

CONNECTIONS BETWEEN WORK FUNCTIONS, SCHOTTKY BARRIERS, AND HETEROJUNCTION BAND OFFSETS

G. PAASCH (1) and E. von FABER (2)

(1) *Zentralinstitut für Festkörperphysik und Werkstofforschung
Dresden, PSF, D- O-8027 Dresden*

(2) *Technische Hochschule Ilmenau, PSF 327, D- O-6300 Ilmenau*

Abstract

Work functions, Schottky barriers, and heterojunction band offsets are the subject of numerous publications. Nevertheless their theoretical description is still strongly dominated by several empirical rules and model assumptions. Thus the connection between different and sometimes converse approximations and the relations between the three quantities are indeed not well established. Therefore here several exact relations will be formulated and discussed. For the work function it is shown that there exists a unique splitting into a surface dipole and a volume contribution including the definition of the reference energy level. The latter is most important for interfaces. For Schottky barriers exact relations are given containing the contact dipole and its splitting into interface and space charge contributions. It is emphasized that their determination cannot be decoupled in general. Finally, the corresponding exact relations for heterojunctions are used for the approximate calculation of band offsets. For 19 heterojunctions the mean deviation from the experimental values is significantly smaller than for the values of the Tersoff theory.

Contents

1. Introduction
2. Work function
 - A. Basic equations
 - B. Surface and volume contributions
 - C. Uniqueness of the cut-off plane
 - D. Metals
 - E. Semiconductors
3. Contact dipole and equilibrium condition
4. Metal-semiconductor contact and the Schottky barrier
 - A. Definition of the Schottky barrier
 - B. Some general relations for the Schottky barrier
 - C. Relations of the Cowley & Sze type
5. Heterojunction and band offset
 - A. Definition of the band offset
 - B. Microscopic concepts
 - C. Dipole due to charged gap states and band offset

- D. Evaluation of the band offset
- E. Band offset and electron affinity

1. Introduction

Work functions, Schottky barrier heights, and semiconductor heterojunction band offsets are the subject of several reviews [1-6] being concentrated on different aspects but there still dominate empirical rules and simplifying model assumptions rather than a unique theoretical concept. It is just the aim of our paper to review the general theoretical description and thereby to improve it by establishing the relation between the three quantities exactly and to demonstrate what really the approximations are underlying well known models. Basic concepts used are (i) the uniqueness of splitting the density at interfaces into volume and surface dipole contributions, (ii) a unique definition of the reference level for the chemical potential for all three quantities, (iii) the splitting of the semiconductor surface dipole into an atomic scale dipole and the band bending, (iv) the consequent use of the exactly formulated equilibrium condition, and (v) the local neutrality level of semiconductor gap states formulated with less artificial assumptions than before. For Schottky barriers a number of exact relations, some of them new, enable us to analyse the connections between different models and empirical rules. For semiconductor heterojunctions the same achieved and in addition, a new formula is derived, containing as limiting cases both Tersoff's [7] and Frenseley & Krömer's [8] results. For nineteen heterojunctions the mean and maximum deviation from experiment is significantly smaller than in Tersoff's theory.

2. Work function

A. Basic equations

The work function ϕ [1,9] is the minimum energy required to bring an electron from the volume of the solid into the vacuum at a large distance. From this definition in a one-electron picture one obtains

$$\phi = v - \mu \quad (2.1)$$

where v is the potential energy of the electron outside and μ is the (electro)chemical potential both relative to the same energy zero. (For semiconductors states at μ are not occupied and (2.1) has to be interpreted statistically.) Actually, this definition includes many body effects, especially the image force arising from the attraction between the negatively charged electron and the remaining solid having a positive charge. Indeed, far from the surface one has [9]

$$v = \varphi(\infty) \quad (2.2)$$

where $\varphi(\infty)$ is the potential energy due to the electrostatic potential of the solid. The density functional theory [1,9,10] yields

$$\mu = \varphi(\vec{r}) + \delta G[n] / \delta n(\vec{r}) \quad (2.3)$$

where $G[n]$ is the functional of kinetic, exchange and correlation energy of the inhomogeneous electron gas. Averaging (2.3) over the volume of the solid (notation $\langle \dots \rangle$) yields

$$\mu - \langle \varphi \rangle = \langle \delta G[n] / \delta n(\vec{r}) \rangle \equiv \bar{\mu} \quad (2.4)$$

Then the work function (2.1) can be expressed as

$$\phi = [\varphi(\infty) - \langle \varphi \rangle] - \bar{\mu} \quad (2.5)$$

Many body effects are contained in (2.5) only via $G[n]$ so that because of the universality of this functional they are regarded exactly. With (2.5) one obtains a separation into surface dipole and volume contributions in the case of the jellium model [11] only as shown already in [1]. Therefore, inclusion of the ionic potentials requires a more detailed description.

B. Surface and volume contributions

Principally by choosing the energy zero one can split (2.1) arbitrarily into two contributions. But a unique separation into surface and volume contribution is required. Let $\varrho(\vec{r}) = \varrho(\vec{r} + \vec{R})$ be the translational invariant total charge density of the bulk. Then the charge density of the solid with surface (mathematical the $x - y$ -plane) can be written as

$$\varrho_s(\vec{r}) = \varrho_c(\vec{r}) + \delta\varrho_s(\vec{r}) , \quad \varrho_c(\vec{r}) = \varrho(\vec{r}) \Theta(z_1 - z) . \quad (2.6)$$

$\delta\varrho_s(\vec{r})$ contains all deviations from the cut-off density $\varrho_c(\vec{r})$. For the determination of the cut-off plane $z = z_1$ the neutrality of $\delta\varrho_s$ is not sufficient as shown below. Fig. 1 shows schematically the potentials φ_c and $\delta\varphi_s$ arising from (2.6) after parallel averaging. Whereas $\delta\varphi_s$ gives rise to the surface dipole

$$P = \delta\varphi_s(\infty) - \delta\varphi_s(-\infty) \quad (2.7)$$

the potential of the cut-off density has for $z \geq z_1$ the value φ_{co} . Therefore one has for (2.2)

$$\varphi(\infty) = P + \varphi_{co} \quad (2.8)$$

and from (2.1) and (2.2) one gets a separation into surface dipole contribution P and a "bulk" contribution $-\mu^b$

$$\phi = P - \mu^b , \quad \mu^b \equiv \mu - \varphi_{co} . \quad (2.9)$$

Introducing

$$\Delta\varphi_{co,m} \equiv \varphi_{co} - \langle \varphi \rangle \quad (2.10)$$

and using (2.4) one can rewrite (2.9) as

$$\phi = P + \Delta\varphi_{co,m} - \bar{\mu} \quad (2.11)$$

where $-\bar{\mu}$ according to (2.4) is a pure bulk contribution and $\Delta\varphi_{co,m}$ is the difference between the potential of the cut-off density just outside and the mean value of the potential inside (Fig. 1). Although one can often find the first equation (2.9) (also for contacts) actually the second equation (2.9) defining μ^b exactly is missing. In addition the cut-off plane z_1 is not yet determined. It should be mentioned that $\Delta\varphi_{co,m}$ depends on the orientation of the cut-off plane.

