On the theory of heterojunction band offsets

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

(1) Technische Hochschule Ilmenau, PSF 327, D–O–6300 Ilmenau

(2) Zentralinstitut für Festkörperphysik und Werkstofforschung Dresden,
PSF, D–O–8027 Dresden

[PACS: 73.40L]

Abstract

The calculation of heterojunction band offsets was treated in a large number of papers. Most of them are based on assumptions which are not sufficiently justified. We present a general theory for the band offsets allowing for a deeper understanding of principles determining band offsets. At first an exactly defined equilibrium condition is used. Secondly both the band bending and the microscopic charge redistribution at the interface are taken into account by considering the dependence of the charging of the gap states on the band bending. The theory is applied to 19 heterojunctions. The mean absolute and the maximum deviations from the experimental values are significantly smaller than the corresponding values of the Tersoff theory. The latter and other previous theories are contained in our formulation as special cases.

1. Introduction

Physics and device applications of heterojunction band discontinuities are the subject of numerous publications (for reviews see [1,2,3]). The band alinement and the corresponding valence and conduction band offsets are intimately connected with the charge transfer between the two semiconductors on both sides of the heterojunction. This charge transfer is determined by the equilibrium condition together with properties of the interface as interface states and bulk properties including doping. Due to the large number of influences in the theories on band offsets till now usually only one or some of these influences are assumed to be dominating. But agreement with the experimental values was achieved by Tersoff [4]. He proposed that the band offset is determined by the requirement that the effective midgap (or neutrality) levels have to coincide. Then a charge transfer on an atomic length scale and the corresponding dipole layer are neglected. A large group of papers is based in principle on the assumption of equalizing reference levels for the band energies of both semiconductors and the occurrence of a microscopic dipole layer due to the difference of the reference levels [1]. These theories (e. g. [5,6,7]) differ from one another by quite different assumptions on the reference levels, and usually also the dipole is neglected. The agreement with experiment is worse than in the Tersoff theory. In one case the result coincides with the electron affinity rule of Anderson [8]. But actually in this case and in related theories [9,10] in addition pure surface quantities are introduced which in fact are absent at the heterojunction. This also holds for the more fundamental theory of Mailhiot & Duke [11].

[†]to whom correspondence should be addressed

In this paper a more general formulation of the theory of band offsets is presented. It allows for at first a classification of the former theories, shows the connection between them, and gives at the same time a deeper understanding in the mechanism determining the band lineup. Calculated band offsets for 19 heterojunctions show a significantly smaller mean and maximum deviations from experiment than all previous theories.

In section 2 from the equilibrium condition the magnitude of the required total interface dipole is calculated. This includes a unique definition of a reference level for the bulk energies. Also, a separation of the influence of doping is possible. In section 3 it is shown that the band offset is determined by the splitting of this interface dipole into a contribution of the band bending in both semiconductors and a second one arising from charge redistribution on an atomic length scale. The connection with former theories and relations between them are demonstrated. In section 4 the contribution of charged gap states to the microscopic interface dipole is determined. We prove that contrary to Tersoff [4] the magnitude of this contribution has to be considered. A final expression for the band offset is derived in section 5. It takes into account especially the dependence of the charging of gap states on the band bending as an important feedback mechanism stabilizing the band offset. Numerical examples are presented showing the success achieved with this formulation. Separately, in section 6 the connection with the electron affinity rule is discussed.

2. Equilibrium condition

As the starting point we consider a unique system consisting of two half infinite solids. The difference in the (electro–) chemical potentials has to be equalized by a charge transfer resulting in thermodynamic equilibrium. Hence, the initial difference in the chemical potentials is equal to the interface dipole P_0 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 already 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 free surface dipoles which are absent at the contact [12,13,14]. So, one has to remove the surface dipoles before calculating the "initial" difference in chemical potentials [15,16].

For one half–infinite solid the charge distribution without the surface dipole is a cutted (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 [17]. After parallel averaging the cut off density has an electrostatic potential which is bulk-like inside and which has to be constant outside the cut-off plane $z = z_{co}$. Therefore, just this cut-off potential is the appropriate reference level necessary to determine the initial difference in chemical potentials. Hence, we have to calculate the chemical potential relative to the cut-off potential φ_{co} outside :

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

(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 that z_{co} is determined by the bulk charge density $\varrho(\vec{r})$ alone. First, all charge densities are averaged parallel to the interface (mathematically 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}. But from (2.2) and Poisson's equation one has the electric field strength

$$\mathcal{E}(z) = \frac{1}{\varepsilon_o\, a} \int_{z-a}^{z} dz'\, z'\, \varrho(z') . \tag{2.3}$$

From (2.3) one obtains that outside the cut off density

$$\varrho_{co}(z) = \varrho(z)\, \Theta(z_{co} - z) \tag{2.4}$$

the electric field (2.3) vanishes only if the first moment of the bulk density $\varrho(z)$ becomes zero:

$$\int_{z_{co}-a}^{z_{co}} dz\, z\, \varrho(z) = 0 . \tag{2.5}$$

The cut-off plane z_{co} is uniquely determined by (2.5) for a given surface orientation and, for a more complicated crystal structure, a given top atomic layer. Having determined the cut-off plane z_{co} using (2.5) one can calculate the cut-off potential φ_{co}. Its value relative to the averaged bulk electrostatical potential $<\varphi(z)>$ will be denoted by

$$\Delta\varphi_{co,m} \equiv \varphi_{co} - <\varphi> \tag{2.6a}$$

and equals the second moment of the total charge density

$$\Delta\varphi_{co,m} = -\frac{e}{2\,\varepsilon_o\, a} \int_{z_{co}-a}^{z_{co}} dz\, z^2\, \varrho(z) > 0 \tag{2.6b}$$

where (2.2) (2.5) were used and

$$<\varphi> = \frac{1}{a} \int_{z_{co}-a}^{z_{co}} dz\, \varphi(z) . \tag{2.6c}$$

