RAPID COMMUNICATIONS

PHYSICAL REVIEW B VOLUME 30, NUMBER 8 15 OCTOBER 1984

Theory of semiconductor heterojunctions: The role of quantum dipoles

J. Tersoff*
AT&T Bell Laboratories, Murray Hill, New Jersey 07974
(Received 30 May 1984)

At any semiconductor heterojunction there is an interface dipole associated with quantum-mechanical tunneling, which depends on the band "lineup" between the two semiconductors. When the interface dipolar response dominates, the actual band discontinuity must be close to that unique value which would give a zero interface dipole. A simple criterion is proposed for this zero-dipole lineup, which gives excellent agreement with experimental band lineups. The close connection between heterojunction band lineups and Schottky barrier formation is emphasized.

Semiconductor heterojunction interfaces exhibit interesting and useful electronic properties associated with the discontinuity in the local band structure at the interface. As a result, such heterostructures have become important as a basis for novel devices. However, the fundamental understanding of their electronic structure is still far from satisfactory.[1,2]

The most important single property of a semiconductor heterojunction is the band "lineup," i.e., the relative position in energy of the band gaps in the two semiconductors. This lineup determines the conduction- and valence-band discontinuities, and hence the effective barrier for electron or hole transport across the interface.

This paper presents a theory for the band lineup at ideal semiconductor interfaces. The central idea is that there is in general a dipole at the interface, associated with gap states induced by the band discontinuities. This dipole depends on the band offset, and tends to drive the band lineup toward that value which would give zero dipole. A simple estimate of this zero-dipole lineup gives excellent agreement with experimental band lineups for a number of heterojunction interfaces.

Figure 1 illustrates schematically the local band discontinuities at a heterojunction, for two possible cases (which are discussed in detail below). If the semiconductors are doped, there is also band bending on a length scale ~ 1000 Å. However, such band bending can be treated semiclassically, and is not of interest here.

Experimentally, the interface properties often depend on growth conditions, so that relatively reliable experimental values for band lineups are available only for a few systems.[1] Such effects may be due to imperfections (e.g., high densities of misfit dislocations[3]) which extend beyond the interface region, and so are outside the scope of microscopic theories of interface electronic structure. Theoretical attempts to calculate band lineups for ideal interfaces have had mixed success,[1,2] with the most realistic calculations being typically less successful than some model approaches.

The simplest theories of band lineup have supposed that the problem consists simply of relating the bands of the two semiconductors to a common absolute energy scale.[1,4-6] Such an approach assumes that no significant additional dipole is formed at the interface. Harrison[4] in particular has argued that this is the case, and has obtained reasonably good predictions of band offsets on this basis.

In general, however, the interface induces states in the gap of one or both semiconductors, analogous to the so-called "metal-induced gap states" (MIGS) at a metal-semiconductor interface.[7,8] As with MIGS, bulk electronic states in one semiconductor which fall energetically in the band gap of the other tunnel a few angstroms into the latter. The mere presence of these gap states is enough to generate an interface dipole.

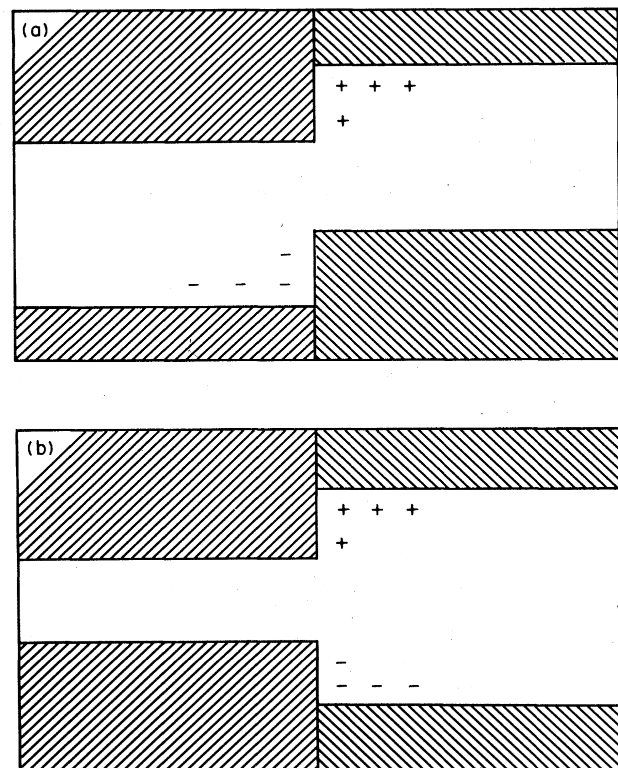

FIG. 1. Two simple examples of the relation between the band lineup and interface dipole. The band gap is shown schematically vs position normal to interface. Crosshatching shows projected bulk bands. Net charge associated with gap states is shown schematically as (+) (electron deficit) and (-) (electron excess), where states at the bottom and top of the gap are occupied or unoccupied, respectively (see text). (a) A single semiconductor in which a band discontinuity is artificially induced, e.g., by an external step potential. (b) An interface between two semiconductors, both with "symmetric" valence and conduction bands (i.e., same electron and hole effective mass, etc.), but with unequal band gaps.

