On the theory of Schottky barrier heights

E. von Faber (1) and **G. Paasch** ** (2)

(1) Technische Hochschule Ilmenau, PSF 327, O-6300 Ilmenau, F. R. G.

(2) Zentralinstitut für Festkörperphysik und Werkstofforschung Dresden, PSF, O-8027 Dresden, F. R. G.

September 20, 1991

Abstract

There are several approaches to determine Schottky barrier heights. Mostly, properties of the free surfaces were used to predict this quantity. In other theories the Schottky barrier height is related to characteristics of the semiconductor only. The shortcomings connected with these procedures are discussed and illustrated numerically. A general theory for the Schottky barrier height is presented allowing for a deeper understanding of Schottky barrier formation. Using this theory (i) the influence of doping on the barrier height is treated more straightforwardly than before. Then, (ii) it is shown why previous theories failed to predict the dependence of the barrier height on the metal even for ionic semiconductors, and (iii) it is explained why theories calculating the barrier height directly from the semiconductor band structure are successful mainly for the noble metals. (iv) In addition the formation of Ohmic contacts is considered.

1. Introduction

The metal–semiconductor contact was the first semiconductor interface which got technological importance resulting in a large number of papers on Schottky barrier formation (for reviews see [1,2,3,4]). The first theoretical models were given by Schottky [5] and Mott [6]. Both authors stated that the rectifying behaviour of the contact is caused by a space–charge layer connected with the occurence of a band bending and a corresponding barrier for the charge carriers as a result of the initial equalizing of the chemical potentials during the formation of the contact. But measured barrier heights are not influenced by the metal as much as supposed by these early theories. According to Bardeen [7] this stabilization can be traced back to intrinsic surface states of the semiconductor. On the other hand, Heine [8] stressed that true surface states do not exist at the metal–semiconductor interface. They are changed by tailing off the metal wavefunctions into the semiconductor band gap. Heine's fundamental idea was further developed by Yndurain [9], Tejedor, Flores & Louis [10], Louie, Chelikowsky & Cohen [11], Tersoff [12], and others.

Stimulated by new experimental observations especially since 1977 there is an increasing interest especially in the interface charge redistribution on an atomic length scale. In particular, high reactivities, interdiffusion between the metal and the semiconductor, and induced defect centers at the semiconductor side were suggested to determine the Schottky barrier height alone [1]. The

---

** to whom correspondence should be addressed

discussion on the origin and the nature of the microscopic interface charges is still in progress. However, metal induced gap states do exist in any case. In addition, besides this microscopic charge redistribution there is a band bending in the semiconductor. This band bending also contributes to the total interface dipole which is built up at the interface in order to achieve equilibrium. This was clearly depicted by Cowley & Sze [13] already in 1965. So, the decomposition of the total interface dipole into a microscopic contribution and the band bending turns out to be the most serious problem in determining Schottky barrier heights. However, another problem, the calculation of the magnitude of the required total interface dipole, determined by the equilibrium condition was not solved completely. This also holds for the hitherto most fundamental theory of Duke & Mailhiot [14].

In this paper we present an equilibrium condition being both exact and applicable. From this condition the total interface dipole can be calculated without the need of utilizing the jellium model and without using surface properties (Section 2). Contrary, in the literature the interface dipole is always determined using surface properties which in fact are absent at the interface. We discuss the problems connected with this procedure. In addition, theories [12] calculating the barrier height directly from the semiconductor band structure are examined critically. In Section 3 the equilibrium condition is related to the barrier height. The band bending and the microscopic dipole layer contributions to the total interface dipole are coupled due to the semiconductor gap states. Then a new expression for the barrier height is derived and related to those known from the literature. In Section 4 the influence of doping is analysed, and in Section 5 the influence of the metal on the barrier height is dicussed in detail. In Section 6 the formation of Ohmic contacts is considered. The results presented here show the importance both to correctly account for the equilibrium condition and to determine the interface dipole directly from the bulk charge densities and not using surface quantities.

2. Equilibrium condition

As the starting point we consider a unique system consisting of two half infinite solids. The difference in (electro–)chemical potentials has to be equalized by the charge transfer resulting in thermodynamic equilibrium. Hence, the initial difference in chemical potentials is equal to the interface dipole P_o arising from this charge transfer. But the (electro–)chemical potential depends on the energy zero and the problem is to find out the appropriate reference level for the chemical potentials. In principle this is known for a long time. Nevertheless, the theory is not well established as will be shown here. Having two solids with real surfaces the initial difference in chemical potentials is given by the contact potential u_K defined as the difference of the work functions. But the contact potential is not suitable for describing the charge transfer producing the contact properties because the work functions contain surface dipoles which are absent at the contact [7,8,9]. So, one has to remove the surface dipoles before calculating the "initial" difference in chemical potentials [10,14].

For one half–infinite solid the charge distribution without the surface dipole is a cut off (total) bulk charge density $\varrho_{co}(\vec{r}) = \varrho(\vec{r})\Theta(z_{co} - z)$ where $\varrho(\vec{r})$ is the charge density of the infinite system. After parallel averaging the cut off density has an electrostatic potential that is bulk-like inside and which has to be constant outside the cut-off plane $z = z_{co}$. Therefore, just this cut-off potential represents the appropriate reference level necessary to determine the initial difference in chemical potentials [15]. Hence, we have to calculate the chemical potential relative to the cut-off

potential φ_{co} outside:

$$\mu^b = \mu - \varphi_{co} . \tag{2.1}$$

The definition of the chemical potential μ^b is unique only for a unique definition of the cut-off plane z_{co}. It will be shown here that z_{co} is determined by the bulk charge density $\varrho(\vec{r})$ alone. First, all charge densities are averaged parallel to the interface (mathematical the $x - y$−plane). Let a be the periodicity length of the bulk density in z−direction. Then the bulk density $\varrho(z)$ has the following properties for all z:

$$\int_{z-a}^{z} dz' \, \varrho(z') = 0 , \qquad \varphi(z) = \varphi(z - a) . \tag{2.2}$$

