PHYSICAL REVIEW B VOLUME 16, NUMBER 6 15 SEPTEMBER 1977

Theory of the energy-band lineup at an abrupt semiconductor heterojunction*

W. R. Frensley and H. Kroemer

Department of Electrical Engineering and Computer Science, University of California, Santa Barbara, California 93106

(Received 31 May 1977)

We present a refined model for the prediction of the energy-band lineup at an abrupt semiconductor heterojunction. The position of the energy bands with respect to the electrostatic potential is calculated by a self-consistent pseudopotential, for the bulk semiconductors. The lineup of the electrostatic potentials is then calculated from an ionic model in which the ionic charges are determined by the electronegativities of the atomic species. The resulting band lineups are independent of the crystallographic orientation of the heterojunction. They are also generally consistent with experimental data.

I. INTRODUCTION

The semiconductor heterojunction is rapidly becoming an important part of electron-device technology. The unique ability of heterojunctions to confine charge carriers to limited regions of a device has made possible the development of efficient semiconductor injection lasers.[1] Heterojunction technology has also produced periodic layered structures on a scale sufficiently small that quantum effects lead to novel optical and electronic properties.[2,3]

A central question in the study of heterojunctions concerns the energy-band discontinuity at the junction. In an earlier paper[4] (hereafter referred to as I) we presented the results of a first attempt to predict the energy-band lineup from properties of the bulk semiconductors. The approach taken was to divide the problem into two parts: A calculation of the energy-band structure relative to the periodic electrostatic potential of the bulk semiconductor, and a simple model of the charge distribution near the interface, to determine the relative lineup of the electrostatic potentials of the two semiconductors. In the present work we have divided the problem in the same way. In Sec. II we discuss a more refined model of the charge distribution near the heterojunction, which allows for charge transfer between the semiconductors, determined by the relative electronegativities of the constituents of those semiconductors. In Sec. III we present refined band-structure calculations. We have also extended the calculations to a number of technologically important semiconductors which were not considered in I.

The relationships between the various energies and potentials to be discussed are illustrated in Fig. 1. The bulk energy bands are referred to the mean interstitial potential $\overline{V}_i$. This is defined as

$$\overline{V}_i \equiv \tfrac{1}{2}[V_{es}(\vec{r}_1) + V_{es}(\vec{r}_2)], \tag{1}$$

where

$$\begin{aligned} \vec{r}_1 &= a(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}), \\ \vec{r}_2 &= a(\tfrac{3}{4}, \tfrac{3}{4}, \tfrac{3}{4}), \end{aligned} \tag{2}$$

assuming that the atomic postions are $a(0,0,0)$ and $a(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4})$. Here V_{es} represents only the *electrostatic* part of the total periodic potential. The conduction- and valence-band-edge energies E_c and E_v are expressed relative to $\overline{V}_i$. As the bandgap lies below $\overline{V}_i$, these energies are negative. The difference between the $\overline{V}_i$'s on the two sides of the heterojunction is what we call the *dipole potential* V_D,

$$V_D \equiv \overline{V}_i(1) - \overline{V}_i(2). \tag{3}$$

The band-edge discontinuities are then given by

$$\Delta E_c = E_c(2) - E_c(1) - V_D, \tag{4}$$

$$\Delta E_v = E_v(1) - E_v(2) + V_D, \tag{5}$$

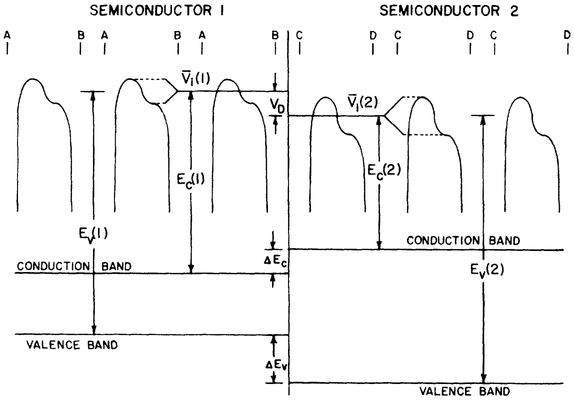

FIG. 1. Definition of the principal quantities used in the model. The figure is a diagram of energy vs position along a line normal to a (111) heterojunction between semiconductor 1 (compound *AB*) and semiconductor 2 (*CD*). The curved lines represent the electrostatic potentials. The mean interstitial potentials $\overline{V}_i$ are derived from them and used to define the band-edge energies E_v and E_c, and the dipole potential V_D.

where 1 and 2 indicate the narrower and wider gap semiconductors, respectively. Note that these are the band-edge discontinuities which would be measured very near an ideally abrupt junction, where the effects of Fermi-level equalization due to doping would be negligible.

II. MATCHING SCHEME

In I we introduced the mean interstitial potential as the reference potential in our band-structure calculations. We conjectured that heterojunction band lineups could be approximated by simply assuming that the $\overline{V}_i$'s line up across the junction. We now wish to demonstrate that the mean interstitial potential arises naturally within an ionic model of the solid and that such a model allows us to calculate the dipole potential V_D.

The model we consider is one in which the microscopic charge distribution of a semiconductor is assumed to be a linear superposition of spherical ions. If the electron density at the interstices (which are one nearest-neighbor distance, $\frac{1}{4}a\sqrt{3}$, from the ions) is negligible, the potential at these points is the same as if the ions were point charges. Within this model it is easily demonstrated that the average interstitial potential is equal to the potential at infinity. Consider a large but finite crystal bounded by *nonpolar faces*. (The requirement of nonpolar faces insures that the crystal will be electrically neutral and there will be no macroscopic electric fields due to the different charges on different faces.) Clearly there will be some unique relationship between the interstitial potential deep within the crystal and the potential at infinity. Now perform that symmetry operation on the crystal which interchanges the ionic species. As the average interstitial potential is symmetric under this operation, it remains unchanged, but, for purposes of calculating the potential, such an operation is equivalent to multiplying the ionic charges by −1, which would multiply the potential by −1. Therefore, the average interstitial potential must be equal to zero, or the potential at infinity. This argument breaks down when polar faces are present, because the macroscopic electric fields produced by such faces are not invariant under the interchange of ionic species.