 ©1984 The American Physical Society

For a given system, there exists a unique band lineup which gives a zero interface dipole. The actual lineup will not, in general, coincide with this "canonical" lineup; however, any deviation from this position gives rise to an interface dipole, which acts to drive the lineup towards the canonical value. If, as is argued below, a small displacement from the canonical lineup gives rise to a large restoring dipole, the actual lineup will be forced very close to the canonical position. Then the lineup in the *absence* of interface dipoles, which plays the central role in most previous theories, becomes relatively unimportant here.

In understanding interface dipoles, the conceptual starting point is the case of a single gap state in one dimension. Such a state may be associated with a surface, interface, or defect. The properties of gap states have been studied extensively.[9–13] Any state in the band gap is necessarily a mixture of valence- and conduction-band character; moreover, there is a sum rule on the local density of states, so that the gap state takes its spectral weight from the local valence and conduction bands, in proportion to its wave-function character. Charge neutrality occurs if the valence band is completely filled, and the conduction band completely empty. Therefore, occupying a state in the gap leads to excess net charge locally, in proportion to its degree of conduction character. Leaving the gap state empty gives a local charge deficit, in proportion to the state's valence character.

If the state lies near the bottom of the gap, filling it corresponds to only a slight excess charge, since it typically has only a little conduction character. Leaving that state empty, however, results in a charge deficit of almost one electron (i.e., almost one hole in the valence band). Conversely, filling a state high in the gap gives a large excess charge, while leaving it empty gives a small local charge deficit.

It is not hard to see how, even when there are no states at the Fermi level, changing the band lineup can give a net dipole. Two particularly simple (though artificial) cases are illustrated in Fig. 1. Consider first an interface between two semiconductors which are identical, except for an overall shift in energy. In other words, the band structures and wave functions are the same, but the zero Fourier components of the two potentials (and, hence, the electron affinities) differ by an amount V. This is equivalent to a single homogeneous semiconductor in the presence of an external step potential of height V.

According to theories which ignore the interface dipole, the band discontinuity should be precisely V. In other words, the potential step is treated as unscreened. However, in reality, the band discontinuity induces gap states and associated charges on both sides of the "interface," as shown schematically in Fig. 1(a). The resulting dipole acts to cancel the potential step. From electrostatics, one knows that the induced local dipole reduces the step by a factor of ϵ, the bulk dielectric constant. The lineup is then within $\epsilon^{-1}V$ of the canonical lineup which would give zero induced dipole (in this case the trivial homogeneous lineup). This is not to say that the induced dipole is nearly zero, but only that a very small deviation from the canonical lineup is needed to provide the screening, since ϵ is large.

At a real heterojunction between two different semiconductors, the analogy to the response of a homogeneous semiconductor to a step potential is still qualitatively correct; the effect of gap states at the interface will be to screen any deviation from the canonical lineup by a characteristic dielectric constant. For covalent and III-V semiconductors, this represents an order of magnitude reduction in the deviation. Thus, the dipole response indeed dominates the difference in electron affinity. Since dipole-free theories[4–6] give lineups typically within ~ 0.5 eV of the canonical lineups tabulated below, the screened deviations from the canonical lineups should be only ~ 0.05 eV, comparable to the numerical accuracy of the calculations here.

Another simple example is the case of "symmetric" valence and conduction bands, where the bands are mirror reflections (with respect to energy) across the center of the gap. In that case the condition of zero dipole requires that the bands of the two semiconductors be aligned symmetrically, i.e., that the centers of the gaps coincide. In that way the charges induced by gap states cancel, as illustrated in Fig. 1(b). Again, any deviation from this lineup results in a restoring dipole. Numerical calculations for model one-dimensional interfaces[14] confirm that this effect can be comparable in magnitude to the Fermi-level pinning by MIGS at a metal-semiconductor interface, and that both mechanisms drive the lineup towards the canonical position.