Thus, the neutrality requirement alone does not fix z_{co}. However, using (2.2) and Poisson's equation one obtains that the potential $\varphi_{co} = \varphi(z \geq z_{co})$ outside the cut off density $\varrho_{co}(z) = \varrho(z) \Theta(z_{co} - z)$ is constant only if the first moment of the bulk density $\varrho(z)$ vanishes:

$$\int_{z_{co}-a}^{z_{co}} dz \, z \, \varrho(z) = 0 . \tag{2.3}$$

The cut-off plane z_{co} is uniquely determined by (2.3) for a given surface orientation and possibly a given top atomic layer for a more complicated crystal structure. Having determined the cut-off plane z_{co} using (2.3) one can calculate the cut-off potential φ_{co}. Its value relative to the averaged bulk electrostatical potential $< \varphi(z) >$ will be denoted by

$$\Delta\varphi_{co,\,a} \equiv \varphi_{co} - <\varphi> \qquad \text{with} \qquad <\varphi> = \frac{1}{a} \int_{z_{co}-a}^{z_{co}} dz \, \varphi(z) . \tag{2.4a}$$

and equals

$$\Delta\varphi_{co,\,a} = -\frac{e}{2\varepsilon_o a} \int_{z_{co}-a}^{z_{co}} dz \, z^2 \, \varrho(z) > 0 \tag{2.4b}$$

where (2.2)(2.3) were used. The sign of $\Delta\varphi_{co,\,a}$ in (2.4b) arises from the fact that the positive ionic charge is more localized than the electron density. Now one can rewrite (2.1) as

$$\mu^b = \overline{\mu} - \Delta\varphi_{co,\,a} \tag{2.5a}$$

with

$$\overline{\mu} \equiv \mu - <\varphi> \tag{2.5b}$$

being a pure bulk property. Hence, the chemical potential relative to the potential outside the cut off density μ^b can be calculated from bulk properties using (2.3) to (2.5).

In order to calculate the chemical potentials μ_s^b for semiconductors (lower index "s") besides (2.5a) a further splitting of this quantity is advantageous. The energetic distance of the chemical potential from the conduction band edge in the bulk

$$\zeta \equiv \varepsilon_c - \mu = \varepsilon_v + E_g - \mu \tag{2.6}$$

depends strongly on temperature and doping. Together with (2.6) one obtains for (2.1)

$$\mu_s^b = \varepsilon_v^b + E_g - \zeta \tag{2.7a}$$

where

$$\varepsilon_v^b = \varepsilon_v - \varphi_{co} \tag{2.7b}$$

is the valence band edge relative to the cut-off potential. In (2.7a) doping and temperature influence mainly ζ alone. So, the quantity ε_v^b is analogue with the volume contribution μ_m^b of metals (lower index "m") with respect to its calculation: No gap problem of the density functional theory occurs in determining the band structure and no doping influences the result. Beyond that the definition of ε_v^b will be very helpful later. Frensley & Krömer [16] calculated the valence band edges relative to the mean interstitial electrostatic potential $\varepsilon_v^{FK} = \varepsilon_v - \overline{v_i}$. As a rough estimate $\overline{v_i} \simeq \varphi_{co}$ can be used and hence one has

$$\varepsilon_v^b \simeq \varepsilon_v^{FK} = \varepsilon_v - \overline{v_i} . \tag{2.8}$$

It should be mentioned that the original idea of Frensley & Krömer giving rise to calculate ε_v^{FK} is very different from the derivation given above.

Now we turn back to contacts. The total system in equilibrium consists of two sticked cut off densities $\varrho_{co,\,s}(z)$ and $\varrho_{co,\,m}(z)$ and the interface dipole layer charge density $\delta\varrho_0(z)$:

$$\varrho_0(z) = \delta\varrho_0(z) + \varrho_{co,\,s}(z) + \varrho_{co,\,m}(z) \tag{2.9a}$$

with

$$\varrho_{co,\,s}(z) = \varrho_s(z)\,\Theta(z_{co,\,s} - z) \tag{2.9b}$$

$$\varrho_{co,\,m}(z) = \varrho_m(z)\,\Theta(z - z_{co,\,m}) . \tag{2.9c}$$

(The semiconductor is situated within the left half space ($z < 0$), and the metal is on the right hand side ($z > 0$)). Before equilibrium is reached ($\delta\varrho_0 = 0$) the difference in chemical potentials of the two solids is given by the difference of their volume contributions μ^b since the boundary condition for the electrostatic potential requires $\varphi_{co,\,s} = \varphi_{co,\,m}$ (Fig. 1a). This difference has to be compensated by the interface dipole (Fig. 1b)

$$P_0 = -\mu_m^b + \mu_s^b \tag{2.10}$$

connected with the occurence of the deviation $\delta\varrho_0(z)$ of the total charge density from the sticked cut off density by

$$P_0 = -\frac{e}{\varepsilon_0} \int\limits_{-\infty}^{+\infty} dz\, \delta\varrho_0\, z . \tag{2.11}$$

Eqs. (2.9) to (2.11) represent the equilibrium condition. In addition mechanical equilibrium has to be taken into consideration. But this influences contrary to the jellium calculation of Mailhiot & Duke [17] only the shape of the charge transfer $\delta\varrho_0(z)$ and the position of the ions in the semiconductor relative to the ions in the metal and not the value of P_0. Reconstructions (if any) have to be considered calculating $\delta\varrho_0(z)$ whereas P_0 is fixed by the bulk systems including interface orientations (and top layers).