The sign of $\Delta\varphi_{co,m}$ in (2.6b) arises from the fact that the positive ionic charge is more localized than the electron density. Now one can rewrite (2.1) as

$$\mu^b = \overline{\mu} - \Delta\varphi_{co,m} \tag{2.7a}$$

where

$$\overline{\mu} \equiv \mu - <\varphi> \tag{2.7b}$$

is a pure bulk property. Hence, the potential outside the cut off density needed in (2.1) can be calculated from bulk properties using (2.5) to (2.7).

In order to calculate the chemical potentials μ^b for semiconductors besides (2.7a) a further splitting of this quantity is advantageous. The energetic distance of the chemical potential from the conduction band edge in the bulk

$$\zeta \equiv \varepsilon_c - \mu = \varepsilon_v + E_g - \mu \tag{2.8}$$

depends strongly on temperature and doping. Together with (2.8) one obtains for (2.1)

$$\mu^b = \varepsilon_v^b + E_g - \zeta \tag{2.9a}$$

where

$$\varepsilon_v^b = \varepsilon_v - \varphi_{co} \tag{2.9b}$$

is the valence band edge relative to the cut-off potential. In (2.9a) doping and temperature influence mainly ζ alone. In addition, no gap problem of the density functional theory occurs in determining ε_v^b. A further advantage of using ε_v^b becomes clear below.

Frensley & Krömer [5] calculated the valence band edge relative to the mean interstitial electrostatic potential $\varepsilon_v^{FK} = \varepsilon_v - \overline{v}$. As a rough estimate $\overline{v} \simeq \varphi_{co}$ can be used and hence one has

$$\varepsilon_v^b \simeq \varepsilon_v^{FK} = \varepsilon_v - \overline{v} . \tag{2.10}$$

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 $\varrho_{co1}(z)$ and $\varrho_{co2}(z)$ and the interface dipole layer charge density $\delta\varrho_o(z)$:

$$\varrho_o(z) = \delta\varrho_o(z) + \varrho_{co1}(z) + \varrho_{co2}(z) \tag{2.11a}$$

with

$$\varrho_{co1}(z) = \varrho_1(z) \, \Theta(z_{co1} - z) \tag{2.11b}$$

$$\varrho_{co2}(z) = \varrho_2(z) \, \Theta(z - z_{co2}) . \tag{2.11c}$$

Before thermodynamic equilibrium is reached ($\delta\varrho_o = 0$) the difference of the chemical potentials is given by the difference of their volume contributions μ^b of the two solids since the boundary condition for the electrostatic potential requires $\varphi_{co1} = \varphi_{co2}$ (Fig. 1a). This difference has to be compensated (Fig. 1c) by the interface dipole

$$P_o = -\mu_2^b + \mu_1^b \tag{2.12}$$

connected with the occurence of the deviation $\delta\varrho_o(z)$ of the total charge density from the sticked cut off densities (Fig. 1b) by

$$P_o = -\frac{e}{\varepsilon_o} \int\limits_{-\infty}^{+\infty} dz \, \delta\varrho_o \, z . \tag{2.13}$$

Eqs. (2.11) to (2.13) represent the equilibrium condition. In addition, mechanical equilibrium has to be considered. But contrary to the jellium calculation of Mailhiot & Duke [11] it influences only the shape of the charge transfer $\delta\varrho_o(z)$ and the position of the ions in the solid 1 relative to the ions in solid 2 and not the value of P_o. Reconstructions (if any) have to be considered in calculating $\delta\varrho_o(z)$ whereas P_o is fixed by the two bulk systems including interface orientations (and the top layers).

With the notation $\Delta E \equiv E_2 - E_1$ (lower indicees 1,2 for the two half spaces left and right, respectively) and using (2.9a) we obtain a new expression determining the interface dipole P_o (2.12)

$$P_o = -\Delta\varepsilon_v^b - \Delta E_g + \Delta\zeta . \tag{2.14}$$

This equation represents the equilibrium condition for semiconductor (hetero)junctions in connection with (2.11) and (2.13) [18].