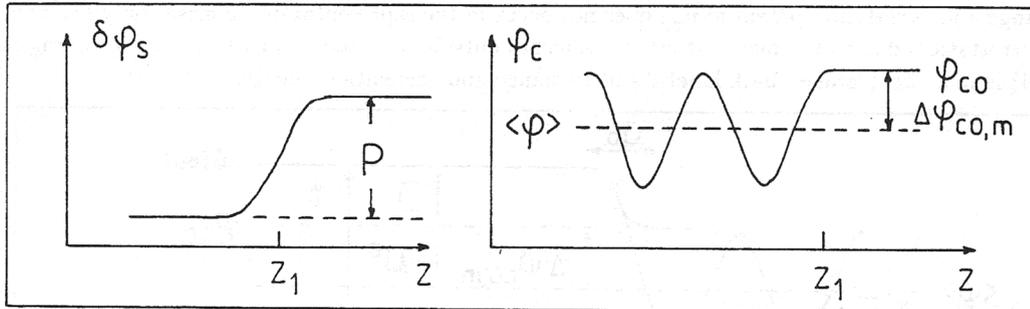


Fig. 1. Parallel averaged surface dipole potential $\delta\varphi_s$ and bulk cut-off potential φ_c .
Other symbols explained in the text.

C. Uniqueness of the cut-off plane

The definition of the bulk chemical potential μ^b (being most important for contacts) and the expressions (2.9) and (2.11) are unique only if the definition of the cut-off plane z_1 is. The neutrality of the surface dipole (after parallel averaging $\int dz \delta\varrho_s = 0$) is not sufficient to determine z_1 . But z_1 is determined by the bulk charge density $\varrho(\vec{r})$ alone. Let a be the periodicity length in z -direction. Then, after parallel averaging the cut-off plane is determined by the following conditions:

$$\int_{z_1-a}^{z_1} dz \rho(z) = 0, \quad \varphi(z_1-a) = \varphi(z_1) = \varphi_{co}, \quad \left. \frac{d\varphi}{dz} \right|_{z_1} = 0. \quad (2.12)$$

Using (2.12) as boundary conditions in the Poisson equation one obtains

$$\int_{z_1-a}^{z_1} dz z \rho(z) = 0 \quad (2.13)$$

that means vanishing of the first moment of the density in the top layer. The cut-off plane z_1 is uniquely determined by (2.13) for a given surface orientation and eventually a given top atomic layer for a more complicate crystal structure. In addition one obtains

$$\Delta\varphi_{co,m} = -\frac{e}{2\varepsilon_0 a} \int_{z_1-a}^{z_1} dz z^2 \rho(z) > 0. \quad (2.14)$$

The sign arises from the fact that the ionic positiv charge is more localized than the electron density.

Usually (2.13) and (2.14) have been overlooked in the literature. But only with these equations the expressions (2.9) (2.11) for the work function are well defined. Accordingly there is a surface dipole contribution P (2.7), the contribution $\Delta\varphi_{co,m}$ (2.10) and (2.14) arising from the cut-off bulk density determined by (2.13) and the pure bulk contribution $-\bar{\mu}$ (2.4).

D. Metals

In this case the extension of the surface dipole layer is determined by the Thomas-Fermi screening length, roughly the Bohr radius a_0 . The corresponding dipole will be denoted by $P_m = D_m$ and the work function is

$$\phi_m = D_m - \mu_m^b. \quad (2.15)$$

Fig. 2 shows schematically the quantities defined in the preceding sections. Apparently the exchange and correlation potential v_{xc} does not occur in the representation (φ arises only from the electrostatics) due to the fact that (i) v_{xc} vanishes outside, $v_{xc}(\infty) = 0$ and (ii) $-\bar{\mu}$ according to (2.4) is just the averaged bulk kinetic and exchange and correlation energy.

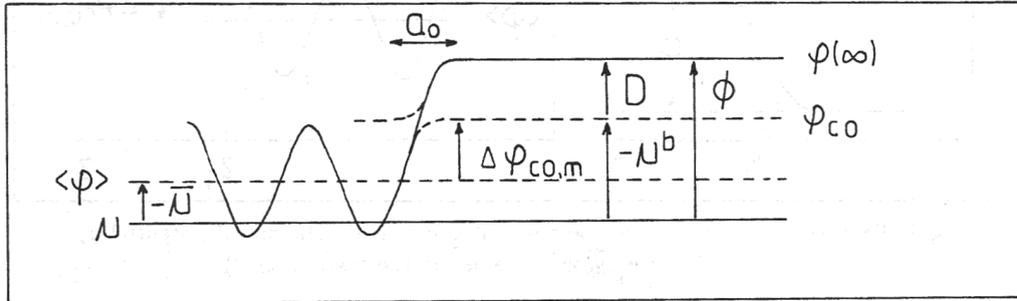


Fig. 2. Energy diagram at a metal surface. Symbols defined in the text.

E. Semiconductors

Again there occurs a decrease of the electron density from the bulk into the vacuum over roughly the length a_0 (and surface reconstruction) giving a dipole D_s . But due to the possibility of charged surface states one has in addition a space charge layer with an extension of the Debye screening length $l_D \gg a_0$. The resulting band bending V_s is a part of the total surface dipole

$$\phi_s = P_s - \mu_s^b, \quad P_s = D_s + V_s. \quad (2.16)$$

Fig. 3 shows the energy diagram in the l_D -scale. In this scale neither P_s nor D_s occur. Fig. 3 shows also the electron affinity χ_s and the photoemission threshold $\phi_{PE} = \chi_s + E_g$. Due to the

