

Theory of Surface States

VOLKER HEINE*

Bell Telephone Laboratories, Murray Hill, New Jersey

(Received 22 January 1965)

The properties of metal-to-semiconductor junctions and of free semiconductor surfaces are usually explained on the basis of surface states. The theory of the metal contacts is discussed critically, because strictly speaking localized surface states cannot exist in such junctions. However, it is shown that virtual or resonance surface states can exist which behave for practical purposes in the same way. They are really the tails of the metal wave functions rather than separate states. In the past, the length of this tail has often been ignored. Some estimates of its length are made and its consequences pointed out. A semiquantitative discussion is given of various recent data, including the effect of an oxide layer on barrier height, the variation of barrier height with the metal, the work function of a free surface at high doping, and the effect of a cesium layer on the work function.

I. INTRODUCTION

THIS paper is in the nature of a commentary and elaboration of the theory of metal-to-semiconductor junctions and of free semiconductor surfaces, with a discussion of some of the more recent data. As is well known, when a contact is made between a metal and a semiconductor, there is a potential barrier Φ . The Fermi level E_F at the surface of a semiconductor lies in the energy gap at a point,¹

$$\xi = E_g - \Phi \quad (1)$$

above the top of the valence band [Fig. 2(a)] where E_g is the gap width. The remarkable thing is that ξ is constant to within about 0.2 eV independent of:

- the degree of doping^{2,3} which swings the bands in the bulk material by something of the order of 1 eV;
- the metal that is used,⁴⁻⁸ some differing from the semiconductor by several volts in the work function ϕ ;
- the crystallographic orientation of the surface^{8,9};
- whether there is a monolayer or so of oxygen or oxide between the metal and semiconductor^{3,8};
- any bias voltage applied across the junction^{4,8};
- whether one dispenses with the metal completely and has a free surface, at least in the case of silicon, again either atomically clean or with an oxide layer.^{10,8} Furthermore, ξ is roughly a constant fraction ~ 0.3 of

* Permanent address: Cavendish Laboratory, Free School Lane, Cambridge, England.

¹ J. Bardeen, Phys. Rev. **71**, 717 (1947).

² C. R. Crowell, S. M. Sze, and W. G. Spitzer, Appl. Phys. Letters **4**, 91 (1964).

³ R. J. Archer and M. M. Atalla, Ann. N. Y. Acad. Sci. **101**, 697 (1963).

⁴ C. R. Crowell, J. C. Sarace, and S. M. Sze, Trans. AIME (to be published).

⁵ C. R. Crowell, W. G. Spitzer, L. E. Howarth, and E. E. La Bate, Phys. Rev. **127**, 2006 (1962).

⁶ C. A. Mead and W. G. Spitzer, Phys. Rev. **134**, A713 (1964); Phys. Rev. Letters **10**, 471 (1963).

⁷ W. G. Spitzer and C. A. Mead, J. Appl. Phys. **34**, 3061 (1963).

⁸ C. R. Crowell (private communication and to be published); D. Kahng, Bell System Tech. J. **43**, 215 (1964); S. M. Sze, C. R. Crowell, and D. Kahng, J. Appl. Phys. **35**, 2534 (1964).

⁹ F. G. Allen and A. B. Fowler, J. Phys. Chem. Solids **3**, 107 (1957). See also F. G. Allen and G. W. Gobeli, J. Appl. Phys. **35**, 597 (1964); Y. Takeishi and H. D. Hagstrum (to be published).

¹⁰ F. G. Allen and G. W. Gobeli, Phys. Rev. **127**, 150 (1962).

the energy gap in a whole range of group IV and III-V semiconductors.^{6,7} Of course there are exceptions to all these statements, but by and large this is the story.

All these facts receive a natural explanation in terms of Bardeen's theory,¹ namely, that there are surface states on the semiconductor in the band gap. A density of even a fraction of a surface state per surface atom per eV is quite sufficient to absorb any required extra charge into the band of surface states without the Fermi level in this band moving very much. Thus E_F at the surface of the semiconductor is effectively pegged to some fixed value of ξ . A few of the data are shown in Fig. 1.

At the time Bardeen's theory was first advanced, the semiconductor-metal contacts were made by pressing the metal against the semiconductor, each coated with an oxide film. Now, however, many of the junctions are made by cleaving the semiconductor in a stream of the metal vapor used to deposit the metal. The effective partial pressure of the condensing metal is ten or more times the residual partial pressure of gas,^{7,8} so that there is intimate atomic contact between the metal and semiconductor without any intervening oxygen or other impurity layer. Under such conditions it appears at first

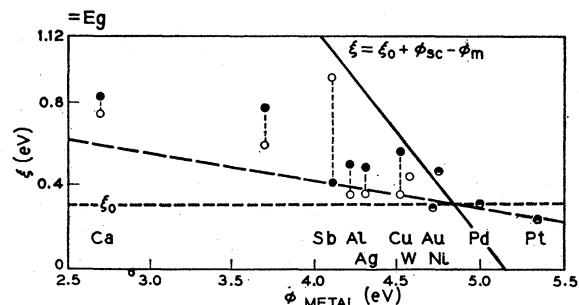


FIG. 1. Position of Fermi level at the surface (ξ) above the valence-band edge for metal-to-silicon contacts. ξ_0 is the value for a free-silicon surface. ξ is given by $E_g - \Phi$, where Φ is the measured barrier height. White circles refer to intimate contact between metal and semiconductor, black circles to the case of a thin intervening oxide layer. Circles half-black are used where the two values coincide within experimental error. The data points second from the left are for magnesium. (After Ref. 8. Data from Refs. 2-8.)

