. A continuous distribution is also expected for surface states on clean surfaces of the type discussed by Shockley, Goodwin, and others.[8] In the analysis to follow we assume that the surface levels are continuously distributed in energy. Many of the same general conclusions would follow, however, if discrete levels had been assumed.

We will show that if there is a relatively high density of surface states, there may be a double layer on the free surface of a semi-conductor re-

[7] W. E. Meyerhof, "Contact potential difference in crystal rectifiers," Technical Report No. 5, Univ. of Pennsylvania, BuShips contract NObs-34144, Aug. 10, 1946; see Phys. Rev. **71**, 727 (1947).

[8] Surface states with energies in the "forbidden" band were studied first by I. Tamm, Physik. Zeits. Sowjetunion **1**, 733 (1932), and are often called Tamm states. Later theoretical work was done by R. H. Fowler, Proc. Roy. Soc. **A141**, 56 (1933), S. Rijanow, Zeits. f. Physik **89**, 806 (1934), A. W. Maue, Zeits. f. Physik **94**, 717 (1935), E. T. Goodwin, Proc. Camb. Phil. Soc. **35**, 205 (1939), W. G. Pollard, Phys. Rev. **56**, 324 (1939), and W. Shockley, Phys. Rev. **56**, 317 (1939). The last of these is the most critical, and shows under what conditions surface states are occupied in a normal crystal.

sulting from a surface charge caused by electrons in surface states and a space charge of opposite sign extending into the crystal to a depth of 10^{-6} to 10^{-4} cm. The double layer tends to make the work function independent of the impurities in the interior. If contact is made to a metal, the contact potential difference is compensated largely by a true surface charge rather than by space charge, so that the height of the barrier is largely independent of the metal.

NATURE OF CONTACT

Before discussing the effect of surface states, we will give a brief review of the nature of the contact between metal and semi-conductor.

At a contact between any two electronic conductors there is an electric double layer which adjusts the potential of one relative to the other for equilibrium conditions. This potential difference may be taken as a measure of the strength of the double layer.

Some care must be taken in defining the absolute value of the potential. There is, of course, no way of measuring the potential inside a metal or semi-conductor. The potential just outside the surface can be measured, but the difference between the potential inside and outside the surface depends on the double layer at the surface. Only differences in potential between different parts of a conductor can be observed directly.

Nevertheless, it is desirable to consider the potential in the interior. The true time-average electrostatic potential is periodic, with the period of the lattice. The macroscopic potential, Φ, may be defined as the space average of the actual potential over a region large compared with atomic dimensions. This definition is somewhat arbitrary; another definition might yield a different value for the strength of the double layer at a contact or at a surface. However, only differences in double layers have direct significance, and these are independent of the way the potential in the interior is defined.

The general theory of equilibria[9] shows that when two electronic conductors are in contact and in thermal equilibrium, the electrochemical potentials, μ_1 and μ_2, must be the same in both:

$$\mu_1 = \mu_2 \text{ in equilibrium.} \tag{1}$$

In the usual picture of an electronic conductor, μ is just the Fermi level which determines the probability, p, that a state of energy E is occupied.

$$p = 1/(1+\exp[(E-\mu)/kT]). \tag{2}$$

The electrochemical potential in a region may be defined thermodynamically, in a way which is independent of any particular model, by the equation:

$$\mu = \frac{\partial}{\partial n}(U-TS),$$

where n is the number of electrons in the region, and U and S are the energy and entropy. The partial differentiation corresponds to a reversible change in which the volume and temperature are held constant. The internal energy U includes an electrostatic potential energy, $-e\Phi$, per electron. The choice of the zero or reference level from which this potential energy is measured is arbitrary, and there is the same arbitrary choice in the definition of μ.

It is convenient to introduce a quantity which depends on the chemical constitution of the material, and on the electron density, but not on the electrostatic potential. This quantity, ζ, which is called the chemical potential, to distinguish it from the electrochemical potential, is defined by

$$\zeta = \mu - \varphi, \tag{3}$$

where

$$\varphi = -e\Phi$$

is the electrostatic energy per electron.