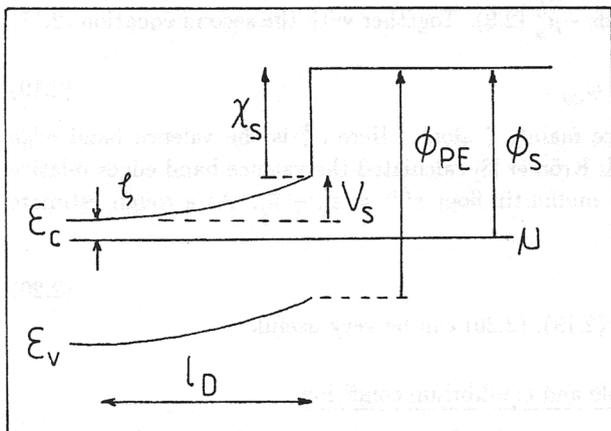


Fig. 3. Energy diagram for a semiconductor surface with conduction and valence band edge ϵ_c, ϵ_v , respectively.

possibility to measure ϕ_{PE} [12] (electron escape depth small) they should be defined relative to the band edges at the surface as shown here. Sometimes there was confusion on this question [13]. Now the work function can be written also as

$$\phi_s = \zeta + V_s + \chi_s, \quad \zeta \equiv \epsilon_c - \mu \quad (2.17)$$

which can be useful for phenomenological considerations and for obtaining V_s by measuring ϕ_s and ϕ_{PE} . But V_s cannot be calculated without the knowledge of the surface states. Further, from (2.17) a comparison with (2.16) using (2.9) yields that

$$\chi_s = D_s - (\epsilon_c - \varphi_{co}) \quad (2.18)$$

is determined by the atomic scale dipole D_s and the position of the band edge ϵ_c relative to the cut-off potential outside. On an a_0 -scale the band edges near the surface are practically constant (Fig. 4). All energies except the chemical potential and $\varphi(\infty)$ are shifted at the surface by V_s .

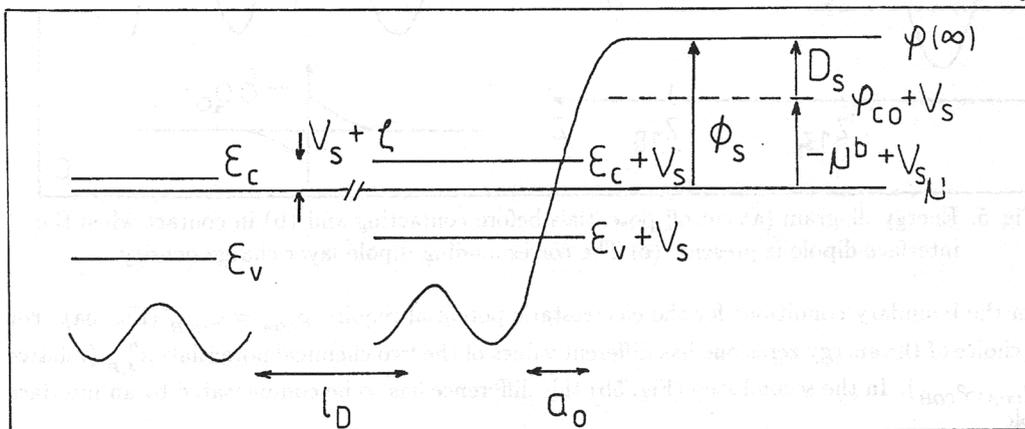


Fig. 4. Energy diagram at a semiconductor surface on an atomic scale.

Note that one has $(\varphi_{co} + V_s) - \mu = -\mu_s^b + V_s$ in accordance with (2.16). Thus on this scale contrary to Fig. 3 the band bending does not occur explicitly but only through the already mentioned

shift. This makes clear that it is not possible adequately to represent all quantities relevant for the work function in a diagram in only one of the two scales. This is the reason for a large number of misunderstandable representations in the literature especially for the more complex contact problems when semiconductors are involved [2,14,20].

In (2.16) besides the dipole P_s one needs $-\mu_s^b$ (2.9). Together with the second equation (2.17) the latter is

$$\mu_s^b = \varepsilon_v^b + E_g - \zeta, \quad \varepsilon_v^b \equiv \varepsilon_v - \varphi_{co}. \quad (2.19)$$

In (2.19) doping and temperatur influence mainly ζ alone. Here ε_v^b is the valence band edge relative to the cut off potential. Frenley & Krömer [8] calculated the valence band edges relative to the mean electrostatic potential in the muffin-tin floor $\varepsilon_v^{FK} = \varepsilon_v - \bar{v}_i$. As a rough estimate $\bar{v}_i \simeq \varphi_{co}$ can be used and hence one has

$$\varepsilon_v^b \simeq \varepsilon_v^{FK} = \varepsilon_v - \bar{v}_i \quad (2.20)$$

in (2.19). Especially for contact problems (2.19), (2.20) can be very useful.

3. Contact dipole and equilibrium condition

In the case of a contact between two materials A and B the surface dipoles of the free surfaces do not exist. But the total system consists of cut-off systems plus an interface dipole layer charge density $\delta\rho_0$ whereby the cut-off values φ_{coA} and φ_{coB} are uniquely determined by the two bulk densities as in section 2. In a Gedankenexperiment the contact can be built up in two steps. In the first step one sticks the two cut-off systems without changes of their charge distribution.

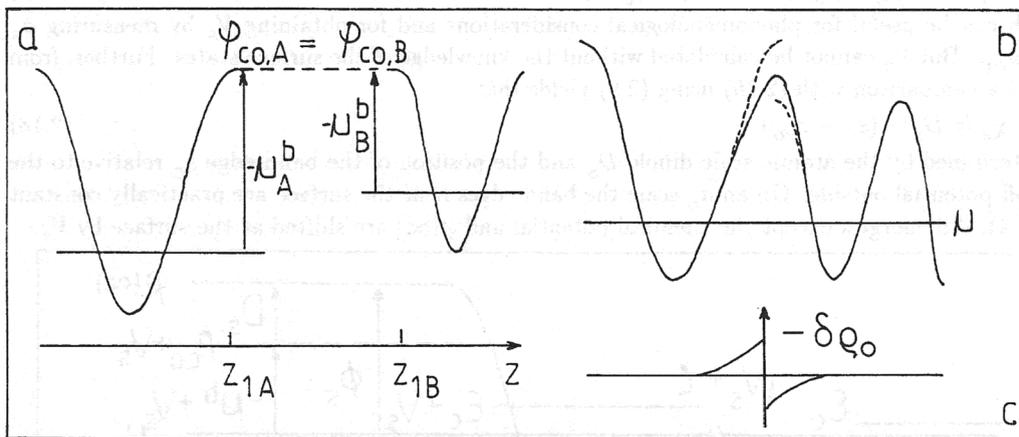


Fig 5. Energy diagram (a) cut-off potentials before contacting and (b) in contact when the interface dipole is present. (c) The corresponding dipole layer charge density.