From (2.10) and (2.7a) one obtains a new equation for the interface dipole:

$$P_0 = -\mu_m^b + \varepsilon_v^b + E_g - \zeta . \tag{2.12}$$

Whereas ζ depends strongly on doping and temperature, the gap E_g is a volume property, and μ_m^b and ε_v^b depend on the respective value of φ_{co} and hence on the orientation of the interface, but they are completely determined by the two bulk electronic structures. This interface dipole is realized as follows. At the interface there will be a charge redistribution on an atomic length scale producing a microscopic contribution D_o. But due to the possibility of charged interface states one has in addition a space–charge layer with an extension of the Debye screening length $l_D \gg a_o$ (a_o: Bohr radius). Its total band bending V_o also contributes to P_o:

$$P_o = V_o + D_o . \tag{2.13}$$

This equation together with (2.12) represent the equilibrium condition for metal-semiconductor contacts keeping in mind both (2.9) and (2.11). (According to (2.11) the band bending is positive if the bands are bent upwards.)
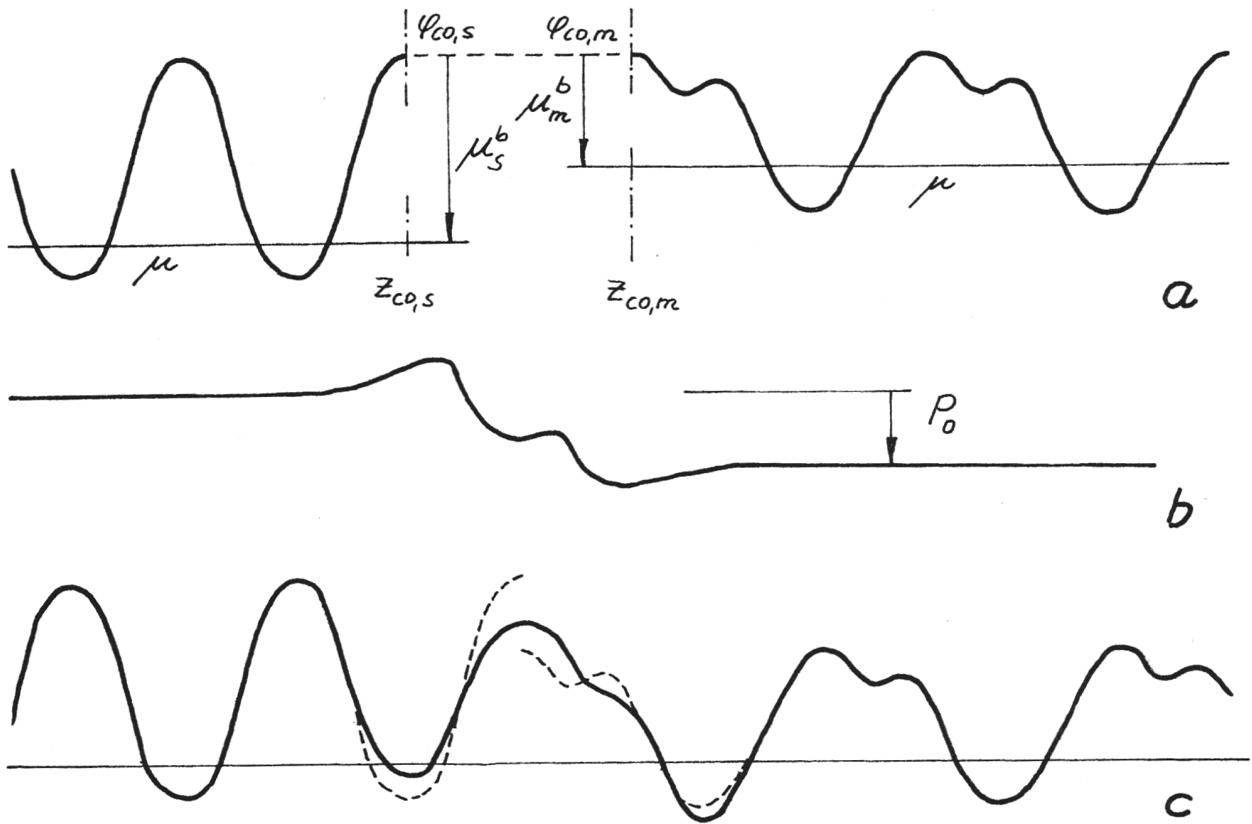

FIG. 1: Electrostatic potentials (a) of the two cut off densities $\varrho_{co,s}$ and $\varrho_{co,m}$, (b) of the interface dipole layer charge density $\delta\varrho_o$, and (c) of the total charge density ϱ_o in thermodynamic equilibrium. – (Note that in this picture, contrary to Schottky contacts, the density $\delta\varrho_o$ is extended only microscopically.)

At the end of this section three approaches of the literature will be discussed. At first, in the literature [5,13] often the difference of the work functions

$$u_K = \phi_m - \phi_s \tag{2.14}$$

is used instead of P_o. Here the surface dipoles are neglected (see below). The correct splitting of the work function ϕ into surface dipole P and volume contribution is given by [18]

$$\phi = P - \mu^b . \tag{2.15}$$

Inserting (2.15) into (2.10) one obtains

$$P_o = \phi_m - \phi_s - P_m + P_s \, . \tag{2.16}$$

But this equation is valid only if the surface orientations at the free surface and at the contact are the same in each case. Sometimes [10,17] (2.16) is used to determine P_o. But then three interfaces have to be analysed. This introduces additional difficulties which are absent in the direct calculation of P_o presented above (see Section 5 for more details). The interface dipole was also written as [14,17]

$$P_o = -\overline{\mu}_m + \overline{\mu}_s \tag{2.17}$$

which is valid for the jellium model only. A correct and complete description of the interface dipole including a correct definition of reference levels for the chemical potentials and a description of the charge densities causing the quantities involved was missing in the past. With the eqs. (2.3) to (2.5) and (2.7) one can calculate P_o from bulk properties without the restriction on the jellium model.