Let φ_e be the potential energy of an electron just outside the surface of a conductor. This potential is to be evaluated at a point where the image potential is negligible, but at a distance from the surface small compared with macroscopic dimensions. The work function of the surface is then:

$$\chi_s = \varphi_e - \mu. \tag{4}$$

This is the energy required to take an electron from the conductor and place it at rest at a point, defined as above, just outside the surface.

[9] An excellent discussion is given by R. H. Fowler and E. A. Guggenheim, *Statistical Thermodynamics* (Cambridge University Press, England, 1939), Chap. XI. The author is indebted to Dr. C. Herring for the method of presentation used above.

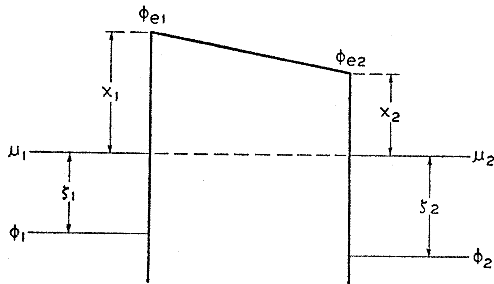

FIG. 2. Diagram to illustrate double layers at a contact between two conductors in equilibrium. There is a double layer of strength $\varphi_{e1}-\varphi_0$ at the surface of conductor 1, a double layer of strength $\varphi_{e2}-\varphi_2$, in the opposite direction at the surface of conductor 2, and a double layer of strength $\varphi_{e2}-\varphi_{e1}=-(\chi_1-\chi_2)$ formed by surface charges on the two conductors. The total strength, $\varphi_2-\varphi_1$, is independent of the surface double layers.

The work function depends on the double layer at the surface. By replacing μ by $\zeta+\varphi$, according to Eq. (3), the work function may be expressed in the form:

$$\chi_s=\varphi_e-\varphi-\zeta, \tag{5}$$

which shows explicitly that the work function is the sum of the energy required to take an electron through the surface double layer, $\varphi_e-\varphi$, and a body term, $-\zeta$, independent of the surface.

It should be emphasized that the separation of the work function into a surface term and a body term depends on the precise way the potential energy, φ, in the interior of the conductor is defined. We have suggested above that the electrostatic potential Φ be defined as space average of the actual potential. Another definition, might yield, for example, a smaller value for $\varphi_e-\varphi$ and a correspondingly larger value for $-\zeta$. Differences in double layers, such as might yield different work functions for different crystal faces of the same material, are not affected by the arbitrary definition of the inner potential. Since the double layer involves a potential difference, the value is independent of the reference level, or zero of potential, which is also arbitrary.

If two electronic conductors are in contact, and in thermal equilibrium, the electrochemical potentials must be the same in both. Thus,

$$\zeta_1+\varphi_1=\zeta_2+\varphi_2,$$

or

$$\varphi_2-\varphi_1=\zeta_1-\zeta_2. \tag{6}$$

The strength of the double layer at the contact, $\varphi_1-\varphi_2$, is equal to the difference between the chemical potentials. Since the chemical potentials depend only on the internal constitutions of the two conductors, the double layer is independent of the work functions of the two surfaces before they are brought into contact.

The situation when there is a small gap between the two conductors in equilibrium is illustrated in Fig. 2. There is a double layer of strength $\varphi_{e1}-\varphi_1$ at the surface of conductor 1, a double layer of strength $\varphi_{e2}-\varphi_2$, in the opposite direction, at the surface of conductor 2, and a double layer of strength $\varphi_{e2}-\varphi_{e1}=-(\chi_1-\chi_2)$, formed by surface charges on the two conductors. The total strength of the double layer is:

$$(\varphi_{e1}-\varphi_1)+(\varphi_{e2}-\varphi_{e1})-(\varphi_{e2}-\varphi_2)=\varphi_2-\varphi_1. \tag{7}$$

The surface double layers drop out, as they should according to Eq. (6).