Then the boundary conditions for the electrostatic potential require $\varphi_{coA} = \varphi_{coB}$ (Fig. 5a). For this choice of the energy zeros one has different values of the two chemical potentials $\mu_{A,B}^b$ (relative to $\varphi_{coA}, \varphi_{coB}$). In the second step (Fig. 5b) this difference has to be compensated by an interface dipole

$$P_0 = P_{BA} = -\mu_B^b + \mu_A^b \quad (3.1)$$

connected with the occurrence of the deviation $\delta\rho_0$ of the total charge density from the stucked cut-off densities by

$$P_o = -\frac{e}{\epsilon_o} \int dz z \delta \rho_o(z) \quad (3.2)$$

($\delta \rho_o(z)$ after parallel averaging). Although the idea of equation (3.1) can be found in principle in the work of Bardeen [15] the quantities in (3.1) (3.2) were not uniquely defined in the past. Bardeen even assumed that the splitting into surface and bulk contributions is arbitrary. Duke & Mailhot [16] used in (3.1) Fermi levels relative to the mean potential $\langle \varphi \rangle$ in the volume (in our notation $-\bar{\mu}$) which is valid for the jellium model only. It should be noted that the cut-off densities are uniquely determined by (2.13) for a given surface orientation with the bulk densities. Therefore the "bulk" contributions $-\mu_{A,B}^b$ are determined by (2.9) - (2.14). Then the interface dipole P_o is fixed by (3.1). It will be created by the density $\delta \rho_o$ (3.2) defined as the deviation from the cut-off densities in equilibrium.

More specifically, in the next section the metal-semiconductor interface will be considered. But before doing this we note that often [17-21] the difference of the work functions

$$u_{BA} = \phi_B - \phi_A \quad (3.3)$$

is used instead of P_{BA} . Then with (2.9) and (3.1) one obtains

$$u_{BA} = P_B - P_A + (-\mu_B^b + \mu_A^b) = P_B - P_A + P_{BA} \quad (3.4)$$

(only if the cut-off orientations at the free surfaces are the same as at the interface). Equation (3.4) shows clearly, that the contact potential defined by (3.3) is a property which is not determined by the interface alone.

4. Metal-semiconductor contact and Schottky barrier

A. Definition of the Schottky barrier

According to section 3 at a metal-semiconductor contact there occurs the interface dipole (3.1) determined by the difference of the two "bulk" contributions $-\mu^b$ (2.9). As in the case of the semiconductor surface the interface dipole consists of a contribution D_o arising from a charge redistribution on an atomic scale at the interface and the band bending V_o in the semiconductor. Thus

$$P_o = -\mu_m^b + \mu_s^b, \quad P_o = V_o + D_o \quad (4.1)$$

The energy diagram in the Debye length scale including the interface and the two surfaces is shown in Fig. 6. As in the case of the semiconductor surface in this Figure one can show only the band

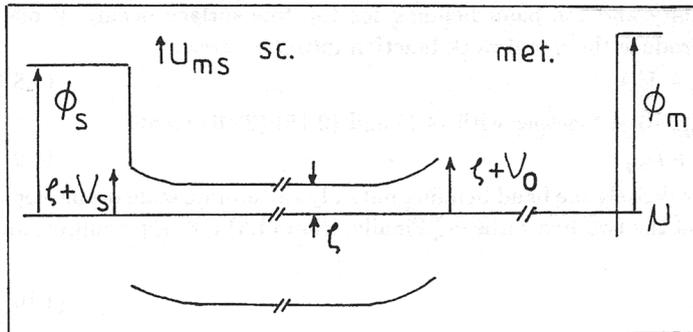


Fig. 6. Energy diagram of the metal-semiconductor contact with the two surfaces in the Debye length scale.

bending V_o and not the values P_o or D_o . For sake of completeness in this Figure also the contact potential (3.4) is shown. But this is not of direct relevance for the contact region. With (2.19) one can rewrite the first equation (4.1) as

$$P_o = -\mu_m^b + \varepsilon_v^b + E_g - \zeta. \quad (4.2)$$

Whereas ζ depends 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 bulk electronic structure.

The Schottky barrier according to the methods to measure it [2,22,23] 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 a n-type semiconductor (Fig. 6) this barrier is just

$$\phi_b = \zeta + V_o \quad (4.3)$$

whereas for p-material one has

$$\phi_b^p = E_g - \zeta^p - V_o^p. \quad (4.4)$$

But the barrier measured via transport is lowered both for electrons and holes due to the image force correction by about [23]

$$\Delta V_{im} = \left[\left(\frac{e^2}{4\pi\varepsilon_{HF}\varepsilon_o} \right)^2 \frac{2e^2 N_{A,D}}{\varepsilon_s \varepsilon_o} (|V_o| - kT) \right]^{1/4} \quad (4.5)$$

where $N_{A,D}$ is the effective doping density and ε_{HF} and ε_s are the high-frequency and static dielectric function, respectively. The main theoretical problem is the following. The fundamental equilibrium condition (4.2) determines P_o and its splitting (4.1) into the two contributions V_o and D_o being the most serious problem since only V_o enters the barrier. We note that the incompatibility of the two scales (cf. Section 2 E) causes much confusion in graphical representations of the contact in the literature [2,14,20].

B. Some general relations for the Schottky barrier

In the past there were numerous attempts to establish phenomenologically connections of the barrier height ϕ_b with quantities visible from experiments [4,23]. Therefore it seems to be sensible to formulate some exact relations connecting the barrier with other quantities. At first it would be trivial to denote the position of the Fermi energy in the gap relative to the valence band edge at the contact by E_o (Fig. 6) giving

$$\phi_b = E_g - E_o. \quad (4.6)$$

Further, one can combine (4.3) with (2.17) and obtains

$$\phi_b = \phi_s - \chi_s + (V_o - V_s) \quad (4.7)$$

where the work function, the affinity and the band bending for the free surface occur. If one inserts (3.3) into (4.7) one can introduce the metal work function into the barrier

$$\phi_b = \phi_m - \chi_s - u_{ms} + (V_o - V_s). \quad (4.8)$$

But then the contact potential u_{ms} (3.4) together with (4.1) and (2.15) (2.16) yields

$$\phi_b = \phi_m - \chi_s - (D_o - D_s + D_m). \quad (4.9)$$

This expression does not contain explicitly the band bending but only the atomic scale dipole contributions both at the interface and the two free surfaces. Finally, from (4.3) and the equilibrium condition (4.1) (4.2) we obtain

$$\phi_b = -\mu_m^b + \varepsilon_v^b + E_g - D_o \quad (4.10)$$

containing the two cut-off "bulk" values $-\mu_m^b$ and ε_v^b , the gap and as the interface property the atomic scale dipole contribution D_o .