Both these examples illustrate the remarkable fact that the relative band positions are "pinned" by the interface electrostatics, even though there are no states at the Fermi level. Dielectric screening plays a role here similar to that attributed to metallic screening in the treatment of Schottky barriers.[8]

Real semiconductors have complicated band structures, so the lineup condition for zero dipole is not obvious. Clearly one must occupy the primarily valencelike states on both sides of the interface, while leaving the conductionlike states empty, so as to achieve local charge neutrality throughout. At some effective midgap energy E_B, the states in the gap are on the average nonbonding in character. States higher or lower in the gap have, respectively, more conduction or valence character on the average. The energy E_B thus plays a role analogous to that of the Fermi level in metals, as discussed in Ref. 8. A reasonable estimate of the zero-dipole lineup is, therefore, to align E_B for the respective semiconductors. This reduces to the symmetric lineup in the case of symmetric valence and conduction bands discussed above.

If one of the semiconductors is replaced with a metal, the heterojunction becomes a Schottky barrier. Then the band lineup suggested above reduces to aligning E_B in the semiconductor with the metal Fermi level, as in Ref. 8. Thus heterojunction band lineups and Schottky barrier heights are here treated within a single unified approach. For both types of systems, the agreement with experiment obtained below is at least as good as any other theoretical treatment to date.

The effective midgap point E_B is calculated exactly as in Ref. 8. One begins by defining the cell-averaged real-space Green's function (restricted to propagation by a lattice vector).

$$G(\vec{R},E)=\int d^3r \sum_{nk} \frac{\psi^*_{nk}(\vec{r})\psi_{nk}(\vec{r}+\vec{R})}{E-E_{nk}} = \sum_{nk} \frac{e^{i\vec{k}\cdot\vec{R}}}{E-E_{nk}} , \tag{1}$$

where $\vec{k}$ is the Bloch wave vector, n the band index, and ψ_{nk} and E_{nk} the corresponding wave function and energy. Then E_B is the energy where valence and conduction bands contribute equally to $G(R,E)$ in (1) (typically with opposite

TABLE I. Semiconductor "midgap" energy E_B, and Fermi-level positions at metal-semiconductor interfaces, relative to valence maxima (eV).

	E_B	$E_F(\mathrm{Au})$[a]	$E_F(\mathrm{Al})$[a]
Si	0.36	0.32	0.40
Ge	0.18	0.07	0.18
AlAs	1.05	0.96	
GaAs	0.70	0.52	0.62
InAs	0.50	0.47	
GaSb	0.07	0.07	
GaP	0.81	0.94	1.17
InP	0.76	0.77	

[a]Reference 18.

TABLE II. Valence-band discontinuities at selected[a] heterojunctions (eV).

	Experiments	Theory	Difference
AlAs/GaAs	0.19[b]	0.35	0.16
InAs/GaSb	0.51	0.43	−0.08
GaAs/InAs	0.17	0.20	0.03
Si/Ge	0.20	0.18	−0.02
GaAs/Ge	0.53	0.52	−0.01

[a]Reference 1.
[b]However, see text and Refs. 1, 19, and 20.

sign). In one dimension this corresponds to the branch point in the (complex) energy bands.[9,10] In three dimensions E_B has no such precise meaning, but provides a convenient criterion for the energy at which the gap states, on the average, cross over from primarily valence to conduction character.

Equation (1) requires the band structure E_{nk} as input. This is calculated as in Ref. 8, with a linearized augmented-plane-wave method.[15] The conduction bands are rigidly shifted to give the correct band gap, following Baraff and Schlüter.[16]

The calculated position of E_B with respect to the valence maximum is given in Table I for a number of covalent and III-V semiconductors. (For GaSb the effect of spin-orbit splitting is included in an approximate way.) Also given for each semiconductor is the Fermi level at interfaces with Au and (where available) Al, based on Schottky barrier measurements. According to Ref. 8, as well as the arguments above, the Fermi level at a metal-semiconductor interface should be pinned at E_B, to within the ~ 0.1-eV variation of barrier height with metal. (This variation can be understood as deriving from the electronegativity difference between different metals.[17]) Experimental values in Table I are from the classic tabulation of Sze.[18] While more recent measurements are available, a critical evaluation of barrier-height data is outside the scope of this paper.

Table II gives the most reliably known band lineup results for semiconductor heterojunctions, according to a recent review by Kroemer.[1] (Calculations of E_B have not yet been carried out for II-VI semiconductors, and so those are excluded.) Theoretical valence-band discontinuities inferred directly from results of Table I are also given. The excellent agreement between the experimental results and the theory described here, shows at the very least that available data are consistent with the assumption that quantum-mechanical dipoles are the dominant factor determining heterojunction band lineups (as well as Schottky barrier heights).

The quantitative comparison of theory and experiment in Table II must be made with some caution. The band lineup even for the extensively studied AlAs/GaAs system remains controversial[19,20] (see especially Ref. 19). Also, calculated band structures are only reliable to 0.1–0.2 eV in general. Any agreement between theory and experiment better than that in Table II would probably be fortuitous.

Note that in the present approach, the band discontinuities could be estimated by taking the difference in E_F rather than E_B in Table I. The resulting predictions are only slightly less accurate than the theoretical values given in Table II, though obtained without any calculation.