If the surfaces are in very intimate contact, so that the electron space charges of the two surfaces overlap, it is not possible to divide the total double layer into different parts which depend on the work functions of the individual surfaces. All that can be said is that the strength of the total double layer is $\varphi_2-\varphi_1$. The work functions of the surfaces before they are brought into contact play no role.[10]

USUAL PICTURE OF RECTIFYING CONTACT

If one of the conductors is a semi-conductor, and the other is a metal, it has generally been assumed that as they are brought into contact, no surface charge forms on the semi-conductor.[11] Instead, there is a space charge near the surface which may extend to a depth of the order 10^{-6} to 10^{-4} cm. If the gap is small compared this distance, but is still large enough so that there is

[10] H. Y. Fan, Phys. Rev. 61, 365 (1942); **62**, 388 (1942), has attempted to calculate the electron space charge distribution in the double layer at the contact between two metals, and at the contact between a metal and a semi-conductor. Although his calculations are based on a highly idealized model, they may serve to give a rough picture of the charge density in the contact region. He does not consider the effect of surface states on the semi-conductor.

[11] N. F. Mott, Proc. Camb. Phil. Soc. **34**, 568 (1938) gives a discussion of the contact between a metal and an insulator or a semi-conductor. A brief analysis is given of the effect of a surface charge resulting from electrons in surface states on the contact between a metal and an insulator. Mott assumes a discrete energy level for the surface states.

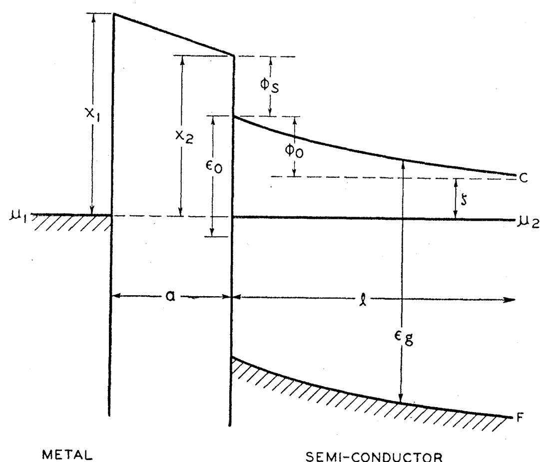

FIG. 3. Energy level diagram for metal semi-conductor contact illustrating notation used in text. The Fermi levels are μ_1 and μ_2 and the work functions χ_1 and χ_2. The lowest state of the conduction band is denoted by C and the highest level of the filled band by F. If the surface states are filled to an energy ϵ_0 below the conduction band there is no net charge on the surface atoms.

no appreciable overlap, most of the energy drop $\chi_1-\chi_2^0$ will occur in the semi-conductor rather than in the gap. This situation is illustrated in Fig. 1b. The space charge raises the energy at the surface of the semi-conductor by an amount $\chi_1-\chi_2^0$. This is the usual picture which indicates that the height of the potential barrier should be equal to the difference in work functions, and is that which is discussed in the introduction.

This picture is deficient in two respects:

(a) The contact may be so intimate that a division of the double layer into one at the surface of the metal, one at the surface of the semi-conductor, and one caused by space charge may not be possible. There will be a double layer at the immediate interface, and another due to space charge, but it is not possible to say how the total strength is divided between them. A cruder way of stating this is that the work functions of the two surfaces are modified by the contact. However, if the contact is intimate, the separate work functions have no meaning.

(b) There may be electronic states localized on the surface of the semi-conductor, so that the surface atoms can become charged. An appreciable field may then exist in the gap. Part of the total drop from metal to semi-conductor will occur across the gap and part across the space charge region. If the density of surface states is sufficiently high, part of the double layer at the *free* surface of a semi-conductor may be formed by a surface states charge compensated by a space charge. The height of the potential barrier of a rectifier formed from the material will then be determined in part by the normal space charge of the free semi-conductor and in part by the work function of the metal.

FREE SURFACE OF SEMI-CONDUCTOR

We will first consider the free surface of a semi-conductor, and then discuss the rectifying contact. The notation to be used is illustrated in the energy level diagram of Fig. 3. The lowest state of the conduction band and the highest state of the filled band of the semi-conductor are indicated, with an energy gap ϵ_g. It is assumed that the distribution of surface states is such that the surface states give no net charge if the states are filled to an energy ϵ_0 below the conduction band.[12] Since the Fermi level cuts the surface above the level determined by ϵ_0, the surface as shown will be negatively charged, this charge resulting from electrons in states between ϵ_0 and the Fermi level. The picture applies to an excess semi-conductor.[13] In the body of the