Now the relations (4.3), (4.6) - (4.10) for the barrier (n-type semiconductor) will be discussed. In the definition (4.3) the band bending V_o is the only contact property whereas in (4.10) there

is only D_o describing the interface. But both V_o and D_o are only a part of the total dipole. As a result of the screening properties of Poisson's equation there exist a strongly non-linear dependence between the quantities occurring in the exact relations (4.7) - (4.10). Therefore, non of these equations expresses a linear relation between ϕ_b and one of the other (except V_o) quantities (as e.g. ϕ_m in (4.9)). Formally (4.6) is an equation as used to describe pinning [24-27]. But (4.6) is identical with the definition (4.3) whereas pinning means that the value E_o depends only on the semiconductor and not on the metal. A large number of papers attempts to estimate metal independent values E_o for special contacts [2-4]. Similarly, (4.7) is reduced to

$$\phi_b \approx \phi_s - \chi_s \quad \text{for} \quad V_o \approx V_s, \quad (4.11)$$

that means if the band bending at the interface is the same as at the surface. Then again the barrier does not depend on the metal. This is e.g. the case for Si-contacts [12,23] with some metals. The Schottky-Mott theory [17,18] assumes that one has the flat band case $V_s = 0$ at the free surface and that the contact potential equals the band bending at the interface $V_o = u_{ms}$ contrary to (3.4). Then from (4.8) one has

$$\phi_b \approx \phi_m - \chi_s. \quad (4.12)$$

The correct condition which leads to (4.12) is seen from (4.9), namely $D_o = D_s - D_m$, that means a superposition of the atomic scale dipoles but not necessarily the flat band case. Equations (4.7) to (4.9) contain not only interface but also surface properties. But theory should prefer expressions describing the contact alone as (4.3), (4.6) and (4.10).

Measured barrier values are compiled in several papers [2,4,22,23] and there exist several attempts to systematize them phenomenologically. One of these expressions is

$$\phi_b = S' \phi_m + C' \quad (4.13)$$

where metal properties are involved only via the metal work function. In the Mott version [17,18] one has $S' = 1$. According to [28] only for CdS with $C' = 4$ eV reasonable results are obtained. A systematic application [20,29] of (4.13) shows mainly that for ionic semiconductors S' is large and hence the metal (or its work function) influences significantly the barrier whereas for covalent semiconductors ϕ_m is less important (S' small), see also [30]. With the relation

$$\phi_b = S E_m + C \quad (4.14)$$

introduced by Mead [22] where E_m is the metal electronegativity and S and C depend only on the semiconductor in some cases the data can be described better. Also in this cases one obtains $S \leq 0.2$ for covalent semiconductors and larger S for ionic semiconductors [31,32]. There does not exist a serious theoretical foundation for the connection between the atomic property E_m and the bulk chemical potential $-\mu_m^b$ entering (4.10). Empirical relations connecting E_m and ϕ_m can be found in [33].

There exist further numerous attempts to express the barrier by semiconductor quantities alone [2-4]. Indeed for several semiconductors in contact with different metals this is possible. According to [34] the Fermi level at the interface for group IV semiconductors and for III-V semiconductors (with zincblende structure) should lie $E_g/3$ over the valence band edge and hence

$$\phi_b = \frac{2}{3} E_g \quad \text{and} \quad \phi_b^b = \frac{1}{3} E_g. \quad (4.15)$$

On the other hand according to Tersoff [35] for the same semiconductors with p-doping the Fermi level lie near the middle of the indirect gap. Other authors [36] claim that the position of the Fermi level is determined only by the anion electronegativity. Although such empirical rules are well suited to systematically describe barriers for different groups of contacts they give only little insight in the mechanism determining the barrier.

C. Relations of the Cowley & Sze type

A special approximation derived originally by Cowley & Sze [20] was often used and modified [37,38]. Here we will derive an expression of this type based on the formulation given above and with a minimum of additional assumptions. Our starting point are the exact equations (4.10) or (4.9) and further the atomic scale interface dipole is considered. Formally one can expand D_o as

$$D_o = D_{on} + \alpha (V_o - V_{on}) + \dots \quad (4.16)$$

where the index n denotes a situation deviating slightly from the actual equilibrium condition. By conveniently defining this state the linear expansion coefficient α will be determined. It is assumed that the state n is realized if the exponentially decaying semiconductor gap states are occupied such that the missing of the valence band states due to the boundary condition at the interface are compensated (local neutrality). The distance of the highest occupied gap state from the valence band edge will be defined as ϕ_n . Then $V_o - V_{on}$ is just the difference between the actual Fermi energy to the valence band, $E_g - \phi_b$ (see Fig. 6) and ϕ_n ,

$$V_o - V_{on} = E_g - \phi_b - \phi_n \equiv \delta E. \quad (4.17)$$

For a smooth 2D density of gap states $N_o(E)$ the number of charged gap states per unit area is then $n_{cs} = N_o(E_f) \delta E$. Exponentially decreasing gap states [39] with decay length q^{-1} yield the contribution

$$D_{o,cs} = -\frac{4\pi\epsilon^2}{\epsilon_s(q)} \frac{1}{2q} n_{cs}, \quad n_{cs} = N_o(E_f) \delta E, \quad \epsilon^2 \equiv \frac{e^2}{4\pi\epsilon_o} \quad (4.18)$$

to D_o . Here the dielectric constant for the wave number q has to be used (for a refinement see [14]). The charged gap states will be screened in the metal by a charged layer with an extension of roughly $\beta^{-1} \approx a_o$ Bohr radius) and their density \mathcal{N}_m gives

$$D_{o,m} = \frac{4\pi\epsilon^2}{\beta} \mathcal{N}_m. \quad (4.19)$$

Further \mathcal{N}_{sc} is screened by the space charge layer with density \mathcal{N}_d . But since in practically all cases \mathcal{N}_d ($\leq 10^{11} \text{ cm}^{-2}$) is negligible one has $\mathcal{N}_m \approx -\mathcal{N}_{sc}$. Then one can add (4.18), (4.19) as the deviation (4.16) (4.17)

$$D_o - D_{on} = D_{o,cs} + D_{o,m} = \bar{\alpha} \delta E \quad (4.20)$$

$$\alpha = 4\pi\epsilon^2 \delta N_o(E_f), \quad \delta = \left(\frac{1}{2q\epsilon(q)} + \frac{1}{\beta} \right). \quad (4.21)$$

