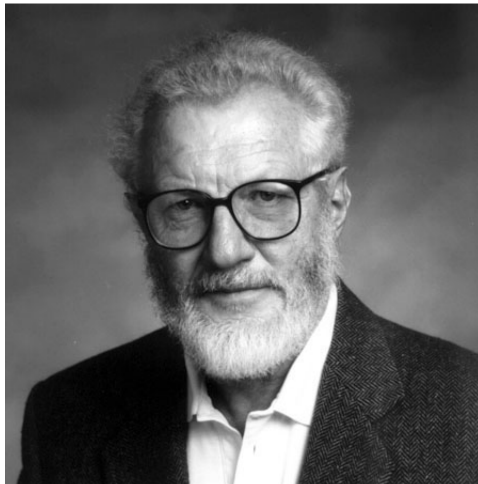

Herbert Kroemer
(1928)

17

Band diagrams of heterostructures

17.1 Band diagram lineups

In a semiconductor heterostructure, two different semiconductors are brought into physical contact. In practice, different semiconductors are "brought into contact" by epitaxially growing one semiconductor on top of another semiconductor. To date, the fabrication of heterostructures by epitaxial growth is the cleanest and most reproducible method available. The properties of such heterostructures are of critical importance for many heterostructure devices including field-effect transistors, bipolar transistors, light-emitting diodes and lasers.

Before discussing the lineups of conduction and valence bands at semiconductor interfaces in detail, we classify heterostructures according to the alignment of the bands of the two semiconductors. Three different alignments of the conduction and valence bands and of the forbidden gap are shown in ***Fig***. 17.1. ***Figure*** 17.1(a) shows the most common alignment which will be referred to as the ***straddled alignment*** or "Type I" alignment. The most widely studied heterostructure, that is the GaAs / $Al_xGa_{1-x}As$ heterostructure, exhibits this straddled band alignment (see, for example, Casey and Panish, 1978; Sharma and Purohit, 1974; Milnes and Feucht, 1972). ***Figure*** 17.1(b) shows the ***staggered lineup***. In this alignment, the steps in the valence and conduction band go in the same direction. The staggered band alignment occurs for a wide composition range in the $Ga_xIn_{1-x}As$ / $GaAs_ySb_{1-y}$ material system (Chang and Esaki, 1980). The most extreme band alignment is the ***broken gap alignment*** shown in ***Fig***. 17.1(c). This alignment occurs in the InAs / GaSb material system (Sakaki *et al*., 1977). Both the staggered lineup and the broken-gap alignment are called "Type II" energy band alignments.

At the semiconductor interface of the heterostructure, the energies of the conduction and valence band edges change. The magnitudes of the changes in the band-edge energies are critically important for many semiconductor devices.

© E. F. Schubert

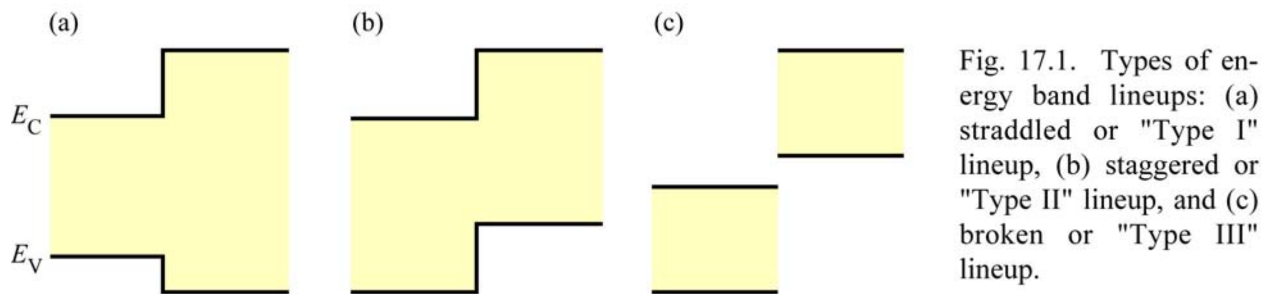

Fig. 17.1. Types of energy band lineups: (a) straddled or "Type I" lineup, (b) staggered or "Type II" lineup, and (c) broken or "Type III" lineup.

There have been numerous attempts and models to predict and calculate the energy ***band offsets*** in semiconductor heterostructures (Anderson, 1962; Harrison, 1977, 1980, 1985; Frensley and Kroemer, 1977; Kroemer, 1985; Ruan and Ching, 1987; Van de Walle, 1989; Van de Walle and Martin, 1986; Tersoff 1984, 1985, 1986; Harrison and Tersoff, 1986). The different models have been reviewed by Kroemer (1985) and by Ruan and Ching (1987). The authors showed that the agreement between the theoretical and experimental band offsets varies for the different approaches. However, none of the theoretical approaches can reliably predict the band offsets of all semiconductor heterostructure combinations. Here, we restrict ourselves to a few empirical rules and fundamental theoretical concepts which will be useful for the understanding of heterojunction band discontinuities.