3. The Schottky barrier

At metal-semiconductor contacts one is interested in the relative position of the band structures of the two materials near the interface ($z \simeq 0$). Mostly, the chemical potential lies within the band gap at a distance ϕ_o above the valence band edge. ϕ_o is given by (Fig. 2)

$$E_g - \zeta = V_o + \phi_o \, . \tag{3.1}$$

The Schottky barrier according to the methods to measure it [4,19] is the minimum energy for an electron on the metal side to cross the interface into the bulk of the semiconductor. In the case of depletion layers (with monotoneous band bending) the barrier is determined by ϕ_o. For n-type semiconductors (Fig. 2) the barrier is just

$$\phi_b = E_g - \phi_o = \zeta + V_o \tag{3.2}$$

whereas for p-material one has

$$\phi_b^p = \phi_o^p = E_g - \zeta^p - V_o^p \, . \tag{3.3}$$

But the barrier measured via transport is lowered both for electrons and holes due to the image force correction ΔV_{im} [4]. The most serious theoretical problem is the decomposition of the interface dipole P_o (2.12) into the two contributions V_o and D_o (2.13) since only V_o enters the barrier (3.2). This does not change if V_o in (3.2) is replaced by D_o using (2.12) (2.13) giving

$$\phi_b = E_g - \phi_o = -\mu_m^b + \varepsilon_v^b + E_g - D_o \tag{3.4}$$

since V_o and D_o are coupled as will be discussed in the following. Before doing this we note that the incompatibility of the two length scales for V_o and D_o caused sometimes confusion in graphical representations of the contact in the literature [1,10,13]. Whereas D_o can be seen only as a part of the crystal potential, V_o is a macroscopic potential drop.

Any shift of the position ϕ_o of the chemical potential within the gap (due to a change of the band bending, cf. (3.1)) leads to a change of the charge of the gap states. So, the microscopic contribution D_o and the band bending V_o (or ϕ_o) are coupled. Formally, one can expand D_o as

$$D_o = D_{on} - \alpha(\phi_o - \phi_n) + \cdots \tag{3.5}$$

where the index "n" denotes a situation deviating slightly from the actual equilibrium condition. By conviniently defining this state the linear expansion coefficient α can be determined. It is assumed that the state "n" is realized if the exponentially decaying semiconductor gap states are occupied (up to the level ϕ_n) such that the lack of valence band states due to the boundary condition is compensated (local neutrality, see [20]). According to Tejedor, Flores & Louis [10] $D_{on} = D_e$ originates in the smooth matching of the valence band states.
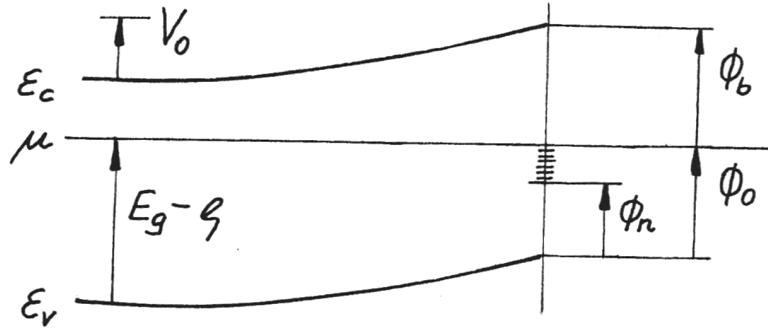

FIG. 2: Energy diagram of a Schottky contact in the scale of the Debye screening length.

Then, if the density of gap states N_{vs} changes only slightly near ϕ_n, the charge per unit area of the gap states is

$$\sigma_{vs} = -e N_{vs} (\phi_o - \phi_n) . \tag{3.6}$$

This exponentially decreasing gap charge [8] with decay length $(2q)^{-1}$ yield the contribution

$$D_{o,vs} = \frac{e}{2q\, \varepsilon(2q)\, \varepsilon_o} \sigma_{vs} \tag{3.7}$$

to D_o. Here the dielectric constant for the wave number $2q$ has to be used [11]. The charged gap states will be screened partly by a charged layer in the metal with an extension of $\beta_m^{-1} \simeq a_o$ giving

$$D_{o,m} = -\frac{e}{\beta_m\, \varepsilon_o} \sigma_m . \tag{3.8}$$

Further σ_{vs} and σ_m are screened by a space–charge depletion layer having the total charge σ_d. But since in as good as all cases $|\sigma_d| < 10^{-4} As/m^2$ one has (see Section 4)

$$\sigma_m + \sigma_{vs} = -\sigma_d \simeq 0 . \tag{3.9}$$

Then one can add D_e, $D_{o,vs}$, and $D_{o,m}$ giving

$$D_o = D_e - \alpha (\phi_o - \phi_n) \tag{3.10}$$

with

$$\alpha = \left(\frac{1}{2q\, \varepsilon(2q)} + \frac{1}{\beta_m}\right) \frac{e^2 N_{vs}}{\varepsilon_o} . \tag{3.11}$$

This approximation was derived originally by Cowley & Sze [13] using additional artifical model assumptions. It was often used and modified [10,11,21].

Inserting (3.10) into (3.4) one obtains for the barrier and for D_o

$$\phi_b = E_g - \phi_o = \frac{1}{1 + \alpha} (\tilde{P} - D_e) + \frac{\alpha}{1 + \alpha} (E_g - \phi_n) \tag{3.12a}$$

$$D_o = \frac{1}{1+\alpha} D_e - \frac{\alpha}{1+\alpha}(E_g - \phi_n - \tilde{P}) \tag{3.12b}$$

with

$$\tilde{P} - D_e = -\mu_m^b + \varepsilon_v^b + E_g - D_e , \tag{3.13}$$

where $\tilde{P}$ is determined only by the two bulk band structures. Eq. (3.12a) differs from those known from the literature [4,9,10,11,13,21] just by the term $\tilde{P} - D_e$ (3.13). An expression as (3.12b) for the microscopic dipole contribution was not derived before. Of course, with (3.12a) also the band bending is given as $V_o = \phi_b - \zeta$ (3.2). If ionic relaxation at the interface will be important the corresponding contribution to the dipole must be included in D_e. It should be noted that the eqs. (3.12) (3.13) contain only three approximations: (i) The charge of the gap states is linearized (3.6). (ii) In (3.10) a third term is omitted being proportional both to the total charge $-\sigma_d$ (3.9) and the effective extension of the immediate interface region. (iii) There is a principal uncertainty in splitting up the total interface dipole P_o (2.13). This uncertainty is just the change of the band bending over the extension of the microscopic dipole layer.