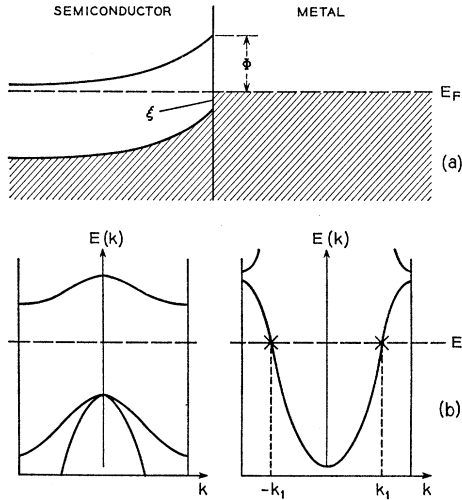


FIG. 2. (a) Definition of ϕ and ξ at semiconductor-metal junction. (b) An energy E of a surface state in the band gap of the semiconductor corresponds to two propagating Bloch functions $k_1, -k_1$ in the metal.

sight unreasonable to talk about surface states at all. For an energy E below E_F in the gap of the semiconductor, the solutions of the Schrödinger equation will decay exponentially in the semiconductor but propagate as Bloch states on the metal side of the junction to form the ordinary volume states of the metal.^{11,12} This follows from simple considerations of matching the wave functions at the boundary as shown in Fig. 2. Let us assume the x axis perpendicular to the surface. For some value of $\mathbf{k} = \mathbf{k}_1$ parallel to the surface, e.g., $k_y = k_z = 0$, we have the bands shown in Fig. 2(b). At energy E , the exponential solution in the semiconductor can always be joined onto the two Bloch states with wave vector $k_1 = k_1, -k_1$ in the metal, the existence of the two k 's ensuring that both ψ and its derivative can be matched at the boundary. Thus for energies in the semiconductor band gap the volume states of the metal all have tails in the semiconductor; there are actually no surface states.

Our first task is therefore to discuss in Sec. II how the Bardeen explanation in terms of surface states may be rescued. Although the exact form of the band of surface states is not known, there are some semiquantitative statements one can make. For example, the form of the exponential tail in the semiconductor is determined by the ordinary volume band structure. In Sec. III we make some estimates of its length and discuss its relation to some of the experimental data.

II. MATCHING THE WAVE FUNCTION

Let us first assume that k_{11} , the component of \mathbf{k} parallel to the boundary plane, is conserved as an elec-

¹¹ V. Heine, *Surface Sci.* **2**, 1 (1964).

¹² V. Heine, *Proc. Phys. Soc. (London)* **81**, 300 (1962); P. J. Price, *Report of the International Conference on the Physics of Semiconductors* (Institute of Physics and the Physical Society, London, 1962), p. 99.

tron crosses the junction. In reality there is sufficient disorder at the boundary for diffuse scattering to be probably a better approximation than the specular assumption made here. But the effect of scattering is to scramble states of the same energy E , leaving invariant any total property such as the density of states or the charge distribution $\psi^*\psi$ summed over all states in some energy interval dE . Thus the device of considering states of definite k_{11} leads to correct results for total or average properties.

As a first step we treat the valence states of the semiconductor and the conduction electrons of the metal as free electron gases, which is a good approximation in view of the small band gap.^{11,13,14} Because of the higher density of four electrons per atom in the semiconductor than the metal, the potential and Fermi surfaces in the two regions are as shown in Fig. 3. The important point is that in this simple model all semiconductor states with k_{11} greater than $A'A''$ do not penetrate into the metal. Their k_{11} is sufficiently large that their perpendicular kinetic energy

$$E_{\perp} = E - \hbar^2 k_{11}^2 / (2m) \quad (2)$$

is negative in the metal, i.e., they decay exponentially on the metal side of the boundary. The value of the Fermi momentum k_F for silicon is compared with that of some typical metals in Table I. We have not counted the d electrons in the case of transition metals because they have such different wave functions with small value at the cell edge that one does not expect them to match at all well onto the silicon wave functions. Let us concentrate for the present on the top part of the table. We note that a large fraction f of the high-energy states of the silicon, given by

$$f = 1 - k_{F,m} / k_{F,sc} \quad (3)$$

do not propagate into the metal. The same applies to any surface states in the band gap formed when the

TABLE I. Fermi momenta k_F of metals (m) compared with silicon (sc).

	$k_{F,m} / k_{F,sc}$
sodium	0.50
palladium, ^a platinum ^a	0.55
calcium, nickel ^a	0.61
silver, gold	0.66
magnesium, copper	0.75
antimony	0.93
aluminum	0.97

^a Assuming half an s - p electron per atom.

¹³ We append a selection of general articles which contain more detailed documentation. V. Heine, in *Proceedings of the Ninth Conference on Low Temperature Physics, 1964* (to be published); J. M. Ziman, *Advan. Phys.* **13**, 89 (1964); L. Kleinman and J. C. Phillips, *Phys. Rev.* **125**, 819 (1962); M. H. Cohen and V. Heine, *ibid.* **122**, 1821 (1961).

¹⁴ D. Brust, *Phys. Rev.* **134**, A1337 (1964).

