

**TOPIC 12: CONTACTS AND SURFACES**READING ASSIGNMENT

"Metal-Semiconductor Contacts and Semiconductor Surfaces", from *Solid State and Semiconductor Physics*, J.P. McKelvey

SUPPLEMENTARY REFERENCE

"Introduction", from *Semiconductor Surfaces and Interfaces*, W. Mönch

LECTURE PROGRAM

1. Metal-Semiconductor Contacts in Equilibrium (16.1)
2. Metal-Semiconductor Contact Rectification (16.2)
3. Surfaces States (16.3)
  - a. Phenomenological
  - b. Tamm states
  - c. Surface inversion
  - d. Surface accumulation
  - e. Metal-Surface-Semiconductor contact
4. Potential, Field and Charge with a Semiconductor Surface Layer
5. a. Surface Conductivity, Field Effect and Surface Mobility  
b. Properties of Actual Semiconductor Surfaces

## CHAPTER 16

# METAL-SEMICONDUCTOR CONTACTS AND SEMICONDUCTOR SURFACES

## 16.1 METAL-SEMICONDUCTOR CONTACTS IN EQUILIBRIUM

When two metals having different work functions are brought into contact with one another, a brief transient current flow takes place which transfers electrons from the metal with the larger Fermi energy to the one with the smaller Fermi energy, thereby generating an equilibrium *contact potential* difference between the two. This process is shown schematically in Figure 16.1. At (a) two metals having different work

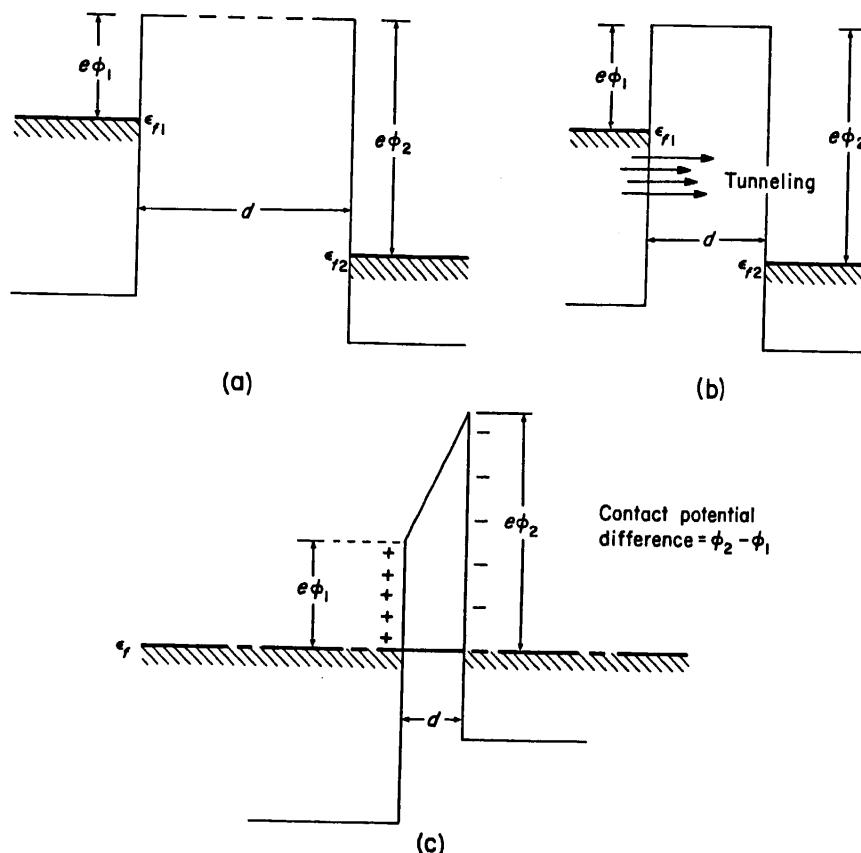


FIGURE 16.1. Successive stages in the establishment of equilibrium between two metals with different work functions.

functions are separated by a distance  $d$  which is assumed to be quite large, possibly of the order of a centimeter. Under these circumstances the Fermi levels of the two metals do *not* coincide, and the system is *not* in equilibrium. To achieve equilibrium between the two metals, electrons would have to *tunnel* through the potential barrier from the metal having the higher Fermi energy to the one with the lower Fermi energy until a condition of equilibrium in which the two Fermi levels coincide is attained. This cannot be accomplished while the metals are widely separated, as at (a), because the potential barrier is so high and thick that the tunneling probability is negligible. Equilibrium might only be established after millions of years under these conditions.

As the two metals are brought closer together, as in (b), the thickness of the tunneling barrier becomes less and less, until finally the tunneling probability becomes sufficiently large that electrons penetrate the barrier, flowing from left to right in the diagram. The physical separation  $d$  which is necessary to bring this state of affairs about is a few times the interatomic distance. The current of electrons creates a negative charge in the metal on the right, and the excess of positive ions which are left behind on the left gives rise to a positive charge in that region. Accordingly, a field is set up which raises the potential energy of the electrons on the right with respect to those on the left until at length the Fermi levels of the two metals coincide. Under these circumstances the system is in equilibrium and the net tunneling current drops to zero, since now tunneling in either direction is equally probable. The transfer of the electrons which flowed from left to right, however, has established a potential difference of magnitude  $\phi_2 - \phi_1$  between the two metals. This voltage is called the contact potential difference between the two metals. The number of electrons which are transferred from one metal to the other in order to establish this potential difference is so small compared with the total number of free electrons in either substance that the relative values of free electron population are practically unaffected.

Somewhat the same line of reasoning may be used to infer the properties of a metal-semiconductor contact. The semiconductor differs from a metallic substance, however, in that an electric field may exist within the interior of a semiconductor. For this reason, the contact potential drop between the metal and semiconductor may take place *within* the semiconductor rather than at the contact interface. In the simplest possible instance, what may happen is illustrated in Figure 16.2 for the case of a contact between a metal and an *n*-type semiconductor crystal, where the work function of the metal,  $\phi_M$ , is larger than the work function  $\phi_{SC}$  associated with the semiconductor. As in Figure 16.1, a process is envisioned wherein the distance  $d$  between the two substances is decreased until at length this distance is small enough so as to permit electrons to tunnel freely through the barrier. The field which arises due to the contact potential difference exists now largely within the semiconductor; the potential energy of an electron at rest at the bottom of the conduction band in the interior of the crystal thus differs from the potential energy of such an electron at the surface by the amount  $e(\phi_M - \phi_{SC})$ , and as a result the conduction and valence band edges are shifted with respect to the Fermi level as illustrated. The positive space charge density in the surface region, due to the excess concentration of ionized donor atoms over the electron population (in conjunction with the electrons which tunneled through to the metal) is just such as to produce a field sufficient to sustain the potential difference  $\phi_M - \phi_{SC}$  between the two materials. The situation within the semiconductor is not essentially different from that which was discussed in connection with the space charge layers of a *p-n* junction.

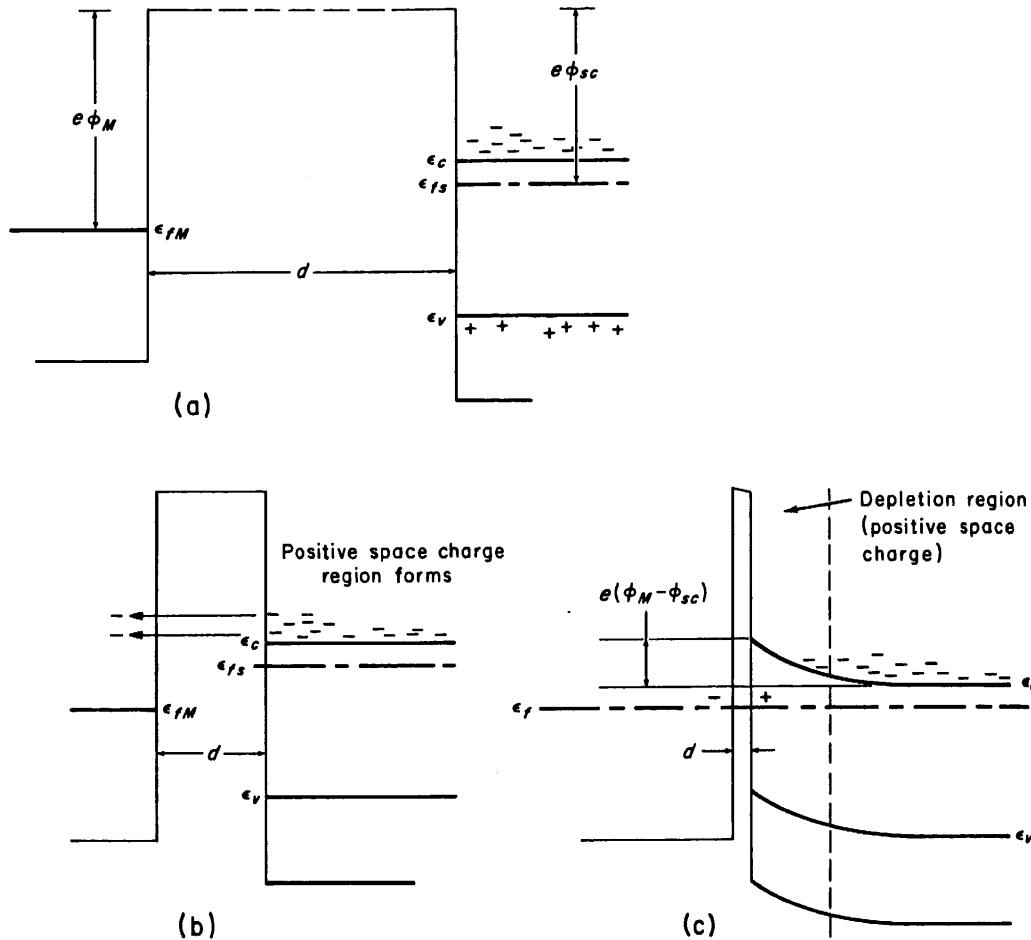


FIGURE 16.2. Successive stages in the establishment of equilibrium between an *n*-type semiconductor and a metal having a greater work function.

In this example, the net carrier density near the surface of the semiconductor is reduced from its bulk equilibrium value, and the surface layer is referred to as a *depletion region*. If  $\phi_M - \phi_{sc}$  is sufficiently large the bands may be shifted with respect to the Fermi level to such an extent that next to the surface the valence band is nearer the Fermi level than the conduction band, the material just adjacent to the surface then becoming in effect *p*-type. The surface is then said to be inverted in conductivity type, and the surface *p*-type region is called an *inversion region*.

It is important to note that in the present example a potential barrier of height  $e(\phi_M - \phi_{sc})$  is formed at the surface. The formation of this potential barrier and the existence of the depletion region are the basis of the explanation of how a metal-semiconductor contact rectifier operates, as we shall see later. One should observe that in general the depletion region is quite thick. As a matter of fact the thickness of the surface space charge region is essentially the same as that of a *p-n* junction space charge region having an internal potential difference  $\phi_M - \phi_{sc}$ . There is therefore no possibility (except under extreme circumstances) for electrons ever to be able to tunnel through such a barrier. On the other hand, the very narrow barrier of thickness  $d$ , as shown in Figures 16.1(c), 16.2(c) and 16.3, is envisioned as being ultimately so thin as to permit electrons to tunnel freely through it. In two materials

which are joined in intimate contact,  $d$  is zero, and the only barrier which remains is the surface potential barrier of the semiconductor.

In Figure 16.4 the situation which arises at the contact interface between a metal and an *n*-type semiconductor where  $\phi_{SC}$  is greater than  $\phi_M$  is shown. Here the semiconductor must acquire a negative charge, the metal a positive one, and the bands

FIGURE 16.3. An inversion layer at the interface between a metal and an *n*-type semiconductor.

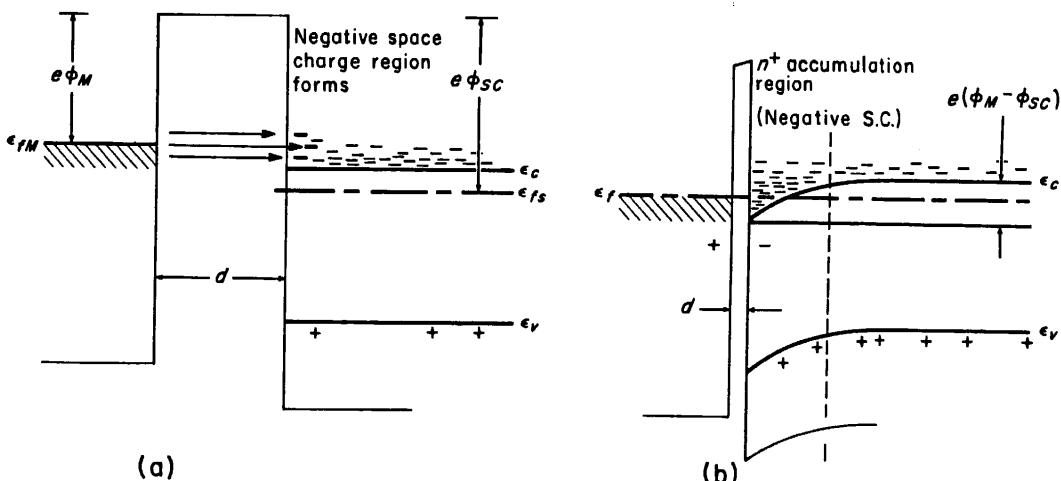
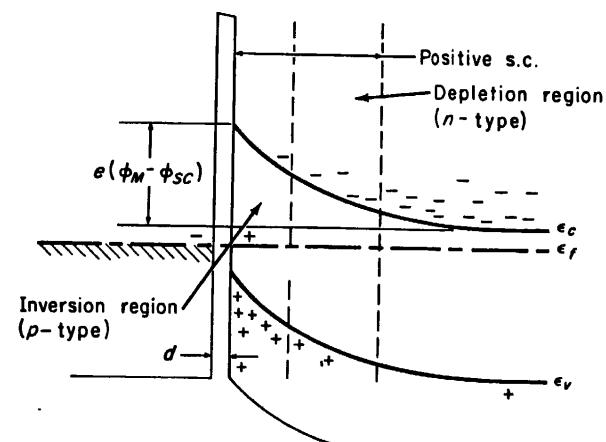


FIGURE 16.4. Successive stages in the establishment of equilibrium between an *n*-type semiconductor and a metal having a smaller work function.

shift downward at the surface. As a result, instead of a potential barrier, an *accumulation region* in which the electron concentration is greater than the concentration of ionized donor atoms is formed. The excess electron concentration in the surface accumulation region gives rise to the negative space charge necessary to support the contact potential difference between the two substances. The contact effects which arise between metals and *p*-type semiconductors may be discussed in very much the same way as those associated with *n*-type semiconductors. It may thus be shown that a depletion or inversion region, with an accompanying potential barrier, is formed at the interface between a *p*-type semiconductor and a metal whose work function is smaller than that of the semiconductor, and an accumulation layer is formed between a *p*-type semiconductor and a metal having a larger work function. The verification of these conclusions is assigned as an exercise.