The suggestion that heterojunction band lineups and Schottky barrier heights are correlated has been made previously, but on the basis of radically different arguments. Spicer *et al.*[21] had suggested that at metal-semiconductor interfaces, the Fermi level in the semiconductor is pinned by intrinsic defects. Katnani and Margaritondo[22] pointed out that were this the case, then such defect pinning at heterojunction interfaces might also account for the band lineups. This would imply Fermi-level pinning, however, which is not observed. More recent experiments,[23,24] demonstrate that the band lineup at Ge-GaAs interfaces is not determined by such defect pinning. These studies also suggest that native defects do not play a crucial role in Schottky barrier formation.

In contradiction to previous assertions,[4] simple estimates based on dielectric screening suggest that the interface dipole is the dominant factor determining band lineup. In conjunction with a simple criterion for the zero-dipole band lineup, this view leads to quantitative predictions of *both* heterojunction band offsets *and* Schottky barrier heights. These predictions are typically accurate to ~ 0.1 eV; they require only one number (E_B) for each semiconductor, which depends only on the bulk band structure; and they involve no auxiliary hypotheses about interface structure, or the presence or absence of native or extrinsic defects.

Stimulating discussions with D. R. Hamann, M. Schlüter, D. E. Aspnes, and F. Capasso are gratefully acknowledged.

*Present address: IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y. 10598.

[1]H. Kroemer, in *Proceedings NATO Advances Study Institute on Molecular Beam Epitaxy and Heterostructures, Erice, Sicily, 1983;* edited by L. L. Chang and K. Ploog (Martinus Nijhoff, The Netherlands, in press).
[2]J. Pollmann and A. Mazur, Thin Solid Films **104**, 257 (1983).
[3]J. M. Woodall, G. D. Pettit, T. N. Jackson, C. Lanza, K. L. Kavanagh, and J. W. Mayer, Phys. Rev. Lett. **51**, 1783 (1983).
[4]W. A. Harrison, J. Vac. Sci. Technol. **14**, 1016 (1977).
[5]W. R. Frensley and H. Kroemer, Phys. Rev. B **16**, 2642 (1977); J. Vac. Sci. Technol. **13**, 810 (1976).
[6]R. L. Anderson, Solid State Electron. **5**, 341 (1962).
[7]V. Heine, Phys. Rev. A **138**, 1689 (1965).
[8]J. Tersoff, Phys. Rev. Lett. **52**, 465 (1984).
[9]W. Kohn, Phys. Rev. **115**, 809 (1959).

[10]J. J. Rehr and W. Kohn, Phys. Rev. B **9**, 1981 (1974); **10**, 448 (1974).

[11]R. E. Allen, Phys. Rev. B **20**, 1454 (1979).

[12]J. A. Appelbaum and D. R. Hamann, Phys. Rev. B **10**, 4973 (1974).

[13]F. Claro, Phys. Rev. B **17**, 699 (1978).

[14]J. Tersoff (unpublished).

[15]D. R. Hamann, Phys. Rev. Lett. **42**, 662 (1979).

[16]G. A. Baraff and M. Schlüter, Inst. Phys. Conf. Ser. No. **59**, 287 (1981).

[17]S. G. Louie, J. R. Chelikowsky, and M. L. Cohen, Phys. Rev. B **15**, 2154 (1977), and references cited therein.

[18]S. M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1969).

[19]R. C. Miller, A. C. Gossard, D. A. Kleinman, and O. Munteanu, Phys. Rev. B **29**, 3740 (1984). These results for alloy heterojunctions, if extrapolated to the pure AlAs/GaAs interface, imply a valence-band discontinuity somewhat larger than given by the theory here, whereas the experimental value given in Table II, also based on extrapolation, is smaller than the theory.

[20]J. R. Waldrop, S. P. Kowalczyk, R. W. Grant, E. A. Kraut, and D. L. Miller, J. Vac. Sci. Technol. **19**, 573 (1981).

[21]W. E. Spicer, I. Lindau, P. R. Skeath, C. Y. Su, and P. W. Chye, Phys. Rev. Lett. **44**, 420 (1980); W. E. Spicer, P. W. Chye, P. R. Skeath, C. Y. Su, and I. Lindau, J. Vac. Sci. Technol. **16**, 1422 (1979).

[22]A. D. Katnani and G. Margaritondo, Phys. Rev. B **28**, 1944 (1983).

[23]P. Chiaradia, A. D. Katnani, H. W. Sang, Jr., and R. S. Bauer, Phys. Rev. Lett. **52**, 1246 (1984).

[24]H. Brugger, F. Schaffler, and G. Abstreiter, Phys. Rev. Lett. **52**, 141 (1984).