Inserting (4.20) with (4.17) in (4.9) or (4.10) one can solve for the barrier as

$$\phi_b = \frac{1}{1+\alpha} \left(\phi_{b1} - D_{on} + \alpha (E_g - \phi_n) \right) \quad (4.22)$$

$$\phi_{b1} \equiv -\mu_m^b + \epsilon_v^b - E_g = \phi_m - \chi_s + D_s - D_m. \quad (4.23)$$

The α -dependence of (4.22) was derived in [20] but $\phi_{b1} - D_{on}$ was not given correctly there. Tejedor, Flores & Louis [14] derived (4.22) using a special model. But the first equation (4.23) is missing there. For small α the determination of ϕ_{b1} and of the value D_{on} corresponding to a non-equilibrium situation would be most difficult. But one can estimate α (4.21) using $(2q)^{-1} \approx 1.5 \dots 6.6 \text{ \AA}$ and $N_o(E_f) \approx (2 \dots 5) \cdot 10^{18} \text{ m}^{-2} \text{ eV}^{-1}$ [14,37,38] and $\epsilon(q) \approx 1.5 \dots 3$ [38]. Then we obtain $\alpha \approx 4 \dots 50$ where for covalent or narrow-gap semiconductors one has $\alpha > 10$ and only for ionic semiconductors smaller values occur. For $\alpha \gg 1$ (4.22) is reduced to $\phi_b \approx E_g - \phi_n$ approaching a pinning situation. The problem would then be the determination of ϕ_n which depends on the matching of the gap states with the metal wave functions. On the other hand from (4.22) (4.23) an apparent proportionality $\phi_b \sim (1+\alpha)^{-1} \phi_m + \dots$ is obtained resembling (4.13).

As in (4.13) the proportionality factor $(1 + \alpha)^{-1}$ is large for ionic semiconductors with smallest α . But also the other terms in (4.22) depend nonlinearly on ϕ_m which is the reason that (4.13) can be useful only accidentally.

5. Heterojunctions and band offsets

A. Definition of the band offset

At semiconductor-semiconductor contacts also an interface dipole P_o is built up in order to equalize the chemical potentials. With the notation for differences $\Delta E \equiv E_2 - E_1$ (3.1) yields using (2.19) a condition for P_o

$$P_o = -\Delta\epsilon_v^b - \Delta E_g + \Delta\zeta. \quad (5.1)$$

The interface dipole consists of the difference of the band bending $V_{o1} - V_{o2}$ (Fig. 7) and an additional contribution D_o due to charge redistribution on an atomic scale at the interface

$$P_o = D_o - \Delta V_o \quad (5.2)$$

where band bending is positive if the bands are bent upwards (Fig. 7). In many papers concerning microscopic theories which will be discussed in Section B no band bending and no equilibrium condition occur.

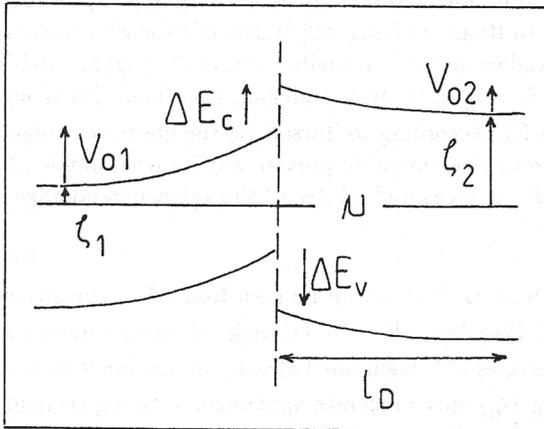


Fig. 7. Energy schema of a heterojunction.

As in the case of Schottky contacts one is interested in determining the relative position of the band structures of the two materials near the interface. The quantity characteristic for heterojunctions is the band offset. The band offsets are defined by

$$\Delta E_v = E_{v2} - E_{v1}, \quad \Delta E_c = \Delta E_v + \Delta E_g. \quad (5.3)$$

where the band edges are measured near the interface on the same energy scale (Fig. 7). Now the constant chemical potential can serve as an energy zero so that (cf. Fig. 7)

$$\Delta E_v = \Delta\zeta - \Delta E_g + \Delta V_o. \quad (5.4)$$

Thus, the band offset can be calculated already if the so called "built-in potential" ΔV_o is known. On the other hand comparing (5.2) and (5.4) one can eliminate both the band bending V_{o1} and V_{o2} . Together with (5.1) one gets

$$\Delta E_v = \Delta\epsilon_v^b + D_o. \quad (5.5)$$

From (5.5) one may suppose that the only interface property influencing the band offset is the dipole D_o at the interface. Nevertheless the original problem of determining D_o from (5.2) remains since the band bending and the dipole D_o are indeed coupled.

B. Microscopic concepts

Most of the concepts do not start with thermodynamic equilibrium and alignment of chemical potentials due to band bending (Fig. 7 and (5.4)). So, only the interface region extended over several Å is considered. First "absolute" reference levels R_i are introduced. Then the valence band edge ε_{vi} relative to this reference level R_i is calculated for each semiconductor

$$\varepsilon_{vi} = E_{vi} - R_i. \quad (5.6)$$

After equalizing the reference levels R_1 and R_2 appropriate to

$$D_o = R_2 - R_1 \quad (5.7)$$

where D_o is a dipole due to charge redistributions at the interface the difference of the valence band edges ΔE_v (band offset) is obtained with (5.6) (5.7) as

$$\Delta E_v = \varepsilon_{v2} - \varepsilon_{v1} + D_o. \quad (5.8)$$

The "absolute" valence band edges ε_v should be determined by bulk properties only. Using as the reference the intrinsic Fermi level as was done by Adams & Nussbaum [40] one gets $\Delta E_v = \Delta\zeta_i - \Delta E_g + D_o$ in contradiction to (5.4). Other microscopic concepts did not consider the chemical potentials. Harrison [41] constructed the crystal potential simply by superimposing atomic orbitals. Then the potential at infinity is regarded as reference and the valence band edges calculated by tight-binding theory have to be inserted in (5.8) neglecting any dipole D_o . Some corrections were made subsequently. Due to Ruan & Ching [42] Harrison's theory coincides with the electron affinity rule established by Anderson [19]. Actually, setting $R = \varphi(\infty)$ yields $\varepsilon_v = -\phi_{PE}$ and hence $\Delta E_v = -\Delta\chi_s - \Delta E_g$ for $D_o = 0$. But Anderson's original idea is not a microscopic concept as shown below (Section E). According to Tersoff [7] the effective midgap energies $R = E_B$ of the two semiconductors have to be identical to prevent any induced dipole D_o produced mainly by charged gap states. Since E_B is measured relativ to the valence band edge ($\varepsilon_v = -E_B$) the valence band offset is given by

$$\Delta E_v = -\Delta E_B. \quad (5.9)$$

Actually Tersoff's assumption $D_o = 0$ did not lead to (5.9) as can be seen from (5.5) since then $\Delta\varepsilon_v^b$ would be the band offset contrary to (5.9). Frenley & Krömer [8] took the mean interstitial potential \bar{v}_i as reference. Their valence band energies ε_v^{FK} mentioned already in Section 2 used to predict the band offset in (5.8) yield (neglecting D_o) only moderate agreement with experiment

$$\Delta E_v = \Delta\varepsilon_v^{FK}. \quad (5.10)$$

They also estimated a correction D_o but this did not improve their results.