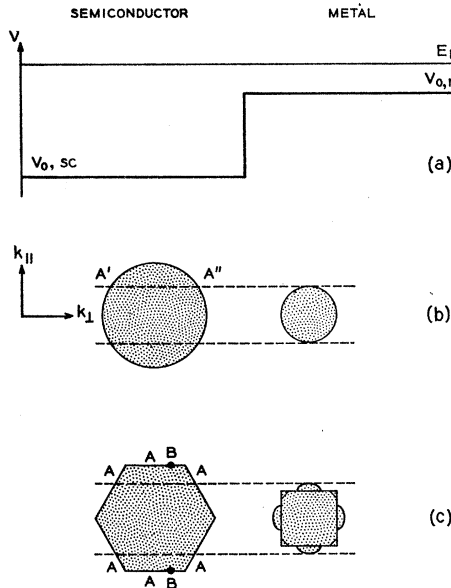


FIG. 3. (a) The potential, (b) the Fermi spheres, and (c) the electron distributions in the Jones zones, in the nearly free-electron model.

Fermi distribution is enclosed in the Jones zone. (We reserve the word Brillouin zone for the reduced zone, and use Jones zone for any zone constructed in extended \mathbf{k} space out of prominent reflection planes.) A surface state at B , if one exists, is made out of plane waves at B and B' , with a small imaginary component in k_1 which does not affect the behavior parallel to the boundary. Thus along the regions AAA of the zone [Fig. 3(c)] genuinely localized surface states are possible in this simple approximation. The remainder would propagate into the metal as discussed in connection with Fig. 2.

This argument can be taken a step further by considering the actual band structure of the semiconductor. The crystal face most commonly chosen for making junctions is the (111) cleavage plane. The surface state is related to the band edges of a section $E(k_1)$ of the volume band structure for fixed \mathbf{k}_{11} , these edges in almost all cases occurring at $k_1=0$. The wave function of the surface state is therefore a mixture of the volume states at $\mathbf{k}=\mathbf{k}_{11}$, $k_1=0$ with of course the slow exponential decay into the semiconductor superimposed. The plane of the vector \mathbf{k}_{11} is defined by the three directions $(10\bar{1})$, $(1\bar{1}0)$, $(01\bar{1})$ at 120° , and a typical \mathbf{k}_{11} about halfway between the center and the edge of the Brillouin zone is $(2\pi/a)(\frac{1}{3}, 0, -\frac{1}{3})$. The Bloch states at the gap are linear combinations of plane waves, in the plane-wave pseudopotential representation of band structures,^{11,13} with \mathbf{k} differing from $(2\pi/a)(\frac{1}{3}, 0, -\frac{1}{3})$ by various reciprocal lattice vectors. The lowest ones relevant at the gap are

$$\frac{2\pi}{a}(-5/3, 0, -\frac{1}{3}) \quad \text{and} \quad \frac{2\pi}{a}(\frac{1}{3}, 0, 5/3). \quad (4)$$

These have $k_{11}=0.79k_F$. In addition, the wave functions contain an admixture of

$$\frac{2\pi}{a}(-\frac{2}{3}, \pm 1, \frac{2}{3}), \quad \frac{2\pi}{a}(\frac{4}{3}, \pm 1, \frac{2}{3}), \quad \frac{2\pi}{a}(-\frac{2}{3}, \pm 1, -\frac{4}{3}), \quad (5)$$

with $k_{11}/k_F=0.79, 0.30, 1.08$ in pairs. Moreover, they have quite a high k_1 , varying between $0.37k_F$ and $1.10k_F$. Thus for all the top group of metals in Table I, we have to a very good approximation surface states in the semiconductor which decay exponentially on the metal side.

Strictly speaking, however, there is some coupling to the states in the metal as envisaged in the discussion in Sec. I. It comes through having some admixture of the wave with $k_{11}=0.30k_F$ in (5) and to a lesser extent through the admixture of higher waves in the metal by Bragg reflections. One ought to discuss the quantity

$$\rho(E) = \int \psi^* \psi n(E) dv, \quad (6)$$

where $n(E)$ is the density of states in the metal and the volume integration is over all the semiconductor. This is the total charge residing in the semiconductor from the tails of the metal wave functions per unit range dE . For a true surface state of energy E_s , it is just a delta function at E_s containing one electron. In our situation, since the coupling between semiconductor and metal is weak, one has the conventional Breit-Wigner situation,¹⁵ and $\rho(E)$ goes through a resonance peak around E_s , the total again being one electron. Following the customary nomenclature, we shall term these "resonance" or "virtual" surface states.

We return now to the opposite case such as aluminum in the bottom part of Table I. Here k_F in the metal is comparable with that in the semiconductor and greater than k_{11} for a typical wave function in the band gap, e.g., the waves (4). For an energy in the gap, the E_1 [Eq. (2)] in the metal is positive and we can only talk about the density of charge $\rho(E)$ in the semiconductor tails of the metal wave functions. For a particular \mathbf{k}_{11} the question of matching the tails to the metal volume states becomes a one-dimensional problem of the idealized form shown in Fig. 4. We choose as our model a weak one-dimensional potential with lattice constant b , and consider the first band gap in a nearly free-electron band structure. The algebra of matching wave functions in the metal of type $\sin(kx-\beta)$ to those in the semiconductor which have the form

$$e^{ax} \sin(\pi x/b - \gamma) \quad (7)$$

for all energies in the band gap, has been performed. The calculation involves three steps: (a) determining $q(E)$ and $\gamma(E)$ in the gap using a two-plane-wave

¹⁵ L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon Press, Inc., London, 1958), p. 440.