## 16.2 METAL-SEMICONDUCTOR CONTACT RECTIFICATION

The potential barrier at a metal-semiconductor interface where a depletion or inversion layer is present behaves as a rectifier in somewhat the same way as does a *p-n* junction in the interior of a semiconductor crystal. Since the concentration of carriers within the depletion region is much smaller than in the other parts of the system, any externally applied voltage will tend to appear primarily across this high-resistivity layer. The applied voltage will thus tend either to increase or reduce the effective barrier height, as shown in Figure 16.5 for the case of an *n*-type semiconductor and a metal

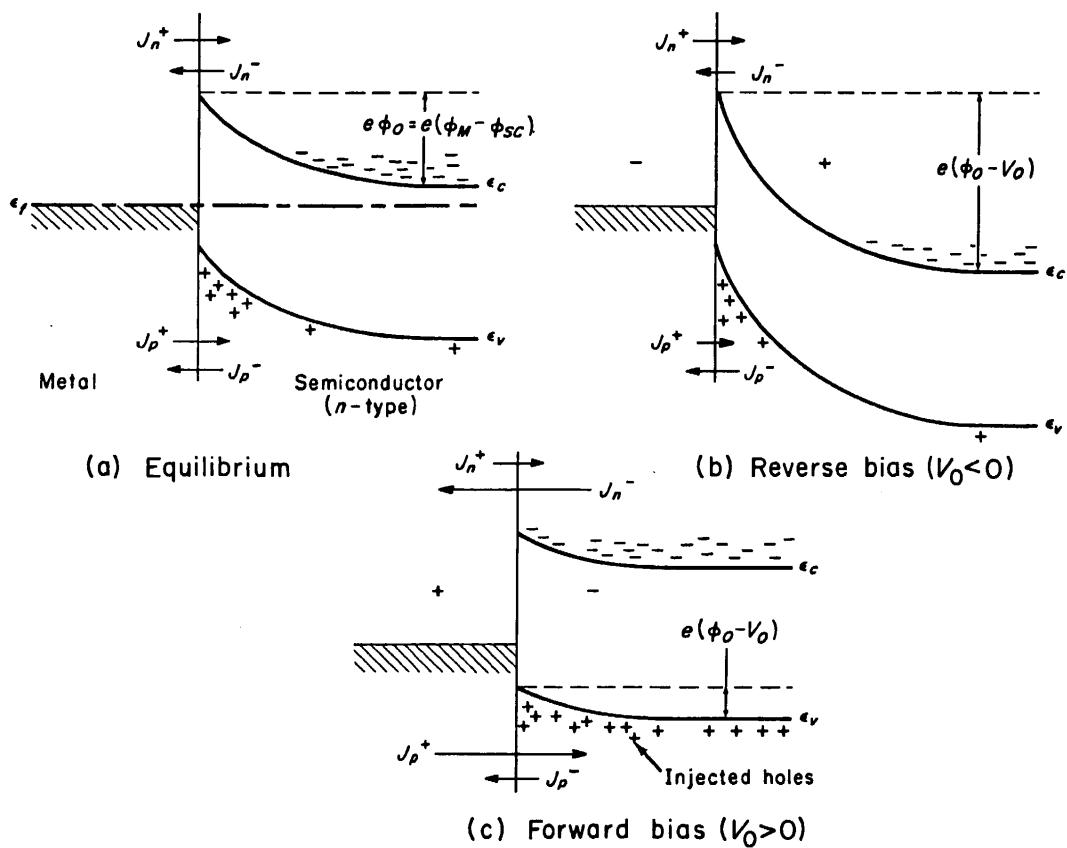


FIGURE 16.5. Potential diagrams illustrating the operation of a metal-semiconductor contact rectifier (a) in the equilibrium state (b) under reverse bias, and (c) under forward bias.

having a larger work function. At (a), in equilibrium, a certain fraction of the electrons in the conduction band of the semiconductor will have sufficient energy to surmount the surface potential barrier. These electrons will cause an electron current  $J_n^-$  to flow to the left. An electron flux  $J_n$  arising from the small fraction of electrons in the metal which can surmount the barrier will also flow to the right from the metal to the semiconductor. According to the principle of detailed balance, these two fluxes must be *equal* in magnitude in the equilibrium state; under these circumstances there is no net electron current. In a similar way, there are two hole fluxes  $J_p^+$  and  $J_p^-$ . The former consists of holes which are generated at the semiconductor surface (by electrons from the valence band occupying occasional empty electronic states in the electron

distribution of the metal) having sufficient energy to overcome the potential barrier and enter the interior of the semiconductor. The latter is made up of holes from the interior of the semiconductor which diffuse to the surface and disappear there by extracting an electron from the Fermi distribution of the metal. Again, at equilibrium, these two fluxes are equal and opposite and no net hole current flows.

When an external voltage is applied to the contact, the situation is as shown in Figure 16.5(b) or (c). The two fluxes  $J_n^-$  and  $J_p^+$  are clearly governed by the number of carriers present in the semiconductor possessing enough energy to get over the surface potential barrier, which, when an external voltage is applied, is of height  $e(\phi_0 - V_0)$ , where

$$\phi_0 = \phi_M - \phi_{sc} \quad (16.2-1)$$

is the difference between the work functions. This number is simply proportional to  $e^{-e(\phi_0 - V_0)/kT}$ . The other two fluxes,  $J_p^-$  and  $J_n^+$ , are dependent only upon the number of thermally generated holes available in the interior of the semiconductor, and the number of electrons from the metal which can surmount the barrier from the left (where the barrier height is fixed), respectively. These two fluxes are functions only of temperature and the material parameters of the two substances, and are independent of the barrier height  $e(\phi_0 - V_0)$ . We may thus express  $J_n^-$  and  $J_p^+$  in the form

$$J_n^- = -J_{n0} e^{-e(\phi_0 - V_0)/kT} \quad (16.2-2)$$

and

$$J_p^+ = J_{p0} e^{-e(\phi_0 - V_0)/kT}, \quad (16.2-3)$$

where  $J_{n0}$  and  $J_{p0}$  are constants expressing the gross fluxes of electrons and holes, of whatever energy, which are initially incident at equilibrium upon the surface barrier. Since  $J_n^+$  and  $J_p^-$  must equal  $-J_n^-$  and  $-J_p^+$ , respectively, when there is no applied voltage, we may write these fluxes as

$$J_n^+ = J_{n0} e^{-e\phi_0/kT} \quad (16.2-4)$$

and

$$J_p^- = -J_{p0} e^{-e\phi_0/kT} \quad (16.2-5)$$

The total fluxes of electrons and holes are then

$$J_n = J_n^+ + J_n^- = -J_{n0} e^{-e\phi_0/kT} (e^{eV_0/kT} - 1) \quad (16.2-6)$$

and

$$J_p = J_p^+ + J_p^- = J_{p0} e^{-e\phi_0/kT} (e^{eV_0/kT} - 1), \quad (16.2-7)$$

while the total electrical current density is simply

$$I = e(J_p - J_n) = I_0 (e^{eV_0/kT} - 1), \quad (16.2-8)$$

where

$$I_0 = e(J_{p0} + J_{n0}) e^{-e\phi_0/kT} \quad (16.2-9)$$

The metal-semiconductor rectifier thus has the same current-voltage dependence as the  $p-n$  junction, except that the saturation current is different. The gross fluxes

of electrons and holes approaching the barrier in equilibrium from the right and left, respectively, are easily expressed<sup>1</sup> as

$$J_{n0} = \frac{n_0 \bar{c}}{4} \quad (16.2-10)$$

and  $J_{p0} = \frac{p_s \bar{c}}{4} = \frac{p_0 \bar{c}}{4} e^{e\phi_0/kT}, \quad (16.2-11)$

where  $p_s$  refers to the concentration of holes in the valence band at the metal surface. Using these results, the saturation current may now be written in the form

$$I_0 = \frac{e\bar{c}}{4} (p_0 + n_0 e^{-e\phi_0/kT}). \quad (16.2-12)$$

In the reverse bias condition, the current flow is small, approaching  $I_0$  at reverse voltages large compared with  $kT/e$ . Under these circumstances, the depletion region increases in thickness with increasing reverse bias just as does the space charge region of a *p-n* junction. The physical processes which govern the depletion layer thickness are essentially the same in both cases. The current which flows consists almost entirely of holes "collected" from the interior of the semiconductor by the contact and the few electrons from the metal which get over the barrier from the left in Figure 16.5(b). Under forward bias, the barrier height is lowered, and the number of electrons in the conduction band which can overcome the barrier is now much larger than in the equilibrium or reverse bias conditions. A large electron flow from semiconductor to metal results. In addition to this flow of electrons, a flow of holes *into* the interior of the semiconductor will be observed, because now the potential barrier to the holes which are generated at the surface when electrons in the valence band of the semiconductor fall into unfilled electronic states in the Fermi distribution of the metal is *also* lowered. This amounts simply to *injection* of minority carriers by the metal contact in the forward-bias condition.

The above calculations have been made for the case of a contact between an *n*-type semiconductor and a metal having a larger work function. The case of a *p*-type semiconductor and a metal having a smaller work function can be treated in essentially the same way. In this case, however, it is found that the forward direction of rectification is obtained when the semiconductor is positive with respect to the metal, which is just the opposite of the result obtained above. Under these conditions *electrons* are injected into the *p*-type semiconductor when it is positive with respect to the metal. The verification of these results is assigned as an exercise. Due to the absence of a potential barrier, no rectification is obtained at a contact where an accumulation layer is formed. An essentially ohmic contact will therefore be obtained at an interface

---

<sup>1</sup> It should be noted in this connection that there can be no fields nor concentration gradients nor generation and recombination of electrons and holes as such in the metal, and for this reason the situation differs in certain essentials from that discussed in Section 13.1 in connection with the *p-n* junction. Also, it is correct to use the expressions (16.2-10) and (16.2-11) only when the depletion region is thin enough so that essentially no diffusion takes place within it. Since the surface barrier thickness is ordinarily about the same as the electron and hole mean free paths, this condition may sometimes not be fulfilled in practice.

where there is no barrier (i.e.,  $\phi_M = \phi_{Sc}$ ) or where an accumulation layer is present. The results obtained above serve to verify the observations about the properties of point-contact rectifiers which were made in connection with the discussion of the Haynes-Shockley experiment in Section 10.4.

The predictions of this theory of metal-semiconductor contacts are not, unfortunately, in all respects in agreement with experiment.<sup>2</sup> In particular, these calculations indicate that the rectification properties of metal-semiconductor contacts should be extremely sensitive to the difference in work functions between the metal and the semiconductor, while, in fact they are found to be nearly independent of this quantity. This discrepancy was explained by Bardeen<sup>3</sup> in 1947, who introduced the notion of *surface states* at the contact interface. This subject will be discussed in the next section. In other respects, the theory was more successful. The predicted current-voltage characteristic is in fairly good agreement with experiment, and, in particular, the effects associated with electron and hole injection (as illustrated, for example, by the Haynes-Shockley experiment) are in complete accord with experimental observations.

### 16.3 SURFACE STATES AND THE INDEPENDENCE OF RECTIFYING PROPERTIES OF WORK FUNCTION

According to the results of the preceding section, the characteristics of metal-semiconductor rectifying contacts should be critically dependent upon the difference in the work functions of the semiconductor and the metal. In particular, if the work functions of the two substances are such as to produce an accumulation layer, no rectification at all would be expected. The experimental results of Meyerhof,<sup>2</sup> on the contrary, showed that in the case of silicon-metal contacts the rectification properties of the contact were practically *independent* of the work function difference, and, in fact, were much the same for all metals. This situation was resolved by Bardeen,<sup>3</sup> who assumed that there could be localized electronic states associated with the surface lying in the forbidden energy region between the conduction and valence bands of the semiconductor.

The fundamental justification for the existence of these *surface states* goes back to the work of Tamm,<sup>4</sup> who showed that if a periodic square-well potential such as that associated with the Kronig-Penney crystal model were terminated on one side by a surface potential barrier such as that illustrated in Figure 16.6, there would be (in addition to the usual allowed energy bands of the Kronig-Penney model) discrete allowed levels within the forbidden energy regions corresponding to wave functions which are localized near the surface. These surface levels were studied in much more detail in a later article by Shockley.<sup>5</sup> According to the calculations there should be one surface state for each surface atom. In addition to this, localized surface levels, discrete or continuously distributed, may be expected to arise from impurity atoms, oxide layers and structural imperfections at the surface. The surface properties of an

<sup>2</sup> See, for example, W. E. Meyerhof, *Phys. Rev.*, **71**, 727 (1947).

<sup>3</sup> J. Bardeen, *Phys. Rev.*, **71**, 717 (1947).

<sup>4</sup> I. Tamm, *Physik. Z. Sowjetunion*, **1**, 733 (1932).

<sup>5</sup> W. Shockley, *Phys. Rev.*, **56**, 317 (1939).

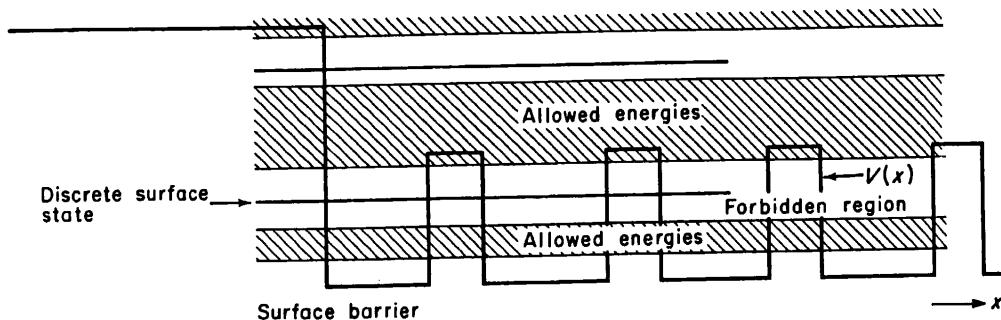


FIGURE 16.6. The formation of localized Tamm states in the forbidden energy region at the surface of a one-dimensional crystal.

individual crystal will then depend upon the density of these surface levels and upon their distribution in energy.

The presence of a considerable density of surface states will result in the formation of a surface depletion layer or accumulation layer within the semiconductor even in the absence of an external metallic contact. To understand how this may come about, let us examine in detail a specific model consisting of an *n*-type semiconductor crystal having a substantial number of acceptor-type surface states (which are charged when occupied and neutral when empty) whose density and distribution in energy within the forbidden region are as illustrated in Figure 16.7. In the absence of any metallic

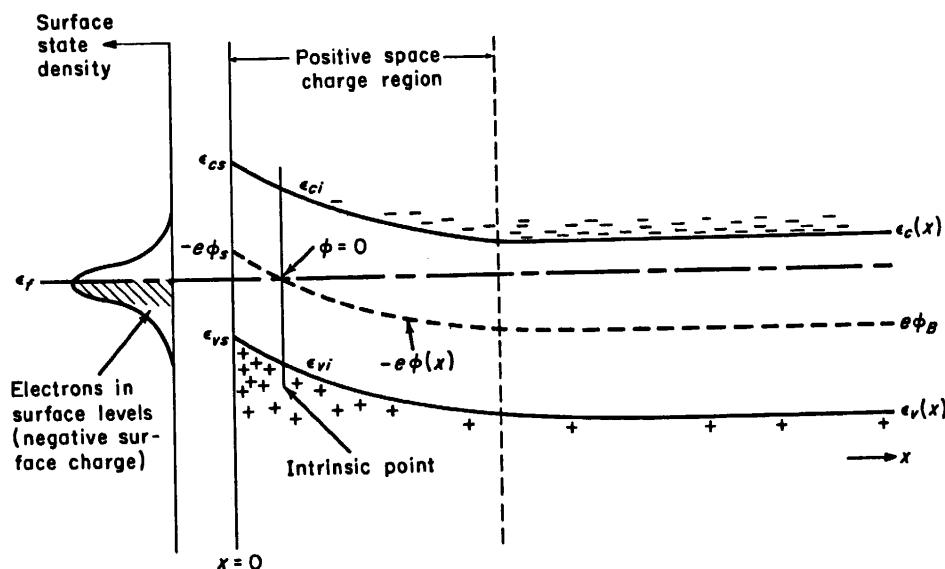


FIGURE 16.7. The formation of a surface inversion layer at a free semi-conductor surface through the interaction of surface states with the interior of the crystal.

contact at the surface, the surface states will be filled up to the Fermi level<sup>6</sup> creating a layer of negative charge at the surface which repels electrons within the conduction

<sup>6</sup> This statement refers, of course, to the state of affairs at absolute zero. At higher temperatures, some states above  $\epsilon_f$  will be occupied and some states below empty according to the Fermi distribution law.

band of the semiconductor from the surface and which therefore leaves a positive surface space charge layer arising from uncompensated donor ions. In this manner a depletion region and a surface potential barrier are set up by the equilibrium between the bulk semiconductor and the surface states, quite independently of any external contact. If the crystal as a whole is electrically neutral, the total charge contained within the space charge region in the semiconductor must equal in magnitude the total charge associated with electrons in surface states. This fact enables one, as we shall see later, to calculate the value of the surface potential  $\phi_s$  within the semiconductor just adjacent to the surface, and, in fact, the entire potential distribution within the semiconductor surface layer. It is clear that if the surface states are primarily of the donor type (charged when empty and neutral when occupied by electrons) an accumulation layer rather than a depletion layer will be formed on an *n*-type crystal.