In the literature in the derivation of ϕ_b just given eq. (2.16) was used instead of (2.12). Then together with $\phi_s = \zeta + \chi_s + V_s$ [22] and $P_s = V_s + D_s$ [18] one has

$$\tilde{P} - D_e = \phi_m - \chi_s - D_j , \qquad D_j \equiv D_m - D_s + D_e . \tag{3.14}$$

Here χ_s is the electron affinity, and V_s is the total band bending both at the free surface. $P_m = D_m$ and D_s are the atomic-scale surface dipole of the metal surface and the microscopic part of the semiconductor surface dipole, respectively. Eq. (3.14) is correct. But here three interfaces with correct surface orientations have to be analysed precisely. Contrary, in (3.13) μ_m^b and ε_v^b are determined by the bulk electronic structure as shown in Section 2. Although (3.14) can be found in the work of Tejedor, Flores & Louis [10] it was first derived starting with bulk properties (cf. (2.17)) by Duke & Mailhiot [14] but under some limiting assumptions. In both papers the model of Smith [23] was used to calculate D_m and D_s and a similar one appropriate to bimetallic interfaces to calculate D_e. Probably they got D_j to be negligible for this reason. Then with $D_j \simeq 0$ (one "major result" in [17]) one obtains from (3.14) the result of Cowley & Sze [13], (3.12a) with

$$\tilde{P} - D_e = \phi_m - \chi_s , \tag{3.15}$$

originally derived using u_K instead of P_o (and $V_s = 0$). In the following we demonstrate the advantage of using (3.13) instead of (3.14) (3.15) [24].

4. The influence of doping on the barrier height

Empirically it is well known that the barrier height ϕ_b and hence the distance ϕ_o of the chemical potential from the valence band edge (Fig. 2) is independent of the doping level. This means especially $\phi_o = \phi_b^p$ and can be expressed with (3.2), (3.3) as the common relation

$$\phi_b + \phi_b^p = E_g . \tag{4.1}$$

If the barriers are measured in transport experiments the right hand side of (4.1) is $E_g - (\Delta V_{im} + \Delta V_{im}^p)$. Previous derivations of (4.1) used special assumptions as the doping independence of the electron affinity [4] or the pinning condition $N_{vs} \to \infty$ [9]. Other rather complicated model calculations [17,25] led to a limiting doping density $N_{A,D} \simeq 10^{17} cm^{-3}$ for the validity of (4.1).

Here a rather general derivation of (4.1) will be presented. Eq. (3.4) shows that ϕ_b can depend on the doping level only via the microscopic dipole barrier D_o. The dependence of D_o on doping is possible principally (i) by the relative position of the metal and the semiconductor band structures at the interface which is described by ϕ_o and (ii) by the total charge $-\sigma_d$ at the interface compensating the space–charge in the semiconductor (boundary conditions). The space–charge is extended over the Debye–length $l_D = (kT\,\varepsilon\,\varepsilon_o/e^2 N_{A,D})^{1/2}$ with the static dielectric function ε and is connected with the band bending V_o. On the other hand the neutralizing interface charge has an effective extension of roughly $\delta_{eff} = (2q\,\varepsilon(2q))^{-1}$. Hence the resulting potential drop due to σ_d is roughly

$$V_o \frac{\delta_{eff}}{l_D} = \Delta\phi_o \qquad (4.2)$$

giving rise to a shift of ϕ_o. It is negligible if $\Delta\phi_o \ll 0.1 V_o$. This leads to the limit $N_{A,D} \ll 4 \cdot 10^{18} cm^{-3}$ (in accordance with [17,25]). Under this condition D_o is independent of the total space–charge σ_d and can depend only on ϕ_o. The second equation (3.4) is then

$$D_o(\phi_o) - \phi_o = -\mu_m^b + \varepsilon_v^b \qquad (4.3)$$

where the function between D_o and ϕ_o is not affected by the doping. Here the right hand side does not depend on doping. Since the left hand side depends only on ϕ_o, this quantity itself and hence (with (3.2)) also the barrier ϕ_b is independent of doping. Having shown the validity of (4.1), it is seen together with (3.2) that although the interface dipole P_o (2.12) depends on ζ (depending on doping) the band bending V_o also depends on ζ in the same way, and $V_o + \zeta = \phi_b$ is constant. This rather general result shows that indeed for a doping below the given limit the contact properties can be calculated by considering the microscopic interface region only which is done in supercell calculations [11].

5. The influence of the metal on the barrier height

One important parameter in the relation (3.12a) for the barrier height is α (3.11). To estimate α we use the values for N_{vs} and $(2q)^{-1}$ from Tejedor, Flores & Louis [10] and from Louis, Yndurain & Flores [21], and $\varepsilon(2q) \simeq 2.5$ [11]. In addition for the metallic decay length roughly $\beta_m^{-1} = 0.5$ Å [15] is used. Resulting values for α are given in Tab. 1. It is seen that α increases with decreasing gap width E_g.

For values $\alpha \geq 10$ one could expect from (3.12a) that

$$\phi_b \simeq E_g - \phi_n \qquad (5.1)$$

is already a good approximation. The barrier height ϕ_b is then determined by the neutrality level ϕ_n alone which itself depends on the matching of the gap states with the metal wave functions. According to its definition usually ϕ_n is assumed to be a property of the bulk semiconductor band structure alone. Then, (5.1) would be independent of the metal. Although (5.1) was often used in this sense [12] the experimental values show a dependence on the metal. Even for the narrow-gap semiconductors with a large α (i. e. $\alpha^{-1} \ll 1$), ϕ_b differs by 0.1 eV or more for junctions with different metals. Hence, either $(\tilde{P} - D_e)/\alpha$ is not negligible in (3.12a) or ϕ_n depends on the metal. In the following the influence of the metal on ϕ_b via $(\tilde{P} - D_e)$ will be discussed. Thereby the advantage of using (3.13) instead of (3.14) (3.15) will be demonstrated.