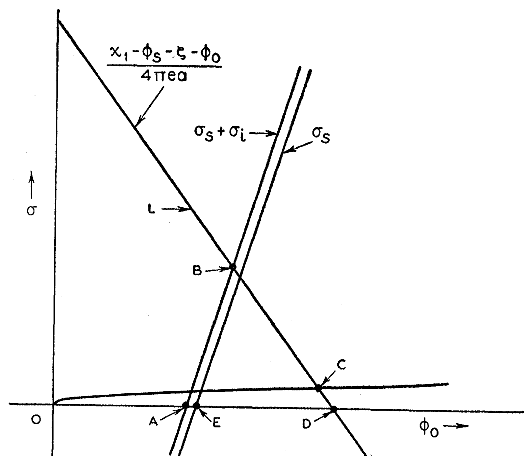

FIG. 4. Schematic diagram showing how the potential rise ϕ_0 is determined from the density of surface levels and the contact potential difference (see text).

[12] We refer to the net charge of the surface states as "surface states" charge to distinguish it from the total charge, including space charge, in the surface layer, which we call "surface charge."

[13] The case of an excess semi-conductor seems to be easier to visualize than that of a defect semi-conductor. All results derived for one case, of course, apply to the other with obvious changes in signs of the charges.

semi-conductor, the Fermi level is an energy ζ below the conduction band. The space charge region extends for an approximate distance l into the semi-conductor, giving a potential energy rise φ_0 at the surface. An energy φ_s is required to remove an electron from the lowest state of the conduction band near the surface to a point just outside the semi-conductor. The work function, χ_2, depends on φ_s as well as on φ_0:

$$\chi_2=\varphi_s+\varphi_0+\zeta. \qquad (8)$$

The amount and extent of the space charge inside the free surface is determined by the density of surface levels. For zero external field, the positive space charge raises the potential at the surface by an amount just sufficient to give a compensating negative surface charge. The larger the positive space charge region, the larger is φ_0, and the smaller is the negative surface states charge. For some φ_0 the two will be equal in magnitude. This is the equilibrium value.

These relations are indicated in a schematic way in Fig. 4, which shows the variation of σ_i, the total space charge per unit area, of σ_s, the surface states charge, and of the surface charge, $\sigma_s+\sigma_i$, with φ_0. The other curves on the diagram will be referred to later in connection with the discussion of a metal contact. For a uniform space charge in the boundary layer, σ_i is proportional to the square root of φ_0 (see the appendix). The plot of σ_s has a steep slope corresponding to a fairly high density of surface states. It passes through zero (point E in Fig. 4) when

$$\varphi_0=\epsilon_0-\zeta, \qquad (9)$$

because, according to the definition of ϵ_0 the surface states will then be filled up to the level corresponding to zero surface states charge. The point A represents a value of φ_0 for which $\sigma_s+\sigma_i=0$, which corresponds to a neutral surface. It can be seen that if the density of surface states is high, this value of φ_0 will be close to that given by Eq. (9).

For the limiting case of a very high density of surface states, the line σ_s becomes vertical, points A and E coincide, and Eq. (9) must be satisfied; hence, according to Eq. (8):

$$\chi_2=\varphi_s+\epsilon_0. \qquad (10)$$

The work function is determined entirely by the surface, and is independent of the position of the Fermi level in the interior.

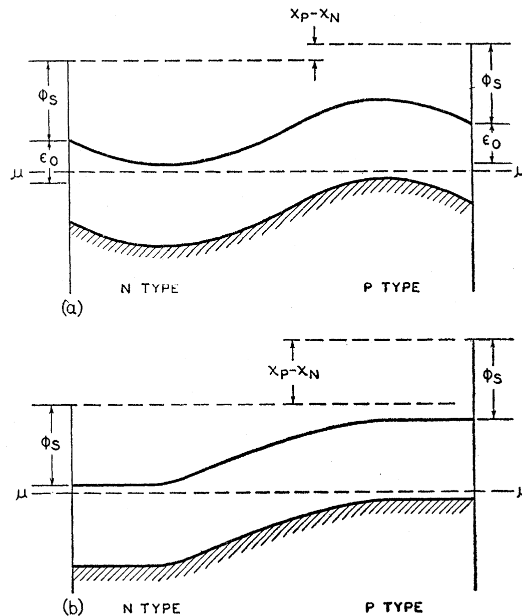

FIG. 5. Energy level diagram showing contact potential difference between two ends of a semi-conductor which changes from N-type to P-type along its length. (a) High density of surface states, small contact potential difference. (b) No surface states, large contact potential difference.