In order to understand the precise effect of the surface states, consider now the effects which take place when a metallic substance is allowed to approach such a semiconductor surface region. We shall discuss specifically the case where the metal is one having a smaller work function than the *n*-type semiconductor, where a non-rectifying *accumulation* layer was formed in the absence of surface states. The situation is illustrated in Figure 16.8. At (a) the distance  $d$  between the metal and semiconductor

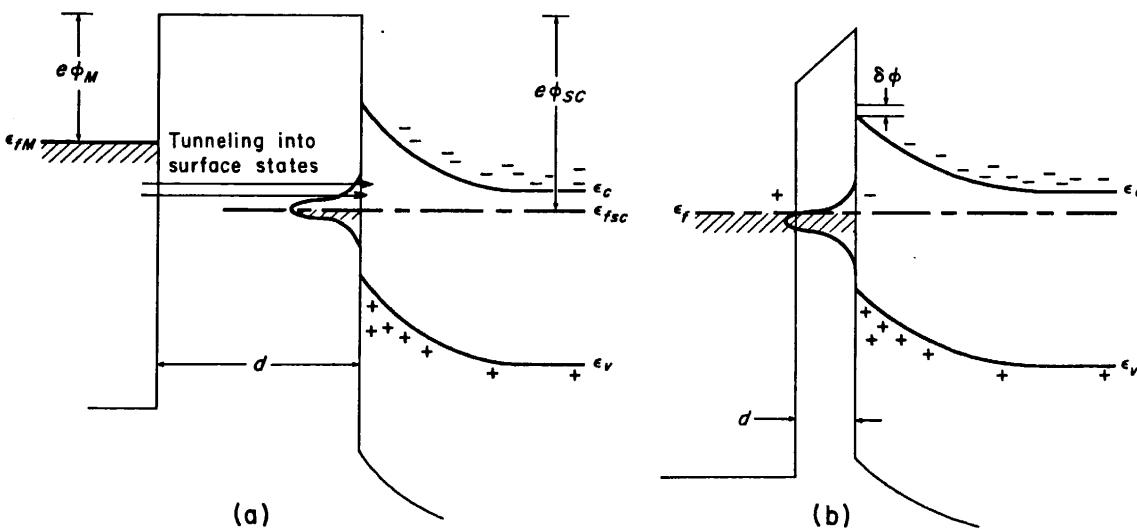


FIGURE 16.8. Diagram illustrating successive stages in the establishment of equilibrium between a metal and a semiconductor with surface states. If the surface state density is reasonably large the interior space charge distribution is barely affected.

has been decreased to the point where tunneling of electrons may begin. The electrons now tunnel from the metal to the surface states, and if the density of surface levels is large enough, the surface states themselves will accommodate all the electrons which are necessary to set up an electric field large enough to equalize the electron potential energies at the respective Fermi levels, the final state of the system being as at (b). The situation inside the semiconductor is essentially the same as at (a), except that the Fermi level is a bit higher with respect to the surface state distribution, and this has resulted in a change of surface barrier height  $\delta\phi$  which, however, may be quite small if the surface state density is appreciable. The depletion layer and surface potential barrier still remain, and the rectifying properties of the contact, which, after all, are

governed primarily by just these characteristics, are practically invariant. It is an easy matter to verify the fact that, given a situation which is the same except that the work function of the metal exceeds that of the semiconductor, the barrier height change  $\delta\phi$  is of opposite sign, the tunneling of electrons from surface states to the metal having lowered the Fermi level with respect to the surface state distribution.<sup>7</sup> Again, the rectifying properties are virtually unaffected. The rectification characteristics are thus seen to be practically independent of the relative values of the two work functions, in agreement with what is observed experimentally.

Another experimental proof of the existence of surface states is afforded by the so-called field effect experiment, in which a semiconductor is made to function as one of the plates of a parallel plate condenser, as shown in Figure 16.9. In this drawing

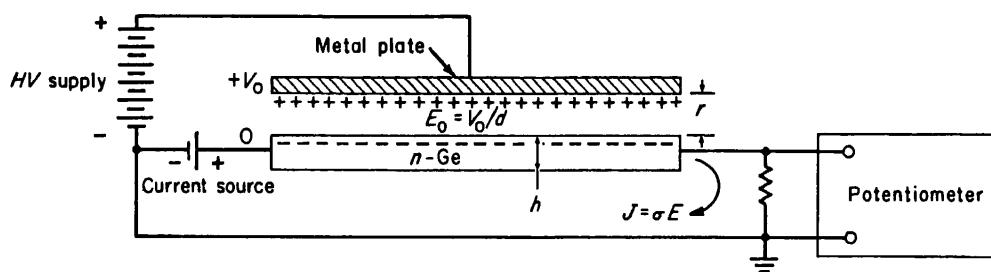


FIGURE 16.9. Schematic diagram of the field effect experiment and associated circuitry.

a sample of *n*-type germanium is used as the negative condenser plate, and as large a negative surface charge as possible is induced by connecting a high voltage supply between the plates. Since the capacitance of a parallel plate condenser is just  $A/4\pi r$ , where  $A$  is the area of the plates and  $r$  their separation, a charge of magnitude

$$q = CV_0 = \frac{V_0 A}{4\pi r} \quad (16.3-1)$$

will be induced in the semiconductor when a potential difference  $V_0$  is applied. If there were no surface states, this charge would consist solely of mobile free electrons in the conduction band at the surface. The number of induced charge carriers,  $\Delta N$  is simply  $q/e$ , or,

$$\Delta N = \frac{V_0 A}{4\pi e r} \quad (16.3-2)$$

This change in the number of mobile free electrons would cause a change in sample conductivity  $\Delta\sigma$  which would amount to

$$\Delta\sigma = e\mu_n \Delta n = e\mu_n \frac{\Delta N}{V} \quad (16.3-3)$$

where  $V$  is the volume of the sample. Writing the volume  $V$  as the product of the

<sup>7</sup> If the surface states are all at the same energy (i.e., if their distribution is essentially a  $\delta$ -function) there will be no shift in barrier height at all.

area  $A$  and the sample thickness  $h$ , and substituting (16.3-2) for  $\Delta N$ , one may obtain, finally,

$$\Delta\sigma = \frac{V_0\mu_n}{4\pi hr} \quad (16.3-4)$$

as the change in measured conductivity of the sample due to the presence of induced surface charge.

When the experiment was performed, however, it was found that the observed conductivity change was only about one-tenth the value predicted by this expression.<sup>8</sup> The result of the experiment was interpreted by Bardeen to indicate that about 90 percent of the excess induced charge goes into the surface states rather than into the conduction band of the semiconductor, and that electrons in the surface states are *immobile* rather than free.

## 16.4 POTENTIAL, FIELD AND CHARGE WITHIN A SEMICONDUCTOR SURFACE LAYER

Let us now consider in detail the fields and potentials within the surface region which are set up when surface states are present. We shall assume, in general, that there is an applied field  $E_0$  outside the semiconductor surface, which is generated by some experimental field-effect apparatus such as that shown in Figure 16.9. The situation within the crystal will then be as represented in Figure 16.7. Note that the origin of the potential function  $\phi(x)$ , which can be chosen arbitrarily, is defined such that at the point where the material is intrinsic (i.e., where  $p_0 = n_0 = n_i$ )  $\phi$  is zero. This is the same choice which was made in the discussion of the  $p$ - $n$  junction in Section 12.3, and indeed there are many points of similarity between the two examples. Under these circumstances the variation of the conduction and valence band edge energies with distance are as given by (12.3-6) and (12.3-7), while the carrier densities at each point within the band are represented by (12.3-4) and (12.3-5). The carrier densities may be expressed as

$$n(x) = U_c e^{-(\epsilon_c(x) - \epsilon_f)/kT} = U_c e^{-(\epsilon_{ci} - \epsilon_f)/kT} e^{e\phi(x)/kT} \quad (16.4-1)$$

$$\text{and } p(x) = U_v e^{-(\epsilon_f - \epsilon_v(x))/kT} = U_v e^{-(\epsilon_f - \epsilon_{vi})/kT} e^{-e\phi(x)/kT}. \quad (16.4-2)$$

$$\text{However, since } U_c e^{-(\epsilon_{ci} - \epsilon_f)/kT} = n_i = U_v e^{-(\epsilon_f - \epsilon_{vi})/kT},$$

according to (9.3-6) and (9.3-15), these may be written simply as

$$n(x) = n_i e^{e\phi(x)/kT} \quad (16.4-3)$$

$$p(x) = n_i e^{-e\phi(x)/kT}. \quad (16.4-4)$$

---

<sup>8</sup> W. Shockley and G. L. Pearson, *Phys. Rev.*, **74**: 223 (1948).

If all the donors and acceptors are assumed to be ionized, Poisson's equation takes the form

$$\begin{aligned} \frac{d^2\phi}{dx^2} &= -\frac{4\pi e}{\kappa} (p(x) - n(x) + N_d - N_a) \\ &= \frac{4\pi en_i}{\kappa} \left( 2 \sinh \frac{e\phi(x)}{kT} - \frac{N_d - N_a}{n_i} \right). \end{aligned} \quad (16.4-5)$$

It is hereby assumed, of course, that the symmetry of the system is such that the potential varies only along the  $x$ -direction. If the surface is a plane and if the sample is a uniform one this condition will be realized in practice. If now, we represent the potential by a dimensionless variable  $u$ , where

$$u(x) = e\phi(x)/kT, \quad (16.4-6)$$

then (16.4-5) becomes

$$\frac{d^2u}{dx^2} = \frac{2}{L_{D_i}^2} \left( \sinh u(x) - \frac{N_d - N_a}{2n_i} \right), \quad (16.4-7)$$

where  $L_{D_i}$  is the intrinsic Debye length as defined by (10.2-49). This can be written in a form which is even simpler by noting that far in the interior of the semiconductor, at very large values of  $x$ ,

$$\lim_{x \rightarrow \infty} n(x) = n_0 = n_i e^{e\phi_B/kT} \equiv n_i e^{u_B}$$

whereby

$$u_B = \ln \frac{n_0}{n_i}. \quad (16.4-8)$$

Here  $\phi_B$  (and thus  $u_B$ ) refer to the electrostatic potential far in the interior of the crystal, as shown in Figure 16.7. Expressing the quantity  $u_B = \ln(n_0/n_i)$  in terms of  $N_d - N_a$  by (9.5-23) and recalling that  $\ln(x + \sqrt{1 + x^2}) = \sinh^{-1} x$ , we may finally write (16.4-7) as

$$\frac{d^2u}{dx^2} = \frac{2}{L_{D_i}^2} (\sinh u - \sinh u_B). \quad (16.4-9)$$

It should also be noted that since  $n_0 p_0 = n_i^2$ ,

$$u_B = \ln \frac{n_0}{n_i} = \ln \frac{n_i}{p_0} = -\ln \frac{p_0}{n_i}. \quad (16.4-10)$$

It is possible to integrate (16.4-9) once by letting

$$F = -\frac{du}{dx} = -\frac{e}{kT} \frac{d\phi}{dx} = -\frac{eE}{kT} \quad (16.4-11)$$

where  $E(x) = -d\phi/dx$  is the electric field. Then

$$\frac{d^2u}{dx^2} = -\frac{dF}{dx} = -\frac{dF}{du} \frac{du}{dx} = F \frac{dF}{du}. \quad (16.4-12)$$

Substituting this into (16.4-9) and integrating from  $F = 0$  (corresponding to  $x = \infty$ , far in the interior, where clearly the field must vanish) to an arbitrary interior point, we may obtain the relation

$$F = -\frac{du}{dx} = \frac{2}{L_{Di}} \sqrt{\cosh u - \cosh u_B - (u - u_B) \sinh u_B}, \quad (16.4-13)$$

which connects the value of the field at any point to the value of the potential. This can be integrated once again, this time between the limits  $x = 0$  (corresponding to the surface, where  $u = u_s = e\phi_s/kT$ ) and an arbitrary interior point to give

$$x = -\frac{L_{Di}}{2} \int_{u_s}^u \frac{du}{\sqrt{\cosh u - \cosh u_B - (u - u_B) \sinh u_B}}. \quad (16.4-14)$$

The relationship between the field and potential in the space charge layer just at the surface can be obtained by substituting  $u = u_s$  into (16.4-13), to give

$$F_s = -\left(\frac{du}{dx}\right)_s = -\frac{eE_s}{kT} = \frac{2}{L_{Di}} \sqrt{\cosh u_s - \cosh u_B - (u_s - u_B) \sinh u_B}. \quad (16.4-15)$$

Unfortunately, Equation (16.4-14) cannot in general be integrated in closed form to give an explicit relation describing  $u$  as a function of  $x$ . There is, however, one case in which an exact analytic solution can be obtained, corresponding to a material in which  $N_d - N_a = 0$ , whereby  $n_0 = n_i$  and  $u_B = 0$ . The bulk semiconductor crystal is therefore intrinsic far in the interior, and this particular situation will be referred to as the *bulk-intrinsic* case. Setting  $u_B = 0$  in (16.4-13) it is easy to show that for this particular case

$$F = -\frac{du}{dx} = \frac{2\sqrt{2}}{L_{Di}} \sinh \frac{u}{2}. \quad (16.4-16)$$

this can now be integrated as described above, and the resulting equation solved for  $u$  to give

$$u(x) = 4 \tanh^{-1} \left( e^{-\frac{x\sqrt{2}}{L_{Di}}} \tanh \frac{u_s}{4} \right). \quad (16.4-17)$$

The relation between field and potential in the space charge layer at the surface of the crystal is, for this example, clearly

$$F_s = -\frac{eE_s}{kT} = \frac{2\sqrt{2}}{L_{Di}} \sinh \frac{u_s}{2}. \quad (16.4-18)$$

The surface fields may be related to the surface state charge density, the interior space charge per unit surface area and an external field  $E_0$ , which one may envision as being applied as shown in Figure 16.9, by applying Gauss's electric flux theorem to cylindrical volumes whose axes are normal to the semiconductor surface, as shown in

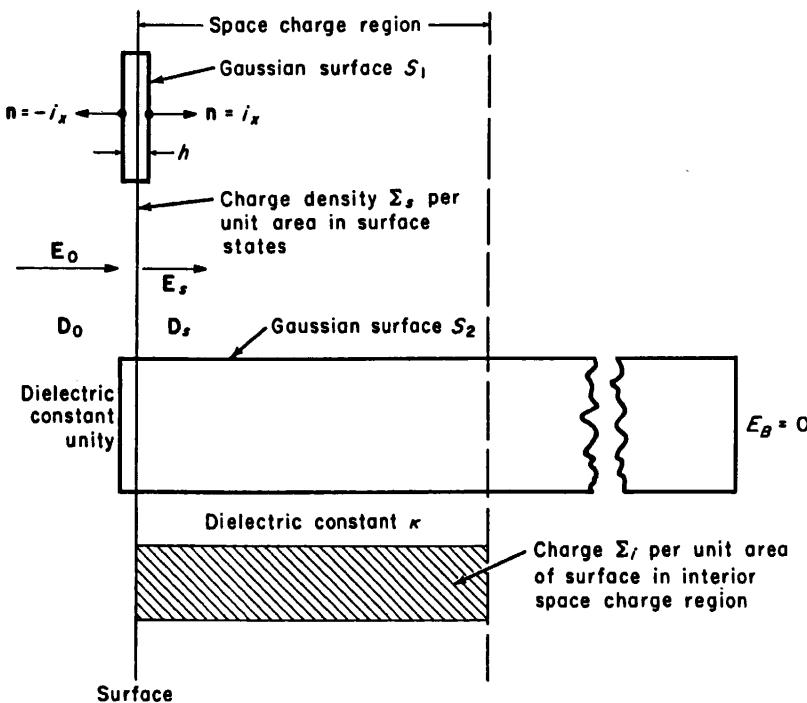


FIGURE 16.10. Diagram illustrating the construction of Gaussian surfaces and notational usage for the calculations of Section 16.4.