At the end of this section three former approaches will be discussed. The correct splitting of the work function ϕ into surface dipole P and volume contribution is given by [19]

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

In the literature [8,20,21] often the difference of the work functions

$$u_K = \phi_2 - \phi_1 \tag{2.16}$$

is used instead of P_o. Actually, together with (2.15) and (2.12) this is

$$u_K = P_2 - P_1 - \mu_2^b + \mu_1^b \tag{2.17a}$$

$$u_K = P_2 - P_1 + P_o. \tag{2.17b}$$

Thus, u_K contains besides P_o the two surface dipoles. Furthermore, (2.17b) itself is valid only if the surface orientations at the free surface and at the contact are the same in each case. Sometimes [9,11,15] (2.17b) is used to determine P_o. But then three interfaces have to be analysed. This introduces additional difficulties. Likewise for the interface dipole [11,16]

$$P_o = -\overline{\mu_2} + \overline{\mu_1} \tag{2.18}$$

was used. This is valid for the jellium model only. A correct and complete description of the interface dipole was missing in the past. Even in attempts starting from cut off densities to describe surfaces [17] the eqs. (2.5) to (2.7) have not been given.
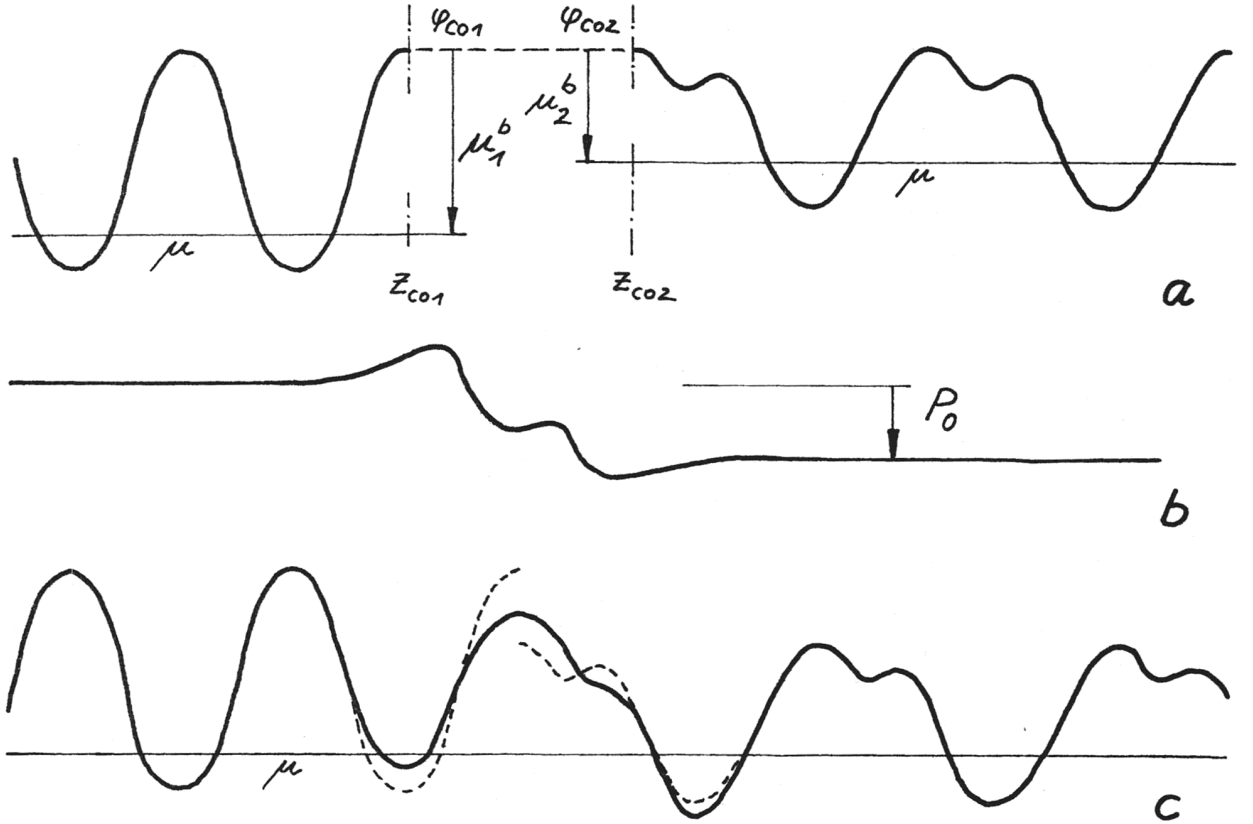

FIG. 1: Electrostatic potentials (a) of the two cut off densities ϱ_{co1} and ϱ_{co2}, (b) of the interface dipole layer charge density $\delta\varrho_o$, and (c) of the total charge density ϱ_o in thermodynamic equilibrium. – (Note that in this picture, contrary to heterojunctions, the density $\delta\varrho_o$ is extended only microscopically.)

3. The band offset

At semiconductor heterojunctions one is interested in the relative position of the band structures of the two materials near the interface. The corresponding quantities are the band offsets defined by

$$\Delta E_v = E_{v2} - E_{v1}, \qquad \Delta E_c = \Delta E_v + \Delta E_g \qquad (3.1)$$

where the band edges are measured near the interface on the same energy scale. The schematic band diagram is shown in Fig. 2. Due to charged interface states there will be a band bending on both sides of the mathematical interface. Since in thermodynamic equilibrium the chemical potentials are equalized the chemical potential can serve as an energy zero for determining the band offsets. Then the valence band edges at the interface E_{vi} depend on the total band bending: $E_{vi} = \zeta_i - E_{gi} + V_{oi}$. (Here the band bending V_o is positive always if the bands are bent upwards in direction to the mathematical interface.) Then we obtain (see also Fig. 2)