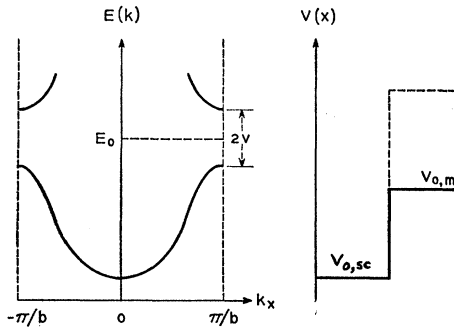


FIG. 4. Band structure of semiconductor, and potential $V(x)$ in the semiconductor (sc) and metal (m), shown on the same energy scale. The value of $V_{0,m}$ shown corresponds to positive E_1 in the metal, the dashed value to negative E_1 . The latter corresponds roughly to a free surface.

approximation, (b) matching the wave functions at the boundary, and (c) integrating $|\psi|^2$ over the tail in the semiconductor to obtain the total charge contained in it. The following approximations were made:

$v = \text{half the band gap,}$

$$\ll E_0 - V_{0,m} = \hbar^2 k_m^2 / 2m \approx E_1, \text{ in the metal,} \quad (8)$$

$\ll E_0 - V_{0,sc} = \hbar^2 \pi^2 / (2b^2 m) \approx E_1$, in the semiconductor, where E_0 , $V_{0,m}$, and $V_{0,sc}$ are defined in Figs. 3 and 4. v is chosen positive corresponding to "crossed bands" in the sense of Shockley,¹⁶ and the boundary between metal and semiconductor was taken at $x = \frac{1}{2}b$ halfway between the "atoms." The result is that $\rho(E)$ is proportional to

$$\frac{1}{(1 - \epsilon^2)^{1/2}} \frac{1}{(\alpha^2 + 1) + (\alpha^2 - 1)\epsilon}, \quad (9)$$

where

$$\epsilon = (E - E_0)/v, \quad \alpha^2 = (E_0 - V_{0,sc}) / (E_0 - V_{0,m}). \quad (10)$$

It is plotted in Fig. 5, and the total charge for energy integrated over the gap is exactly one electron. Although (9) has a mild infinity at each band edge, this is a

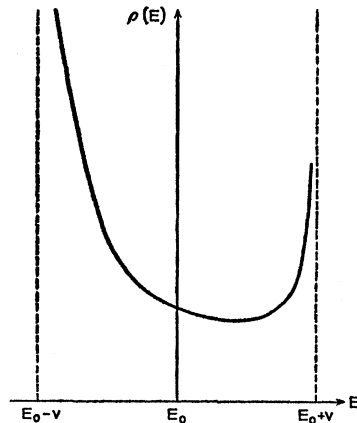


FIG. 5. Density of charge per unit energy in the semiconductor tails of the metal wave functions.

general feature of one-dimensional band structures which disappears on integrating over k_{11} and the integrated charge involved is intrinsically small. It is the second factor in (9) which controls where the bulk of the contribution to the total charge comes from. In our case it is a maximum at the bottom of the band gap. It is precisely here, at an energy

$$E \text{ (surface state)} = E_0 - v + 2v/\alpha^2 \quad (11)$$

just above the bottom of the gap, that we get a surface state if we raise $V_{0,m}$ as high above E_0 (shown dotted in Fig. 4) as we have so far taken it below E_0 . Incidentally, if we take the boundary at the point $x = \frac{1}{4}b$, which perhaps corresponds a bit more closely to the spacing of atomic layers in the (111) direction because of the two atoms per unit cell, both the maximum of $\rho(E)$ and the surface state come nearly at the center of the gap: It is difficult to say whether the model is too crude for this to have any physical significance. We conclude therefore that the tails of the wave functions behave rather like a localized surface state, both in the total charge and approximately in its distribution in energy in the gap. That is, the resonance has become considerably broader but is still recognizable, as seen near the bottom of the gap in Fig. 5.

We have therefore demonstrated how it can be possible for the system to be relatively stable against the application of different metals with or without an oxide layer, i.e., against different boundary conditions. Since it is the most important property, it is perhaps worth commenting how it may be looked at from various points of view. Firstly, in the limit of small band gaps, q in (7) becomes small and the length of the tail of the wave function long compared with the Fermi wavelength $\lambda/(2\pi)$. Under such conditions we might expect a theorem to hold for $\rho(E)$ analogous to that for blackbody radiation, namely that the energy density per unit frequency range is constant except within λ of the walls, independent of boundary conditions. A second factor is the Shockley theorem¹⁶ and the rapid variation of ψ'/ψ with energy on the semiconductor side so that the surface-state energy is insensitive¹¹ to relatively small changes in boundary condition ψ'/ψ on the other side. Thirdly, a comment about "dangling bonds." In accordance with the Shockley theorem, the existence of surface states has everything to do with the formation of covalent bonds in the sense of the crossing of bands, i.e., the fact that the potential tends to heap up charge between the atoms rather than at the atoms. But the broken orbital on the actual surface atom is not very important, at least so the experimental evidence indicates. A model which depended sensitively on what happens to the dangling orbital would not have the insensitive properties mentioned in (a) to (f) in Sec. I, particularly as regards surface orientation. Similar evidence comes from the electrical behavior of dislocations which does not correlate at all with the presence or

¹⁶ W. Shockley, Phys. Rev. 56, 317 (1939).