Figure 16.10. Considering at first the Gaussian surface  $S_1$  at the top in that diagram, Gauss's theorem<sup>9</sup> requires that

$$\int_{S_1} \mathbf{D} \cdot \mathbf{n} da = 4\pi q \quad (16.4-19)$$

where  $\mathbf{D}$  is the electric displacement vector and  $\mathbf{n}$  is the outward normal. Clearly on the lateral surfaces of the cylindrical volume  $\mathbf{n}$  and  $\mathbf{D}$  are mutually perpendicular, and the dot product vanishes. On the ends of the cylinder, we obtain

$$\int_S [\mathbf{D}_s \cdot \mathbf{i}_x + \mathbf{D}_0 \cdot (-\mathbf{i}_x)] da = (D_s - D_0)A_1 = 4\pi\Sigma_s A_1 \quad (16.4-20)$$

where  $A_1$  is the area of the end of the cylindrical surface and  $\Sigma_s$  the charge density per unit area in surface states. Since the height of the cylinder may be assumed to be infinitesimally small, the amount of charge within the surface contributed by that part

<sup>9</sup> See, for example, J. R. Reitz and F. J. Milford, *Foundations of Electromagnetic Theory*, Addison-Wesley Publishing Co., Reading, Mass. (1960), p. 77.

of the interior volume space charge which is included therein may be neglected. Since  $D_0 = E_0$  and  $D_s = \kappa E_s$ , the result may be written as

$$\kappa E_s - E_0 = 4\pi \Sigma_s. \quad (16.4-21)$$

Gauss's theorem may now be applied in the same way to surface  $S_2$ ; since the field far in the interior vanishes, the result will be

$$-E_0 = 4\pi(\Sigma_s + \Sigma_i) = 4\pi\Sigma_t \quad (16.4-22)$$

where  $\Sigma_i$  is the charge per unit surface area in the *interior* space charge region and  $\Sigma_t$  represents the *total* charge per unit surface area. By substituting the value given by (16.4-21) for  $\Sigma_s$  into (16.4-22), it is easily seen that the relation between the interior field strength  $E_s$  at the surface and the total interior space charge density  $\Sigma_i$  is just

$$E_s = -\frac{4\pi\Sigma_i}{\kappa}. \quad (16.4-23)$$

If the density and distribution of surface states is known, and if the externally applied field  $E_0$  is given, the value of the surface potential  $u_s$  may be calculated. Once this is known, the boundary conditions on the interior potential are established and the interior potential and field and the interior space charge density may be determined unambiguously. To illustrate this more clearly, let us consider the bulk-intrinsic case where there are  $N_s$  acceptor-type surface states per unit surface area, all at energy  $\epsilon_s$ . The surface state density ( $\text{cm}^{-2}$ ) may then be represented as

$$g_s(\epsilon) d\epsilon = N_s \delta(\epsilon - \epsilon_s) d\epsilon. \quad (16.4-24)$$

The energy level associated with the surface states is located at some value which is fixed relative to the conduction and valence band edges at the surface. Since the latter vary with the value of the surface potential so also does the surface state energy  $\epsilon_s$ ; in fact, referring to Figure 16.11, it is clear that

$$\epsilon_s = \epsilon_{s0} - e\phi_s, \quad (16.4-25)$$

where  $\epsilon_{s0}$  is the energy difference between the surface level and the intrinsic point in the forbidden gap. Then, since in the bulk-intrinsic case  $\epsilon_f = 0$ , we may write  $n_s$ , the number of electrons in surface states per unit surface area as

$$n_s = \int_{\epsilon_c}^{\epsilon_v} f_0(\epsilon) g_s(\epsilon) d\epsilon = N_s \int_{\epsilon_c}^{\epsilon_v} \frac{\delta(\epsilon - (\epsilon_{s0} - e\phi_s))}{1 + e^{(\epsilon - \epsilon_f)/kT}} d\epsilon = \frac{N_s}{1 + e^{(\epsilon_{s0} - e\phi_s)/kT}}. \quad (16.4-26)$$

The surface charge density  $\Sigma_s$  is then

$$\Sigma_s = -en_s = -eN_s f_0(\epsilon_{s0} - e\phi_s), \quad (16.4-27)$$

and from (16.4-21) and (16.4-18) an equation for the interior surface field  $E_s$  can be

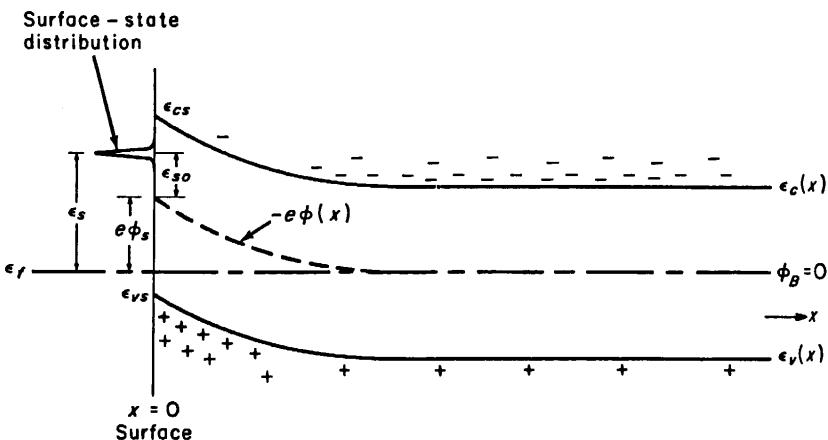


FIGURE 16.11. Potential diagram and notational usage for the calculation of surface potential as a function of applied field discussed in Section 16.4.

written in the form

$$E_S = \frac{4\pi\Sigma_S}{\kappa} + \frac{E_0}{\kappa} = \frac{-4neN_S}{\kappa} \frac{1}{1 + e^{\epsilon_{SO}/kT} e^{-us}} + \frac{E_0}{\kappa} = \frac{2\sqrt{2kT}}{eL_{Di}} \sinh \frac{us}{2}. \quad (16.4-28)$$

whence [recalling (10.2-49)],

$$\frac{1}{1 + e^{\epsilon_{SO}/kT} e^{-us}} = \frac{E_0}{4\pi e N_S} - \frac{2\sqrt{2n_i L_{Di}}}{N_S} \sinh \frac{us}{2}. \quad (16.4-29)$$

In principle this equation may be solved for  $u_s$  to determine the surface potential. Actually, however, the algebraic difficulties in arriving at an analytical solution are so great<sup>10</sup> that for our purposes a graphical solution, as illustrated in Figure 16.12 is more informative.

In that figure the expressions on the left-hand side of (16.4-29) and on the right-hand side of that equation are shown plotted separately as functions of  $u_s$ . At the intersection of the two curves, the functions are equal, and the  $u_s$  coordinate of that point represents the real root of (16.4-29). The variation of the value of  $u_s$  as a function of applied field may be studied by noting that as  $E_0$  is varied, the  $\sinh$  curve is translated up or down parallel to the  $f(u_s)$ -axis. For  $E_0 = 0$ ,  $u_s$  is invariably negative, corresponding to a positive value for  $-e\phi_s$ . This is to be expected, since the electrons in the surface states create a positive surface space charge layer, hence a region from which electrons are depleted, as in Figure 16.11, leading to a positive value for  $-e\phi_s$ . With the surface state densities usually associated with germanium and silicon surfaces, it is found that the surface potential energy usually may be varied through several  $kT$  units on either side of the zero-field value by the application of experimentally realizable values of  $E_0$  in a field-effect experiment such as that shown in Figure 16.9. Once  $u_s$  is determined as a function of the known quantities  $N_S$ ,  $\epsilon_{SO}$  and  $E_0$  by this

<sup>10</sup> By substituting  $x = e^{-us}$  into (16.4-29) this equation may be shown to reduce to a cubic equation in  $x$ , which has only one real root.

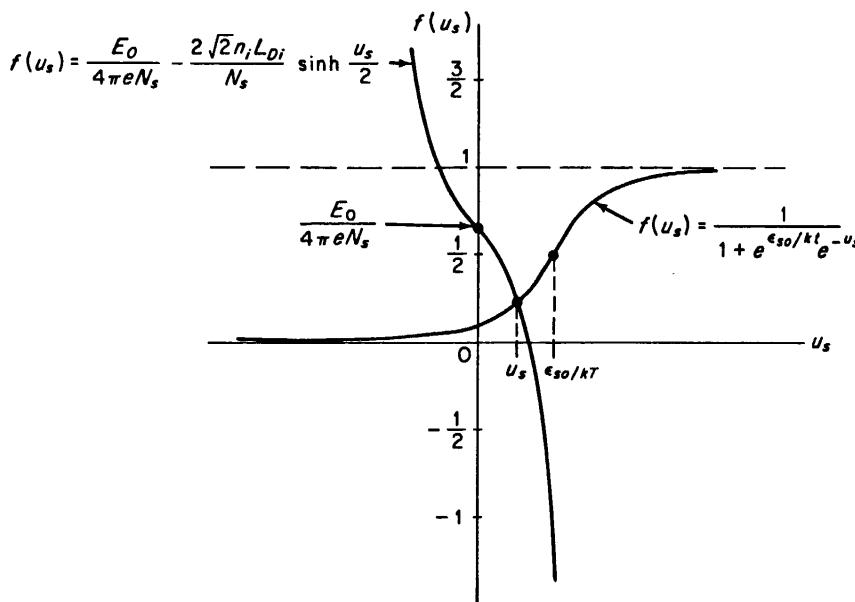


FIGURE 16.12. Diagram illustrating the graphical solution of equation (16.4-29) and the variation of surface potential with applied field and other variables.

method, the precise form of the interior potential follows from (16.4-17), the interior field  $E_s$  at the surface from (16.4-28), the surface state charge density from (16.4-27) and the interior charge density from (16.4-23). The same line of approach may be used with other than bulk-intrinsic samples, although the mathematical work is much more involved.

In actual practice, unfortunately, one ordinarily wishes to reverse this procedure, starting with experimentally measured values of surface and interior charge density as a function of applied field, deriving the corresponding values for  $E_s$  and  $u_s$  as functions of field, and *inferring* from these data the density of surface states and their distribution in energy. This problem is much more difficult, and the answer can usually be obtained only by trying various *assumed* surface state models, modifying them as needed until a satisfactory fit to the experimental data is obtained when the calculation described above is carried out. Some of the subtleties involved in these field-effect measurements and the conclusions which have been drawn from the experimental results will be discussed in the next section.

## 16.5 SURFACE CONDUCTIVITY, FIELD EFFECT, AND SURFACE MOBILITY; PROPERTIES OF ACTUAL SEMICONDUCTOR SURFACES

In the field effect experiment of Figure 16.9, the surface conductance arising from an increase or decrease of the carrier density within the interior space charge region may be measured as a function of applied field. The total induced charge density  $\Sigma$ , may

be inferred from (16.3-1) to be  $V_0/(4\pi r)$ . From these data one may calculate the variation of the surface potential with applied field, and thus (using the procedures outlined in the preceding section) obtain information about the density and distribution in energy of the surface states.

Using the bulk-intrinsic case again as an example, it is clear that if  $\phi_s$  is zero there will be no surface space charge layer and the bulk value of conductivity will prevail even in the surface region. If  $\phi_s$  is positive (thus  $-e\phi_s$  negative) the bands as shown in Figure 16.11 will be bent downward near the surface, and electrons which are thermally generated in the interior will slide down the potential energy barrier and accumulate near the surface, while holes will be repelled from the surface region. In effect, the surface will be *n*-type. The electron concentration near the surface, as a function of  $\phi$ , will be given by (16.4-3), the hole concentration by (16.4-4). It is clear that if  $\phi_s$  is reasonably large the surface electron concentration may be many times what it is in the bulk, and hence the electrical conductivity in the surface layer will be much larger than in the bulk. An increase in sample conductance will therefore result. Likewise, if  $\phi_s$  is negative, the bands will be bent upward, and the resulting accumulation of holes at the surface may lead to a surface layer conductivity which is much larger than the bulk value. It is apparent that the conductance of the sample will reach a minimum near<sup>11</sup>  $\phi_s = 0$  and increase as  $\phi_s$  departs from this value in either direction. The exact values of  $\phi(x)$  and thus  $n(x)$  and  $p(x)$  and the conductivity can be calculated as functions of  $\phi_s$  by the methods outlined in Section 16.4, and therefore an exact plot of sample conductance *versus*  $\phi_s$  (or *versus*  $\Sigma_i$ , which is related to  $\phi_s$  through (16.4-23) and (16.4-18)) may be constructed. In making these calculations, it must be recognized that carriers in a surface layer are subject to a scattering process which does not affect carriers in the bulk, namely surface scattering. The effect of surface scattering in space charge layers has been considered by Schrieffer,<sup>12</sup> who adopted a simple linear potential model to approximate the actual surface potential. It was found that the surface mobility was reduced from the bulk value by a factor depending on the surface potential. In pure germanium, this factor ranges from unity for  $\phi_s = 0$  to about 0.4 at the largest values of  $\phi_s$  which are realizable experimentally. The value of the factor also depends somewhat upon the bulk carrier density.<sup>13</sup>

Such a plot is shown in Figure 16.13 for a near-intrinsic sample of germanium.<sup>14</sup> In the same diagram is plotted the experimentally determined value of conductance change as a function of the total induced charge  $\Sigma_i$ . Consider first some particular observed value of conductance change. The corresponding values of  $\Sigma_i$  and  $\Sigma_t$  can be read off the curves, and since the conductance change is produced entirely by the

<sup>11</sup> The minimum is not exactly at the point  $\phi_s = 0$  due to the fact that the electron and hole mobilities are not equal.

<sup>12</sup> J. R. Schrieffer, in *Semiconductor Surface Physics* (R. H. Kingston, Ed.) University of Pennsylvania Press, Philadelphia (1957), p. 55.

<sup>13</sup> This can be understood by recalling that in samples where the bulk impurity concentration is large the thickness of the surface space charge region is small, and *vice versa*. In relatively pure crystals, the surface space charge layer thickness will generally be much greater than the mean free path for scattering in this region, and the effect of the surface scattering mechanism upon the total surface layer conductance will be insignificant. As the impurity density increases, the entire thickness of the space charge region may become of the order of, or less than, the mean free path. Under these circumstances the contribution of surface scattering to the overall mobility associated with surface layer conductance may be very important.