TABLE I. Ion parameters.

Element	Z	Q	α (a_0^{-1})	γ (a_0^{-1})	V_0 (Ry a_0^3)	q (e)	X_{Ph}
Al	13	10	4.04	1.7	43	0.72	1.18
Si	14	10	4.52	2.25	44	⋯	1.41
P	15	10	4.99	2.5	41	1.32	1.64
S	16	10	5.47	2.6	41	1.40	1.87
Zn	30	28	3.33	2.0	25	0.79	0.91
Ga	31	28	3.64	2.0	28	0.89	1.13
Ge	32	28	3.94	2.06	36.8	⋯	1.35
As	33	28	4.22	2.2	42	1.22	1.57
Se	34	28	4.49	2.25	44	1.30	1.79
Cd	48	46	3.02	1.8	46	0.58	0.83
In	49	46	3.24	2.0	52	0.75	0.99
Sb	51	46	3.63	1.8	56	1.03	1.31
Te	52	46	3.81	1.8	58	1.13	1.47

Nevertheless, we can apply this model to a semiconductor heterojunction by conceptually bringing together two semiconductors. If we do not allow any charge redistribution between the two, the vacuum levels, and therefore the average interstitial potentials, should line up. Thus, within this very naive model of the free surface, our interstitial-potential matching scheme is equivalent to the electron-affinity rule for heterojunction band lineups proposed by Anderson.[5]

The crude scheme described above suffers from a rather severe shortcoming. If it is applied to heterojunctions on polar faces, it predicts macroscopic electric fields on one side or another of the heterojunction, depending on the configuration of the other faces of the crystal. We can overcome this problem if we think of the ionic charge as being due to a transfer of charge between nearest neighbors. Then, any region of the crystal containing many atoms will be electrically neutral. The amount of charge transferred between a given pair of nearest neighbors should be determined by some parameter of the ionic species, and a natural parameter for this purpose is the electronegativity. We have previously[6] derived effective ionic charges from our band-structure calculations and shown that these can be expressed as the difference between parameters of the ionic species, which show a good correlation with the Phillips electronegativities.[7] These parameters will be denoted q and are listed in Table I.[8] The effective charge of an ion in the bulk semiconductor AB is given by

$$e^* = q(B) - q(A) \tag{6a}$$

or

$$e^* = 0.76e[X_{\mathrm{Ph}}(B) - X_{\mathrm{Ph}}(A)] . \tag{6b}$$

(We can use either the q's or the Phillips electronegativities X_{Ph} to calculate the charge on the ion. As our band calculations give no information on the relative electronegativities of Si or Ge, we will use the Phillips electronegativities for heterojunctions involving these semiconductors, and we will use the q parameters for all other heterojunctions, because they should in principle be more accurate.) To apply this to a heterojunction, we assume that that the charge on a given ion is given by Eq. (6),

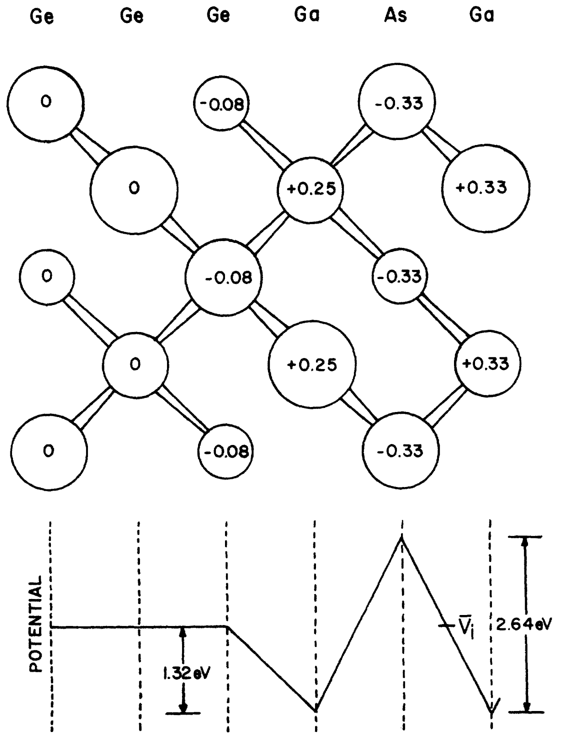

FIG. 2. Ionic model for a (100) Ge-GaAs heterojunction. The atomic positions and effective ionic charges are shown above. Below is a diagram of the plane-averaged potential.

but we average the q's of the nearest neighbors. This is equivalent to assuming a charge transfer of $\frac{1}{4}[q(B)-q(A)]$ between any pair of nearest neighbors.

Consider the case of an unreconstructed Ge-GaAs heterojunction on a (100) Ga face, as shown in Fig. 2. The charge on the Ga ions adjacent to the interface is

$$e^*_{\mathrm{Ga}} = 0.76e[\tfrac{1}{2}X_{\mathrm{Ph}}(\mathrm{As}) + \tfrac{1}{2}X_{\mathrm{Ph}}(\mathrm{Ge}) - X_{\mathrm{Ph}}(\mathrm{Ga})] = 0.25e\,. \tag{7}$$

Similarly, the charge on the Ge atoms at the interface is

$$e^*_{\mathrm{Ge}} = 0.76e[\tfrac{1}{2}X_{\mathrm{Ph}}(\mathrm{Ga}) - \tfrac{1}{2}X_{\mathrm{Ph}}(\mathrm{Ge})] = -0.08e\,. \tag{8}$$

The As ions near the interface, not being bonded to anything but Ga ions, will have the bulk GaAs charge of $-0.33e$.

Given the location and magnitude of the ionic charges, we can easily calculate the electrostatic potential. As we are only interested in dipole moments across the heterojunction, we may average the charge density over the plane *parallel* to the interface. We are then dealing with sheets of surface charge and the potential differences between them. Since we wish to express the dipole moment at the heterojunction as the difference between the average interstitial potentials in the bulk materials, we must find some way to identify those levels in such a calculation. Applying the same sort of interchange-of-ions argument as used above, we see that the average interstitial potential is the plane-averaged potential at a point halfway between any two adjacent atomic planes. In the Ge-GaAs case, there is one-half the charge transfer from Ga to Ge as there is from Ga to As. Thus the charge transfer from Ga to Ge brings the Ge potential up to the GaAs interstitial potential. In other words, the dipole across the interface (measured between interstitial potentials) is in this case negligible.