On the other hand, if the density of surface states is small, $\sigma_s+\sigma_i$ is approximately equal to σ_i and the condition for a neutral surface leads to a small value for φ_0. In the limiting case of vanishing surface states charge, $\varphi_0=0$, and the work function is

$$\chi_2=\varphi_s+\zeta, \qquad (11)$$

which of course *does* depend on the position of the Fermi level in the interior of the semi-conductor.

Some semi-conductors can be made either excess or defect, or N- or P-type,[4] depending on the nature and concentration of impurities. The distribution of impurities in a single sample may be such as to make one end N-type and the other end P-type. Silicon and germanium are examples of materials which may behave in this way. For an N-type conductor, ζ is small; for a P-type conductor ζ is almost equal to the energy gap ϵ_g, which is the order of one electron volt for these materials.

If the density of surface states is low, a large difference in work function between N- and

P-type is to be expected. The difference will be almost equal to the energy gap ϵ_g. For a high density of surface states, the difference in work function will be small. Figures 5a and b show schematic energy level diagrams of these two limiting cases. Both show the variation in potential along, and the work functions at the two ends of a sample which changes from N- to P-type from left to right along its length. Figure 5a is that corresponding to a high density of surface states. There is a space charge region at each end. The potential at the left is raised so as to bring the energy level corresponding to a neutral surface close to the Fermi level. The potential at the right is similarly lowered. The difference in work functions, $\chi_P - \chi_N$, is small. Figure 5b is the case corresponding to a small density of surface states. There are no space charge regions at the ends, and the difference in work functions is large.

A difference in work function can be detected experimentally as a difference in contact potential. Meyerhof[7] has in this way measured the differences in work functions of various samples of N- and P-type silicon and has investigated the effect of surface treatment on these differences. He finds that the difference between N- and P-type samples is about 0.25 ev, which is considerably smaller than the energy gap of about 1.1 ev. It is quite possible that there is a high density of surface states, either on the pure material or resulting from surface films or impurities, which is sufficient to account for this discrepancy.

We have so far discussed the density of surface states in a purely qualitative way. It is of interest to make an estimate of the density required to produce an appreciable space charge layer at the free surface of a semi-conductor. The density depends on the density of charged centers, or donors, in the boundary layer. If there are N such centers per unit volume, and if the thickness of the boundary layer (assumed uniform) is l, we have for the total space charge per unit area:[1]

$$\sigma_i = eNl. \tag{12}$$

The potential energy rise at the surface is

$$\varphi_0 = (2\pi/\kappa)e^2Nl^2, \tag{13}$$

where κ is the dielectric constant.

Let the number of surface states in the energy interval $d\epsilon$ be $nd\epsilon/\epsilon_0$. We assume, for simplicity that n is a constant, which is of the order of the total number of surface levels per unit area with energies in the gap. The change in surface states charge density corresponding to a potential energy rise φ_0 at the surface is then

$$\Delta\sigma_s = en\varphi_0/\epsilon_0. \tag{14}$$

The condition for an appreciable space charge layer is that φ_0 be of the order of magnitude of ϵ_0 when $\Delta\sigma_s = \sigma_i$. This requires that n be the order of Nl. Taking, for example, $\kappa \sim 15$, and $N \sim 10^{17}$, Eq. (13) gives $l \sim 10^{-5}$, so that

$$n \sim Nl \sim 10^{12} \text{ states/cm}^2. \tag{15}$$

This density corresponds to about one surface state per thousand surface atoms. Other things being fixed, the limiting density of surface states is proportional to the square root of the density of charged centers in the boundary layer.

METAL SEMI-CONDUCTOR CONTACT

As a metal surface approaches that of a semi-conductor, the electrostatic field in the gap increases. There is a surface charge σ_M on the metal surface, and a charge of equal magnitude and opposite sign divided between a surface charge, σ_s, and a space charge, σ_i (the latter has been defined as the total space charge per unit area of surface). Thus

$$\sigma_M = -(\sigma_s + \sigma_i). \tag{16}$$

In the case of the metal, the surface charge causes only a very slight change in work function. However, the charge on the semi-conductor may cause appreciable changes in its work function, so that its value χ_2 will differ appreciably from the zero field value χ_2^0. These changes can be analyzed with the aid of Figs. 3 and 4.