<sup>14</sup> This particular example is discussed by W. L. Brown, W. H. Brattain, C. G. B. Garrett and H. C. Montgomery in *Semiconductor Surface Physics* (R. H. Kingston, Editor), University of Pennsylvania Press, Philadelphia (1957), p. 115.

space charge region, the value of  $\phi_s$  on the computed curve corresponding to this value of conductance change is the actual value of the surface potential. The horizontal distance between the two curves corresponds to the surface state charge density. The variation of  $\phi_s$  with the surface state charge density  $\Sigma_s$  is thus directly obtainable, and a surface state model which reproduces this observed variation may be sought.

It is found experimentally that the gaseous ambient in which the sample is immersed has a profound effect upon the surface state structure. The zero-field value of the surface potential may, in fact, be varied over a wide range simply by changing the nature of the atmosphere surrounding the sample. For example, it has been found<sup>12</sup> that the presence of ozone in the gaseous ambient produces, in the case of germanium, a large negative value of  $\phi_s$  and thus a highly *p*-type surface, while a moist air ambient produces a large positive value of  $\phi_s$ , hence a strongly *n*-type surface.

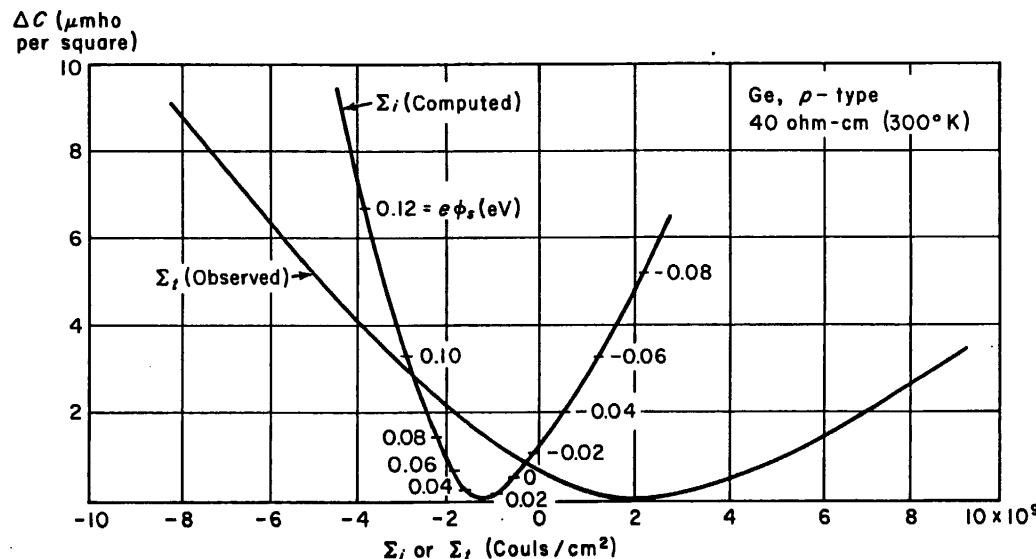


FIGURE 16.13. Observed and computed values of surface charge as a function of surface potential and applied field. [After W. L. Brown, W. H. Brattain, C. G. B. Garrett, and H. C. Montgomery in *Semiconductor Surface Physics*, R. H. Kingston, Editor, University of Pennsylvania Press, Philadelphia (1957), p. 115].

Another readily observed phenomenon is the slow relaxation of the field effect. When a voltage is applied in the experiment shown in Figure 16.9 a change in surface potential and thus a change in interior space charge density is observed immediately, but the surface potential gradually (over a period which, depending upon the surface treatment and temperature, ranges from seconds to hours or days) decays to a value not far removed from the zero-field value. This effect has been explained by assuming that there are two distinct classes of surface states, a group of "fast" states which interact with the interior space charge region very quickly, and a group of "slow" states which may interact only very slowly with the interior of the crystal. The fast states are usually regarded as residing at the surface of the semiconductor crystal, while the slow states are thought to be associated with the outside surface of the oxide layer which is ordinarily present on germanium or silicon samples. This view of the situation is supported by the fact that the relaxation time for the field effect becomes

very long when thick oxide layers are present. The magnitude of the observed "fast" field effects, and the fact that the surface potential almost always relaxes to a value not far from the zero-field value after a long time suggests that the density of fast states on germanium is typically of the order of  $10^{11}$  or  $10^{12} \text{ cm}^{-2}$ , while the slow state density is much higher, perhaps nearly  $10^{15} \text{ cm}^{-2}$ . The effects associated with changes in gaseous ambient are thought to involve primarily the slow surface states, leaving the fast states practically unaffected. The fact that the observed density of fast surface states is much less than the value predicted by the theories of Tamm and Shockley leads one to the conclusion that the fast states cannot be directly identified with Tamm states, at least in the simple form envisioned in those theoretical investigations.

It is apparent also from the preceding results that the surface states and the surface potential distribution must be largely responsible for determining the recombination probability for excess carriers incident upon the surface, and thus for establishing the surface reflection coefficient and the surface recombination velocity. In particular, it is easily seen that the measured values of the surface reflection coefficient do not represent the "real" reflection coefficient of the surface, but rather express the *combined* effect of the surface itself and the associated space charge region upon the recombination of excess carriers in the surface region. It is to be expected, therefore, that the surface recombination velocity should be affected by anything (for example, applied fields, temperature, gaseous ambient) which can change the surface potential, as well as anything which may change the density and distribution of surface levels themselves. A detailed treatment of surface recombination based upon the Shockley-Read recombination theory has been undertaken by Stevenson and Keyes.<sup>15</sup> This theory assumes that the surface states may act as Shockley-Read traps, and predicts a relation between surface recombination velocity and surface potential which is in general agreement with the results of experimental measurements. The results appear to indicate that it is the fast surface states rather than the slow ones which are of primary importance in determining the surface recombination velocity.

## EXERCISES

1. Using the line of reasoning developed in Section 16.1 discuss in detail the nature of an ideal contact, involving no surface states, between a metal and a *p*-type semiconductor.
2. Derive the current-voltage relation for a rectifying contact between a metal and a *p*-type semiconductor ( $\phi_{sc} > \phi_M$ ). Assume that there are no surface states.
3. Discuss the nature of the surface space charge layer which is formed when donor-type surface states (neutral when occupied by electrons, positively charged when empty) are present at the surface of a semiconductor.
4. Show explicitly, using the line of approach adopted in Section 16.3, that the rectifying properties of a metal-semiconductor contact for which  $\phi_M > \phi_{sc}$  are essentially the same as those which are observed when  $\phi_M < \phi_{sc}$  whenever a surface state distribution resulting in a depletion layer is present at the semiconductor surface.
5. Compute the percentage conductivity change to be expected, if no surface states are present, when a potential difference of 1000 V is applied across an air gap of 0.1 mm to a sample of *n*-type germanium containing  $10^{14}$  electrons per  $\text{cm}^3$  at equilibrium which is 0.1 mm thick.

---

<sup>15</sup> D. T. Stevenson and J. R. Keyes, *Physica*, **20**, 1941 (1954).

6. Reproduce the mathematics of the derivation of the electric field, potential and integrated interior space charge density associated with the bulk-intrinsic case.

#### GENERAL REFERENCES

- W. H. Brattain and J. Bardeen, "Surface Properties of Germanium," *Bell Syst. Tech. J.*, **32**, 1 (1953).
- D. R. Frankl, *Electrical Properties of Semiconductor Surfaces*, Pergamon Press, London (to be published 1966).
- R. H. Kingston (editor) *Semiconductor Surface Physics*, University of Pennsylvania Press, Philadelphia (1957).
- R. H. Kingston, "Review of Germanium Surface Phenomena," *J. Appl. Phys.*, **27**, 101 (1956).
- C. G. B. Garrett and W. H. Brattain, "Physical Theory of Semiconductor Surfaces," *Phys. Rev.*, **99**, 376 (1955).
- T. B. Watkins, *The Electrical Properties of Semiconductor Surfaces*, in *Progress in Semiconductors*, Heywood and Co., London (1960), Vol. 5, p. 7.
- J. N. Zemel (editor) *Semiconductor Surfaces*, Pergamon Press, London (1960).

# 1. Introduction

134

135

137

139

144

151

155

155

157

161

167

168

171

172

176

176

179

183

187

195

39

The remarkable electronic and structural properties of semiconductor surfaces and interfaces result from the existence of surface and interface states, respectively. Surface states on clean surfaces originate from dangling bonds and on adsorbate-covered surfaces from bonds between adsorbate and semiconductor-surface atoms. At abrupt metal-semiconductor interfaces, the wavefunctions of those metal electrons, which energetically overlap the semiconductor band gap, decay exponentially into the semiconductor. These tails represent metal-induced interface states. This concept also applies to semiconductor heterostructures and semiconductor-insulator interfaces. Surface and interface states above the bulk valence-band maximum may become charged. Surface charge neutrality then requires the existence of space-charge layers which penetrate from the surface or interface into the semiconductor.

Atoms on ideally terminated semiconductor surfaces exhibit broken bonds which contain one quarter of the number of the respective valence electrons. Adatoms and surface dimers reduce the density of broken bonds and, by this, the total energy of the surface. Further lowering of the total energy is achieved when surface states are either empty or occupied by two electrons. Adsorbates form chemical bonds with semiconductor surface atoms and, by this, saturate their dangling bonds. In both cases the resulting surface band-structures are semiconductor-like.

On both clean and adsorbate-covered surfaces, the saturation of dangling bonds is generally accompanied by local deformations of bond angles while the bond lengths remain almost unchanged. The associated strain energy increases the total energy of the surface. Stable surface structures result when the strain energy is overcompensated by the energy gain which results from a reduction of the density of dangling bonds and/or their saturation.

## 1.1 Historical Remarks

The development of semiconductor surface and interface science is intimately correlated with the evolution of semiconductor device physics. The beginning of both may be dated back to 1874 when *Braun* discovered and first described rectifying behavior of metal-semiconductor contacts. He used metal sulfides such as chalcopyrite, iron pyrite, galena and fahlore and observed 'daß der

## 2 1. Introduction

*Widerstand derselben verschieden war mit Richtung, Intensität und Dauer des Stromes. Die Unterschiede betragen bis zu 30pCt. des ganzen Werthes*<sup>1</sup>. Schuster [1874] confirmed these observations of unipolar conductivity, as he called this deviation from Ohm's law, '*in a circuit composed entirely of copper wires, joined together by means of binding-screws*'. He reported that '*cleaning the ends of the wire with the knife generally destroyed the effect but this unilateral conductivity generally appeared when the wires had some rest*'. Schuster not only confirmed Braun's findings but also discovered cupreous oxide as a new semiconductor. At the same time, Adams and Day [1876] also observed rectification with platinum-selenium contacts.

It took another 30 years before metal-semiconductor contacts were applied as solid-state rectifiers in the early days of radio telegraphy and, later on, in the rapidly growing field of broadcasting. The initially used *cat's whisker* rectifiers consisted of a metal point pressed against a piece of usually lead sulfide and were devices of the kind first described by Braun. They were replaced by plate rectifiers based on cupreous oxide and, later on, selenium. Eventually, such devices were widely applied in power rectification. The successful development of solid-state rectifiers exclusively rested on empirical approaches and is marked by patents issued to Bose in 1904 and to Grondahl in 1925. A physical understanding of the basic phenomena in metal-semiconductor contacts, on the other hand, was still lacking until 1938.

Already Braun [1874, 1876] carried out a number of experiments to find a physical explanation for the rectifying properties of metal-semiconductor contacts. He excluded thermoelectric as well as polarization effects and concluded that the *anomalous phenomena* observed are intimately related to the contact itself and have to originate from a very thin surface layer of extremely high resistance. He hesitated to attribute his observations to '*films of air which, even when they are not electrically conductive, might play a secondary, quasi catalytic role*' since too many of his results clearly contradicted this opinion.

In 1929, Schottky and Deutschmann scanned the potential gradient along cupreous oxide rectifiers, which were biased in reverse direction, by using needle-shaped probes. They demonstrated that the voltage drop occurs across a narrow *blocking layer* at Cu-Cu<sub>2</sub>O interfaces. Furthermore, they measured the differential capacitance of such rectifiers as a function of applied voltage and modeled the barrier layers by a parallel-plate capacitor. From the capacitance measured with no bias applied, they estimated the thickness of the blocking layer to be approximately  $3 \times 10^{-5}$  cm. Since the width of the barrier increases as a function of reverse bias, they definitely excluded transport by tunneling. Furthermore, they already speculated on the possible existence of a negative space charge in the blocking layer of cupreous oxide rectifiers.

<sup>1</sup> He observed that their resistance differed with the direction, intensity and duration of the current. The differences amounted up to 30 per cent of the total value.

## 1.1 Historical Remarks

At that time, they were extremely close to an understanding of the rectifying properties of metal-semiconductor contacts. For several reasons, however, Schottky's explanation that space-charge layers, which are depleted of mobile carriers, exist on the semiconductor side of such interfaces had to wait for publication until 1938. First of all, Wilson published his quantum theory of semiconductors only in 1931. Already at the end of that year, Schottky and Peierls corresponded on the depletion-layer concept as Schottky mentioned in one of his papers published in 1939. Their considerations remained, however, inconclusive since at that time the sign of the Hall coefficient of cupreous oxide was erroneously determined to be negative and it took some time until Fritsch [1935] eventually arrived at the correct and positive sign.

The rectifying properties of Schottky contacts, as metal-semiconductor contacts are generally labeled to honor Schottky's contribution to their understanding, are determined by the respective *work function of the majority carriers*, as Schottky originally called the barrier height. This quantity equals the distance in energy between the Fermi level and the edge of the respective majority carrier band, i.e., the conduction band for *n*- and the valence band for *p*-type semiconductors, at the interface. By using a most simple approach Schottky [1940] and Mott [1938] found the barrier height at metal-semiconductor interfaces as the difference between the work function of the metal and the electron affinity of the semiconductor in contact. Schweikert [1939] was the first to test this rule with metal-selenium rectifiers. His data, which are contained in one of Schottky's publications [1940], revealed a linear relationship between barrier heights and metal work-functions but the slope parameter was less than unity as postulated by the Schottky-Mott rule.

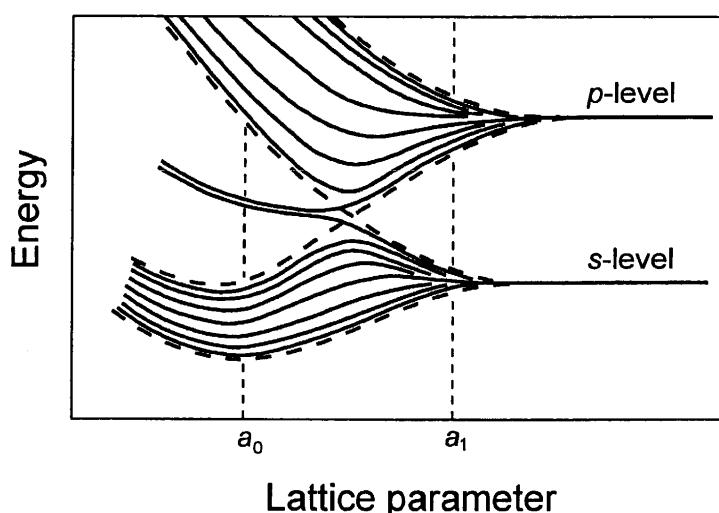
The developments of semiconductor surface physics and of the transistor are also intimately related. As described by Brown [1953], a conducting channel or, in other words, an inversion layer across the *p*-type base of an *n-p-n* transistor, for example, will cause a short-circuit between its emitter and collector. Such surface excess of minority carriers, which are electrons in the case considered here, is induced by positively charged centers which may be electronic interface states at the semiconductor-oxide interface or defects in the oxide layer.