The above calculation can be generalized to a heterojunction between arbitrary semiconductors AB and CD, for interfaces on the three principal low-index planes [(100), (110), and (111)]. In all cases the results are the same, if the ideal atomic geometry is assumed. The interface dipole is given by (mks notation)

$$V_D = (1/4\epsilon_0 a)\{\tfrac{1}{2}[q(A)+q(B)] - \tfrac{1}{2}[q(C)+q(D)]\}\,, \tag{9a}$$

or

$$V_D = (0.76e/4\epsilon_0 a)\{\tfrac{1}{2}[X_{\mathrm{Ph}}(A)+X_{\mathrm{Ph}}(B)] - \tfrac{1}{2}[X_{\mathrm{Ph}}(C)+X_{\mathrm{Ph}}(D)]\}\,, \tag{9b}$$

where a is the cubic lattice constant of the (lattice-matched) semiconductors. A detailed derivation is given in the Appendix. This result is intuitively pleasing because it simply says that the dipole is proportional to the difference between the average electronegativities of the semiconductors. The direction of the dipole is such as to raise the energy bands of the more electronegative semiconductor.

Eq. (9) suggests the definition of an "electronegativity potential" V_{en} for each compound,

$$V_{\mathrm{en}}(AB) = (1/8\epsilon_0 a)[q(A)+q(B)]\,, \tag{10a}$$

or

$$V_{\mathrm{en}}(AB) = (0.76e/8\epsilon_0 a)[X_{\mathrm{Ph}}(A)+X_{\mathrm{Ph}}(B)]\,. \tag{10b}$$

The heterojunction dipole is then simply the difference between the V_{en}'s. These parameters, as derived from both the q's and the Phillips electronegativities, are given in Table II.

Unfortunately, the dipole is a small difference between relatively large electronegativity potentials. Thus, small relative uncertainties in the electronegativities lead to a rather large uncertainty in the dipole. This is particularly true for Al, where there is a significant disagreement between the results obtained from the Phillips electronegativities and our q values.

TABLE II. Semiconductor band energies. The "electronegativity potentials" V_{en} used in calculating the interface dipole are also given. All energies are in eV.

Semiconductor	E_v	E_c (direct gap)	E_c (indirect gap)	E_g (calculated)	V_{en} (from q parameters)	V_{en} (from Phillips electronegativity)
Si	−3.16	−0.21	−2.10	1.06	···	8.93
Ge	−3.25	−2.39	−2.51	0.74	···	8.21
AlAs	−3.96	−1.21	−2.00	1.96	7.75	8.35
AlSb	−3.94	−1.66	−2.35	1.59	6.45	6.98
GaP	−4.12	−1.26	−1.92	2.20	9.17	8.74
GaAs	−3.96	−2.44	···	1.52	8.44	8.21
GaSb	−3.89	−3.04	···	0.85	7.13	6.88
InP	−4.58	−3.15	···	1.43	7.98	7.70
InAs	−4.38	−3.98	···	0.40	7.36	7.26
ZnS	−5.34	−1.53	···	3.81	9.12	8.80
ZnSe	−5.07	−2.33	···	2.74	8.34	8.19
ZnTe	−4.74	−2.51	···	2.23	7.12	6.71
CdS	−5.42	−3.11	···	2.31	7.69	7.97
CdSe	−5.29	−3.53	···	1.76	7.03	7.44
CdTe	−4.90	−3.41	···	1.49	5.97	6.10

III. BAND-STRUCTURE CALCULATIONS

The present calculations were done within the self-consistent pseudopotential approach described in I, in which the potential is constructed so that the electrostatic potential can be calculated separately from the pseudizing and exchange-correlation potentials.

The ionic pseudopotential consists of the electrostatic potential and an adjustable pseudizing potential. The distribution of core electrons is assumed to be of the form

$$\rho_{\text{core}}(r) = (Q\alpha^2 e/4\pi r)e^{-\alpha r}, \tag{11}$$

where Q is the number of core electrons and α is a parameter characterizing the size of the ion. It is matched to atomic Hartree-Fock calculations[9] by setting

$$\alpha^2 = 6/\langle r^2\rangle, \tag{12}$$

where $\langle r^2\rangle$ is the second radial moment of the Hartree-Fock core electron distribution. Equation (11) leads to a core electrostatic potential of

$$V_{\text{es}}(r) = -(e^2/4\pi\epsilon_0 r)[(Z-Q)+Qe^{-\alpha r}], \tag{13}$$

where Z is the nuclear charge. The pseudizing potential is taken to be

$$V_{\text{ps}}(r) = (Ze^2/4\pi\epsilon_0 r)\exp[-(Q/Z)^{1/2}\alpha r] + (V_0\gamma^3/(2\pi)^{3/2})\exp(-\gamma^2 r^2/2). \tag{14}$$

The first term has the effect of smearing the nuclear charge, and the second provides the repulsive core. The only two freely adjustable parameters in the calculation are V_0 and γ. The total ionic pseudopotential is combined with the Hartree potential and a local exchange potential to obtain the total pseudopotential. The valence electron distribution is calculated from the wave functions and the calculations are iterated until self-consistency is achieved.