Linear superposition of atomic-like potentials

We first discuss the model of the linear superposition of atomic-like potentials developed by Kroemer (1975, 1985). He pointed out that the problem of theoretically understanding the relative alignment of bands is the problem of determining the relative alignment of the two periodic potentials of the two semiconductors forming the heterostructure. Once the periodic potential of a semiconductor or of a heterostructure is known, the energy bands can be calculated.

The periodic potential of a semiconductor can be viewed as a linear superposition of the overlapping atomic-like potentials as shown in ***Fig***. 17.2. Near the atomic nuclei, the atomic-like potentials resemble the potentials inside the free atoms. However, a reconfiguration of the valence electrons occurs when initially isolated atoms form a lattice of atoms. The atomic potentials in a solid state atomic lattice will be different from the atomic potentials of isolated atoms. Therefore, the potentials in a solid-state lattice are designated as atomic-*like* potentials.

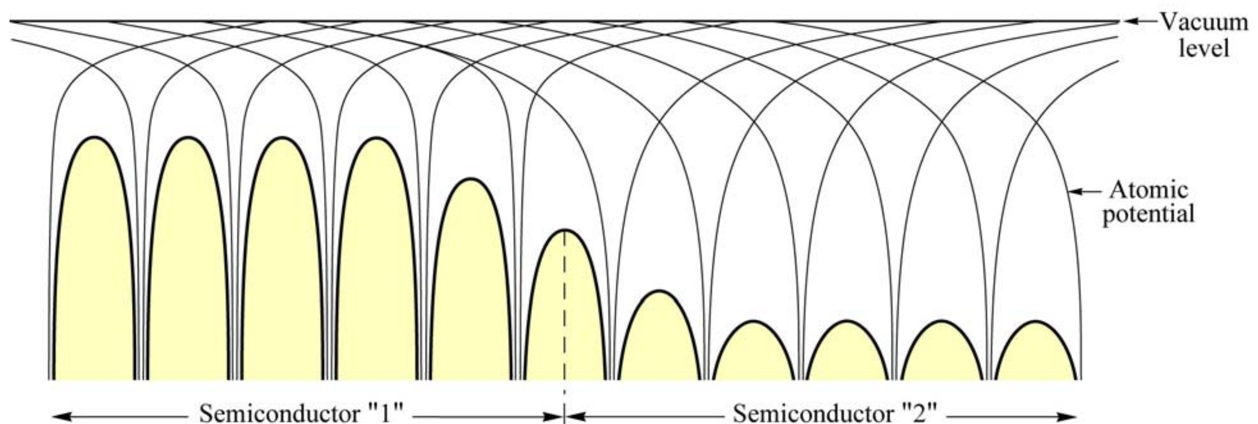

Fig. 17.2. Atomic potentials in the vicinity of two semiconductors "1" and "2". Within each semiconductor, all atomic potentials are identical. The resulting crystal potential is obtained by the superposition of all atomic potentials.

© E. F. Schubert

In the simplest atomic theory of band lineups, the unmodified atomic-like potentials would be superimposed throughout the entire structure. In the intimate vicinity of the interface, the potential would contain contributions from atoms from both sides of the interface, as shown in ***Fig***. 17.2. However, deep inside either of the two semiconductors, the periodic potential would be unaffected by the atomic-like potentials of the other semiconductor. In such a model, the lineup of the periodic potentials is well defined. The band lineups are then also well defined, and the only problems are those of the computational technique used to calculate the bandstructure from the periodic potential. Although the model of the superposition of atomic-like potentials is very instructive, the ability of this model to predict offsets between semiconductors is very limited (Kroemer, 1985).

We next consider the transition region between the two semiconductors, namely the abruptness of this transition. Atomic and atomic-like potentials are *short-range* potentials. They decay exponentially and have completely vanished after only a few inter-atomic distances, as schematically shown in ***Fig***. 17.2. As a result of the short-range nature of the atomic potential, the transition region in which the potential has intermediate values will be very thin, *i. e*. at most just a few atomic layers thick. Assuming that the bands closely follow the periodic potential, the transition of bands from the bulk structure in one semiconductor to the bulk structure in the other semiconductor will also occur within a very thin layer. The model of the linear superposition of atomic-like potentials therefore demonstrates that the transition region for chemically abrupt interfaces is very thin, namely just a few atomic layers thick. The free carrier de Broglie wavelength is much longer than the transition region. Therefore, *the potential and band transition region at the interface between two semiconductor can be considered to be abrupt* for chemically abrupt semiconductor interfaces. In other words, the electronic transition between two semiconductors is (nearly) as abrupt as the chemical transition.