$$\Delta E_v = \Delta \zeta - \Delta E_g + \Delta V_o. \qquad (3.2)$$

Thus, the band offset can be calculated already if the so called "built-in potential" ΔV_o is known.
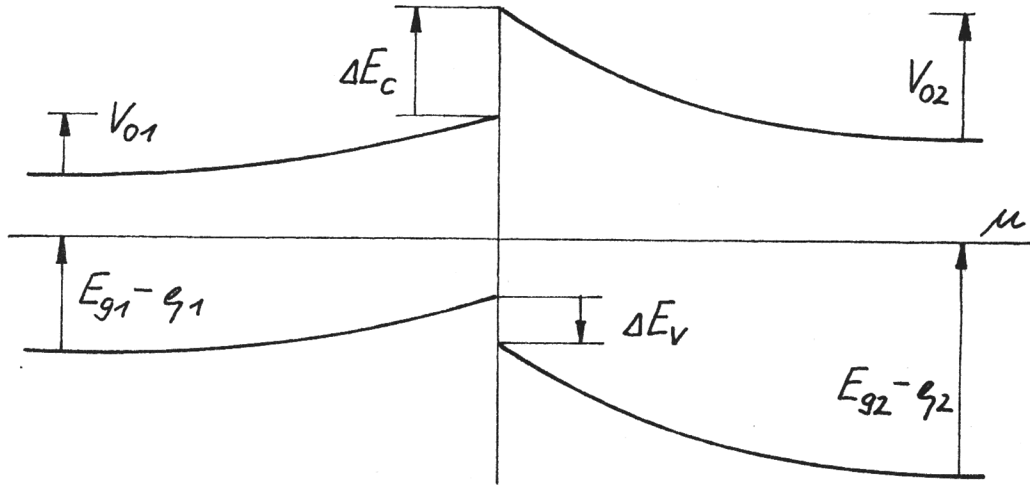

FIG. 2: Energy diagram of a heterojunction in the scale of the Debye screening length.

In order to equalize the chemical potentials an interface dipole P_o is built up (2.14). The interface dipole P_o consists of the difference of the band bending $V_{o1} - V_{o2}$ and an additional contribution D_o due to charge redistribution on an atomic scale at the interface:

$$P_o = D_o - \Delta V_o. \qquad (3.3)$$

Comparing (3.2) to (3.3) one can eliminate both the band bending V_{o1} and V_{o2}. Together with (2.14) one gets

$$\Delta E_v = \Delta \varepsilon_v^b + D_o. \qquad (3.4)$$

In this expression contrary to (3.2) the band bending does not occur explicitly. Apparently, from (3.4) one could conclude that the only (self-consistent) interface property influencing the band offset is the dipole D_o at the interface. Nevertheless, the original problem of determining D_o from (3.3) remains since the band bending and the dipole D_o are indeed coupled as will be explained in detail in the next section. Before doing this we want to discuss some microscopic concepts in the literature in which band bending and thermodynamic equilibrium were not consistently considered.

Most of the concepts do not start with thermodynamic equilibrium and alinement of chemical potentials due to band bending. So, only the interface region extended over several Å is considered. Typically, first "absolute" reference levels R_i are introduced. Then the valence band edge E_{vi} relative to this reference level R_i is calculated for each semiconductor:

$$\varepsilon_{vi} = E_{vi} - R_i .\tag{3.5}$$

After equalizing the reference levels R_1 and R_2 appropriate to

$$D_o = R_2 - R_1 \tag{3.6}$$

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 (3.5) (3.6) as

$$\Delta E_v = \varepsilon_{v2} - \varepsilon_{v1} + D_o .\tag{3.7}$$

The "absolute" valence band energies ε_{vi} should be determined by bulk properties only.

One example for such a procedure is the work of Adams & Nussbaum [7]. They used the intrinsic Fermi-level as reference. Since the valence band energies ε_v were calculated in the bulk one obtains $\Delta E_v = \Delta \zeta^i - \Delta E_g + D_o$ contrary to (3.2) and (3.4). This shows clearly that in this approach band bending is omitted. In the original paper D_o was neglected at all. A further development of the theory was given by Unlu & Nussbaum [22]. Other microscopic concepts did not consider the chemical potentials. Harrison [6] constructed the crystal potential simply by superimposing atomic orbitals. Then the potential at infinity was regarded as reference and the valence band energies calculated by tight-binding theory had to be inserted in (3.7) neglecting any dipole D_o. Some corrections were made subsequently (see [1]). According to Ruan & Ching [10] Harrison's theory coincides with the electron affinity rule established by Anderson [8]. Actually, measuring the valence band edges at free surfaces relative to the potential outside the free surface would give $\varepsilon_v = -\phi_{PE}$ where ϕ_{PE} is the photoelectric threshold [23]. Together with (3.7) one gets the electron affinity rule $\Delta E_v = -\Delta \phi_{PE}$ for $D_o = 0$. But Anderson's original idea is not a microscopic concept as will be shown in section 6.

Frensley & Krömer [5] assumed that the mean interstitial potentials $R = \bar{v}$ are the appropriate reference levels. Their valence band energies ε_v^{FK} mentioned already above (2.10) used to predict the band offset in (3.7) yield (neglecting D_o) only moderate agreement with experiment

$$\Delta E_v = \Delta \varepsilon_v^{FK} .\tag{3.8}$$

They also estimated a correction D_o determined by differences of averaged electronegativities but this did not improve their results. Nevertheless, if D_o would be taken into account correctly this approach is equivalent to the correct formulation (3.4) within the approximation (2.10).