The parameters determining the ionic pseudopotentials are given in Table I. These pseudopotentials were tested, apart from the band-structure calculations, by calculating the energy levels of an electron in an isolated ionic pseudopotential. These energies were compared to experimental ionic term values and were generally found to be in agreement to within 5%.[10]

The accuracy of the band-structure calculations may be briefly summarized. Apart from AlAs, CdS, and CdSe, the agreement between calculated and experimental bandgaps is quite reasonable. If we exclude those three compounds, the average error for the first direct gap at Γ is only 0.02 eV. The $\Gamma_{15v}-\Gamma_{15c}$ is generally correct to 0.2 eV. The indirect gaps from Γ to Λ and Δ are in error by 0.2 to 0.3 eV on the average, but the minimum bandgap of the indirect gap semiconductors is generally accurate to within about 0.1 eV. The $L_{3v}-L_{1c}$ and $L_{3v}-L_{3c}$ gaps are in error by an average of 0.3 and 0.5 eV, respectively. The direct gap at X is consistently too small, by an average of 0.8 eV. AlAs shows a somewhat poorer fit, both the first direct and indirect gaps being 0.2 eV too low. The direct gap in CdS is only 0.1 eV low, but the other gaps show errors of 0.5 to 2.5 eV. There appear to be no experimental data on cubic CdSe, but the calculated bandgap is close to that experimentally observed in the hexagonal material.

The above experimental comparison is based largely on electroreflectance data. In fitting the band structures, we assumed that the E_0 peaks are due to transitions at Γ, the E_1 peaks to transitions at L, and the E_2 peaks to transitions at X. Chelikowsky and Cohen[11] have recently reported very accurate nonlocal pseudopotential calculations for several of the semiconductors considered here. Their more elaborate fit to the experimental data largely supports the naive interpretation of the E_0 and E_1 peaks, but in some cases (particularly the phosphides) they find gaps at X a few tenths of 1 eV lower than the E_2 peak. Therefore, at least some of the discrepancy between our calculated energy bands and the experimental data at X may result from an oversimplified interpretation of the spectroscopic data.

The band-edge energies, relative to the average interstitial potential, are given in Table II. The calculations were done without the spin-orbit interaction, but in the table the calculated valence-band energies have been corrected for this by adding one-third of the experimentally observed splitting (Δ_0) to the uncorrected energies. This facilitates comparison with experimental heterojunction data.

Another aspect of the band-structure calculations which deserves comment is the role of correlation effects. The dominant manifestation of such effects at a heterojunction is the dielectric image force.[12] Our approach to the lineup problem will automatically include the image force, if the correlation effects are properly taken into account in the bulk band-structure calculations. Pantelides *et al.*[13] have studied the effects of correlation on Hartree-Fock energy bands and found them to be significant. Our calculations, though, are not true Hartree-Fock calculations, since we include the exchange interaction via a Slater approximation. Phillips and Kleinman[14] noted that such an approximation incorporates the qualitative features of the correlation effects, and more recent calculations have shown that local exchange-correlation potentials seem to be adequate. The surface calculations of Appelbaum and Hamann,[15] in particular, gave an ionization potential for the Si (111) surface which is in good agreement with experiment. Since correlation (image force) effects contribute significantly to the ionization potential, this indicates that the local exchange-correlation approximation does not introduce unacceptable errors into the *absolute* band energies.

The calculations in Ref. 15 employed the full Wigner interpolation formula.[16] In I, we used a pure Slater approximation, but in the present calculations we have adopted a generalized Slater "$X\alpha$" exchange[17] with parameter $\alpha = 0.8$. This leads to an approximation which is closer to the Wigner formula for the semiconductors studied,[18] and has been used successfully in other surface calculations.[19] The effect of this change in α, apart from allowing us to better fit the bandgaps of a larger number of semiconductors, was to shift *all* band energies upward by an approximately constant 2.5 eV. This can be easily explained by noting that the Slater exchange potential, evaluated at the average valence electron density of germanium, for example, is about −12 eV. Changing α from 1 to 0.8 would then make this potential less negative by about 2.4 eV. Thus, while the band energies are sensitively dependent on the exchange parameter α, the heterojunction lineups are not, particularly in the case of lattice-matched heterojunctions, where the participating semiconductors necessarily have the same average valence electron density.

IV. HETEROJUNCTION LINEUPS

The predicted band lineups for selected heterojunctions are given in Table III, both neglecting and including the dipole correction. The band-edge discontinuities ΔE_v and ΔE_c are given by Eqs. (4) and (5). The dipole potential is given by

$$V_D = V_{\text{en}}(1) - V_{\text{en}}(2)\,, \tag{15}$$

where V is the "electronegativity potential," as defined in Eq. (10).

The Si-GaP heterojunction was discussed in I. The band lineup obtained without the dipole potential is essentially identical to our earlier calculation, but taking the dipole into account almost eliminates the conduction-band discontinuity. The few experimental data[20] indicate zero conduction-band discontinuity.

The Ge-GaAs and Ge-ZnSe predictions are also substantially the same as reported in I. The dipole potentials are negligible because the average electronegativities of GaAs and ZnSe are equal to X(Ge). The small conduction-band discontinuity at the Ge-GaAs heterojunction is in agreement with the calculations of Baraff, Appelbaum, and Hamann,[21] and at least the more recent experimental data.[22]

The best experimental data on heterojunction band lineups are those of Dingle *et al.*[2] for the GaAs-$\text{Al}_X\text{Ga}_{1-X}\text{As}$ system. For compositions in the range $X = 0.2$–0.3, they find that the valence-band discontinuity is 0.15 ± 0.03 of the total difference in bandgaps. Neglecting the dipole potential, our calculations give zero valence-band discontinuity, in contrast to the nearly perfect (and probably fortuitous) result obtained in I. If we include the dipole

TABLE III. Predicted band lineups for selected heterojunctions. The dipole is positive if it raises the narrower gap semiconductor with respect to the wider gap one. All energies are given in eV.

	Band lineup neglecting charge-transfer dipole			Band lineup including charge-transfer dipole	
Heterojunction	ΔE_v	ΔE_c	V_D	ΔE_v	ΔE_c
Si-GaP	0.96	0.18	+0.19[a]	1.15	−0.01
Ge-GaAs	0.71	0.07	0.00[a]	0.71	0.07
Ge-ZnSe	1.82	0.18	+0.02[a]	1.84	0.16
GaAs-AlAs	0.00	1.23[b]	+0.69	0.69	0.54[b]
GaAs-ZnSe	1.11	0.11	+0.10	1.21	0.01
InP-CdS	0.84	0.04	+0.29	1.13	−0.25
InAs-GaSb	−0.49	0.94	+0.23	−0.26	0.71
$Ga_{0.48}In_{0.52}As$-InP	0.40	0.09	−0.13	0.27	0.22
GaAs-$Ga_{0.52}In_{0.48}P$	0.38	0.27	−0.13	0.25	0.40

[a] Evaluated using the Phillips electronegativity scale.