The possible existence of states localized at surfaces of solids was first realized and theoretically investigated by Tamm [1932]. He studied a semi-infinite, one-dimensional array of  $\delta$ -function potential barriers. This idealized potential was introduced by Kronig and Penney [1931] for modeling the electronic band structure of crystalline solids shortly before. Tamm obtained one discrete surface level in each of the energy gaps. Maue [1935] then investigated the existence of surface states within the nearly free electron approximation. He solved Schrödinger's equation for complex wave vectors and obtained a continuum of Virtual Gap States (ViGS) as these *solutions* were called later. Provided the potential satisfies certain boundary conditions, surface states derive from the continuum of virtual gap states of the

complex band structure. Later on, *Goodwin* [1939a,b] reproduced *Maué's* results but also obtained conditions for the occurrence of surface states in a tight-binding approach. However, *Goodwin* [1939c] found ‘*some difficulty in coordinating those results obtained by the different approximations employed*’. In his classic contribution, *Shockley* [1939] reconciled the different findings. He investigated ‘*how the surface levels originate from the atomic levels as the crystal is conceived of as being formed by varying the lattice constant from infinity to a finite value*’. He found the distance between the atoms and, above all, the termination of the crystal potential at the surface to be the most important parameters.

*Shockley's* famous *energy diagram for a one-dimensional lattice with eight atoms* is reproduced in Fig. 1.1. With decreasing lattice parameter the atomic energy levels are first broadened until the boundary curves of the bands, which are shown as dashed lines in Fig. 1.1, are crossing and inverted band gaps, to use *Shockley's* notation, are opening due to new interactions. After that crossing of the bands, one energy level each is split off from the lower and the upper band. For each state that appears in a gap, one level vanishes from a bulk band. The respective wavefunctions of the split-off states exhibit surface-state behavior in that they decay exponentially into both vacuum and the semiconductor bulk.

In constructing the energy diagram shown in Fig. 1.1, *Shockley* has used a periodic potential up to the surface, while in his tight-binding approach *Goodwin* assumed a more realistic, asymmetric termination at the surface. Both potentials are shown schematically in Fig. 1.2. By using the potential proposed by *Goodwin*, *Shockley's* treatment gives surface states lying in the gap just near to each band edge for both uncrossed as well as crossed bands. The potential employed by *Maué* is similar to the one preferred by *Shockley*.



**Fig. 1.1.** Energy spectrum for a one-dimensional lattice with eight atoms. After *Shockley* [1939]

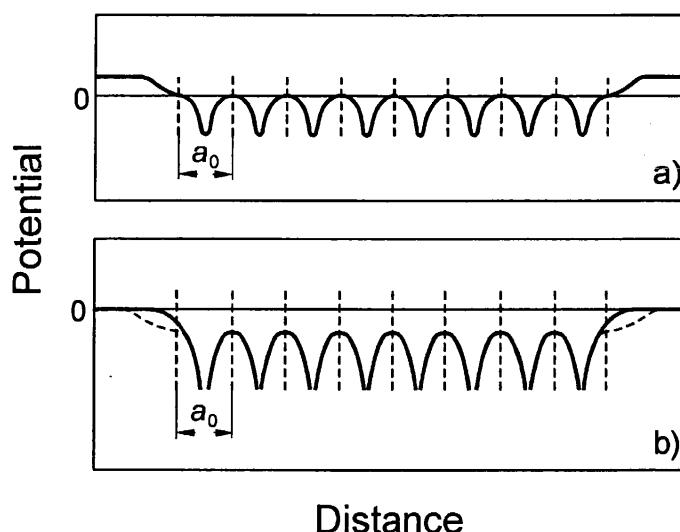
## 1.1 Historical Remarks 5

*Bardeen* [1947] was the first to apply the theoretical concept of surface states to free surfaces and to metal-semiconductor contacts as well. His premise was the condition of charge neutrality at surfaces and interfaces

$$Q_{ss} + Q_{sc} = 0, \quad (1.1)$$

i.e., in thermal equilibrium the surface band-bending adjusts such that a net charge  $Q_{ss}$  in surface states is balanced by a space charge  $Q_{sc}$  below the semiconductor surface. The first and most direct test of this concept was performed by *Shockley* and *Pearson* [1948]. They applied electric fields perpendicular to the surfaces of evaporated Ge films and measured the change of the film conductance. Only 10% of the induced charge were found to consist of mobile carriers. These results were difficult to explain since the films were polycrystalline. In the light of experiments performed later with single crystals, these early *field-effect* experiments had already established the existence of surface states which trapped most of the charge induced. These investigations also provided the basis for the field-effect transistor which was already proposed in the early 1930s by *Lilienfeld* and by *Heil* and was eventually fabricated by *Kahng* and *Atalla* in 1960.

In his pioneering paper, *Bardeen* [1947] also attributed the deviation of experimental barrier heights in metal-semiconductor contacts from the Schottky-Mott rule to the existence of interface states. *Heine* [1965] identified these interface states as the continuum of *Metal-Induced Gap States* (*MIGS*), as they are called now. As surface states on free surfaces, the *MIG* states also derive from the virtual gap states of the complex semiconductor band structure. It has to be emphasized again that virtual gap states are solutions of Schrödinger's equation only which become of physical significance only when the proper boundary conditions are considered.



**Fig. 1.2.** Model potentials in a one-dimensional lattice as used (a) by *Maué* [1935] and *Shockley* [1939] and (b) by *Goodwin* [1939b]. After *Shockley* [1939]

## 6 1. Introduction

By now it is generally agreed that the continuum of metal-induced gap states primarily determines barrier heights of metal-semiconductor contacts. The net charge in these MIG states is due to a charge transfer across the interface. In generalizing *Pauling's* concept [1939] of the correlation between the *Partial Ionic Character of Covalent Bonds* and the *Relative Electronegativity of Atoms*, the charge transfer across semiconductor interfaces may be described by the difference in electronegativities of the two materials in contact [Mönch 1986b]. With this 'boundary condition', the MIGS model yields a linear relationship between barrier heights and electronegativity differences. Since the experimental data do not follow such a simple trend Mönch [1987, 1988a, 1989] proposed that deviations from barrier heights as predicted by the MIGS-and-electronegativity model are caused by secondary mechanisms which might be fabrication-induced defects, strain or dipoles related to interface structure.

First detailed studies of electronic properties of semiconductor surfaces were reported by *Brattain* and *Bardeen* in 1953. They employed the Kelvin method [Thomson 1881, Lord Kelvin 1898] and measured the contact potential between a Pt reference electrode and etched or sand-blasted Ge surfaces. By controlling the composition of the gaseous ambient, to which the samples were exposed under a bell jar, they were able to change the contact potential systematically by 0.5 V. Provided the work function of the reference electrode remains stable, measured contact potential differences equal variations of the work function of the semiconductor as a function of surface treatment. Even then such data are difficult to analyze since the ionization energy as well as the surface band-bending enter into the work function of semiconductors and both quantities may change due to adsorbate-induced variations of the surface dipole and of the distribution of surface states, respectively. By measuring surface conductivity, on the other hand, ambient-induced changes of surface band-bending are most directly probed. An evaluation of such data needs the electron and hole mobilities as a function of surface band-bending. From his measurements of surface conductivity, *Morrison* [1953] concluded that the contact potential differences reported by *Brattain* and *Bardeen* were almost equal to the variations of surface band-bending that he had found.

Further progress in surface science was intimately correlated with the development of experimental tools and techniques which are suited for studies with clean and intentionally modified surfaces. First of all, UltraHigh Vacuum (UHV) was needed for such investigations<sup>2</sup>. Techniques for obtaining

---

<sup>2</sup> From kinetic gas theory, one obtains the number of atoms or molecules striking a surface of unit area in unit time as

$$\frac{dN_{imp}}{dt} = P / \sqrt{2\pi m k_B T}, \quad (1.2a)$$

where  $P$  is the gas pressure and  $m$  is the mass of the molecules or atoms. For room temperature and the gas pressure in Pa, one obtains

$$\frac{dN_{imp}}{dt} = 1.51 \times 10^{19} P / \sqrt{M} [\text{cm}^{-2} \text{s}^{-1}]. \quad (1.2b)$$

## 1.1 Historical Remarks

UHV were well known from the fabrication of electron tubes. However, pressures below  $10^{-5}$  Pa could not be measured until the late 1950s. At that time, the most sensitive devices available for measuring low gas pressure were hot-filament ionization gauges. The performance of such gauges is limited by photoelectrons which are excited from the ion collector by soft X-rays produced by electrons impinging on the positively biased acceleration grid. In conventional triode gauges, electrons are emitted from a filament which is surrounded by a concentric acceleration grid and an outer concentric ion collector. Such devices have X-ray limits of approximately  $10^{-5}$  Pa. *Bayard* and *Alpert* [1950] lowered the X-ray limit of hot-filament ionization gauges to approximately  $10^{-9}$  Pa by using thin wires as ion collectors and by inverting the positions of ion collector and filament. After the Bayard-Alpert ion gauge was invented, surface science rapidly developed.

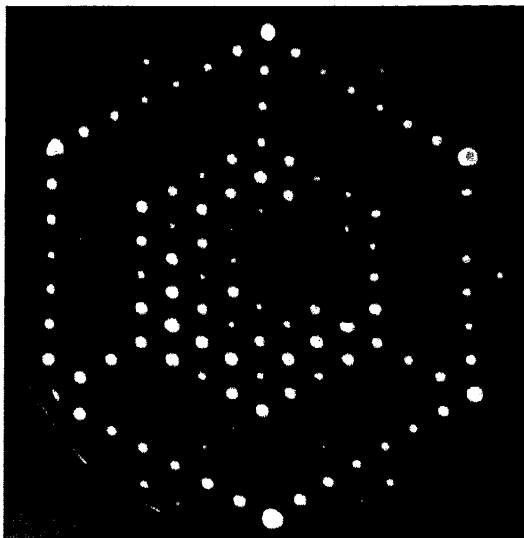
Details on experimental tools and techniques used in surface-science studies are not discussed here. There are a number of profound presentations available as, for example, the books by *Ertl* and *Küppers* [1985], *Woodruff* and *Delchar* [1986], and *Lüth* [1993]. On occasion, the reader will be also referred to overviews on specific experimental techniques.

The first investigations of atomic arrangements on semiconductor surfaces cleaned *in situ* were performed by Farnsworth and coworkers. Their equipment was similar to the one utilized by *Davisson* and *Germer* [1927] in their classic experiments where they demonstrated the wave nature of matter by Low-Energy Electron Diffraction (LEED) at clean nickel surfaces. *Schlier* and *Farnsworth* [1959] detected half-integral diffraction spots with Ge(001) and Si(001) surfaces and seventh-order spots with Si(111) surfaces in addition to the integral-order beams which are expected according to the bulk lattice spacings. At these clean surfaces the unit meshes are obviously larger than in equivalent lattice planes in the bulk. Such surfaces are called to be *reconstructed*. The dimensions of surface unit meshes may be expressed as multiples, m and n, of respective basis vectors in equivalent bulk {hkl} planes [Wood 1964]. Then, surface structures may be classified as (hkl)-m × n. Figure 1.3 displays a LEED pattern of the famous Si(111)-7 × 7 reconstruction.

*Farnsworth* and his coworkers recorded their diffraction patterns by using a movable Faraday cup. Even with an automated system this was a slow and tedious procedure. A huge step forward was made by *Scheibner* et al. [1960] who were the first to realize a display-type LEED system. Such a design with postacceleration of diffracted electrons to excite a fluorescence screen was

---

Here  $M$  is the molecular or atomic weight of the incident molecules and atoms, respectively. At a pressure of  $1.3 \times 10^{-4}$  Pa of oxygen, there are  $3.5 \times 10^{14}$  oxygen molecules impinging per  $\text{cm}^2$  per second. The number of surface sites on solids amounts to approximately  $10^{15}$  per  $\text{cm}^2$ . Provided a surface is clean initially and the sticking probability is one, all surface sites will be covered by oxygen atoms after exposures for approximately 1.5 seconds. Reducing the pressure by a factor of  $10^4$  to  $1.3 \times 10^{-8}$  Pa, i.e., to UHV, increases this time interval to approximately 4 hours.



**Fig. 1.3.** Low-energy electron diffraction (LEED) pattern of the Si(111)- $7 \times 7$  structure

already proposed by *Ehrenberg* in 1934 shortly after *Davisson* and *Germer*'s pioneering work. Further improvements came from *Lander* et al. [1962] who replaced the flat screen used by *Scheibner* et al. by a spherical one. This design avoids any distortions of LEED patterns.

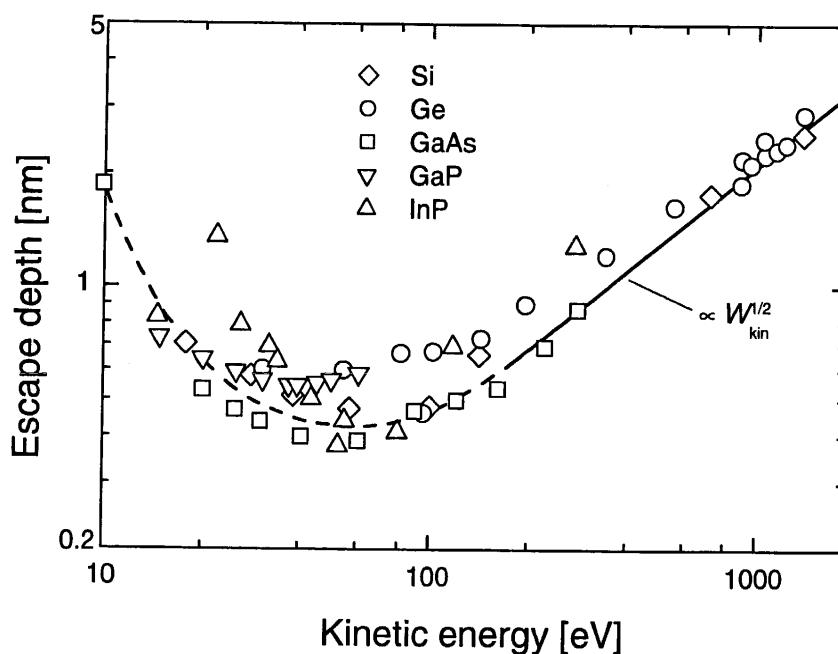
Such LEED optics are designed to operate in the retarding field mode in order to suppress the background of inelastically scattered electrons and, therefore, they may also be used for measurements of energy distribution curves of electrons backscattered from surfaces [Palmberg 1967]. In continuation of earlier work by *Lander* [1953] and *Harris* [1968], *Weber* and *Peria* [1967] used a LEED optics for observing electron-excited Auger transitions on clean and cesiated Si(001) and Ge(001) surfaces. Until X-ray Photoelectron Spectroscopy (XPS) was applied in surface studies Electron-excited Auger Electron Spectroscopy (E-AES) was the method routinely used for determinations of surface composition. A further step forward was made when *Palmberg* et al. [1969] applied a Cylindrical Mirror Analyzer (CMA) as an Auger electron spectrometer. CMAs and Concentric Hemispherical Analyzers (CHA) are the most widely used electron spectrometers in surface science.