Van de Walle and Martin (1986) calculated the atomic potentials of Si and Ge in Si / Ge heterostructures. The calculation indeed confirmed that the transition region from the Ge bulk periodic potential to the Si bulk periodic potential is very thin, namely just two monolayers thick. Assuming that the energy bands closely follow the periodic potential, the transition region from the Ge bulk band diagram to the Si bulk band diagram is also just a few atomic monolayers thick. Hence, the periodic potential and energy band calculations of Van de Walle and Martin clearly confirm the assumption of Kroemer that the transition region in chemically abrupt semiconductor heterostructures are just a few monolayers thick.

The electron affinity model

The electron affinity model is the oldest model invoked to calculate the band offsets in semiconductor heterostructures (Anderson, 1962). This model has proven to give accurate predictions for the band offsets in several semiconductor heterostructures, whereas the model fails for others. We first outline the basic idea of the electron affinity model and then discuss the limitations of this model.

The band diagram of a semiconductor-vacuum interface is shown in ***Fig***. 17.3. Near the surface, the n-type semiconductor is depleted of free electrons due to the pinning of the Fermi level near the middle of the forbidden gap at the semiconductor surface. Such a pinning of the Fermi level at the surface occurs for most semiconductors. The energy required to move an electron from the semiconductor to the vacuum surrounding the semiconductor depends on the initial energy of the electron in the semiconductor. Promoting an electron from the bottom of the conduction band to the vacuum beyond the reach of image forces requires work called the ***electron affinity*** χ. Lifting an electron from the Fermi level requires work called the ***work function*** W, which is defined the same way in semiconductors as it is in metals. Finally, raising an electron from the top of the valence band requires the ***ionization energy*** E_i. This energy is

measured by photoionization experiments, in which semiconductors are illuminated by monochromatic light with a variable wavelength. The longest wavelength at which photoionization occurs defines the ionization energy.

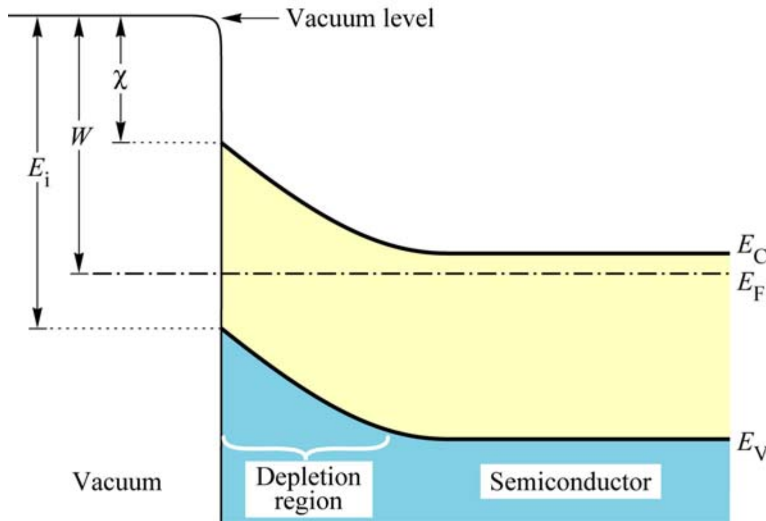

Fig. 17.3. Electron affinity χ, work function W, and ionization energy E_i of a semiconductor. The electron affinity is measured from the bottom of the conduction band at the semiconductor surface, the work function from the Fermi level, and the ionization energy from the top of the valence band at the surface.

Next consider that two semiconductors are brought into physical contact. The two semiconductors are assumed to have an electron affinity of χ_1 and χ_2 and a bandgap energy of E_{g1} and E_{g2}, respectively, as illustrated in ***Fig***. 17.4. Near-surface band bending and the effect of image forces have been neglected in the figure. The electron affinity model is based on the fact that the energy balance of an electron moved from the vacuum level to semiconductor "1", from there to semiconductor "2", and from there again to the vacuum level must be zero, that is $\chi_1 - \Delta E_c - \chi_2 = 0$ or

$$\boxed{\Delta E_c = \chi_1 - \chi_2} \tag{17.1}$$

The valence band discontinuity then follows automatically as

$$\Delta E_v = E_{g2} - E_{g1} - \Delta E_c \tag{17.2}$$

Note that Eqs. (17.1) and (17.2) are valid only if the potential steps caused by atomic dipoles at the semiconductor surfaces and the heterostructure interfaces can be neglected. In this case, the knowledge of the electron affinities of two semiconductors provides the band offsets between these two semiconductors. Shay *et al*. (1976) concluded that the influence of dipole layers at semiconductor surfaces change the values of the electron affinity by about only 1%. Therefore, the authors argued, the electron affinity rule is indeed applicable to semiconductor heterostructures.