[b] Measured between the Γ minimum of GaAs and the Γ minimum of AlAs. This is more appropriate for a comparison with experimental data taken on heterojunctions involving $Al_XGa_{1-X}As$ in the direct-gap composition range.

potential, our predictions become considerably poorer. Using the q's, we find $V_D = 0.69$ eV, or a valence-band discontinuity of 0.56 of the total. Using the Phillips electronegativities, we find $V_D = -0.14$ eV, which goes the wrong way. The dipole potential for this system is ultimately dependent on the relative electronegativities of aluminum and gallium. Phillips finds Al slightly more electronegative than Ga, while we find it considerably less electronegative, but one should bear in mind our poor band structure for AlAs. The truth presumably lies somewhere in between. It seems reasonable to assume that AlAs is somewhat more ionic than GaAs (the bandgap is larger, for example), but our band-structure calculations probably overestimate the ionicity. The difficulty is probably in our neglect of the effects of nonlocality of the pseudopotential. One would expect nonlocal effects to be important for elements from the Si row of the periodic table.[23]

Another point that needs to be mentioned here is that our calculations are appropriate for heterojunctions between pure semiconductors. In the above discussion we implicitly assumed that we could derive predictions for a heterojunction involving an alloy system by simply linearly interpolating the band energies for the pure compounds. Such a procedure is somewhat hazardous in view of the known nonlinearities in the dependence of bandgap on composition,[24] but lacking an adequate treatment of randomly disordered systems within the self-consistent pseudopotential scheme, this is as meaningful as any other procedure.

The InP-CdS heterojunction has been studied experimentally by Shay *et al.*[25] They find a staggered lineup, with $\Delta E_c = -0.56$ eV. Our calculations, without the dipole potential, give almost continuous conduction bands. However, when the charge-transfer effects are included, we also get a staggered lineup with $\Delta E_c = -0.25$ eV. It should be noted that our calculations were done for zinc-blende-structure CdS, whereas the experimental junctions were grown with wurtzite material. This difference, along with the inherent uncertainties in the electronegativities, and the poor CdS band structure, perhaps explains the remaining discrepancy.

In some ways the most interesting heterojunction prediction is that for the InAs-GaSb system. If we do not include the dipole effects, we find the *conduction* band of InAs about 0.1 eV *below* the *valence* band of GaSb. Similar lineups result from the predictions of Harrison[26] and the electron-affinity rule.[27] If we include charge-transfer effects via our ionic model, we find a dipole potential sufficient to raise the conduction band of InAs about 0.1 eV above the valence band of GaSb. This is still an extremely staggered band lineup, and this heterojunction is likely to have some very interesting transport properties, possibly including interband tunneling or electron-hole pair generation via the Auger effect.[28]

There is a great deal of current technological interest in heterojunctions involving the quaternary alloy $Ga_XIn_{1-X}As_YP_{1-Y}$.[29] We may use linear interpolation to derive the band energies for these alloys. We have chosen two representative heterojunctions from the many that could be fabricated within this alloy system. The ternary solid solution for $Y=1$, $Ga_{0.48}In_{0.52}As$ is lattice matched to

InP, so that a heterojunction between the two is representative of the junctions that could be grown on InP substrates. Including the dipole effects, we find about equal valence- and conduction-band discontinuities. For the GaAs-$Ga_{0.52}In_{0.48}P$ heterojunction we find a conduction-band discontinuity slightly larger than the valence-band discontinuity.

V. DISCUSSION

A. Procedure for predicting heterojunction lineups

We have presented a systematic procedure for the calculation of the energy-band lineup at an abrupt semiconductor heterojunction. The procedure requires only a knowledge of the energy-band structures of the participating semiconductors, and it does not invoke any properties—empirical or theoretical—of the free surface, nor does it require a self-consistent calculation of the electronic structure in the immediate vicinity of the interface. The accuracy of the method is believed to reflect directly the quality of the band structures. Given sufficiently accurate band structures and a sufficiently sophisticated model of the charge distribution near the interface, it should be capable of serving as an accurate tool for the quantitative prediction of band lineups at heterojunctions for which no experimental data yet exist.

The procedure itself consists of three steps: (i) the self-consistent calculation of the energy-band structures and the microscopic electrostatic potentials, (ii) the determination of an effective potential at infinity, and (iii) the estimate of the interface dipole shift by which the effective potentials at infinity are displaced relative to each other. We review these steps in more detail.

1. Band structure

We believe two aspects of our calculations to be absolutely essential for accurate band lineups: self-consistency between band structures and electrostatic potentials, and methodological consistency in the treatment of different atomic species and semiconductors. By the latter we mean perfect consistency in the treatment of the ionic cores, exchange effects, etc. One would certainly introduce unacceptable systematic errors if one were to use core data from different sources for different atomic species, or different exchange-correlation approximations for different semiconductors, or different analytical forms for the pseudopotentials. (Because of these considerations, one should particularly not try to combine the results of the present work with those of I, in view of the different exchange approximations.) Given these twofold consistencies, the choice of the calculational method is probably a purely practical question. We have used a pseudopotential approach, but other methods, such as OPW (orthogonalized plane wave), APW (augmented plane wave), or LCAO (linear combination of atomic orbitals)[26] calculations should serve as well if carried to comparable accuracies in the self-consistent band structures. We prefer the pseudopotential approach because of its overall simplicity, and because it lends itself particularly well to a self-consistent treatment. In addition, it can be easily upgraded to higher accuracy by increasing the number of adjustable parameters, leading to potentially very accurate results for a modest computational effort.[11]

In the present paper, we have implemented this general approach only in a rather low-order approximation, in terms of a local pseudopotential using only two adjustable parameters per atomic species. While the results obtained in this way are very gratifying, the band structures themselves are still too crude to be entirely satisfactory, or to allow one to place a *quantitative* reliance on the heterojunction lineups thus derived. The least satisfactory band structure (considering both the fit to experimental data and the technological importance of the material) is that of AlAs. Unfortunately, this places the greatest uncertainty on the lineup of the experimentally best-known heterojunction, AlAs-GaAs. The difficulty is probably attributable to the absence of *d* electrons in the Al-ion core, which makes a local pseudopotential a poor approximation. Thus, we are convinced that for quantitatively accurate lineup results, nonlocal pseudopotentials should be employed. Because of the importance of methodological consistency, this should be done for all atomic species of interest. We have not undertaken such a refined calculation.