Pioneering work on electronic properties of atomically clean silicon surfaces, which were prepared *in situ* by cleavage, was done by *Allen* and *Gobeli* [1962]. They used a Kelvin probe and also measured the spectral dependence of the total photoelectric yield. The exact energy distribution of the surface states could not be obtained from the experimental data acquired by using these experimental techniques. Shortly afterwards, essential features of the surface band structure on cleaved Si(111) surfaces were provided by optical measurements. *Chiariotti* et al. [1968, 1971] were the first to observe direct optical transitions between surface states on cleaved germanium and silicon {111} surfaces by using multiple internal reflection of infrared light. At the same time, *Müller* and *Mönch* [1971] detected indirect optical transitions from bulk states at the valence-band maximum to empty surface states on

cleaved Si(111) surfaces by using photoconductance spectroscopy. These observations already provided essential elements of the band structure of surface states on clean cleaved Si(111) surfaces.

Wavevector- or  $k$ -resolved surface band structures were eventually determined by using Angle-Resolved Ultraviolet Photoemission Spectroscopy (ARUPS) and angle- or  $k$ -resolved Inverse PhotoEmission Spectroscopy (ARIPIES, KRIPIES). Photoemission spectroscopy was widely applied after new sources for ultraviolet light and soft X-rays became available. Differentially-pumped, windowless discharge lamps equipped with monochromators are laboratory line-sources which may be routinely operated to yield photons between 16.8 and 40.8 eV while a continuum of synchrotron radiation became accessible from electron storage rings. In papers published one after the other in 1972, *Eastman* in cooperation with *Grobman* and *Wagner* together with *Spicer* first reported energy distribution curves of electrons photoemitted from cleaved Si(111) surfaces. It took another decade until *Dose* et al. [1981, 1982] were the first to report on empty surface states on GaAs(110) surfaces observed by utilizing Inverse PhotoEmission (IPES) or Bremsstrahlung Isochromat Spectroscopy (BIS).

The mean free path of electrons and thus the escape depth of, for example, photoexcited electrons strongly varies as a function of electron energy. Figure 1.4 displays experimental data for silicon, germanium, and some III-V compounds. For photoemission from core levels, surface- and bulk-sensitive



**Fig. 1.4.** Electron escape depth as a function of kinetic electron energy:  $\diamond$  data for Si from *Flitsch* and *Raider* [1975] and *Garner* et al. [1979],  $\circ$  data for Ge from *Szajman* et al. [1978] and *Gant* and *Mönch* [1981],  $\square$  data for GaAs from *Eastman* et al. [1980] and *Bertness* et al. [1988],  $\nabla$  data for GaP from *McLean* and *Ludeke* [1989],  $\triangle$  data for InP from *Kendelewicz* et al. [1987] and *Bertness* et al. [1988]

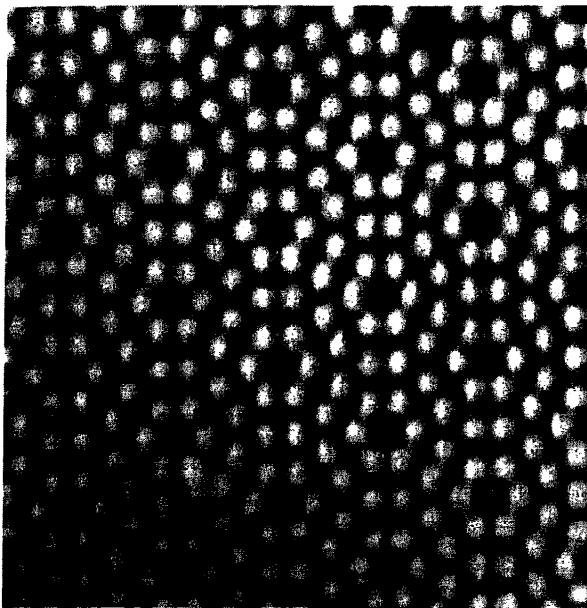
## 10      1. Introduction

conditions may thus be achieved by properly adjusting the photon energy. This is most easily accomplished when synchrotron radiation is used. In the course of such studies, *Eastman* et al. [1980] detected pronounced surface core-level shifts on clean cleaved {110} surfaces of III-V compound semiconductors. On these surfaces, this effect is due to a difference in Madelung energies at the surface and in the bulk [*Mönch* 1986a] while surface core-level shifts observed with elemental semiconductors indicate a charge transfer at the surface.

A major breakthrough in semiconductor surface science occurred after the invention of the Scanning Tunneling Microscope (STM) by *Binnig* et al. [1982a,b] and its extension to scanning tunneling spectroscopy. Already a year later, *Binnig* et al. [1983] achieved atomic resolution with an STM in a study of the most complicated Si(111)- $7 \times 7$  reconstruction. Until then, the large number of atoms per  $7 \times 7$  unit mesh had remained a big obstacle for obtaining the atomic positions of this reconstruction. This challenge provoked numerous structure proposals but the solution had to wait until some of the details of this structure were *seen* by using an STM which had atomic resolution in real space. Figure 1.5 displays an STM image obtained with an Si(111)- $7 \times 7$  surface. *Binnig* et al. were the first to observe that each  $7 \times 7$  unit mesh contains a deep corner hole and 12 protrusion. Some years earlier in 1976, *Harrison* already proposed Si adatoms to be at least one important ingredient of the Si(111)- $7 \times 7$  reconstruction. On Si(111) surfaces, a Si adatom ties up three dangling bonds and thus drastically reduces the number of dangling bonds and by this the electronic band-structure energy. Consequently, *Binnig* et al. identified the protrusions observed as Si adatoms. This was a tremendous step forward towards the hitherto accepted Dimer-Adatom-Stacking fault (DAS) model of the most complicated Si(111)- $7 \times 7$  reconstruction which was eventually proposed by *Takayanagi* et al. [1985b] a few years later.

On semiconductor surfaces, scanning tunneling microscopy probes empty and occupied surface states depending on the polarity of the voltage applied between sample and tip. Therefore, this technique may also be used to map electronic surface properties with, in favorable cases, atomic resolution. The potential of Scanning Tunneling Spectroscopy (STS) was convincingly first demonstrated by *Hamers* et al. [1986b] again in a study of the complicated Si(111)- $7 \times 7$  reconstruction. The technique of STS nicely supplements both angle-resolved photoemission and inverse photoemission spectroscopy.

Another real-space technique is X-ray Standing Waves (XSW) which was increasingly applied in the recent past. Standing wave fields are produced in front of a single crystal when Bragg diffraction occurs. By tilting the crystal through a Bragg reflection, the phase of the diffracted wave with respect to the incident one changes by  $\pi$  so that the position of the antinodal planes of the standing waves varies between coincidence with the diffraction planes and midway between them. Making use of this effect and the fact that ab-



**Fig. 1.5.** Topographic image of a  $7 \times 7$ -reconstructed Si(111) surface as recorded by using a scanning tunneling microscope with a bias voltage of +2 V applied to the sample. From Wiesendanger et al. [1990]

sorption of X-rays will excite, for example, photoelectrons, Auger electrons or X-ray fluorescence, the distance of adatoms normal to the diffraction planes of the substrate may be determined from the intensity variation of the specific excitations selected as the sample is tilted through a Bragg reflection. The respective yield from the adatoms will be largest when an antinodal plane coincides with the position of the adatoms. The technique of XSW was developed by Batterman [1964, 1969] and Golovchenko et al. [1974] for studying sites of foreign atoms in single crystals and was first applied to adatoms by Cowan et al. [1980].

Tremendous progress has also been achieved in computing band structures of surface states as well as arrangements of surface atoms in relaxed  $1 \times 1$  structures and at reconstructed surfaces. Quite a number of different theoretical methods were developed and have been employed<sup>3</sup>. At present, most calculations use either the local-density functional approximation and first-principles pseudopotentials or an ( $s, p, s^*$ ) set of tight-binding parameters [Vogl et al. 1983].

The most severe problem was how to deal with semi-infinite solids since their potentials are no longer periodic in the direction normal to the surface. Most direct approaches will match the vacuum and surface solutions to known bulk solutions. Appelbaum and Hamann [1973] assumed a matching plane between the second and third layer below the surface. Pollmann and Pantelides [1978], on the other hand, applied the scattering theoretic technique, which is a Green's function method, to surfaces and interfaces of

<sup>3</sup> Details on the theoretical concepts used in calculating dispersion curves of surface states on semiconductor surfaces may be found in reviews by Forstmann [1978] and Pollmann [1980] and monographs by Bechstedt and Enderlein [1988] and Lannoo and Friedel [1991].

## 12      1. Introduction

semiconductors. This approach makes full use of the fact that surfaces and interfaces are localized perturbations on crystals. Here, eigensolutions of the unperturbed bulk system are scattered by short-range perturbations which represent the specific surface or interface under study.

Another and most commonly used approach to this problem is provided by the slab method which was introduced by *Hirabayashi* in 1969. Slabs simulate semi-infinite crystals by a small number of layers which has, however, to be sufficiently large so that both surfaces do not interact. An elegant extension of this single-slab method and another way to circumvent the lack of periodicity in semi-infinite crystals is to restore periodic boundary conditions by considering periodic arrays of supercells. This technique was first used by *Schlüter et al.* [1975]. Each of the supercells contains a single slab and some ‘layers’ of vacuum. The number of vacuum layers is chosen such as to minimize the overlap of evanescent waves from surfaces of neighboring slabs. Typical numbers are ten semiconductor and five equivalent vacuum layers per supercell.

The cluster method, on the other hand, deals with a small number of atoms where the bulk is simulated by a termination of dangling bonds with hydrogen atoms. Such approaches generally make use of well-established quantum-chemical methods. Calculations performed by *Goddard et al.* [1978], for example, gave structural surface parameters which are in close correspondence with experimental data. This finding indicates that atomic arrangements at semiconductor surfaces are dominated by rehybridization of orbitals at the surface. Cluster calculations, however, provide no wavevector-resolved properties.

Theoretical approaches, which use the local-density-functional formalism, generally yield band gaps which are typically 30 to 50% less than observed experimentally. *Hybertsen and Louie* [1986] presented a first-principles theory of quasiparticle energies in semiconductors. They evaluated the self-energy operator in the  $GW$  approximation. The self-energy includes the effects of exchange and correlation on single-particle energies and is taken to be the first term in an expansion in terms of the screened Coulomb interaction  $W$  and the electron Green’s function  $G$ . Calculations for cleaved GaAs(110) and Si(111) surfaces reproduced not only the  $k$ -dependence of the occupied but also the empty surface states as well as the widths of the respective band gaps [*Hybertsen and Louie* 1988, *Northrup et al.* 1991a].

Equilibrium structures of crystalline solids may be obtained from calculations of the total energy. For computations of atomic arrangements on semiconductor surfaces, *Chadi* [1978b, 1979a] developed a new approach for a minimization of the total energy. He expressed variations of the total energy associated with atomic displacements as a sum of contributions due to changes in the band-structure energy and an empirical correction term which depends on the fractional change of the distance between nearest neighbors only. *Chadi* computed the band-structure energy by employing a tight-

## 1.2 Surface Space-Charge and Surface States 13

binding approximation. The second term of the total energy then accounts for double counting of electron-electron interactions in the band-structure term and also includes the ion-ion interaction energy. Later on, this empirical tight-binding approach, which gave stable structures in close agreement with experiment, was replaced by first-principles pseudopotential calculations [Ihm et al. 1980].

Meanwhile, these most recently developed theoretical methods were also applied to single adatoms and overlayers of foreign atoms on semiconductor surfaces. As for structural and electronic properties of clean semiconductor surfaces, very good agreement was again obtained with corresponding experimental data. However, these were only case studies which demonstrated the potential of the theoretical methods available and additional theoretical investigations are needed for the many other adsorbates on semiconductors already investigated experimentally.

## 1.2 Surface Space-Charge and Surface States: Some Preliminary Remarks

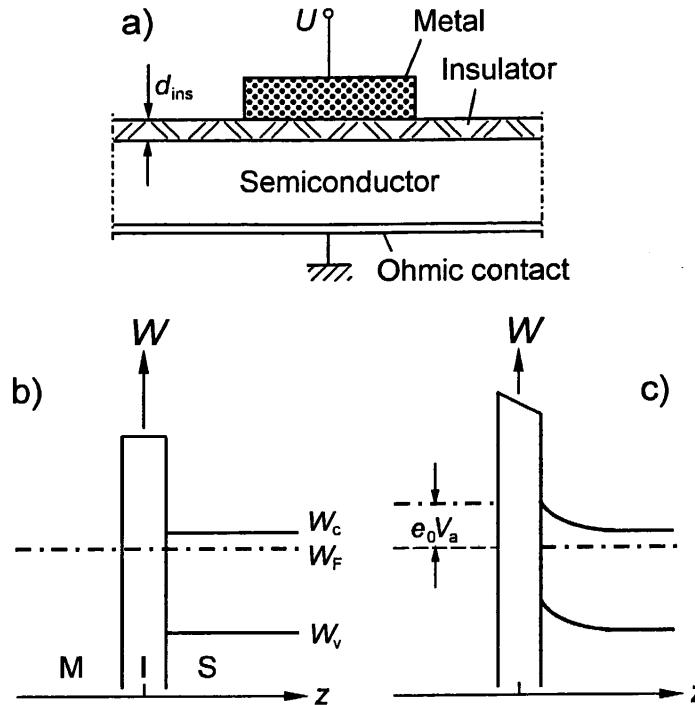
At semiconductor surfaces and interfaces, spatially extended space-charge layers may be present. This behavior distinguishes semiconductors from metals. It is due to the low densities of free carriers in non-degenerately doped semiconductors, which results in large screening lengths, and to the existence of electronic surface and interface states above the valence-band maximum which may become charged. The occurrence of space-charge layers at semiconductor surfaces and interfaces shall be illustrated by two simple examples.

First, an ideal Metal-Insulator-Semiconductor (MIS) structure shall be considered. It is shown schematically in Fig. 1.6a and may be described as a parallel plate capacitor. The semiconductor is assumed to be doped *n*-type and to have no deep impurities and the insulator to contain no charged defects and impurities. In addition, no electronic interface states should exist within the semiconductor band-gap at the insulator-semiconductor interface. This is almost realized in technical SiO<sub>2</sub>/Si interfaces where some 10<sup>9</sup> states per cm<sup>2</sup> and per eV are routinely achieved and the oxide is free of charged centers. When all work functions are identical then the bands of the semiconductor will be flat up to the interface in thermal equilibrium. This is shown schematically in Fig. 1.6b.

A voltage  $V_a$  applied across the insulator of thickness  $d_{ins}$  induces a charge per unit surface area

$$Q_{in} = \epsilon_{ins}\epsilon_0 V_a/d_{ins}, \quad (1.3)$$

where  $\epsilon_0$  and  $\epsilon_{ins}$  are the permittivity of vacuum and the dielectric constant of the insulator, respectively. For an applied voltage of 1 V and 0.1 μm of SiO<sub>2</sub> ( $\epsilon_{ins} = 3.9$ ), (1.3) gives a surface charge of  $3.5 \times 10^{-8}$  As/cm<sup>2</sup> or  $2.2 \times 10^{11}$



**Fig. 1.6a–c.** Ideal metal-insulator-semiconductor (MIS) structure and respective band diagrams in thermal equilibrium and with a negative bias applied to the metal electrode

$e_0/\text{cm}^2$  where  $e_0$  is the elementary charge. With a negative bias applied to the metal electrode with respect to the semiconductor, the surface charge on the semiconductor has positive sign. The induced charge will then be made up by positively charged, static donors and, as a consequence, the surface will become depleted of free and mobile electrons. When all free carriers are neglected in a depletion layer its thickness may be estimated as

$$z_{\text{dep}} = Q_{\text{sc}}/e_0 N_{\text{d}}. \quad (1.4)$$

This approximation is usually referred to as *Schottky's* model [1942]. The space-charge density  $Q_{\text{sc}}$  equals the charge induced per unit area provided no electronic interface states are present in the energy gap of the semiconductor. Otherwise, the induced charge will be distributed between the space-charge layer and interface states, i.e.,

$$Q_{\text{in}} = \delta Q_{\text{sc}} + \delta Q_{\text{ss}}. \quad (1.5)$$

For a donor density  $N_{\text{d}} = 10^{17} \text{ cm}^{-3}$  and the parameters used above, (1.4) gives  $z_{\text{dep}} = 22 \text{ nm}$ . In depletion layers on semiconductors doped *n*-type, the distance from the Fermi level to the conduction-band minimum will be increased or, in other words, the bands will be bent upward. This is shown schematically in Fig. 1.6c. The spatial variation of the band bending may be easily evaluated by solving Poisson's equation. This is discussed in Sect. 2.1.