absence of dangling bonds and can be explained quite well without them.^{17,18} This also makes sense from a theoretical point of view. A wave function cannot be localized more than a Wannier function, and this is spread over several atomic distances in a semiconductor because of the small band gaps; i.e., q in (7) is small. This means that of any localized or surface state, only a small probability $\psi^*\psi$ is concentrated on any one atom or layer of atoms. Consequently that one atom or layer cannot influence the energy very much as we see in the shallow donor and acceptor levels. Of course deep traps do exist too, but they need enormous perturbations like doubly charged zinc or negative copper,¹⁹ and the hybridization energy of a broken bond is not like that. Dangling bonds might tend to form shallow states, but they would be near the valence- and conduction-band edges for each particular $k_{||}$, which means almost all of them would be well hidden in the total valence and conduction bands. Finally surfaces tend to reconstruct atomically so as to join up into conjugated systems, thus eliminating the dangling bonds.

III. DISCUSSION OF THE DATA

We turn first to Fig. 1 and the effect of a thin oxide layer between the semiconductor and metal. If $k_{F,m}$ is small so that the resonance surface state is sharp and only slightly coupled to the metal, the intervention of the oxide layer would just reduce the coupling further, leaving the energy of the resonance and thus ξ unchanged. If $k_{F,m}$ is large and the coupling therefore strong, the tails of the wave functions will suffer a reduction in amplitude in the oxide, by a larger factor than in the same thickness of semiconductor because the band gap in the oxide is larger. Thus the amplitude of the tails in the semiconductor is smaller and E_F has to rise in the band gap to achieve the same total charge in the tails needed for electrical neutrality. The magnitude of the increase in ξ , i.e., the amount that the black circles lie above the white ones in Fig. 1, correlates well with the position of the metal in Table I in accordance with our explanation. Only antimony in Fig. 1 behaves completely anomalously, and we note it behaves normally when there is an intervening oxide layer, but not when in intimate contact with the semiconductor, a point to which we return later.

Having discussed the differences in Fig. 1 between with and without an oxide layer, we turn now to the absolute value of ξ for the intimate contacts (white circles in Fig. 1) and its deviation from ξ_0 , the value at a free surface. The most elementary explanation would run as follows: As we move to the left in Fig. 1, the dipole

$$D = \varphi_{sc} - \varphi_m \quad (12)$$

¹⁷ R. K. Mueller, J. Appl. Phys. **32**, 640 (1961); and (private communication).

¹⁸ V. Heine (to be published).

¹⁹ N. B. Hannay, *Semiconductors* (Reinhold Publishing Corporation, New York, 1959), p. 342.

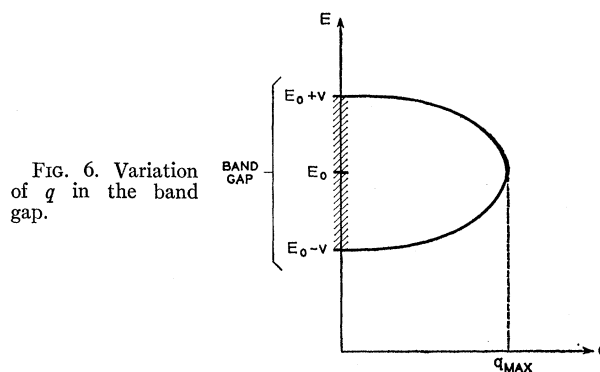


FIG. 6. Variation of q in the band gap.

that has to exist between the surface states and the metal increases, i.e., more electrons flow into the surface states to set up the dipole, and E_F in the surface states has to rise. We have

$$D = 4\pi\sigma t/\epsilon, \quad (13)$$

where $t \approx 1/q$ is the mean separation between the negative charge in the surface states and the positive charge on the metal, ϵ is the dielectric constant, and σ is the charge density per unit area. Further

$$\delta E_F \equiv \xi - \xi_0 = \sigma/\rho(E), \quad (14)$$

where $\rho(E)$ is also expressed per unit area. Thus

$$\xi = \xi_0 + \left[\frac{1}{4\pi\rho(E)(t/\epsilon)} \right] (\varphi_{sc} - \varphi_m). \quad (15)$$

Incidentally, in (12) and (15) it is not quite the real work functions φ_{sc} and φ_m that we want, but the volume contributions to them without the surface dipoles on the free surfaces. It is this which represents the difference in binding of an electron in the two materials. However, since the surface dipoles are small and anyway unknown, we use the measured φ 's.

In order to calculate ξ from (15) we must estimate t/ϵ and $\rho(E)$. We start with t/ϵ , discussing both theoretical and experimental estimates. Theoretically q in (7) varies in the band gap as shown in Fig. 6, with q_{max} being given approximately by^{20,21}

$$q_{max} \approx k_{at\ gap}(v/2E_0), \quad (16)$$

where E_0 is the energy at the gap $\hbar^2 k_{gap}^2/(2m)$ measured from the bottom of the band. Taking again the representative point defined by Eqs. (4) and (5), and a mean vertical band gap $2v$ of $4\frac{1}{2}$ eV to the first two conduction bands,¹⁴ we obtain²²

$$1/q_{max} = 6.0 \text{ \AA}. \quad (17)$$

A calculation following Kane²³ for the center of the Brillouin zone gives 5.5 Å. Since $|\psi|^2$ in the surface

²⁰ W. Kohn, Phys. Rev. **115**, 809 (1959).

²¹ E. T. Goodwin, Proc. Cambridge Phil. Soc. **35**, 205 (1939).