2. Reference potential

We express our band-structure energies with respect to an electrostatic reference potential which is chosen so that, within some model, it is the effective potential at infinity. By "effective potential at infinity" we mean a level, defined *inside* the crystal, which would be equal to the potential far from the crystal if the crystal had a suitably idealized surface. We have shown that, within the spherical-ion model and for a crystal having only nonpolar faces, the mean interstitial potential is equal to the effective potential at infinity. The actual charge distribution is of course tetrahedrally distorted by the covalent bonds, and if one incorporated this distortion into the model, the mean interstitial potential would be somewhat lowered with respect to the effective potential at infinity.

3. Dipole potential

In many ways, the interface dipole potential would be the hardest correction to calculate *exactly* because it is a true interface property. The exact calculation of this potential would require a full self-consistent calculation of the interface itself, such as undertaken by Baraff, Appelbaum, and Hamann[21] for the Ge-GaAs heterojunction. However, it appears that in most cases the dipole shifts are surprisingly small, of the order of a few tenths of 1 eV. Therefore, a very high relative accuracy is not required, and simple estimates based on chemical-bond concepts should do. This is essentially what we have done. Again, our model could be refined. We have already mentioned the tetrahedral distortion of the valence-charge distribution. This could be taken into account by some version of the bond-charge model.[30] There is also the question of refining our charge-transfer parameters q. The value of $q(\text{Al})$ is clearly out of line with the rest of the column-III elements.[6] To use the Phillips electronegativities instead is only a temporary expedient. Ultimately, the charge transfer along the bonds should be determined from self-consistent band-structure calculations, rather than taken from sources that are conceptually different from the purely electrostatic quantities we are after. The same argument applies to the interrelation of the q-parameters between the column-IV elements, the III-V compounds, and the II-VI compounds. Simply adjusting the origins of the different q scales to fit the single Phillips electronegativity scale, as was done in Ref. 6, calls for ultimate replacement in terms of something better. Probably the best approach would be to perform self-consistent band-structure calculations on the II-IV-V_2 and I-III-VI_2 semiconductors, in addition to the III-V and II-VI compounds. This would give information on the charge transfer across I-VI, II-V, III-VI, and IV-V bonds.

B. Relation to electron-affinity concepts

Although our point of departure had been a critique[12] of the electron-affinity rule,[5] we have arrived at a rule that shares one important feature with the electron-affinity rule. The band discontinuities are transitive quantities, that is, they can be expressed as a simple difference between two numbers, each of which is a property of only one of the participating semiconductors. The original[4] model described in I, involving only the band-energy contribution, is necessarily transitive, but as Eq. (10) indicates, even the dipole potential is transitive. This aspect of the dipole potential is somewhat surprising, as it was not obviously put into the assumptions of the ionic model.

A closely related point is the lack of an orientation dependence in the dipole potential. This orientation independence presumably exists only in the model of spherical ions, ignoring any separate bond charges. It would certainly be destroyed by any nonideal atomic geometry at the interface, such as relaxation between atomic planes or periodic reconstruction of the sort that occurs at a free surface. However, any orientation dependence of the band lineup from these sources would certainly be a second-order effect and, hence, would probably be small. The only experimental evidence of an orientation dependence are the old Ge-GaAs data by Fang and Howard.[31] However, in this case, the scatter in the data for a given orientation is of the same magnitude as the difference between orientations, so these data cannot be considered conclusive.

It is instructive to compare our conduction-band energies with measured electron affinities. The electron affinity of GaAs, for example, is 4.07 eV.[27] By comparing our value of $E_c(\text{GaAs}) = -2.44$ eV, we conclude that the mean interstitial potential must lie 1.63 eV below the vacuum level. If the same comparison is made for other semiconductors, the separation between the interstitial and vacuum levels falls between 1 and 2 eV. Such a potential difference may be plausibly attributed to the combined effects of the surface dipole and the tetrahedral distortion of the valence charge.

C. Relation to other work

Until recently, the Anderson model, in which ΔE_c is postulated to be the difference between the free-surface-electron affinities, was widely accepted. However, as one of us has pointed out,[12] this model suffers from a number of limitations. More recent theoretical work on the lineup problem has taken two main approaches. One is to try to express the heterojunction band lineup in terms of the properties of the bulk crystals, as in the present work and that of Harrison.[26] The other approach is to calculate the detailed electronic structure of the interface, as is done by Baraff, Appelbaum, and Hamann,[21] and by Cohen and coworkers.[32]

Harrison has followed an approach that is quite similar to ours. He calculated an absolute energy for the top of the valence band by the linear combination of atomic orbitals (LCAO) method. The reference level for these energies is effectively set by the use of atomic term values for the diagonal elements of the Hamiltonian. Thus the band-edge energies are given with respect to the vacuum level outside a *free atom*. The significance of such a level inside the solid is not entirely clear. Empirically, Harrison's valence-band energies, $E_v(\mathrm{H})$, are related to those from the present work, $E_v(\mathrm{FK})$, by

$$E_v(\mathrm{H}) \simeq E_v(\mathrm{FK}) + 2.0\ \mathrm{eV/Å} \times a - 17.1\ \mathrm{eV}\ , \qquad (16)$$

where a is the lattice constant. Thus, Harrison's heterojunction lineups are generally in agreement with ours (neglecting the dipole potential) for lattice-matched systems.