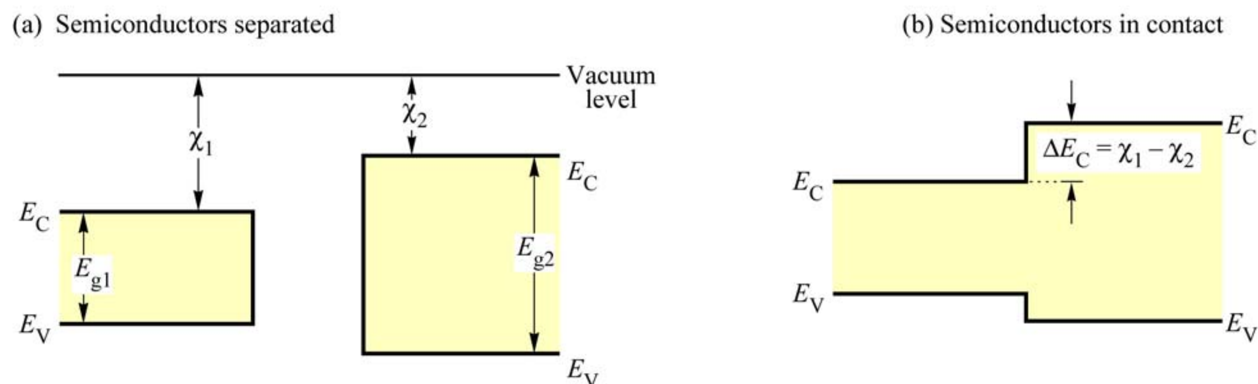

Fig. 17.4. Band diagrams of (a) two separated semiconductors and (b) two semiconductors in contact. The semiconductors have a band gap energy of E_{g1} and E_{g2} and an electron affinity of χ_1 and χ_2.

The electron affinity model has successfully explained the band discontinuities of several semiconductor heterostructures. In the InAs / GaSb material system, the electron affinity rule correctly predicts a broken-gap alignment (Gobeli and Allen, 1966; Kroemer, 1985). The highly asymmetric lineup of InAs / GaAs heterostructures is also predicted well (Kroemer, 1985). In the Si / Ge heterostructure system, the electron affinity model predicts $\Delta E_c = 0.12$ eV and $\Delta E_v = 0.33$ eV in reasonable agreement with experimental data (Kroemer, 1985). Shay *et al*. (1976) and Phillips (1981) used the electron affinity rule to calculate ΔE_c in CdS / InP heterostructures and found excellent agreement with their experimental data.

Despite the reasonable agreement between theory and experiment, the electron affinity model suffers from several conceptual problems which have been pointed out by Kroemer (1985). *First*, surface dipole layers affect the measurement of the electron affinity. Generally, all semiconductor surface undergo surface reconstruction, *i. e*. a rearrangement of atoms on the semiconductor surface in order to reduce the total energy of the semiconductor surface. Such a surface reconstruction includes frequently the outward or inward displacement of surface atoms. As a result, electrostatic dipole layers are formed which will change the measured electron affinity. At semiconductor-semiconductor interfaces, the interface reconstruction will be clearly different than the surface reconstruction. As a consequence, the magnitude of interface dipoles will be different. Therefore, the measurement of χ is influenced by surface effects and the measured values of χ will not be meaningful for semiconductor heterostructures, unless the influence of surface and interface dipoles is negligible small, or if the surface dipoles are identical to the interface dipoles. Both possibilities are unlikely. However, Shay *et al*. (1976) pointed out that the influence of surface dipoles is very small for most semiconductor surfaces. *Second*, electron correlation effects also influence the measured values of the electron affinity (Kroemer, 1985). When one electron is taken from a semiconductor and promoted to the vacuum level, the remaining electrons will rearrange themselves in order to reduce the total energy of the electron system. Such correlation effects are due to coulombic repulsion between electrons but also due to quantum-mechanical exchange effects (essentially the Pauli exclusion principle). Generally, the magnitude of correlation effects is small. Due to the dipole and correlation effects, the applicability of the electron affinity rule is limited to semiconductors in which these effects are negligibly small.

It is useful to recall that the electron affinity model was invoked by Schottky (1938, 1940) explain the barrier heights of metal-semiconductor contacts also called Schottky contacts. Schottky proposed that the barrier height be given by the difference in the work function in the

metal and the electron affinity of the semiconductor, *i. e.* $W - \chi$. However, it is well known, that the Schottky model clearly fails to explain the barrier heights in metal-semiconductor contacts. Subsequently, Bardeen (1947) showed, the important role of interface states whose energy is within the forbidden gap. Bardeen showed that interface dipoles caused by charged interface states determine the barrier height of metal-semiconductor contacts and that the difference $W - \chi$ does not play a significant role. In lattice-matched semiconductor-semiconductor junctions, the influence of interface dipoles cannot be possibly as large as it is in metal-semiconductor junctions. Lattice-matched semiconductor heterostructures have highly ordered atomic transitions between the two semiconductors with relatively little atomic and electronic reconstruction. Therefore, the electron affinity model is expected to provide much better results for semiconductor-semiconductor junctions than it does for metal-semiconductor junctions. This expectation is indeed confirmed by experimental results.