According to Tersoff [4] the effective midgap energies $R = E_B$ of the two semiconductors have to be identical. Any deviation of the two levels at the interface would induce a strong dipole $D_o = D_{vs}$ due to charged gap states which forces the midgap levels E_B together. Since E_B is measured relative to the valence band edge $(\varepsilon_v = -E_B)$ the valence band offset is given by

$$\Delta E_v = -\Delta E_B .\tag{3.9}$$

Tersoff's approach will be discussed in more detail later.

These microscopic concepts do not contain band bending which is a part of the interface dipole (3.3). But comparing (3.7) to (3.4) one sees a similarity. (Note that (3.4) 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 $R = \varphi_{co}$ defined uniquely relative to the band structure. However, from the failure of the Frensley & Krömer attempt [5] one can realize that the contribution D_o is not negligible. On the other hand the success of Tersoff's approach [4] shows that gap states are important. But any dipole D_{vs} produced by these charged gap states can only be calculated if band bending is considered. The fundamental problem of splitting up the interface dipole P_o into space charge contribution ΔV_o and D_o due to charge redistribution on an atomic scale remains both in (3.2) and (3.4).

4. Microscopic interface contribution and charged gap states

Similar as in the case of metal-semiconductor contacts the microscopic interface contribution D_o can be regarded as consisting of two parts [15]:

$$D_o = D_e + D_{vs} \tag{4.1}$$

where D_e is due to the smooth matching of the valence charge densities and D_{vs} is created by charged gap states. Whereas D_e was estimated already by Tejedor, Flores & Louis [15] D_{vs} was not investigated thoroughly in the literature.

For the two semiconductors on both sides of the interface neutrality levels ϕ_{ni} (relative to the valence band edge at the interface) are defined such that one has local neutrality if this level coincides with the chemical potential: $\phi_{ni} = E_{gi} - \zeta_i - V_{oi}$. In general the chemical potential deviates from ϕ_{ni} by

$$\delta E_{Fi} = E_{gi} - \zeta_i - V_{oi} - \phi_{ni}. \tag{4.2}$$

This quantity depends itself on the band bending V_{oi} and gives rise to charged gap states with a charge per unit area [24]

$$\sigma_{vsi} = -e N_{vsi} \delta E_{Fi} \tag{4.3}$$

where N_{vsi} is the density of gap states near ϕ_{ni} [21] in the semiconductor i. Eq. (4.3) is already linearized assuming δE_{Fi} to be small compared to E_{gi}. The total charge

$$\sigma_{vs1} + \sigma_{vs2} = \sigma_{vs} \tag{4.4}$$

gives rise to band bending and space charge layers in both semiconductors. There remains a microscopic dipole layer of charged gap states with the effective (screened) extensions δ_i giving rise to

$$D_{vs} = -\frac{e}{\varepsilon_o} \left[\delta_2 \sigma_{vs2} - \delta_1 \sigma_{vs1} \right]. \tag{4.5}$$

Via (4.3) (4.2) D_{vs} depends on the band bending. One can rearrange (4.5) with (4.3) (4.4) as

$$D_{vs} = -\alpha \left[\delta E_{F1} - \delta E_{F2} \right] - \frac{e}{\varepsilon_o} \left[\frac{\delta_2 N_{vs2} - \delta_1 N_{vs1}}{N_{vs2} + N_{vs1}} \right] \sigma_{vs} \tag{4.6a}$$

with

$$\alpha \equiv \frac{e^2 \left(\delta_1 + \delta_2 \right)}{\varepsilon_o} \bigg/ \left[\frac{1}{N_{vs1}} + \frac{1}{N_{vs2}} \right]. \tag{4.6b}$$

The effective extensions δ_i and the densities of states N_{vsi} describe exponentially decreasing gap states. Because of neutrality $-\sigma_{vs}$ equals the total charge contained in the two space charge layers. That's why this quantity is very small compared both to σ_{vs1} and σ_{vs2}. Therefore, if the gap energies of the two semiconductors are not too much different the second term in (4.6a) can be neglected and D_{vs} is proportional to $[\delta E_{F1} - \delta E_{F2}]$ which is the energetic distance of the two neutrality levels, and the chemical potential is situated in between them. This situation was already supposed by Flores & Tejedor [9] and Tersoff [4] but without the second term of (4.6a) and without the expression (4.6b) for α. Within this approximation the total dipole (4.1) is with (4.6a)

$$D_o = D_e - \alpha \left[\delta E_{F1} - \delta E_{F2} \right] . \tag{4.7}$$

According to (4.2) the dipole contribution depends on the band bending on both sides of the interface. But, using (3.2) one can replace it by the band offset as