TABLE 1 : Gap width, density of gap states, their extension, α from (3.11), ε_v^{FK} and E_B

	E_g (eV)	N_{vs} $(10^{18}eV^{-1}m^{-2})$	$(2q)^{-1}$ (Å)	α	ε_v^{FK} (eV)	E_B (eV)
ZnS	3.60	1.96	1.56	3.99	-5.34	–
GaP	2.27	3.12	2.94	9.45	-4.12	0.81
GaAs	1.43	3.72	4.81	16.32	-3.96	0.50

Contrary to (5.1) in the literature one can find various attempts [2,18] to systematize phenomeno-logically the measured barrier heights in terms of metal properties alone [1,4,19]. One of the frequently used expression is

$$\phi_b = S'\phi_m + C' \tag{5.2}$$

with the metal work function ϕ_m. S' and C' are fit parameters which are independent of the metal. Systematic applications of the assumed dependence (5.2) has shown mainly, that for ionic semiconductors S' is large (order of magnitude is one) and hence the metal (or its work function) influences the barrier significantly whereas for covalent semiconductors ϕ_m is less important ($S' \ll 1$). The transition between these two limits is described to be rather abrupt [26]. To compare (5.2) with the theory we note that apparently from (3.12a) with (3.14) one can obtain a similar dependence

$$\phi_b \propto \frac{1}{1+\alpha} \left(\phi_m - D_m - D_e \right) . \tag{5.3}$$

Terms depending only on the semiconductor are not included in (5.3). Evidently (5.3) is identical with (5.2) with

$$S' = \frac{1}{1+\alpha} \qquad \text{if} \qquad \frac{d}{d\phi_m} \left(D_m + D_e \right) = 0 . \tag{5.4}$$

That means if the microscopic dipoles D_m at the metal surface and D_e at the interface to the semiconductor are such that $D_m + D_e$ does not depend on the metal. This will be the case especially if $D_j = D_m - D_s + D_e$ (3.14) is zero: superposition of D_m and D_s to the interface dipole D_e. Indeed, in the literature [14,17] D_j was calculated to be negligible. But the numerical values for S' calculated from the first equation (5.4) with (3.11) [10,21] are systematically smaller than those fitting the experiment with (5.2). Even for the ionic ZnS eq. (5.4) gives with Tab. 1 $S' = 0.2$ whereas $S'_{exp} > 0.6$. The conclusion in [10,21] was that the calculated values of α are too large. To clarify this controverse situation we notice that instead of (5.3) from (3.12a) with (3.13) one obtains alternatively

$$\phi_b \propto \frac{1}{1+\alpha} \left(-\mu_m^b - D_e \right) . \tag{5.5}$$

Since D_e arises from the smooth matching of the valence densities whereas D_m, D_s occur at the free surfaces one will have $| D_e | \ll | D_m |, | D_s |$. Then, ~~from (5.5) and (5.2)~~ one obtains *

$$\frac{d\phi_b}{d(-\mu_m^b)} \simeq \frac{1}{1+\alpha} , \qquad\qquad S' \equiv \frac{d\phi_b}{d\phi_m} \tag{5.6}$$

* *from (5.5) (neglecting D_e) and (5.2)*

respectively. It is now important that the volume contributions μ_m^b for a large number of metals are expected to vary by about 3 eV [27] whereas ϕ_m varies only by about 1 eV. The work function is stabilized around 4... 5 eV, though $-\mu_m^b$ (due to the attractive ionic potential) varies much stronger (Heine & Hodges [27]). This means

$$\frac{d\phi_b}{d(-\mu_m^b)} \quad < \quad S' \equiv \frac{d\phi_b}{d\phi_m} \tag{5.7}$$

and hence with (5.6)

$$\frac{d\phi_b}{d(-\mu_m^b)} \simeq \frac{1}{1+\alpha} \quad < \quad S' \equiv \frac{d\phi_b}{d\phi_m} \tag{5.8}$$

as mentioned above.

TABLE 2 : For GaAs–metal contacts experimental values for the barrier, the metal volume contribution $-\mu_m^b$ determined from (3.12a) (3.13) with the experimental ϕ_b, and the experimental metal work functions ϕ_m (all values in eV)

		Au	Pt	Ag	Cu	Al
ϕ_b^{exp}	[19]	0.95	0.94	0.93	0.87	0.80
$-\mu_m^b$	(3.12a) (3.13)	3.80	3.63	3.46	2.42	1.21
ϕ_m	[29]	5.30	5.52	4.48	4.66	4.23

Now (3.12a), (3.13) will be applied explicitly, and thereby the validity of (5.7) and (5.8) will be demonstrated numerically. To do this we use

$$\varepsilon_v^b \simeq \varepsilon_v^{FK} = \varepsilon_c^{FK} - E_g , \qquad \phi_n = E_B , \qquad D_e = 0 , \tag{5.9}$$

where E_B is the effective mid-gap level of Tersoff [12,28]. The approximation for D_e has been discussed already. Only for some semiconductors (as GaAs and GaP) values for ε_v^{FK} [16], E_B [12,28], α [10,21] can be found together with measured barrier heights [19]. More important is the lack of knowledge of the reliable values for $-\mu_m^b$. Therefore we proceed as follows. At first we consider GaAs–metal contacts for the respective metal using the experimental barrier heights, work functions, eq. (5.9) and $\alpha = 16.32$ from Tab. 1. One has a linear dependence of the barrier on the resulting values $-\mu_m^b$ (Tab. 2) with a slope $d\phi_b/d(-\mu_m^b) = 0.06$. Further, using the experimental work functions one can determine a mean value for S', second equation (5.6), giving $S' = 0.11$ in accordance with our theoretical inequality (5.7). Next, we use now the values $-\mu_m^b$ just determined (Tab. 2) to calculate with (3.12a), (3.13) the barrier height for GaP–metal junctions ($\alpha = 9.45$). For the five contacts considered in Tab. 3 the deviation from experiment is 0.02 ... 0.2 eV. Alternatively one can use (3.12a) with (3.15) with the experimental values $\phi_m - \chi_s$. The resulting barrier heights deviate from the experiment about 0.02 ... 0.22 eV. By linearly interpolating the dependence of the experimental barrier heights on $-\mu_m^b$ and on the metal work function ϕ_m we obtain

$$\frac{d\phi_b}{d(-\mu_m^b)} \simeq 0.09 , \qquad \frac{1}{1+\alpha} = 0.10 , \qquad S' \equiv \frac{d\phi_b}{d\phi_m} = 0.23 . \tag{5.10}$$

Again these values confirm the theoretical relations (5.7) and (5.8). This result and Tab. 3 demonstrate clearly the advantage of the description developed here. Nevertheless, it should be emphasized that in general there is no linear dependence of the barrier on ϕ_m, D_m, χ_s, or D_s apparently seen in (3.12a) with (3.14) since the dipoles D_e, D_m, D_s are non–linearly connected with the other quantities. But strictly, there is no linear dependence between ϕ_b and $-\mu_m^b$, ε_v^b, and E_g.