These microscopic concepts do not contain band bending which is a part of the interface dipole (5.2). But comparing (5.8) and (5.5) one sees a similarity. (Note that (5.5) was derived using chemical potentials and thermodynamic equilibrium which are both absent in microscopic theories described above.) Thus the microscopic concept is correct provided that one chooses the reference levels φ_{co} defined uniquely relativ to the band structure. However from the failure of the Frenley & Krömer attempt one learns that the contribution D_o is not negligible. On the other hand the success of Tersoff's approach shows that charged gap states are important. But any dipole D_{vs} produced by these charged gap states can only be calculated if band bending is considered.

C. Dipole due to charged gap states and band offset

It will be assumed that the dipole D_o arises from charged gap states only, $D_o = D_{vs}$. Every deviation δE of the chemical potential from a neutrality level ϕ_n will induce a total charge n_{vs} at the surface. Measuring ϕ_n relative to the valence band edge E_v one has

$$E_g - \zeta - V_o = \phi_n + \delta E \quad (5.11)$$

and the induced total charge treated in linear approximation equals

$$n_{vs} = N_{vs} \delta E \quad (5.12)$$

where N_{vs} is the density of gap states near ϕ_n . The whole system containing two interface regions of the semiconductors and two space charge layers with total charge n_d must be neutral

$$n_{d1} + n_{vs1} + n_{vs2} + n_{d2} = 0. \quad (5.13)$$

Introducing effective extensions δ_i of the gap states on both sides of the mathematical interface the potential drop due to charged gap states is

$$D_o = D_{vs} = -4\pi\epsilon^2 \left[\delta_1 n_{vs1} - \delta_2 n_{vs2} \right]. \quad (5.14)$$

Using (5.14) with (5.12), (5.13) and assuming that a small correction due to the total charge ($n_{d1} - n_{d2}$) can be neglected we obtain

$$D_{vs} = -\alpha \left[\delta E_1 - \delta E_2 \right], \quad \alpha \equiv 4\pi\epsilon^2 (\delta_1 + \delta_2) / \left[\frac{1}{N_{vs1}} + \frac{1}{N_{vs2}} \right]. \quad (5.15)$$

Note that the chemical potential is then always situated between the neutrality levels ϕ_n .

The dipole (5.15) in the band offset (5.5) is according to (5.11) proportional to

$$\left[\delta E_1 - \delta E_2 \right] = \Delta\zeta - \Delta E_g + \Delta V_o + \Delta\phi_n = \Delta E_v + \Delta\phi_n \quad (5.16)$$

where in the second step (5.4) is used. Thus indeed the band bending is important for the dipole D_o but it can be eliminated so that the dipole becomes itself proportional to the band offset. Now combining (5.5) with (5.15), (5.16) one can solve for the band offset

$$\Delta E_v = \frac{1}{1+\alpha} \Delta\epsilon_v^b - \frac{\alpha}{1+\alpha} \Delta\phi_n. \quad (5.17)$$

where α (5.15) depends on the contact. It is clearly seen that the actual limit giving Tersoff's result is $\alpha \rightarrow \infty$ whereas $\alpha \rightarrow 0$ yields just a result similar to that of Frenley & Krömer. With (5.17) both the dipole due to the charged gap states (contrary to Frenley & Krömer) and the equilibrium condition (contrary to Tersoff) are taken into account.

D. Evaluation of band offsets

To evaluate (5.17) one needs the bulk valence band edge ϵ_v^b , the neutrality levels ϕ_n and α (5.15) characterising the gap states. Here for a first estimate we use the Frenley & Krömer values ϵ_v^{FK} according to (2.20) and ϕ_n is identified with the effective midgap value E_B of Tersoff. For a first simplified estimate we use a common value α for all contacts. The effective extensions δ_i of the gap states are given by the characteristic length of the virtuals [39] reduced by effective screening [38]. One yields approximately $\delta_i \simeq 1.5 \text{ \AA}$. The density of gap states was selected to be $N_{vs} \simeq 7.5 \cdot 10^{17} \text{ m}^{-2} \text{ eV}^{-1}$ (in between Schottky contacts and free surfaces). Then one has $\alpha \simeq 2$ and

$$\Delta E_v \simeq \frac{1}{3} \left[\epsilon_{v2}^{FK} - \epsilon_{v1}^{FK} \right] + \frac{2}{3} \left[E_{B1} - E_{B2} \right]. \quad (5.18)$$

With the available data we calculated the valence band offsets and their errors presented in Tab. 1. The experimental values are from [5] and the deviation of Tersoff's band offsets is also shown for comparison. (InAs was not considered since $E_B > E_g$.) Our results are significantly better than Tersoff's. Both the averaged and the maximum error were decreased by about 60 meV. This is really important since Tersoff improved the values from the electron affinity rule in average by 36 meV bearing in mind that the maximum error of the electron affinity rule is however 60 meV. Although there is a large scattering [5] of the experimental values the results listed in Tab. 1 indicate the importance to correctly account for the equilibrium condition as done here.