$$\delta E_{F1} - \delta E_{F2} = \Delta E_v + \Delta \phi_n \tag{4.8}$$

and the dipole contribution (4.7) is

$$D_o = D_e - \alpha \left(\Delta E_v + \Delta \phi_n \right) . \tag{4.9}$$

Thus, the dipole contribution (4.9) depends on the band bending but the latter itself is proportional to D_o as seen in (3.4). This coupling between the two quantities represents a negative feedback mechanism which stabilizes the band offset at its equilibrium value. The physics of this mechanism will be discussed now in some more detail. At first, according to (3.2) the band offset is proportional to the band bending difference ΔV_o. Let us now assume a slight increase of ΔV_o and hence of ΔE_v. Then, according to (4.8) the difference of the two neutrality levels (with the chemical potentials in between) increases too: Therefore, the charged gap states in the semiconductor 2 (right) and 1 (left) get a additional positive and negative charge, respectively. Accordingly, the contribution (4.7) decreases and this leads to a decrease of ΔE_v (3.4) contrary to the supposed increase. Some feedback mechanism was also proposed by Flores & Tejedor [9] and by Ruan & Ching [10]. That's why the theory of the latter provided quite reasonable good results. In Tersoff's theory [4] the effect of the gap states was immensely overstated. He assumed that there is no dipole contribution besides D_{vs} ($D_e = 0$). Then, according to his verbal arguments any deviation of the two neutrality levels is prevented by a (too) large dipole D_{vs} leading to the band offset $\Delta E_v = -\Delta \phi_n$ (4.8) (cf. (3.9)). But if, in fact, $\delta E_{F1} - \delta E_{F2}$ is very small as supposed by Tersoff the dipole D_o (4.7) is small too since $\alpha < 10$ ($D_e = 0$, $E_{g1} E_{g2} > 0.7 \, \text{eV}^2$). However, $D_o = 0$ leads with (3.4) to $\Delta E_v = \Delta \varepsilon_v^b$ which in general deviates considerably from $-\Delta \phi_n$, as will be demonstrated with the numerical examples given in the next section.

5. Evaluation of the band offset

In the preceeding sections it has been shown that the band offset ΔE_v and the dipole contribution D_o are given by the two coupled equations (3.4) and (4.9) containing the equilibrium condition and the connection between band bending and charging of the gap states. These two equations can be solved giving

$$\Delta E_v = \frac{1}{1+\alpha} \left[\Delta \varepsilon_v^b + D_e \right] - \frac{\alpha}{1+\alpha} \Delta \phi_n \tag{5.1}$$

$$D_o = \frac{1}{1+\alpha} D_e - \frac{\alpha}{1+\alpha} \left[\Delta \varepsilon_v^b + \Delta \phi_n \right] \tag{5.2}$$

where α is given by (4.6b). These equations contain only three approximations: (i) The expression (4.3) is a linearization. It gives only a minor error since according to our numerical applications usually the δE_{Fi} are smaller than one tenth of the gap energy. (ii) The second term in (4.6a) is neglected. Our estimates have shown this approximation to be justified. Note that for that reason the band offset does not depend on doping in accordance with experimental observations. But even if in some cases this term would not be small compared to the first one it could be added to D_e in (5.1) and (5.2) which itself has only a small influence as will be proved below. (iii) There is a principal uncertainty in deviding the total dipole P_o (2.14) into the band bending and the microscopic contribution (3.3). This uncertainty is just the change of the band bending over the extension of the microscopic dipole layer which is less than 0.01 eV. Therefore, the relations (5.1) (5.2) are practically exact ones. For $\alpha \to \infty$ (infinite density of gap states) (5.1) gives just Tersoff's result (3.9) even if $D_e \neq 0$ but with a finite D_o (5.2). On the other hand for $\alpha \to 0$ (5.1) one gets a result similar to that of Frensley & Krömer (3.8) and then according to (5.2) the only contribution to D_o is just D_e.

To evaluate (5.1) (5.2) one needs the valence band energies ε_v^b (2.9b), the neutrality levels ϕ_n, the quantity α (which is a kind of polarizability) and the dipole D_e due to smooth matching of the valence band states (exactly up to the lower neutrality level in the gap). D_e can be treated approximately as in the work of Tejedor, Flores & Louis [15] and of Duke & Mailhiot [16]. But it will be neglected here (see below). Further for ε_v^b we use the values ε_v^{FK} (2.10) of Frensley & Krömer [5] and for the neutrality levels ϕ_n the effective midgap values E_B [25,26] of Tersoff:

$$D_e = 0 , \qquad \varepsilon_v^b = \varepsilon_v^{FK} \; [5] , \qquad \phi_n = E_B \; [25, 26] . \qquad (5.3)$$

In a first estimate [19] we used a common value α for all contacts. Already in this case an improvement compared to Tersoff's values was achieved. A better estimate without going too much into perticulars of the electronic structure is already possible by using the simplest complex gap band structure which is symmetric relative to the midgap energy. Then within a nearly free electron model (lattice constant a) the minimum decay length of the wave function is

$$\frac{1}{q} = \frac{2 \pi \hbar^2}{m_o E_g a} = \frac{47.8 \; eV \; \text{Å}^2}{E_g \, a} . \qquad (5.4)$$

Hence, the effective extension [27] of the charge in gap states is

$$\delta = \frac{1}{2q \, \varepsilon_{vsr}} \simeq \frac{9.57 \, eV \, \text{Å}^2}{E_g \, a} \qquad (5.5)$$

where $\varepsilon_{vsr} \simeq 2.5$ is the effective dielectric constant appropriate to length $1/2q$. In addition the density of gap states is given by

$$N_{vs} \simeq \frac{2}{\pi \, a^2 \, E_g} . \qquad (5.6)$$

In Tab. 1 we report all data necessary to calculate band offsets using (5.1)–(5.3). The gaps E_g and lattice constants a are taken from Landolt-Börnstein [28] and Ruan & Ching [10]. Then the minimum decaying length $1/2q$ and the effective extension δ of the gap states were calculated using (5.4) and (5.5), respectively. The density of gap states N_{vs} estimated with the help of (5.6), the volume contributions ε_v^{FK} [5] and the midgap energies E_B [25,26] are also given.