According to Fig. 3, the field in the gap between metal and semi-conductor is $(\chi_1 - \chi_2)/ea$. This is produced by the surface charge which gives a field $4\pi(\sigma_s + \sigma_i)$. Equation (8) expresses χ_2 as a function of φ_0. Equating the two expressions for the field leads to:

$$\frac{\chi_1 - \zeta - \varphi_s - \varphi_0}{4\pi ea} = \sigma_s(\varphi_0) + \sigma_i(\varphi_0). \tag{17}$$

The left-hand side of the equation is plotted in Fig. 4 as the line "L," and is a straight line of negative slope. The solution to (17) is represented by the point B, which is the intersection of "L" with the line representing $\sigma_s+\sigma_i$. The line "L" crosses the axis at the point D, which corresponds to $\chi_2=\chi_1$, or

$$\varphi_0=\chi_1-\varphi_s-\zeta. \tag{18}$$

It can be seen that if the density of surface levels is high, so that σ_s and $\sigma_s+\sigma_i$ have steep slopes, the point B will lie close to the point A which gives the value of φ_0 for the free surface. In this case,

$$\varphi_0\sim\epsilon_0-\zeta, \tag{19}$$

and is practically independent of the work function of the metal. This is a possible explanation of the results of Meyerhof, who found that the value of φ_0 for metal points on silicon does not depend very much on the metal used.

In the limiting case of vanishing surface states charge density, $\sigma_s=0$, and the solution of (17) is given by the point C of Fig. 4. The value of φ_0 is then close to that given by the intersection of the line L with the horizontal axis (point D). In this case,

$$\varphi_0\sim\chi_1-\varphi_s-\zeta, \tag{20}$$

and the usual picture of the contact between metal and semi-conductor applies.

As the metal is brought closer to the semiconductor, the slope of the line "L" increases with decreasing "a" and one might expect a transition from a solution near A to one near D with a resulting value of φ_0 given by Eq. (20). However, even if there is only one surface state per hundred surface atoms, the slope of the line σ_s is so great that the solution still lies near point A even when a is reduced to atomic distances, say 3A. This may be verified by making a rough calculation based on Fig. 4. For the value of φ_0 to be largely determined by the semiconductor, the value of φ_0 for the solution B must lie closer to A then to D. The necessary and sufficient condition for this is that the slope of the line σ_s be much greater, in absolute value, than the slope of "L." We have

$$\begin{aligned}\text{slope of } \sigma_s&=m_s=en/\epsilon_0,\\ \text{slope of } L&=-m_L=-1/4\pi ea.\end{aligned} \tag{21}$$

The condition that m_s/m_L be large is

$$m_s/m_L=4\pi e^2an/\epsilon_0\gg1, \tag{22}$$

or

$$n\gg\epsilon_0/4\pi e^2a. \tag{23}$$

Setting for example, $\epsilon_0\sim10^{-12}$ erg and $a\sim3\times10^{-8}$ cm, this condition requires that

$$n\gg10^{13}. \tag{24}$$

Thus if there is appreciably more than about one surface state per hundred surface atoms, the metal work function will have little effect on φ_0. This is about an order of magnitude larger than the surface density required for the existence of an appreciable boundary layer at the free surface.[14]

The theory is worked out in detail for the special case of a uniform Schottky exhaustion layer[1] in the appendix.

CONCLUSIONS

If the density of surface levels with energies in the "forbidden" band is sufficiently high ($>\sim10^{12}/\text{cm}^2$), there will be a double layer at the free surface of a semi-conductor formed from a surface states charge and a space charge of opposite sign. The space charge region is similar to that which exists at a rectifying contact. This double layer tends to make the work function independent of the height of the Fermi level in the interior, and so independent of the impurity content.