Common anion rule

Many compound semiconductor heterostructures consist of two compounds which share a common anion element. For example in AlGaAs / GaAs heterostructures, As is the anion element on both sides of the heterostructure. It is a well established fact that the valence band wave functions evolve mainly from the atomic wave function of anions and the conduction band wave functions evolve mainly from the atomic wave functions of cations (see, for example, Harrison, 1980). Hence, the valence band structure of different semiconductors with the same anion element will be similar. Furthermore, *the valence band offsets of compound semiconductors with the same anion element is generally smaller than the conduction band offset*. This rule is clearly confirmed in the material system $Al_xGa_{1-x}As$ / GaAs where $\Delta E_c / \Delta E_v \approx 2/1$ for direct-gap range of $Al_xGa_{1-x}As$ ($x \leq 0.45$). The common anion rule also works well for GaAs / InAs heterostructures in which $\Delta E_c / \Delta E_v \approx 5/1$ (Kowalczyk *et al.*, 1982).

Harrison atomic orbital model

Harrison (1977, 1980, 1985) developed a theory based on atomic orbitals to predict band offsets in semiconductor heterostructures. Kroemer (1985) compared the Harrison atomic orbital model and other models with experiments and he arrived at the conclusion that the Harrison model gives very good overall agreement with experimental band offsets.

The basis of the Harrison model is the *linear combination of atomic orbitals* of a very small group of atoms which is then used to calculate the band structure. The band structure calculation would be correct if the true atomic-like potentials and energy eigenfunctions of the atoms forming the semiconductor would be known. Because the atomic-like potentials and eigen energies of the atoms in the crystal lattice are unknown, Harrison simply takes as unperturbed atomic energy values the theoretical values of *free* atoms. Hence, the Harrison model is clearly an approximation. In this model, several more approximations are employed for the calculation of the matrix elements coupling the relevant atomic states between nearest neighbors. For further discussion, the reader is referred to the literature (Harrison, 1977, 1980, 1985, Kroemer, 1985).

A comparison between Harrison's theoretical and experimental valence band offsets is shown in ***Fig***. 17.5. The data used in the figure was compiled by Kroemer (1985) except the value for the GaAs / $Al_xGa_{1-x}As$ where $\Delta E_v = 0.32\ \Delta E_g$ has been used, consistent with more recent results (Pfeiffer *et al.*, 1991). ***Figure*** 17.5 displays a very good overall agreement between experiment and the Harrison atomic orbital model.

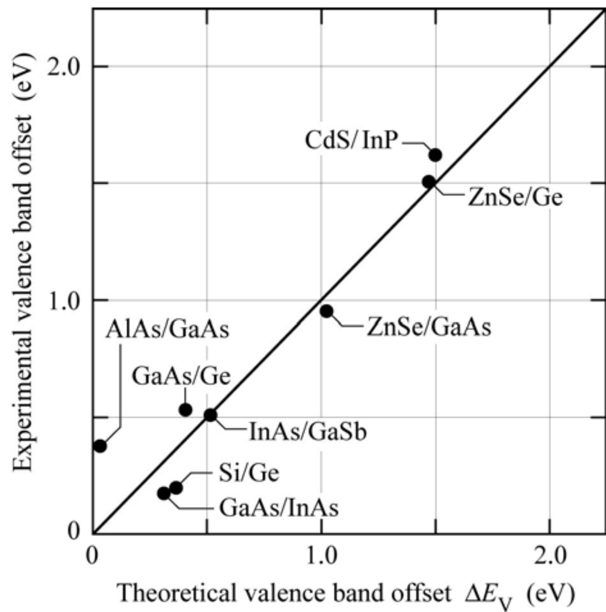

Fig. 17.5. Comparison of experimental valence band offsets with theoretical valence band offsets calculated by the Harrison atomic orbital theory. The AlAs / GaAs value is extrapolated from $Al_{0.30}Ga_{0.70}As$ / GaAs assuming $\Delta E_V / \Delta E_g = 1/3$.

The effective dipole model

As we have already stated above, any dipole charges at the heterointerface will change the heterostructure band discontinuity. These dipole charges are due to the locally different atomic and electronic structure at the heterointerface as compared to the bulk atomic structure of either semiconductor. As a result of the different atomic environment at the heterointerface, valence electrons of atoms at the interface will move from their bulk equilibrium positions to new equilibrium positions. Hence, atomic dipoles are formed due to the new charge distribution at the heterointerface.