²² The estimate below Eq. (5) of Ref. 11 was somewhat too large.

²³ E. O. Kane, J. Appl. Phys. **32**, 83 (1961).

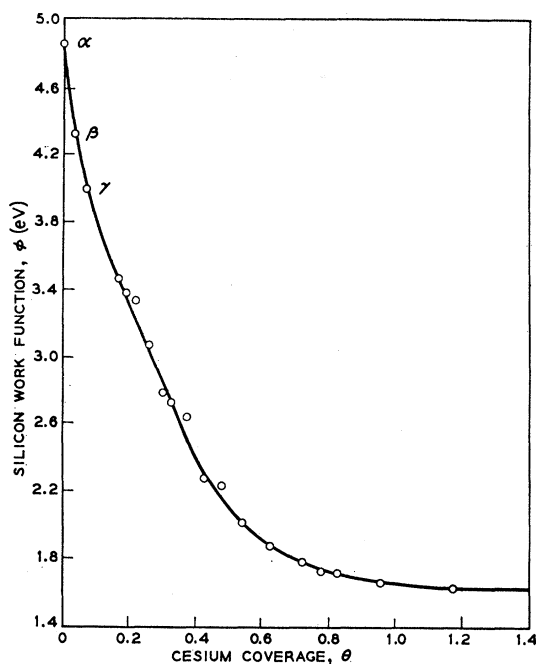


FIG. 7. Variation of silicon work function ϕ with cesium coverage. (After Ref. 24.)

state varies as $\exp(-2qx)$, the center of gravity of the charge is a distance $1/(2q)$ from the surface. We note that our estimates give a *minimum* length of the tail, applicable when the surface state has an energy at the center of the direct gap.

Experimentally the measurement most directly involving t is the change of work function ϕ of silicon when a fraction θ of a monolayer of cesium is deposited (Fig. 7).²⁴ Let us consider the limit of low coverage. The ionization energy of a cesium atom is 3.9 eV, and if we bring it up to the silicon surface the proximity of the dielectric reduces this further. The work function of silicon is 4.8 eV, so that the electron prefers by 0.9 eV to go into surface states at E_F in the semiconductor. The potential and wave function are shown schematically in Fig. 8, and the dipole moment contributed by each cesium atom gives a measure of the length of the tail. We have

$$\phi(0) - \phi(\theta) = 4\pi Q n_0 [(t/\epsilon)_1 + (t/\epsilon)_2] \theta, \quad (18)$$

where Q is the effective charge of the dipole and suffixes 1 and 2 refer to the part of the separation t inside and outside the semiconductor, respectively. Since the diameter 3.2 Å of a cesium ion is comparable with the penetration in the semiconductor, a fraction of the electron will be situated on the ion, and we take $Q \sim \frac{3}{4}e$. The number of sites n_0 is $8 \times 10^{14} \text{ cm}^{-2}$. Turning to the data, what is "low coverage" for our purposes? At the point β the ϕ has already dropped 0.5 eV so that the

²⁴ F. G. Allen and G. W. Gobeli, *Report of the International Conference on the Physics of Semiconductors* (Institute of Physics and the Physical Society, London, 1962), p. 818.

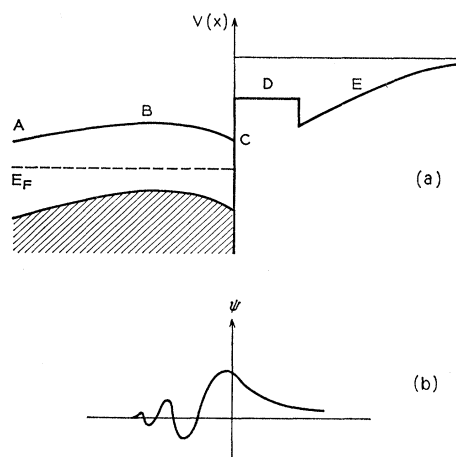


FIG. 8. Potential around Cs^+ ion on the silicon surface. AB are the bands bent by the space-charge layer. The bending in BC represents the Coulomb potential from the Cs^+ ion. DE is the pseudopotential of the ion [V. Heine and I. Abarenkov, *Phil. Mag.* **9**, 451 (1964)].

dipole potential produced at one ion by its neighbors is of this order of magnitude. That is already a considerable perturbation compared with the energy differences depicted in Fig. 8, and indeed the experimental curve appears to be still increasing in slope between α and β . The straight line slopes from β to γ and α to β are, respectively, 9 and 15 eV from which we guess a slope at $\theta=0$ of 25 eV. It would be interesting to have more precise data in this region. (18) now gives

$$(t/\epsilon)_1 + (t/\epsilon)_2 = 2.2 \text{ \AA}. \quad (19)$$

For $(t/\epsilon)_2$ we take $t=1.6 \text{ \AA}$ the radius of the core, and $\epsilon=1$ because the core is tightly enough bound. Using $\epsilon(\text{Si})=11.6$ we obtain for the mean length of the tail in the silicon

$$t_1 = 8 \text{ \AA}. \quad (20)$$

This figure compares satisfactorily with our theoretical estimate $1/(2q) \geq 3 \text{ \AA}$ and a figure $t/\epsilon \leq 1 \text{ \AA}$ (i.e., $t \leq 11 \text{ \AA}$) deduced by Crowell⁸ from the lack of variation of barrier height Φ with applied bias voltage in some junctions.