The total strength of the double layer at a rectifying junction between a metal and semiconductor is fixed by the difference in chemical potentials, and so depends on the body properties of the metal and semi-conductor, and is independent of the work functions of the surfaces before they are brought into contact. The double layer consists of the following parts:

[14] S. Benzer, Phys. Rev. **71**, 141 (1947), has recently reported on the current-voltage characteristic observed when contact is made between two pieces of the same (homogeneous) germanium crystal. He states that "the characteristic observed for both polarities is more the order of the back resistance observed when either piece of the crystal is contacted with a metal; in both directions the negative resistance at high voltage appears, which is typical of the back characteristics of metal-Ge contacts using these alloys." The presence of a space charge layer at the surface of each piece which is little modified by contact would result in a characteristic similar to that of two rectifiers in series opposition. Such boundary layers are to be expected if the density of surface states is the order of that given by Eq. (24) and if the contact is not too intimate. Benzer's results are thus indirect evidence for surface states on germanium crystals.

(1) A double layer of atomic dimensions at the metal surface.

(2) A double layer of atomic dimensions at the semi-conductor surface.

(3) A double layer formed from surface charges on the metal and semi-conductor, both of atomic dimensions.

(4) A double layer formed from a surface charge of atomic dimensions and a space charge extending to a depth of 10^{-6} to 10^{-4} cm into the semi-conductor.

The strengths of the double layers may be estimated in different cases as follows:

(a) If the density of surface levels is sufficiently high ($>\sim 10^{13}/\text{cm}^2$) the double layer (4) will be independent of the metal, and will be the same as that for the free surface of the semi-conductor. The rectification properties will then be largely independent of the work function of the metal. The difference in contact potentials is compensated by (3).

(b) If the density of surface levels is small ($<\sim 10^{13}/\text{cm}^2$), the double layer (3) will be small, and (4) will be determined approximately by the difference in work functions.

(c) If the contact between the metal and semi-conductor is very intimate, it may not be possible to distinguish between the double layers (1), (2), and (3). The metal will tend to broaden the surface levels, but if the broadening is small compared with the energy gap, conclusion (a) will still be valid.

(d) If the broadening of the surface levels by the metal is large, no conclusions about the space charge can be drawn from measurements of contact potential differences.

It is believed that all of these cases may be realized.

The author is indebted to various members of the technical staff of the Bell Telephone Laboratories, particularly to W. Shockley, C. Herring, and W. H. Brattain, for the benefit of numerous discussions concerning the subject matter of this paper, and for helpful comments on the preparation of the manuscript.

APPENDIX

Theory for Schottky Exhaustion Layer

The detailed calculation of φ_0 requires a knowledge of the distribution of space charge in the semi-conductor and its dependence on φ_0. The dependence of surface charge on φ_0 is also required. We will carry through the calculation explicitly only for the case of a Schottky exhaustion layer[1] with a uniform density of charge. This case is particularly simple and brings out the essential features of the problem. For simplicity we also assume that the surface states are uniformly distributed in energy.

Let eN be the positive charge density in the barrier region, assumed constant. Let l be the thickness of the charged layer. Let $nd\epsilon/\epsilon_0$ be the number of surface levels per unit area with energies in the range $d\epsilon$. We assume that n is a constant, independent of energy. For other notation see Fig. 3.

We have the following relations:

(a) The total space charge per unit area is:

$$\sigma_i = Nel. \tag{12}$$

(b) The potential energy at the surface of the semi-conductor is

$$\varphi_0 = \frac{2\pi e^2 N}{\kappa} l^2, \tag{13}$$

where κ is the dielectric constant.

(c) The surface states charge is

$$\sigma_s = -en(\epsilon_0 - \varphi_0 - \zeta)/\epsilon_0. \tag{15}$$

(d) The total surface charge is

$$\sigma_i + \sigma_s = Nel - en(\epsilon_0 - \varphi_0 - \zeta)/\epsilon_0. \tag{25}$$

(e) The work function of the semi-conductor is:

$$\chi_2 = \varphi_s + \zeta + \frac{2\pi e^2 N}{\kappa} l^2. \tag{26}$$

Equating $\sigma_i + \sigma_s$ with $(\chi_1 - \chi_2)/4\pi e a$, we get Eq. (17) from which φ_0 or the thickness of the barrier layer, l, can be determined. With the notation:

$$l_0 = \frac{n}{N}(1 - \zeta/\epsilon_0), \tag{27a}$$

$$l_1 = \kappa\epsilon_0/2\pi e^2 n, \tag{27b}$$

$$l_2^2 = \kappa(\chi_1 - \varphi_s - \zeta)/2\pi e^2 N, \tag{27c}$$

Eq. (17) reduces to:

$$l^2/l_1 + l - l_0 = (l_2^2 - l^2)/2\kappa a, \tag{28}$$

which is a simple quadratic equation for l.