Ruan and Ching (1987) calculated heterostructure band offsets based on (i) the electron affinity model and (ii) by taking into account atomic dipoles at the interface which cause an additional shift of the band discontinuity. The authors pointed out that interface dipoles are neglected in Anderson's electron affinity model. If no net charge is transferred between the two semiconductors forming the heterojunction, then the Anderson model gives the correct band offset. (We do not consider here the difficulties in obtaining the correct electron affinities χ_e, but simply assume that they are known. Ruan and Ching used “average values of those experimental data which are judged to be current and reliable”.) To calculate the charge transfer between the two semiconductors forming the heterojunction, the authors assume that the two valence bands are misaligned, that is the valence band edges of the two semiconductors have different energies. Electrons with an effective mass m^* in the valence band of one semiconductor will tunnel into the forbidden regions of the other semiconductor. The dipole charge is calculated by integrating over the exponentially decaying charge distribution of electrons tunneling into into the forbidden gaps of the adjoining semiconductor. Using this method to calculate the band offsets between semiconductors, Ruan and Ching (1987) calculated nearly all conceivable heterostructure band offsets. A comparison revealed that the theoretical band offsets of Ruan and Ching differs, on average, only about 0.1 eV from the available experimental data.

Experimental data of band offsets between different semiconductors are given in Table 3. The table includes data for elemental as well as binary and ternary compound semiconductors. Tiwari and Frank (1992) used experimental data of band offsets in order to plot the band edges of semiconductors as a function of the lattice constant. The plot, shown in ***Fig***. 17.6 relies on the experimental observation that the band offset from material “A” to material “B” plus the offset from material “B” to material “C” is equal to the band offset from material “A” to material “C”.

This *linearity* of offsets is consistent with the electron affinity model, and this property allows one to predict band alignments of any semiconductor heterostructure.

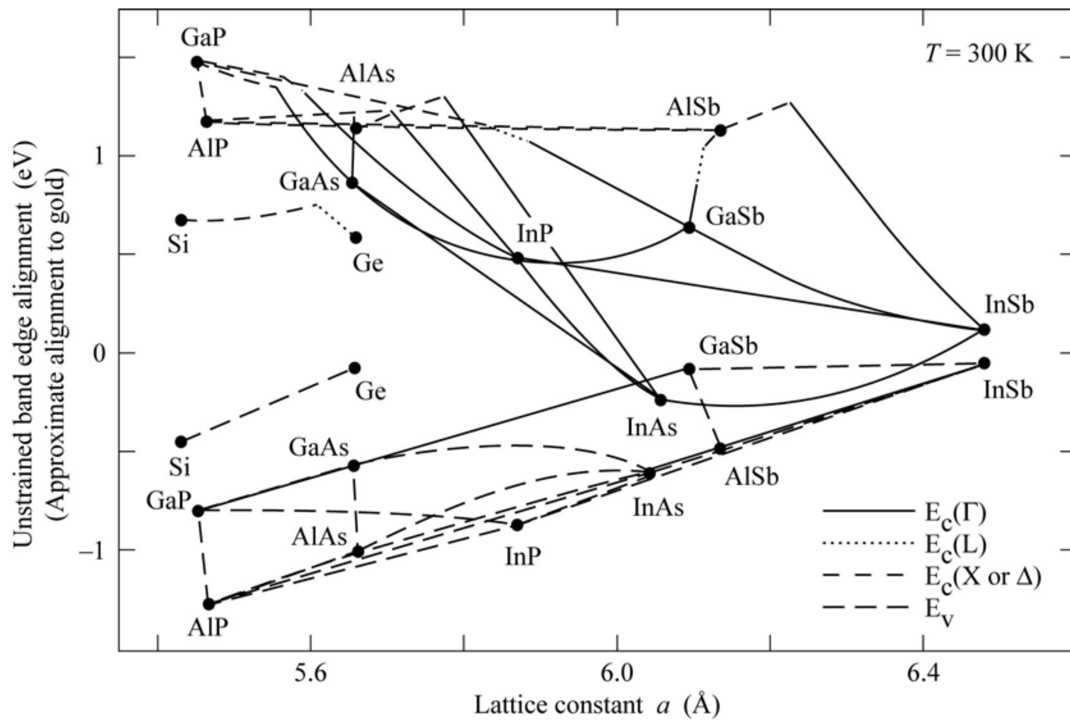

Fig. 17.6. Band edges as a function of the lattice constant. The zero energy point represents the approximate Fermi level (located mostly in the forbidden gap) of the gold / semiconductor Schottky contact (after Tiwari and Frank, 1992).