Baraff, Appelbaum, and Hamann[21] have done a fully self-consistent surface calculation for a system with three atomic layers of Ge on (100) GaAs. They find a very small dipole moment across the heterojunction,[33] leading to a band lineup very similar to our prediction. Moreover, they have derived the effective charge transfer from the Ga to the Ge atomic planes from the variation of the potential as the interplane spacing is changed. Their value of 0.05 electron per atom is in reasonable agreement with the 0.08 electron derived from our ionic model, considering that their value is effectively a dynamic charge, while ours is a purely static one. They also find a band of partially occupied interface states which they attribute to the unsaturated chemical bonds between the Ga and Ge atoms. Our model is of course not capable of treating the interface states, but we find the agreement on the band lineup and charge transfer most gratifying.

Another approach which might be applied to the heterojunction problem is the work of McCaldin, McGill, and Mead on Schottky-barrier heights.[34] They found that the barrier height for holes in semiconductor-gold "junctions" is dependent only on the anion, and is in fact linearly related to the anion electronegativity. We find a similar, though not as strong, correlation in our valence-band energies. Also, our ionic model indicates that both the anion and cation participate equally in determining the dipole potential. Thus, these are two approaches which employ the same sort of chemical concepts, but the relationship between the two is not yet clear.

D. Interface states

Our model does not provide any information on the existence or spectral distribution of localized states at the interface. However, this is not likely to affect the validity of the band lineups. If an interface band exists that can trap free carriers from the bulk semiconductors, the result will be a sheet of charge. This will cause a discontinuity in the electric *field* at the junction, but not in the *potential*. Hence it will not affect the band discontinuities at the heterojunction.

The interface states found by Baraff, Appelbaum, and Hamann are associated with the Ga-Ge bond. These states certainly contribute to the bond charge and the charge transfer between the two atoms. To the extent that our ionic model, or some more sophisticated version of it, describes the charge distribution due to such bonding, it also includes the effects of the interface states.

Note added in proof. We have recently become aware of another electronegativity scale, proposed by St. John and Bloch [Phys. Rev. Lett. 33, 1095 (1974).] It supports our view of the relative electronegativities of Al and Ga. If these electronegativity values are applied to the GaAs-AlAs heterojunction (through a fitting procedure similar to that used in Ref. 6) the resulting V_D raises the valence band of GaAs 0.30 eV above the valence band of AlAs. This is a valence-band discontinuity of 25%, in much better agreement with the experimental data of Ref. 2.

APPENDIX: CALCULATION OF THE DIPOLE POTENTIAL

1. (100) orientation

The geometry is as shown in Fig. 3(a). The mean interstitial potential of compound AB is equal to the plane-averaged potential halfway between atomic planes A and B. Similarly, the mean interstitial potential of CD is that halfway between planes C and D. Thus the dipole potential is

$$V_D = V_{BC} - \tfrac{1}{2}V_{AB} - \tfrac{1}{2}V_{CD}\ . \qquad \text{(A1)}$$

The charge transferred between planes A and B is $\tfrac{1}{2}[q(B) - q(A)]$ per atom, therefore the effective surface charges are

$$\sigma = (1/2A)[q(B) - q(A)]\ , \qquad \text{(A2)}$$

where $A = \tfrac{1}{2}a^2$ is the area per atom. Therefore,

$$V_{AB} = (\sigma/\epsilon_0)\tfrac{1}{4}a = (1/4\epsilon_0 a)[q(B) - q(A)]\ . \qquad \text{(A3)}$$

Similarly,

$$V_{CD} = (1/4\epsilon_0 a)[q(D) - q(C)]\ , \qquad \text{(A4)}$$

$$V_{BC} = (1/4\epsilon_0 a)[q(B) - q(C)]\ . \qquad \text{(A5)}$$

Inserting (A3), (A4), and (A5) into (A1), we get

$$V_D = (1/4\epsilon_0 a)\{\tfrac{1}{2}[q(A) + q(B)] - \tfrac{1}{2}[q(C) + q(D)]\}\ . \qquad \text{(A6)}$$

2. (111) orientation

The geometry is as shown in Fig. 3(b). Again, we have

$$V_D = V_{BC} - \tfrac{1}{2}V_{AB} - \tfrac{1}{2}V_{CD}\ . \qquad \text{(A7)}$$

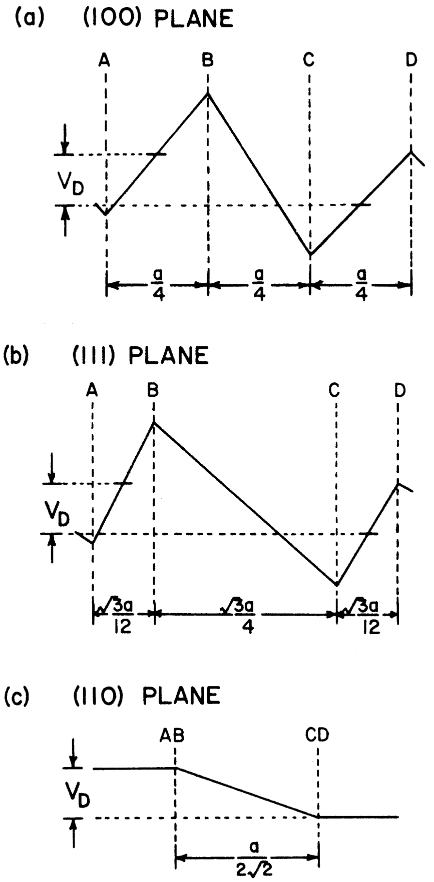

FIG. 3. Plane-averaged potential vs position for heterojunctions on the principal low-index planes.