On metal surfaces the penetration depth of the induced charge is determined by the Thomas-Fermi screening length

## 1.2 Surface Space-Charge and Surface States 15

$$L_{\text{Th}-\text{F}} = [(e_0^2/\varepsilon_0) D_m(W_F)]^{-1/2}, \quad (1.6)$$

where  $D_m(W_F)$  is the number of states per unit volume and unit energy at the Fermi level. For  $D_m(W_F) = 3.45 \times 10^{22} \text{ cm}^{-3}\text{eV}^{-1}$ , which is the experimental value for aluminum, one obtains  $L_{\text{Th}-\text{F}} = 4 \times 10^{-2} \text{ nm}$ . This value amounts to only 14 % of the nearest-neighbor distance of 0.286 nm in aluminum but is by three orders of magnitude smaller than the depletion layer width evaluated for the semiconductor above. The induced charge is thus confined to the very surface of metals but penetrates deeply into non-degenerately doped semiconductors.

This first example examined the formation of surface space-charge layers under non-equilibrium conditions. A second example will deal with space-charge layers at semiconductor surfaces in thermal equilibrium.

One of the most typical electronic properties of a surface is its work function. It is defined as the difference in energy of an electron at rest in vacuum just outside and an electron at the Fermi energy inside the solid, i.e.,

$$\phi = W_{\text{vac}} - W_F. \quad (1.7)$$

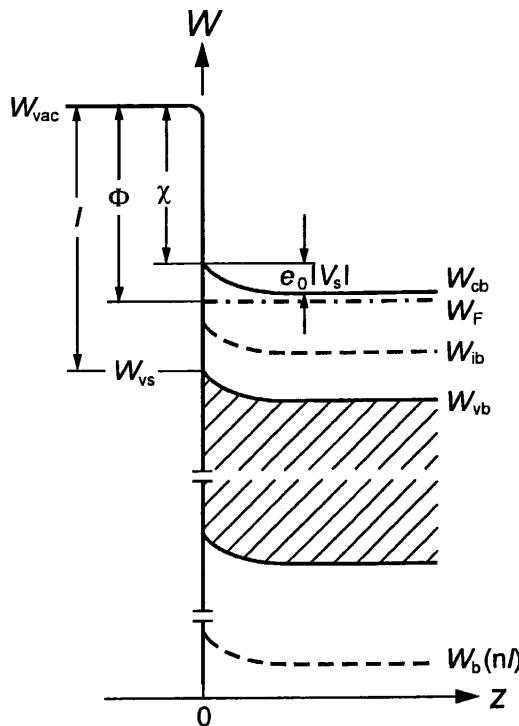
In metals, the Fermi level divides occupied and empty electron energy-levels and, in analogy to atomic notations, the work function of a metal may be considered as its electron affinity as well as its ionization energy. With semiconductors, on the other hand, the filled valence-band and the unoccupied conduction-band states are separated by an energy gap. Electron affinity  $\chi = W_{\text{vac}} - W_{\text{cs}}$  and ionization energy  $I = W_{\text{vac}} - W_{\text{vs}}$  of a semiconductor, which measure the energy difference from the conduction-band bottom and the valence-band top, respectively, to the vacuum level, therefore, differ by the width  $W_g = W_c - W_v$  of the band gap between the valence-band maximum and the conduction-band minimum. By considering the ionization energy in (1.7), the work function of a semiconductor may be conveniently written as

$$\phi = I - (W_F - W_{\text{vs}}). \quad (1.8)$$

The subscript s indicates that the corresponding quantity has to be taken at the surface. In Fig. 1.7 an upward bending of the bands is assumed at the surface of a semiconductor doped *n*-type. This is again the depletion-layer case. Throughout this book the coordinates are chosen as in Fig. 1.7, i.e., the surface is at  $z = 0$  and positive values of  $z$  are pointing into the semiconductor.

The work function of a solid may be experimentally determined by using, for example, the Kelvin method [Thomson 1881, Lord Kelvin 1898] which measures the work-function difference of the surface under study and a vibrating metal probe which is taken for reference. Measurements of the total photoemission yield as a function of the photon energy, on the other hand, give the ionization energies of semiconductors. Therefore, this quantity is also called the photoelectric threshold.

## 16 1. Introduction



**Fig. 1.7.** Energy diagram of an *n*-type doped semiconductor exhibiting an upward band bending at the surface, i.e., a depletion layer (schematically)

Clean (111)-oriented silicon surfaces, which were prepared by cleavage and subsequent heating to at least 650 K in ultrahigh vacuum, shall be considered as an example. Such treatment produces the famous Si(111)- $7 \times 7$  surface structure [Lander et al. 1963]. A typical low-energy electron diffraction (LEED) pattern is displayed in Fig. 1.3. Figure 1.8 shows the room-temperature value of the work function of Si(111)- $7 \times 7$  surfaces to amount to 4.6 eV irrespective of the type of bulk doping [Bachmann 1968]. The ionization energy, on the other hand, measures 5.3 eV [Guichar et al. 1975] so that (1.8) gives the *surface* position of the Fermi level at 0.7 eV above the valence-band maximum irrespective of whether the samples are doped *p*- or *n*-type. Such behavior is commonly referred to as a *pinning of the Fermi level at the surface*. To be somewhat more precise, the margins of experimental error have to be considered in such statements. Measurements using Kelvin probes are accurate to approximately  $\pm 20$  meV. In the case discussed here, the *pinning of the Fermi level at the surface* is thus uncertain within these limits. From the given acceptor and donor densities of the samples considered in Fig. 1.8, the *bulk* positions of the Fermi level are obtained as 0.3 eV and 0.88 eV, respectively, above the valence-band maximum [Richter 1963]. The identical positions of the Fermi level above the valence-band top at both surfaces mean that the bands are bent downward by 0.4 eV and upward by 0.18 eV at the surfaces of the sample doped *p*- and *n*-type, respectively. Since in both cases the energy distance between the Fermi level and the respective majority-carrier band increases towards the surface both samples are depleted of majority carriers beneath their surfaces so that space-charge layers will exist. The space charge is formed by negatively charged acceptors and

## 1.2 Surface Space-Charge and Surface States

17

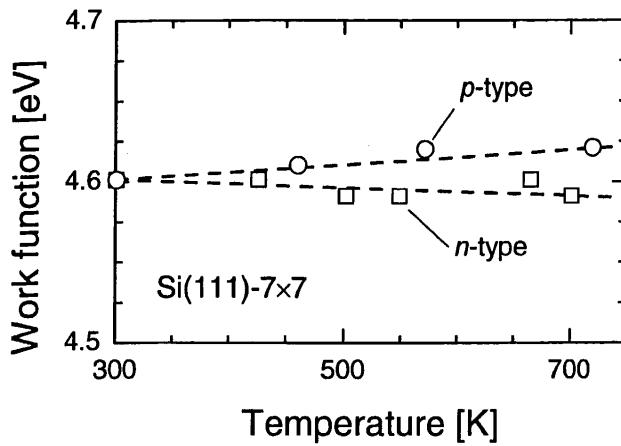


Fig. 1.8. Work function of Si(111)-7 × 7 surfaces as a function of temperature for one sample each doped *p*-type with  $N_a = 4 \times 10^{13} \text{ cm}^{-3}$  and *n*-type with  $N_d = 1.5 \times 10^{16} \text{ cm}^{-3}$ . Data from Bachmann [1968]

positively charged donors at the surface of the sample doped *p*- and *n*-type, respectively.

The experimental data displayed in Fig. 1.8 also show the work function of both samples to be rather insensitive to changes in temperature. This behavior indicates the Fermi level to be pinned at the surface to within approximately 30 meV in the range from room temperature up to 700 K.

Charge neutrality demands that a surface space-charge  $Q_{sc}$  is balanced by an equal amount of charge but of opposite sign. Bardeen [1947] suggested that the compensating charge is found in electronic surface states and he was the first to formulate the condition (1.1) of charge neutrality at semiconductor surfaces. As states in the bulk, surface states have either donor or acceptor character and, depending on their energy position with respect to the Fermi level, they are either neutral or become positively and negatively charged, respectively.

The existence of depletion layers at clean Si(111)-7 × 7 surfaces irrespective of the type of bulk doping, as discussed above, needs an excess  $Q_{ss}$  of positive and negative charge in surface states on samples doped *p*- and *n*-type, respectively. Since at room temperature the Fermi level is found to be pinned to within 30 meV, which are the limits of experimental error for the data presented in Fig. 1.8, the positions of the band edges with respect to the Fermi level need to change only within this energy range in order to turn the net charge in surface states from positive to negative. At Si(111)-7 × 7 surfaces, surface states of both donor and acceptor character are obviously present in sufficient densities in an energy interval of approximately 0.03 eV at 0.7 eV above the valence-band maximum.

Experimental pinning positions of the Fermi level measured with respect to the valence-band top are compiled for various surfaces of elemental and compound semiconductors in Table 1.1. With cleaved Si(111)-2 × 1 and Ge(111)-2 × 1 as well as annealed Si(111)-7 × 7 and Ge(111)-c(2 × 8) surfaces, the Fermi level is always pinned within the bulk band gap. At well-cleaved (110) surfaces of III-V and II-VI compound semiconductors the bands are

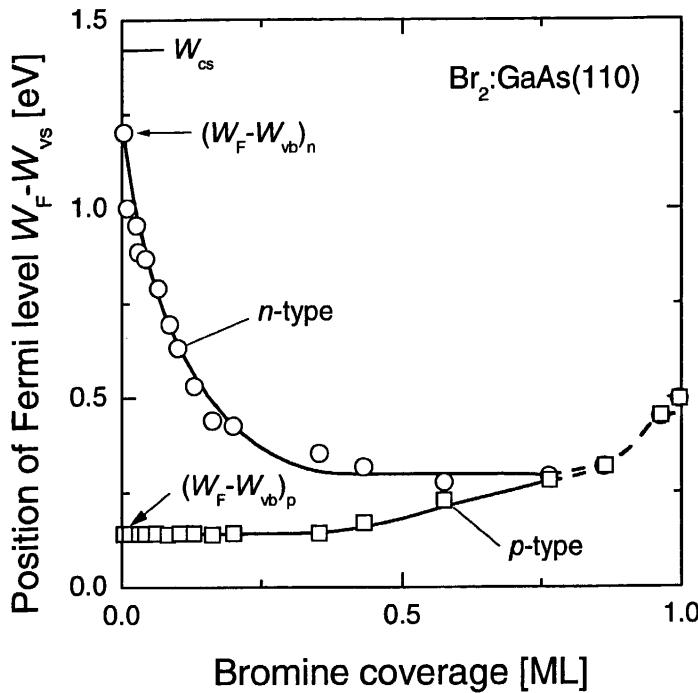
## 18 1. Introduction

**Table 1.1.** Pinning positions of the Fermi level at clean semiconductor surfaces

Semiconductor surface	$W_F - W_{vs}$ [eV]	Reference
C(111)-2 × 1	1.5 ± 0.2	Himpsel et al. [1980b]
Si(111)-2 × 1	0.34 ± 0.08	Allen and Gobeli [1962]
	0.28 ± 0.08	Aspnes and Handler [1966]
	0.36 ± 0.02	Henzler [1967], Mönch [1970]
	0.41 ± 0.05	Fischer [1968]
	0.48 ± 0.04	Sebenne et al. [1975]
	0.40 ± 0.03	Himpsel et al. [1983]
Si(111)-7 × 7	0.70 ± 0.07	Guichar et al. [1975]
	0.78 ± 0.10	Clabes and Henzler [1980]
	0.63 ± 0.05	Himpsel et al. [1983]
	0.65 ± 0.05	Demuth et al. [1986]
Si(100)-2 × 1	0.32	Jeans and Mularie [1971]
	0.34	Himpsel et al. [1980a]
	0.31 ± 0.02	Mönch et al. [1981]
Ge(111)-2 × 1	0.00 ± 0.04	Gobeli and Allen [1964]
	-0.05 ± 0.07	Grant and Webster [1968]
	0.10	Henzler [1968]
	0.00 ± 0.03	v. Wienskowski and Mönch [1971]
Ge(111)-c(2 × 8)	0.105 ± 0.05	Gobeli and Allen [1964]
	0.102	Handler and Eisenhour [1964]
	0.18	Henzler [1968]
	0.095 ± 0.05	v. Wienskowski and Mönch [1971]
AlSb(110)	flat bands	Fischer [1965]
GaP(110)	flat bands	Chiaradia et al. [1989, 1995], Bekacem et al. [1988]
GaP(110) GaAs(110) GaSb(110) InP(110) InAs(110)	flat bands	van Laar et al. [1977]
CdTe(110)	flat bands	Orlowski et al. [1988]

flat up to the surface. At such surfaces the bulk band gap is, therefore, free of any intrinsic surface states. The physical reasons for these observations are discussed in Chap. 7.

During the formation of a clean semiconductor surface, for example, by cleavage, chemical bonds are broken. In a most intuitive but correct picture, surface states on clean semiconductor surfaces originate from such broken or dangling bonds. As substitutional impurities in the bulk, adatoms will



**Fig. 1.9.** Position of the Fermi level above the valence-band maximum as a function of Br-coverage on cleaved *n*- and *p*-GaAs(110) surfaces. Data from Cierocki et al. [1992]

form covalent bonds with surface atoms of the semiconductor substrate. This saturation of dangling bonds removes the clean-surface surface states and replaces them by new, adsorbate-induced surface states. This behavior is most easily followed with adsorption on cleaved {110} surfaces of III-V compound semiconductors. As long as such surfaces are clean their bulk band gaps are free of any intrinsic surface states and, consequently, the bands are flat up to the surface.

Adsorbate-induced surface states within the bulk band gap are then expected to cause surface band-bending. As an example, surface states induced by bromine adatoms on cleaved GaAs(110) surfaces shall be considered. Variations of the surface position of the Fermi level with respect to the valence-band maximum, which are induced by bromine adsorbed on cleaved GaAs(110) surfaces at room temperature, are displayed in Fig. 1.9 [Cierocki et al. 1992]. Throughout this book, adsorbate coverages will be given in monolayers which are scaled by the number of atoms per unit area in an equivalent bulk {hkl} plane. For GaAs(110) surfaces, one monolayer (1 ML) then means  $8.85 \times 10^{14}$  adatoms per  $\text{cm}^2$ . As a function of Br-coverage, the bands remain flat on the sample doped *p*-type while they bend upward or, in other words, a depletion layer forms on the *n*-sample. This behavior is understood by the formation of Br-induced surface acceptors within the bulk band gap. According to the condition of surface charge neutrality (1.1), the positive space charge in the Br-induced depletion layer on *n*-GaAs is compensated by a net negative charge in Br-induced surface states which are consequently of acceptor type. Obviously, their energy position is above the Fermi level in the bulk of the *p*-sample so that they are neutral in this case. On the *n*-sample,

## 20 1. Introduction

the Fermi level becomes pinned at 0.3 eV above the valence-band maximum for coverages larger than 0.2 ML. Above this limit, the space charge and by this the net charge in surface states remains constant and, as a consequence, additional Br-induced surface states are uncharged.