© E. F. Schubert

Material system A / B	E_g^A (eV)	E_g^B (eV)	ΔE_V (eV)	ΔE_V / ΔE_g (absolute value)	Remarks
Si / Ge	1.12	0.67	–0.16 to –0.40	0.35 to 0.89	(a)
Si / GaP	1.12	2.25	+0.80	0.71	(b)
Si / GaAs	1.12	1.42	+0.05	0.17	(b)
Si / GaSb	1.12	0.72	–0.05	0.12	(b)
Si / ZnSe	1.12	2.70	+1.25	0.79	(b)
Si / CdTe	1.12	1.52	+0.75	1.87	(b)
Ge / AlAs	0.67	2.15	+0.92	0.62	(b)
Ge / GaAs	0.67	1.42	+0.25 to +0.65	0.33 to 0.87	(b)
Ge / InP	0.67	1.34	+0.64	0.95	(b)
AlAs / GaAs	2.15	1.42	–0.40	0.55	(c)
$Al_{0.3}Ga_{0.7}As$ / GaAs	1.79	1.42	–0.12	0.32	(d)
AlSb / GaSb	1.61	0.72	–0.4	0.45	(b)
GaAs / InAs	1.42	0.36	–0.17	0.16	(b)
GaAs / ZnSe	1.42	2.70	+0.96 to +1.10	0.75 to 0.86	(b)
GaSb / InAs	0.72	0.36	+0.46	1.28	(b)
InP / CdS	1.34	2.42	+1.63	1.51	(b)
$Al_{0.48}In_{0.52}As$ /	1.45	0.75	–0.21	0.30	(e)
$Ga_{0.47}In_{0.53}As$	1.88	1.42	–0.23	0.50	(f)
$Ga_{0.52}In_{0.48}P$ / GaAs	1.45	1.34	+1.19	1.73	(g)
$Al_{0.48}In_{0.52}As$ / InP	0.75	1.34	+0.40	0.68	(g)
$Ga_{0.47}In_{0.53}As$ / InP					

(a) after Ruan and Ching (1987). Van de Walle and Martin (1986) showed that E_V depends strongly on strain.
(b) after Ruan and Ching (1987)
(c) indirect gap AlAs, after Ruan and Ching (1987)
(d) direct gap $Al_xGa_{1-x}As$, after Pfeiffer *et al*. (1991) and after Menendez *et al*. (1986)
(e) after Peng *et al*. (1986) and after Sugiyama *et al*. (1986)
(f) Rao *et al*. (1987)
(g) after Tiwari and Frank (1992)

Table 17.1: Bandgap energies and valence band offsets of semiconductor heterostructures "A / B". The valence band offset ΔE_V is positive, if the top of the valence band of semiconductor "A" is higher than that of semiconductor "B".

17.2 Boundary conditions at heterointerface

So far we have seen how the energy bands in semiconductors evolve and how these bands align in semiconductor heterostructures. We have also seen that the transition from one band diagram to the band diagram of another semiconductor is very abrupt in a chemically abrupt semiconductor heterostructure. In this section we will discuss the transition of other physical quantities at the heterointerface. These transitions follow a number of rules and these rules are called the *boundary conditions* of the heterointerface. Below, the boundary conditions will be summarized. A heterointerface is schematically illustrated in ***Fig***. 17.7. The interface is located in the plane $z = 0$ of a cartesian coordinate system. The semiconductors "1" and "2" have a dielectric permittivity, magnetic permeability, and effective mass of ε_1, μ_1, m_1^*, and ε_2, μ_2, m_2^*, respectively.

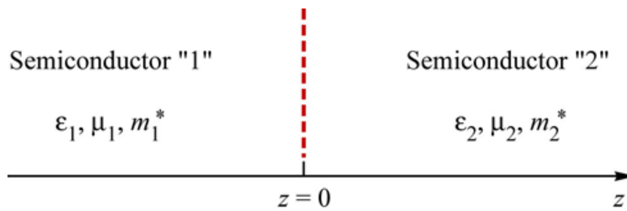

Fig. 17.7. Boundary between two semiconductors "1" and "2" which have a permittivity of ε_1, and ε_2, and an effective mass m_1^* and m_2^*, respectively.

The first boundary condition considered here concerns the Fermi level. *The **Fermi level** is constant across a heterointerface under thermal equilibrium conditions*. The Fermi level is defined as the energy at which electronic states are populated with a probability of one half. We next assume a heterointerface, in which the Fermi level on one side of the heterointerface is, at a given time, different from the Fermi level on the other side of the heterointerface. As a consequence, electrons will transfer from the semiconductor with the higher Fermi level to the semiconductor with the lower Fermi level, where they can occupy states at lower energy. Thus, the Fermi level rises in the semiconductor with the initially lower Fermi level. (Also, the Fermi level decreases in the semiconductor with the initially higher Fermi level.) The transfer of electrons continues, until the Fermi level is the same on both sides of the heterointerface. Thus, under thermal equilibrium conditions, the Fermi level is constant across heterointerfaces.

Four electrodynamic boundary conditions must be satisfied at heterointerfaces. The general boundary conditions for electric and magnetic fields were derived in Chap. 2. For completeness, these boundary conditions are summarized as follows: The ***magnetic boundary condition*** states that the tangential component of the magnetic field, $\mathcal{H}_t$, and the normal component of the magnetic induction, $\mathcal{B}_n$, are constant across interfaces. The ***electric boundary condition*** states that the tangential component of the electric field, $\mathcal{E}_t$, and the normal component of the dielectric displacement, $\mathcal{D}_n$, are constant across interfaces.