TABLE 1: Values for calculating the band offsets (N_{vs} in $10^{18}\ m^{-2}eV^{-1}$)

sem.	E_g (eV)	a (Å)	$1/2q$ (Å)	δ (Å)	N_{vs}	ε_v^{FK} (eV)	E_B (eV)
Ge	0.67	5.658	6.311	2.524	2.968	-3.25	0.18
GaSb	0.72	6.094	5.452	2.181	2.381	-3.89	0.07
Si	1.11	5.431	3.968	1.587	1.944	-3.16	0.36
InP	1.26	5.869	3.235	1.294	1.467	-4.58	0.76
GaAs	1.43	5.642	2.965	1.186	1.399	-3.96	0.50
CdTe	1.44	6.477	2.565	1.026	1.054	-4.90	0.85
AlSb	1.60	6.136	2.437	0.975	1.057	-3.94	0.45
AlAs	2.15	5.661	1.966	0.786	0.924	-3.96	1.05
ZnTe	2.26	6.101	1.735	0.694	0.757	-4.74	0.84
GaP	2.27	5.450	1.934	0.773	0.944	-4.12	0.81
ZnSe	2.67	5.667	1.581	0.632	0.742	-5.07	1.70

Using the data of Tab. 1 we calculated the band offset (5.1) and the microscopic dipole contribution (5.2) for those heterojunctions for which experimental data for comparison are available. The latter are taken from the compilation [1]. The calculated values are given in Tab. 2 together with the values of Frensley & Krömer [5], Tersoff [25,26] and the experimental ones [1]. Frensley & Krömer's result deviate significantly stronger from experiment than Tersoff's. Eq. (5.1) shows that this fact arises from the relatively large values $\alpha \simeq 2 \ldots 11$ ($E_{g1}E_{g2} < 3\,\mathrm{eV}^2$) whereas the Frensley & Krömer limit is just $\alpha = 0$. On the other hand our complete description gives a further clear improvement compared to the Tersoff values. Besides the smaller mean square and maximum deviations from experiment it is important that for the 19 heterojunctions considered in 13 cases our result is better than Tersoff's. It should be mentioned that the mean square and maximum deviations are $0.27\,\mathrm{eV}$ and $0.6\,\mathrm{eV}$ for the electron affinity rule (see below). Thus the improvement obtained here is really important.

The calculated dipole D_O (Tab. 2) in general has the same order of magnitude as the band offset and therefore it is not negligable as resulting from Tersoff's theory. In our calculation $D_e = 0$ was assumed. For $|D_e| < 1\,\mathrm{eV}$ and with the values α (Tab. 2) eqs. (5.1) (5.2) show that this error will be of the order $0.1 \ldots 0.3\,\mathrm{eV}$ which is just the same order as our deviations from the experiment.

Finally, it should be mentioned that the theory outlined here and resulting in (5.1) (5.2) is "non-linear" in so far as the resulting band offsets are not transitive in agreement with experimental investigations [2]. This non-linearity, of course, is a result of the discussed feedback mechanism connecting the equilibrium condition with the coupling between charging of the gap states and band bending. Note that if the surface orientation and the doping of each semiconductor are the same for the contacts considered, the interface dipole P_O is transitive. Nevertheless, its parts ΔV_O and D_O (and therefore the band offset) are not transitive, since α connects properties of the two semiconductors non-linearly.

TABLE 2: Calculated band offsets ΔE_v (5.1), the calculated dipole D_o (5.2) –(see also eqs. (4.6b) (5.3) and Tab. 1)–, and for the sake of comparison the band offsets of Frensley & Krömer (3.8), of Tersoff (3.9), and from experiment [1] (' $*$ ' averaged)   (all values except α in eV)

sc. 2	sc. 1	α	$\Delta\varepsilon_v^b$ [5]	$-\Delta\phi_n$ [25,26]	ΔE_v this	ΔE_v exp.	D_o
Si	Ge	8.74	0.09	-0.18	-0.15	-0.28 *	0.24
	GaP	2.71	0.96	0.45	0.59	0.80	-0.37
	GaAs	4.08	0.80	0.14	0.27	0.05	-0.53
	GaSb	7.30	0.73	-0.29	-0.17	0.05	-0.90
	InP	4.36	1.42	0.40	0.59	0.57	-0.83
	ZnSe	2.16	1.91	1.34	1.52	1.25	-0.39
	ZnTe	2.25	1.58	0.48	0.82	0.85	-0.76
	CdTe	3.23	1.74	0.49	0.79	0.75	-0.95
Ge	AlAs	4.22	0.71	0.87	0.84	0.86 *	0.11
	GaP	4.27	0.87	0.63	0.68	0.8	-0.19
	GaAs	6.38	0.71	0.32	0.37	0.49 *	-0.34
	GaSb	11.25	0.64	-0.11	-0.05	0.2	-0.69
	InP	6.78	1.33	0.58	0.68	0.64	-0.65
	ZnSe	3.39	1.82	1.52	1.59	1.40 *	-0.23
	ZnTe	3.51	1.49	0.66	0.84	0.95	-0.65
	CdTe	4.99	1.65	0.67	0.83	0.85	-0.82
GaAs	AlAs	1.99	0.00	0.55	0.37	0.34 *	0.37
	ZnSe	1.60	1.11	1.20	1.17	1.03 *	0.06
GaSb	AlSb	4.18	0.05	0.38	0.32	0.4	0.27
deviation	m. squ.		0.56	0.21	0.15		
from exp.	max		0.99	-0.37	0.27		