It is more difficult to estimate the density of states $\rho(E)$. As regards the total number of surface states, there is roughly speaking only one number which can be invariant under all the circumstances outlined in Sec. I, and that is 2 electrons (counting both spin directions) per surface atom. One might argue that it should be half this because of counting per unit cell, but, on the other hand, there are at least two bands at the valence- and conduction-band edges likely to give rise to surface states. The question is how spread out in energy they are. Scaling in energy the results of Pugh²⁵ for diamond would give a very high value, $6.4 \times 10^{15} \text{ electrons cm}^{-2} \text{ eV}^{-1}$. However, looking at Fig. 5, one might be more inclined to go to the opposite extreme and assume

²⁵ D. Pugh, *Phys. Rev. Letters* **12**, 390 (1964).

uniform spreading of the "states" in the band gap, at least for the junctions with metals with high k_F . Since these do not behave radically different from other junctions or the free surfaces, we might infer the same true there. In any case such an assumption gives a reasonable low-estimate $\rho_{1,e}(E)$. Adopting a mean band gap of $4\frac{1}{2}$ eV again,¹⁴ we obtain

$$\rho_{1,e}(E) = 3.5 \times 10^{14} \text{ electrons cm}^{-2} \text{ eV}^{-1}. \quad (21)$$

We shall adopt (21) in subsequent calculations, because as we shall see a value much larger would imply very small values of $\xi - \xi_0$. We note that (21) is only twice the "straight-line" estimate of Allen and Gobeli.¹⁰ We reject their detailed model with two very narrow bands for the same reason as above, namely it seems too fine a structure to be retained in junctions with metals with high k_F (though one could also argue that their two bands are placed symmetrically above and below the Fermi level so that any broadening would also be symmetrical and retain the same E_F). Such very dense bands might also be expected to show up in photoemission.²⁶ At dopings less than 10^{19} cm^{-3} , (21) is quite consistent with their data, predicting a curve which bends up 0.02 eV at the p -type end and down 0.03 eV at the n -type end. However, the experimental variation of φ at and beyond 10^{19} cm^{-3} does not fit. At 10^{19} cm^{-3} , the thickness of the space-charge region is 50 Å (p -type) and 90 Å (n -type) which is becoming comparable with the length of the surface-state tails. Thus, the bands bend over the length (20) by about 0.12 eV. It seems reasonable that under conditions of such steep band bending there might be deviations of ~ 0.03 eV from the simple theory as would be required to fit the data. (We note that at even higher dopings the change of φ reverses sign.) At least we suggest this as an alternative interpretation of the data. Measurements on diodes at a lower temperature should give a clearer indication of the density of states applicable to the central range of doping. We shall later come across other deviations from the simple theory under conditions of extreme band bending.

We return at last to Eq. (15) and the data of Fig. 1. We put

$$\frac{t}{\epsilon} = \frac{8.0}{11.6} + \frac{0.5}{1} = 1.2 \text{ \AA}, \quad (22)$$

the first term being taken from (20) and the second term representing the screening distance in the metal.²⁷ With the use of (21) and (22), Eq. (15) becomes

$$\xi = \xi_0 + 0.13(\varphi_{sc} - \varphi_m), \quad (23)$$

which is shown on Fig. 1 as the dashed line. It reproduces the general trend of the data as well as any straight line can. If we accept t as correct to 50%, then $\rho(E)$ cannot

be much less than half of (21) without producing a line that is too steep to represent the data. On the other hand, if filling up the surface states is indeed the explanation of the trend to higher ξ values at lower φ_m , then $\rho(E)$ cannot be much larger than (21) either. However we shall later mention a possible alternative explanation of the high values of ξ in Fig. 1.

We first discuss two other consequences of (21), (22), and filling up the surface states. One is the variation of barrier height with applied bias voltage. Since the surface states are in good electrical contact with the metal, we assume there is no extra dipole field between them. All the charge $\sigma = \epsilon E / 4\pi$ on which the lines of force from the applied field E terminate therefore reside in the surface band, raising its Fermi level by $\sigma / \rho(E)$: i.e., Φ would decrease by this amount to keep E_F constant. Experimentally⁸ a bias of 10^5 V/cm produces no measurable change in Φ , i.e., $\delta\Phi \leq 2 \times 10^{-3}$ eV. The $\delta\Phi$ predicted from (21) for this field is 2×10^{-3} eV, just on the limit of error. Again we find $\rho(E)$ cannot be much lower than (21) without leading to contradictions. A similar effect to bias is the space-charge field produced by doping. High n -type doping produces a strong field which on the present model would lead to a reduction in Φ by 0.01 eV for a $0.2\text{-}\Omega\text{-cm}$ sample ($N_d \sim 9 \times 10^{17} \text{ cm}^{-3}$). However, an *increase* of 0.03 eV in Φ is found,^{2,8} which cannot in any way be reconciled with the mechanism of the present model.