TABLE 1: Experimental and calculated band offsets (eV, ' * ' are averaged values)

<i>sem.2</i>	<i>sem.1</i>	$\Delta\varepsilon_v^{FK}$	$-\Delta E_B$	$\Delta E_v^{(5,18)}$	$\Delta E_v^{exp.}$	$-\Delta E_B - \Delta E_v^{exp.}$	$\Delta E_v^{(5,18)} - \Delta E_v^{exp.}$
Si	Ge	0.09	-0.18	-0.09	-0.28 *	-0.10	-0.19
	GaP	0.96	0.45	0.62	0.8	0.35	0.18
	GaAs	0.80	0.14	0.36	0.05	-0.09	-0.31
	GaSb	0.73	-0.29	0.05	0.05	0.34	0.00
	InP	1.42	0.40	0.74	0.57	0.17	-0.17
	ZnSe	1.91	1.34	1.53	1.25	-0.09	-0.28
	ZnTe	1.58	0.48	0.85	0.85	0.37	0.00
	CdTe	1.74	0.49	0.91	0.75	0.26	-0.16
Ge	AlAs	0.71	0.87	0.82	0.86 *	-0.01	0.04
	GaP	0.87	0.63	0.71	0.8	0.17	0.09
	GaAs	0.71	0.32	0.45	0.49 *	0.17	0.04
	GaSb	0.64	-0.11	0.14	0.2	0.31	0.06
	InP	1.33	0.58	0.83	0.64	0.06	-0.19
	ZnSe	1.82	1.52	1.62	1.41 *	-0.12	-0.22
	ZnTe	1.49	0.66	0.94	0.95	0.29	0.01
	CdTe	1.65	0.67	1.00	0.85	0.18	-0.15
GaAs	AlAs	0.00	0.55	0.37	0.34 *	-0.21	-0.03
	ZnSe	1.11	1.20	1.17	1.03 *	-0.17	-0.14
GaSb	AlSb	0.05	0.38	0.27	0.4	0.02	0.13
averaged deviation:						0.183	0.126
maximum deviation:						0.37	-0.31

E. Band offsets and electron affinity

Since Anderson's theory [19] for heterojunctions is analogous with the Schottky-Mott theory the following discussion will be very similar to the one in section 4B. Starting with (5.5) one obtains with (2.18) (2.19)

$$\Delta E_v = -\Delta\chi_s - \Delta E_g + (D_o + D_{s2} - D_{s1}) . \quad (5.19)$$

According to Anderson [19] the valence band offset is given by $\Delta E_v = -\Delta\chi_s - \Delta E_g$. Thus Anderson's result is valid if $D_o = D_{s1} - D_{s2}$. The only moderate agreement with experiment shows that this is usually not realized. Equation (5.19) can already be found in the work of Flores & Tejedor [43]. But it was first derived by Mailhlot & Duke [13] under some limiting assumptions.

References

1. G. Paasch and M. Hietschold, in Ergebnisse in der Elektronentheorie der Metalle, P. Ziesche and G. Lehmann, eds., Akademie-Verlag, Berlin (1983), p. 403 .
2. F. Bechstedt and R. Enderlein, Semiconductor Surfaces and Interfaces, Akademie-Verlag, Berlin (1987) .
3. F. Flores and C. Tejedor, Journal of Phys. C, 20, 145 (1987).

4. W. Mönch, Reports on Progress in Physics, to be publ.
5. G. Margaritondo and P. Perfetti, in Heterojunction Band Discontinuities: Physics and Device Applications, F. Capasso and G. Margaritondo, eds., North Holland, Amsterdam, Oxford, New York, Tokyo (1987), p. 59 .
6. A. D. Katnani, in Heterojunction Band Discontinuities: Physics and Device Applications, F. Capasso and G. Margaritondo, eds., North Holland, Amsterdam, Oxford, New York, Tokyo (1987), p. 115 .
7. J. Tersoff, Physical Review B, 30, 4874 (1984).
8. W. R. Frensley and H. Krömer, Physical Review B, 16, 2642 (1977).
9. N. D. Lang and W. Kohn, Physical Review B, 3, 1215 (1971).
10. R. G. Parr and L. J. Bertolotti, Journal of Chem. Phys., 87, 2810 (1983).
11. A. J. Bennett and C. B. Duke, Physical Review, 160, 541 (1967) and 162, 578 (1967).
12. F. G. Allen and G. W. Gobeli, Physical Review, 127, 150 (1962).
13. C. Mailhiot and C. B. Duke, Physical Review B, 33, 1118 (1986).
14. C. Tejedor, F. Flores and E. Louis, Journal of Physics C, 10, 2163 (1977).
15. J. Bardeen, Physical Review, 71, 717 (1947).
16. C. B. Duke and C. Mailhiot, Journal of Vac. Sci. & Techn. B, 3, 1170 (1985).
17. N. F. Mott, Proc. Cambridge Phil. Soc., 34, 568 (1938).
18. W. Schottky, Zeitschrift für Physik, 113, 367 (1939).
19. R. L. Anderson, Solid-State Electronics, 6, 341 (1962).
20. A. M. Cowley and S. M. Sze, Journal of Appl. Physics, 36, 3212 (1965).
21. J. M. Woodall and J. L. Freeouf, Journal of Vac. Sci. & Techn., 21, 574 (1981).
22. C. A. Mead, Solid-State Electronics, 9, 1023 (1966).
23. E. H. Rhoderick, Metal-semiconductor Contacts, Oxford University Press, Oxford (1980).
24. J. Tersoff, Journal of Vac. Sci. & Techn. B, 3, 1157 (1985).
25. O. F. Sankey, R. E. Allen, S. F. Ren and J. D. Dow, Journal of Vac. Sci. & Techn. B, 3, 1162 (1985).
26. W. E. Spicer, T. Kendelewicz, N. Newman, K.K. Chin and I. Lindau, Surface Science, 168, 240 (1986).
27. A. Zur, T. C. McGill and C. A. Mead, Physical Review B, 28, 2060 (1983).
28. A. M. Goodman, Journal of Appl. Physics, 35, 573 (1964).
29. A. Thanailakis, Journal of Physics C, 8, 655 (1975).
30. E. v. Faber, Dissertation A, Technische Hochschule Ilmenau, Ilmenau (1990)
31. S. Kurtin, T. C. McGill and C. A. Mead, Physical Review Letters, 22, 1433 (1969).
32. M. Schlüter, Physical Review B, 17, 5044 (1978).
33. W. Gordy and W. J. O. Thomas, Journal of Chem. Physics, 24, 439 (1956).
34. C. A. Mead and W. G. Spitzer, Physical Review A, 134, A713 (1964).
35. J. Tersoff, Physical Review B, 32, 6968 (1985).
36. J. O. McCaldin, T. C. McGill and C.A. Mead, Physical Review Letters, 36, 56 (1976).
37. E. Louis, F. Yndurain and F. Flores, Physical Review B, 13, 4408 (1976).
38. S. G. Louie, J. R. Chelikowsky and M. L. Cohen, Physical Review B, 15, 2154 (1977).
39. V. Heine, Physical Review, 138, A1689 (1965).
40. M. J. Adams and A. Nussbaum, Solid-State Electronics, 22, 783 (1979).
41. W. A. Harrison, Journal of Vac. Sci. & Techn., 14, 1016 (1977).
42. Y.-C. Ruan and W. Y. Ching, Journal of Appl. Physics, 62, 2885 (1987).
43. F. Flores and C. Tejedor, Journal of Physics C, 12, 731 (1979) and 11, L19 (1978).