TABLE 3 : The barrier height ϕ_b for GaP–metal contacts: experimental values and theoretical ones from (3.12a) (3.13) with $-\mu_m^b$ from Tab. 2, and from (3.12a) (3.15) using experimental values for $\phi_m - \chi_s$, respectively. (all values in eV)

		Au	Pt	Ag	Cu	Al
ϕ_b^{exp}	[19]	1.34	1.52	$\simeq 1.26$	1.34	1.14
ϕ_b	(3.12a) (3.13)	1.51	1.49	1.47	1.37	1.26
ϕ_b	(3.12a) (3.15)	1.47	1.49	1.39	1.40	1.36
$\phi_m - \chi_s$	[29,30]	1.52	1.74	0.70	0.88	0.45

At the end of this section the following peculiarity of (3.12a) will be discussed. Eq. (3.12a) yields (5.1) exactly not only in the limit $\alpha \to \infty$ but also for arbitrary α if

$$\tilde{P} - D_e = E_g - \phi_n , \qquad \text{or} \qquad (5.11a)$$

$$\phi_b = -\mu_m^b + \varepsilon_v^b + E_g - D_e = E_g - \phi_n \qquad (5.11b)$$

if (3.13) is used. The values $-\mu_m^b$ can be taken from Tab. 2 and the other values contained in (5.11b) with (5.9) are compiled in Tab. 4. Indeed (5.11b) is valid in some cases as demontrated in Tab. 5. At first it is seen from Tab. 5 that in these cases $\phi_b = E_g - \phi_n$ is fulfilled with accuracy. But the last column shows that indeed this quantity coincides well with $\tilde{P} = -\mu_m^b + \varepsilon_v^b + E_g$ showing that (5.11) is valid even with $D_e = 0$ except in the case of the Si–Cu junction. Another point of view would be the following. The second equation (5.11b) is

$$-\varepsilon_v^b - \phi_n + D_e = -\mu_m^b . \qquad (5.11c)$$

Neglecting D_e one obtains for the left hand side for a large number of semiconductors 2.8 ... 4.1 eV. But for most of the semiconductors this value equals 3.5 eV or more. Such large values for the right hand side $-\mu_m^b$ can occur mainly for noble metals. Therefore, mainly for noble metals (5.11) can be fulfilled and one has indeed (5.1). But this does not signify pinning due to an "infinite" density of states.

TABLE 4 : Valence band energies of Frensley & Krömer [16], effective midgap energies of Tersoff [12,28] and gap energies for some semiconductors

	ε_v^{FK} [16] (in eV)	E_B [12,28] (in eV)	E_g (in eV)
Si	−3.16	0.36	1.11
GaAs	−3.96	0.50	1.43
InP	−4.58	0.76	1.26
GaSb	−3.89	0.07	0.72

TABLE 5 : Independently determined values for eq. (5.11b)

	ϕ_b^{exp} [19] (in eV)	$E_g - \phi_n$ (in eV)	$-\mu_m^b + \varepsilon_v^b + E_g$ (in eV)
Si–Cu	0.75	0.75	0.37
GaAs–Ag	0.93	0.93	0.93
InP–Au	0.49	0.50	0.48
GaSb–Au	0.61	0.65	0.63

6. Ohmic contacts

Often no rectifying behaviour is desired equivalent to negative barriers ϕ_b or ϕ_b^p, respectively. From (3.4) it is seen that for n-type semiconductors ϕ_b is negative if (with $\varepsilon_c^b \simeq \varepsilon_c^{FK}$)

$$D_o > -\mu_m^b - (-\varepsilon_c^{FK}) \qquad (6.1)$$

is valid. Since in this case the chemical potential at the interface lies above ϕ_n and even in the conduction band the microscopic interface dipole D_o is negative. Therefore, according to (6.1) Ohmic contacts are expected only if $-\mu_m^b$ is small and $-\varepsilon_c^{FK}$ is large. Indeed, for semiconductors with $-\varepsilon_c^{FK} < 2.8 \ eV$ no Ohmic contact was observed [19]. In the left part of Tab. 6 the values ε_c^{FK} for three typical semiconductors having $-\varepsilon_c^{FK} > 2.8 \ eV$ are shown. But also for CdS (Tab. 6) no Ohmic contact was observed, although CdS–Al–contacts ($-\mu_m^b$ small) yield already a very small barrier. For InP $-\varepsilon_c^{FK}$ is larger. Accordingly, the contact with Al is Ohmic whereas those with the noble metals ($-\mu_m^b$ large) are rectifying [19]. For InAs $-\varepsilon_c^{FK}$ is largest. But in this case $\phi_n = E_B$ is situated within the conduction band (Tab. 6). Therefore, if $\phi_b < 0$ the dipole D_o can be positive and according to (6.1) Ohmic contacts are formed also for the noble metals ($-\mu_m^b$ large) what is verified by experimental investigations [19].