Whether the semi-conductor or the metal predominates in determining l, and thus the barrier height, depends on the relative magnitudes of the terms on the left and right sides of Eq. (28). The terms on the right side will be negligible if

$$l_1 \ll 2\kappa a \quad (29)$$

and

$$l_0 \gg l_2^2/2\kappa a. \quad (30)$$

The first condition requires that

$$n \gg \epsilon_0/2\pi e^2 a, \quad (31)$$

and the second that

$$n \gg (\epsilon_0/4\pi e^2 a)(\chi_1 - \varphi_s - \zeta)/(\epsilon_0 - \zeta). \quad (31a)$$

These conditions are essentially equivalent to that given by Eq. (23), and lead to the conclusion that $n \gg 10^{13}$ in order that the metal have little influence on the space charge region. In this case of high density of surface states, the equation for the layer thickness reduces to that for the free surface of the semi-conductor:

$$l^2/l_1 + l - l_0 = 0. \quad (32)$$

Equation (32) may, of course, be used to estimate the thickness of the space change region at the free surface of the semi-conductor, regardless of the density of surface states. The limiting case of high density corresponds to

$$l_0 \gg l_1, \quad (33a)$$

or

$$n^2 \gg N\kappa\epsilon_0^2/2\pi e^2(\epsilon_0 - \zeta). \quad (33b)$$

This is essentially the requirement, stated above Eq. (15), that n be large compared with Nl. The approximate solution of Eq. (32) is then

$$l^2 \approx l_0 l_1.$$

The product $l_0 l_1$ is independent of n, and is just the square of the thickness of the Schottky layer for a barrier height

$$\varphi_0 = \epsilon_0 - \zeta. \quad (34)$$

This is the condition that the Fermi level cross the surface near the energy corresponding to zero surface states charge.

In the limiting case of a vanishingly small density of surface states, the thickness of the barrier layer is determined by setting the right-hand side of Eq. (28) to zero. This gives

$$\varphi_0 = \chi_1 - \varphi_s - \zeta. \quad (35)$$

The height of the barrier is equal to the difference in work functions.

Equation (28) may be used for intermediate cases.

PHYSICAL REVIEW VOLUME 71, NUMBER 10 MAY 15, 1947

Contact Potential Difference in Silicon Crystal Rectifiers* **

WALTER E. MEYERHOF***
University of Pennsylvania, Philadelphia, Pennsylvania

(Received February 16, 1947)

The rectifying portion of a crystal rectifier is the contact between a small point made of metal such as tungsten, and a semiconductor such as silicon or germanium containing suitable impurities. The potential energy of a conduction electron near the contact determines the rectifying action of the crystal rectifier. The most important feature of this potential energy, as far as the present paper is concerned, is the height of the potential barrier, which the electrons have to overcome when they pass from the metal to the semiconductor or vice versa. The height of the barrier is called here contact potential difference (c.p.d.), because theoretically it is equal to the difference in the work functions of the substances in contact. The c.p.d. has been measured using both n- and p-type silicon and different metallic contacts. (The c.p.d. can be obtained from the variation of the contact resistance with temperature.) The work function differences (w.f.d.) between the same substances were obtained independently by a parallel plate condenser method (Kelvin method). The results showed no correlation between the c.p.d. and the w.f.d. The c.p.d. is practically independent of the kind of metal used and also of the structure of the silicon surface. These results are in contradiction to the present theoretical model of the silicon crystal rectifier.

* Condensation of a dissertation in Physics presented to the Graduate School of the University of Pennsylvania in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Dissertation published with limited distribution under "University of Pennsylvania, BuShips Contract NObs-34144, Technical Report No. 5, August 10, 1946." Copies available from University of Pennsylvania, Department of Physics, Philadelphia 4, Pennsylvania.

** This work was done under Contracts OEMsr-388 and NObs-34144 between the Trustees of the University of Pennsylvania and the Office of Scientific Research and Development and the Navy Department, Bureau of Ships, respectively, which assume no responsibility for the accuracy of the statements contained herein.

*** Now at the University of Illinois.