The charge transfer between planes A and B is $\frac{3}{4}[q(B)-q(A)]$ per atom, $A=\frac{1}{4}\sqrt{3}a^2$, and the separation of pianes A and B is $\frac{1}{12}\sqrt{3}a$. Therefore,

$$V_{AB}=\frac{1}{\epsilon_0}\frac{\sqrt{3}a}{12}\frac{4}{\sqrt{3}a^2}\{\tfrac{3}{4}[q(B)-q(A)]\}$$

$$=\frac{1}{4\epsilon_0 a}[q(B)-q(A)]. \tag{A8}$$

Similarly,

$$V_{CD}=(1/4\epsilon_0 a)[q(D)-q(C)]. \tag{A9}$$

The charge transfer between planes B and C is $\frac{1}{4}[q(B)-q(C)]$, and the separation is $\frac{1}{4}\sqrt{3}a$, so

$$V_{BC}=(1/4\epsilon_0 a)[q(B)-q(C)]. \tag{A10}$$

As in the previous case, we get

$$V_D=(1/4\epsilon_0 a)\{\tfrac{1}{2}[q(A)+q(B)]-\tfrac{1}{2}[q(C)+q(D)]\}. \tag{A11}$$

3. (110) orientation

This case is somewhat different since the planes are nonpolar,

$$V_D=V_{AB}-V_{CD}, \tag{A12}$$

as illustrated in Fig. 3(c). There is a transfer of charge of $\frac{1}{4}[q(B)-q(C)]$ between the B and C atoms, but there is also a transfer of $\frac{1}{4}[q(D)-q(A)]$ in the opposite direction between the A and D atoms. Therefore the net planar charge is

$$\sigma=(1/A)\{\tfrac{1}{4}[q(B)-q(C)]-\tfrac{1}{4}[q(D)-q(A)]\}$$

$$=(\sqrt{2}/4a^2)\{[q(A)+q(B)]-[q(C)+q(D)]\}. \tag{A13}$$

Since the separation between the planes is $a/2\sqrt{2}$, the potential is

$$V_D=\frac{\sigma}{\epsilon_0}\frac{a}{2\sqrt{2}}=\frac{1}{4\epsilon_0 a}\{\tfrac{1}{2}[q(A)+q(B)]$$

$$-\tfrac{1}{2}[q(C)-q(D)]\}, \tag{A14}$$

thus demonstrating that the dipole potential is the same for the principal low-index planes.

---

*Supported by Army Research Office.

[1] For a recent review see M. B. Panish, Proc. IEEE 64, 1512 (1976).

[2] R. Dingle, A. C. Gossard, and W. Wiegmann, Phys. Rev. Lett. 34, 1327 (1975).

[3] L. Esaki and L. L. Chang, Crit. Rev. Solid State Sci. 6, 195 (1976).

[4] W. R. Frensley and H. Kroemer, J. Vac. Sci. Technol. 13, 810 (1976).

[5] R. L. Anderson, Solid-State Electron. 5, 341 (1962).

[6] W. R. Frensley and H. Kroemer, Appl Phys. Lett. 31, 48 (1977).

[7] J. C. Phillips, *Bonds and Bands in Semiconductors* (Academic, New York, 1973), Chap. 2.

[8] These are the Q parameters of Ref. 6 divided by 8.287 eV Å, so that the effective charges are given directly in units of the electron charge. See Eq. (6) of Ref. 6.

[9] J. P. Desclaux, At. Data and Nucl. Data Tables 12, 311 (1973).

[10] W. R. Frensley, Ph.D. thesis (University of Colorado, 1976) (ARO report, grant DAHC04-74-G-0114, 1976) (unpublished).

[11] J. R. Chelikowsky and M. L. Cohen, Phys. Rev. B 14, 556 (1976).

[12] H. Kroemer, Crit. Rev. Solid State Sci. 5, 555 (1975).

[13] S. T. Pantelides, D. J. Mickish, and A. B. Kunz, Phys. Rev. B 10, 2602 (1974).

[14] J. C. Phillips and L. Kleinman, Phys. Rev. 128, 2098

(1962).

[15] J. A. Appelbaum and D. R. Hamann, Phys. Rev. Lett. 32, 225 (1974).

[16] E. Wigner, Phys. Rev. 46, 1002 (1934).

[17] J. C. Slater, *The Self-consistent Field for Molecules and Solids* (McGraw-Hill, New York, 1974), Chap. 1.

[18] J. A. Appelbaum and D. R. Hamann, Phys. Rev. B 8, 1777 (1973).

[19] M. Schlüter *et al.*, Phys. Rev. B 12, 4200 (1975).

[20] G. Zeidenbergs and R. L. Anderson, Solid-State Electron. 10, 113 (1967).

[21] G. A. Baraff, J. A. Appelbaum, and D. R. Hamann, Phys. Rev. Lett. 38, 237 (1977).

[22] A. R. Riben and D. L. Feucht, Int. J. Electron. 20, 583 (1966).

[23] J. R. Chelikowsky and M. L. Cohen, Phys. Rev. B 10, 5095 (1974).

[24] J. A. Van Vechten and T. K. Bergstresser, Phys. Rev. B 1, 3351 (1970).

[25] J. L. Shay, S. Wagner, and J. C. Phillips, Appl. Phys. Lett. 28, 31 (1976).

[26] W. A. Harrison, J. Vac. Sci. Technol. 14, 1016 (1977).

[27] G. W. Gobeli and F. G. Allen, in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beer (Academic, New York, 1966), Vol. 2, p. 263.

[28] H. Kroemer, Tech. Rep. No. AFAL-TR-65-243, 1965 (unpublished) (Defense Documentation Center, Alexandria, Va). A partial published account of this work is given in A. G. Milnes and D. L. Feucht, *Heterojunctions and Metal-Semiconductor Junctions* (Academic, New York, 1972), pp. 28–29.

[29] See, for example, C. C. Shen, J. J. Hsieh, and T. A. Lind, Appl. Phys. Lett. 30, 353 (1977).

[30] J. C. Phillips, Phys. Rev. 166, 832 (1968).

[31] F. F. Fang and W. E. Howard, J. Appl. Phys. 35, 612 (1964).

[32] W. E. Pickett, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. 39, 109 (1977).

[33] They define the dipole moment as the difference between the spatially averaged pseudopotentials on either side of the junction. Such a definition is meaningful in this case, since they have specified $V_{\text{ion}}(\text{Ge}) = \frac{1}{2}[V_{\text{ion}}(\text{Ga}) + V_{\text{ion}}(\text{As})]$, but is likely to be misleading for general heterojunction pairs. There are many inequivalent ways to define the dipole moment, and one must be careful to fully specify the definition in any given calculation.

[34] J. O. McCaldin, T. C. McGill, and C. A. Mead, Phys. Rev. Lett. 36, 56 (1976).