For p-type semiconductors the barrier ϕ_b^p is negative if

$$D_o < -\mu_m^b - (\varepsilon_v^{FK}) \qquad (6.2)$$

what is obtained using (3.4) (3.3) with $\varepsilon_v^b \simeq \varepsilon_v^{FK}$. Here the chemical potential at the interface is situated within the valence band (below ϕ_n) and D_o is therefore positive. Then, from (6.2) it is seen that Ohmic contacts can only be formed if $-\mu_m^b$ is large and $-\varepsilon_v^{FK}$ is small. Note, that the situation is contrary to n-type semicondutors (6.1) and that $-\varepsilon_v^{FK} > -\varepsilon_c^{FK}$. At the right hand

side of Tab. 6 three semiconductors having smallest values $-\varepsilon_v^{FK}$ are listed. But also for Si (as well as for Ge) having smallest values $-\varepsilon_v^{FK}$ no Ohmic contact is formed [19] since the right hand side of (6.2) can not exceed about 0.8 eV. For GaAs and other semiconductors $-\varepsilon_v^{FK}$ is larger and also no Ohmic contact is observed [19]. But for GaSb the neutrality level $\phi_n = E_B$ is situated near the valence band edge (Tab. 6) and even if the barrier is negative, D_o is positive but small. Therefore, the condition (6.2) can be fulfilled for noble metals. Actually, GaSb–Au contacts are Ohmic [19]. Since the values for $-\mu_m^b$ and $-\varepsilon_v^{FK}$ in (6.2) for most semiconductor–metal junctions in connection with the positive D_o prevent Ohmic contacts, one can propose the following procedure nevertheless to obtain them: One has to decrease D_o by a suitable technology. Such a required negative contribution to D_o is possible if directly at the interface on the semiconductor side acceptors a introduced. This corresponds to the empirical methods to produce Ohmic contacts [4].

TABLE 6 : E_g, E_B, and $-\varepsilon_c^{FK}$ or $-\varepsilon_v^{FK}$ (in eV)

	E_g	E_B	$-\varepsilon_c^{FK}$		E_g	E_B	$-\varepsilon_v^{FK}$
InAs	0.36	0.50	4.02	Si	1.11	0.36	3.16
InP	1.26	0.76	3.32	GaSb	0.72	0.07	3.89
CdS	2.42	–	3.00	GaAs	1.43	0.50	3.96
...			...	...			...

7. Summary

The charge transfer, which determines the properties characteristic to the contact and establishes the thermodynamic equilibrium, equals the change of the total charge density in comparison to the cut off densities and gives rise to an interface dipole uniquely determined by bulk properties only. To this end both the cut-off plane and the reference level for the chemical potential are shown to be uniquely defined. The interface dipole depends on the surface orientations and strongly on doping. We emphasize that our description opens up the of way of calculating it without the need of using the jellium model. The interface dipole consists of the total band bending and an atomic-scale contribution. Both portions are coupled what is clearly seen considering charging gap states. But only the band bending enters the barrier.

The position of the chemical potential within the gap is shown to be independent of doping provided that the dopant density is not too high. Then the barrier height can be calculated by considering the microscopic interface region only notwithstanding the value of the density of gap states. The barrier height does not depend linearly on surface properties (as e. g. the metal work function). Previous theories failed to explain the dependence of the barrier height on the metal even for the more ionic semiconductors. The reasons for that originating in the use of surface properties are discussed in detail and illustrated numerically. In addition, it is explained why approaches relating the barrier height only to the semiconductor band structure are sucessful only for the rather noble metals. The formation of Ohmic contacts is considered at the end.

References

[1] F. Bechstedt and R. Enderlein: Semiconductor Surfaces and Interfaces, Akademie-Verlag, Berlin (1987)

[2] W. Mönch: Reports on Progress in Physics 3, 221 (1990)

[3] F. Flores and C. Tejedor: Journal of Physics C 20, 145 (1987)

[4] E. H. Rhoderick: Metal-semiconductor Contacts, Oxford University Press, Oxford (1980)

[5] W. Schottky: Zeitschrift für Physik 113, 367 (1939)

[6] N. F. Mott: Proc. Cambridge Phil. Soc. 34, 568 (1938)

[7] J. Bardeen: Physical Review 71, 717 (1947)

[8] V. Heine: Physical Review 138, A1689 (1965)

[9] F. Yndurain: Journal of Physics C 4, 2849 (1971)

[10] C. Tejedor, F. Flores and E. Louis: Journal of Physics C 10, 2163 (1977)

[11] S. G. Louie, J. R. Chelikowsky and M. L. Cohen: Physical Review B 15, 2154 (1977)

[12] J. Tersoff: Physical Review Letters 52, 465 (1984)

[13] A. M. Cowley and S. M. Sze: Journal of Applied Physics 36, 3212 (1965)

[14] C. B. Duke and C. Mailhiot: Journal of Vac. Sci. & Techn. B 3, 1170 (1985)

[15] G. Paasch and M. Hietschold: in Ergebnisse in der Elektronentheorie der Metalle, P. Ziesche and G. Lehmann, eds., Akademie-Verlag, Berlin (1983), p. 403

[16] W. R. Frensley and H. Krömer: Physical Review B 16, 2642 (1977)

[17] C. Mailhiot and C. B. Duke: Physical Review B 33, 1118 (1986)

[18] G. Paasch and E. v. Faber: Progress in Surface Science 35, 19 (1990)

[19] C. A. Mead: Solid-State Electronics 9, 1023 (1966)

[20] F. Garcia-Moliner and F. Flores: Introduction to the Theory of Solid Surfaces, Cambridge University Press, Cambridge (1979)

[21] E. Louis, F. Yndurain, and F. Flores: Physical Review B 13, 4408 (1976)

[22] R. K. Swank: Physical Review 153, 844 (1967)

[23] J. R. Smith: Physical Review 181, 522 (1969)

[24] E. v. Faber: Dissertation A, TH Ilmenau, Ilmenau (1990)

[25] A. Zur, T. C. McGill, and D. L. Smith: Physical Review B 28, 2060 (1983)

[26] S. Kurtin, T. C. McGill, and C. A. Mead: Physical Review Letters 22, 1433 (1969)

[27] V. Heine and C. H. Hodges: Journal of Physics C 5, 225 (1972)

[28] J. Tersoff: Surface Science 168, 275 (1986)

[29] J. Hölzl: in Solid State Physics, Springer Tracts in Modern Physics, Vol. 85, Springer-Verlag, p. 45-85 (1979)

[30] Y.-C. Ruan and W. Y. Ching: Journal of Applied Physics 62, 2885 (1987)