This discrepancy leads us to point out that we have in fact been using the idea of a Fermi level in the band of surface states in a very cavalier fashion. The very fact that the charge in these states is spread out over some distance t means that there will be very severe bending of the bands over this distance, as shown in Fig. 9. Indeed the concepts of band bending must not be taken too seriously over such distances since the band states themselves cannot be localized to smaller than the size of the Wannier function which is of the same order²⁰ as t . From (22), over half of the dipole potential difference between a metal and a semiconductor in the junction occurs in the semiconductor. Thus if φ_m is less than φ_{sc} by more than 1 eV, the bands are bent so much that the conduction-band edge dips below the Fermi level [Fig. 9(c)]. This might be the explanation of the good surface conductivity found on silicon when covered with a monolayer of metallic cesium.²⁶ The same should apply for a calcium monolayer. Certainly such bending of the conduction band is compatible with the existence of a barrier. Calcium produces a barrier (Fig. 1), and it would be interesting to know if cesium does too. We come now to an alternative explanation of the trend of ξ in Fig. 1. Basically a higher value of $\rho(E)$, i.e., a less sloping line, would not be incompatible with the data for metals with high $\varphi_m \geq 4$ eV. That leaves calcium and magnesium anomalously high. This might be related to the extreme bending of the bands in the sense that in Fig. 9(c) the length of the tails would be seriously curtailed, at least in the lower half of the gap, necessi-

²⁶ F. G. Allen (private communication).

²⁷ N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys* (Clarendon Press, Oxford, 1936), p. 88.

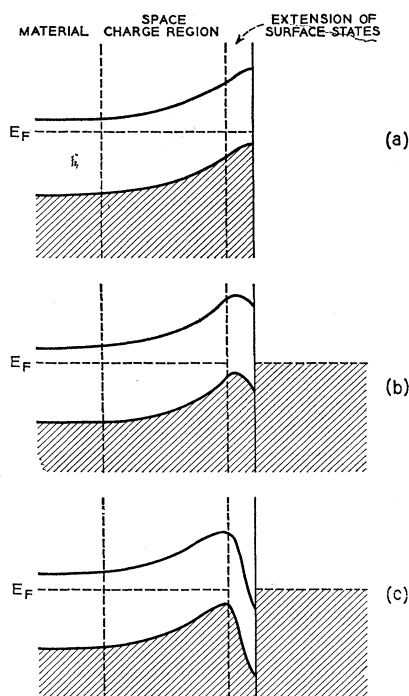


FIG. 9. Surface-state band bending (a) on a free surface, (b) in a metal junction when $\varphi_m < \varphi_{sc}$, (c) in a junction when $\varphi_m \ll \varphi_{sc}$. On the left is the semiconductor with bent valence- and conduction-band edges shown, on the right the metal [cases (b) and (c)].

tating a higher ξ and lower barrier to make up the required total charge. Another effect would be as follows: The charge density $|\psi|^2$ is so high in the tails that the bands bend downward already at a depth considerably greater than t which represents the center of gravity of the charge, which would also tend to lower the barrier. We conclude by reiterating that this surface-state band bending is an inescapable consequence of the surface-state charge being spread out in a tail, but the exact consequences of it for barrier heights are far from clear. In particular we have not been able to use it to provide a convincing explanation for the sign of the change of barrier height with doping.

Our present models say nothing specific about the case of antimony in Fig. 1. Here ξ lies close to the solid sloping line, which represents the ξ that would be observed if the whole of the dipole potential $\varphi_{sc} - \varphi_m$ appeared across the space-charge region rather than between the surface states and the metal. This suggests that the surface layer of the semiconductor is highly conducting and acts as one conducting system with the metal so that no dipole field can exist between them. If the first layer of antimony bonded tetrahedrally to the

silicon and acted as donor centers, it might well produce such a conducting n -type layer: Its thickness would only have to be comparable with (20) for it to be effective.

Finally a comment about the work function of silicon with a monolayer of cesium or oxygen: Since the thickness of the surface states is as large as (20), the charge that has to be transferred per atom to produce the required dipole $|\varphi_{sc} - \varphi_m|$ is only about one-tenth of an electron per atom. Consequently, the electronic condition in a monolayer is very nearly the same as in bulk material. Thus the work function with a monolayer will be close to that of pure cesium or "oxygen metal," respectively. By the latter term we mean that the oxygen is likely to be bonded to the silicon as a close-packed monolayer rather than existing as diatomic molecules. In that case it would be metallic, and we can estimate crudely its work function as 8 eV, the mean of the ionization energy (13.6 eV) and the electron affinity (2.3 eV) of atomic oxygen. As a smaller second-order effect, the existence of the dipole leaves the cesium and the oxygen, respectively, with a small net positive and negative charge, thus increasing and decreasing φ , respectively. We can estimate the order of magnitude by taking one-tenth of the difference between the ionization potential and the electron affinity of the atom, since that represents the Coulomb energy for one extra electron, obtaining 0.3 eV for cesium and 1 eV for oxygen. Thus φ for cesium on silicon might be expected to drop ~ 0.3 eV below the value for bulk cesium at about one monolayer coverage, and then increase again to the latter for thicker layers, of course. Something like that happens for cesium on metals,²⁸ and there appears to be some experimental evidence for it on semiconductors.²⁹ The φ for a monolayer of oxygen on silicon becomes $8 - 1 = 7$ eV, compared with the experimental value²⁶ of 6 eV. For thicker layers φ returns to 4.6 eV, very near the value for a clean surface, which presumably indicates the conversion of the oxygen layer to an electrically inert oxide layer.²⁶

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the kind hospitality of Bell Telephone Laboratories during the summer of 1964, where this work was done. Several useful conversations with Dr. J. Bardeen, Dr. F. G. Allen, Dr. C. R. Crowell, Dr. E. O. Kane, and Dr. C. A. Mead have contributed significantly to the present work, and their help is gratefully acknowledged.

²⁸ J. M. Houston and H. F. Webster, *Advan. Electron. Electron Phys.* **17**, 125 (1962).

²⁹ J. Bardeen (private communication).