6. Band offsets and electron affinity

In order to discuss the electron affinity rule and related concepts two exact equations for the band offset should be derived. The work function ϕ_s of a semiconductor written in terms of measured quantities is given by [29]

$$\phi_s = \chi_s + V_s + \zeta \tag{6.1}$$

where V_s is the total band bending at the surface and the electron affinity χ_s is obtained via the photoemission threshold

$$\phi_{PE} = \chi_s + E_g . \tag{6.2}$$

Replacing ζ in the definition of the band offset (3.2) by the work functions (6.1) and the difference of the latter by the contact potential u_K using (2.16) one obtains

$$\Delta E_v = -\Delta\chi_s - \Delta E_g + u_K + \Delta V_o - \Delta V_s . \tag{6.3}$$

This equation does not contain the equilibrium condition. The correct splitting of ϕ_s into surface and volume contributions is given by [19]

$$\phi_s = V_s + D_s - \mu^b \tag{6.4}$$

where D_s is due to charge redistribution on an atomic scale. Comparing (6.1) to (6.4) (and using (2.9a)) one obtains

$$\chi_s = D_s - \varepsilon_v^b - E_g . \tag{6.5}$$

Inserting (6.5) into (3.4) (just this equation contains the equilibrium condition) the band offset is given by

$$\Delta E_v = - \Delta \chi_s - \Delta E_g + \left(D_o + D_{s2} - D_{s1} \right) . \tag{6.6}$$

Eqs. (6.3) and (6.6) enable us to discuss several attempts to calculate the band offset using properties of free surfaces.

According to Anderson [8] the valence band offset is given by

$$\Delta E_v = - \Delta \chi_s - \Delta E_g = - \Delta \phi_{PE} . \tag{6.7}$$

His original idea can be explained with the help of (6.3). Sticking the two semiconductor surfaces together the initial difference of the chemical potentials u_K (contact potential (2.16)) is equalized by change of the band bending only. So, it is not necessary to require flat bands ($V_s = 0$) at the free surfaces as assumed by Anderson. The corresponding physical background can be extracted from (6.6). From this equation the electron affinity rule (6.7) is obtained for $D_o = D_{s1} - D_{s2}$. That means superposition of the localized surface dipoles D_{s1} and D_{s2} when the two solids are brought together to form the contact. The only moderate agreement of (6.7) with the experiment shows clearly that no superposition of that kind takes place. Although (6.6) can already be found in the work of Flores & Tejedor [9] it was first derived by Mailhiot & Duke [11] starting with bulk properties determining an equilibrium condition. But, for instance, the contributions $\Delta \varphi_{co, m}$ were omitted and flat bands at the free surfaces were assumed. It should be noted that in equations as (6.6) containing measurable quantities of the free surfaces the advantage of using this properties is lost since one needs then the unknown microscopic dipole contributions D_{si} of the free surfaces. To calculate them is as difficult as to calculate the interface dipole.

7. Summary

The theory of band offsets at semiconductor heterojunctions presented here uses the equilibrium condition with properly defined reference levels for the chemical potentials. The dipole layer arising due to equalizing these chemical potentials contains a contribution from the band bendings in the space charge layers and a microscopic dipole layer being dominated by the charging of gap states due to the deviation of the neutrality levels from the chemical potentials. This charging of gap states depends in turn on the band bending. This theory allows for a complete understanding of the mechanism stabilizing the band offset at its equilibrium value. Also the approximations underlying actually previous theories became clear in this way. The results of Frensley & Krömer and of Tersoff are special limits of our theory for vanishing and infinite densities of gap states, respectively. For 19 heterojunctions the calculated band offsets show a mean square deviation from experiment 60 meV less than the Tersoff theory and the absolute value of the maximum deviation is 100 meV smaller. For 13 of the 19 heterojunctions the calculated values are better than Tersoff's and for further two both results deviate only slightly from experiment.

References

[1] 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

[2] 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

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

[4] J. Tersoff: Physical Review B 30, 4874 (1984)

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

[6] W. A. Harrison: Journal of Vac. Sci. & Techn. 14, 1016 (1977)

[7] M. J. Adams and A. Nussbaum: Solid-State Electronics 22, 783 (1979)

[8] R. L. Anderson: Solid-State Electronics 6, 341 (1962)

[9] F. Flores and C. Tejedor: Journal of Physics C 12, 731 (1979) and
C. Tejedor and F. Flores: Journal of Physics C 11, L19 (1978)

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

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

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

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

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

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

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

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

[18] E. von Faber: Dissertation A, TH Ilmenau, Ilmenau (1990)

[19] G. Paasch and E. von Faber: Progress in Surface Science Vol. 35 (1990), p. 19

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

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

[22] H. Unlu and A. Nussbaum: Solid-State Electronics 10, 1095 (1987)

[23] F. G. Allen and G. W. Gobeli: Physical Review 127, 150 (1962)

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

[25] J. Tersoff: Journal of Vac. Sci. & Techn. B 3, 1157 (1985)

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

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

[28] Landolt-Börnstein: New Series, Group III, Vol. 17a+b, 22, Springer, Berlin/Heidelberg/New York (1982)

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