The latter boundary condition, $\mathcal{D}_n$ = const, is now used to derive another "boundary condition", namely the ***charge neutrality condition***. We denote the normal component of the dielectric displacement at the interface as D_{1n} and D_{2n} in semiconductor "1" and "2", respectively. Furthermore we assume that the dielectric displacement vanishes for sufficiently large distances from the interface. Then, using Gauss's equation, the boundary condition $\mathcal{D}_{1n} = \mathcal{D}_{2n}$ can be written as

$$\mathcal{D}_{1n} \quad = \quad \int_{z=-\infty}^{z=0} \rho(z)\,dz \quad = \quad -\int_{z=0}^{z=\infty} \rho(z)\,dz \quad = \quad \mathcal{D}_{2n}\,. \tag{17.3}$$

The equation states that the net charge on one side of the heterointerface $z \le 0$, left-hand side of Eq. (17.3) must be equal to the negative net charge on the other side of the heterointerface

($z \geq 0$, right-hand side of Eq. (17.3)). All charges must be taken into account in Eq. (17.3) including free-carrier accumulation layer charges, free-carrier inversion layer charges, and depletion layer charges. Charges at heterointerfaces are caused by the transfer of carriers from one semiconductor across the heterointerface to the other semiconductor. All of these charges remain in close vicinity of the heterointerface. Thus the condition of charge neutrality of Eq. (17.3) can be stated as: *There is not net charge in the vicinity of heterointerfaces*.

We finally discuss the boundary conditions for the quantum-mechanical wave function at heterointerfaces. The interface is located in the plane $z = 0$ and we are only interested in the z dependence of the wave function $\psi(z)$. In the chapter entitled "Resume of quantum mechanical principles", the boundary conditions for $\psi(z)$ is given by

$$\boxed{\psi_1(z \to -0) \quad = \quad \psi_2(z \to +0)} \tag{17.4}$$

That is, the wave function is continuous at the interface.

The boundary condition for the derivative of the wave function is given by

$$\boxed{\frac{1}{m_1^*} \frac{\mathrm{d}\psi_1}{\mathrm{d}z}\bigg|_{z \to -0} \quad = \quad \frac{1}{m_2^*} \frac{\mathrm{d}\psi_2}{\mathrm{d}z}\bigg|_{z \to +0}} \tag{17.5}$$

The proof of this equation is given in the chapter entitled "Resume of quantum mechanical principles".

17.3 Graded gap structures

In regular semiconductor heterostructures, the chemical transition from one semiconductor to another semiconductor structure is abrupt. In the preceding section, we have seen that the periodic potential and the band diagram are nearly as abrupt as the chemical transition. That is, the transition of the periodic potential and of the band diagram occur within a few atomic layers of a chemically abrupt semiconductor heterostructure. In graded heterostructures, the chemical transition from one semiconductor to another semiconductor is intentionally graded. In this section, the properties of such graded heterostructures are discussed.

Assume two semiconductors "A" and "B" that are chemically miscible. The mixed compound, also called ***semiconductor alloy***, is designated by the chemical formula A_xB_{1-x}, where x is the ***mole fraction*** of semiconductor "A" in the mixed compound. The mole fraction x is also designated as the chemical ***composition*** of the compound A_xB_{1-x}. Most semiconductors of practical relevance are completely miscible. Assume further that the gap energy of semiconductor "A" and "B" are different and that the bandgap energy depends on the composition. The dependence of the forbidden-gap energy on the composition x is usually expressed in terms of a parabolic (linear plus quadratic) dependence. The gap energy of the alloy A_xB_{1-x} is then given by

$$E_\mathrm{g}^\mathrm{AB} \quad = \quad x\, E_\mathrm{g}^\mathrm{A} \;+\; (1-x)\, E_\mathrm{g}^\mathrm{B} \;+\; x(1-x)\, E_\mathrm{b} \tag{17.6}$$

where the first two summands describe the linear dependence of the gap and the summand $x\,(1-x)\,E_\mathrm{b}$ describes the quadratic dependence of the gap. The parameter E_b is called the ***bowing parameter***. For some semiconductor alloys, *e. g.* $(AlAs)_x(GaAs)_{1-x}$, the bowing parameter is vanishingly small. The bandgap of the alloy is then given by

$$E_g^{AB} = x E_g^A + (1-x) E_g^B \quad (17.7)$$

Equations (17.6) and (17.7) are valid for homogeneous bulk semiconductors. However, the validity of the equations is not limited to bulk semiconductors. They also apply to the local bandgap of graded structures. We have seen in the preceding section that the atomic potentials and the energy bands closely follow the composition in a chemically abrupt heterojunction. Accordingly, the band edges and the gap energy will follow the chemical composition of graded semiconductors.

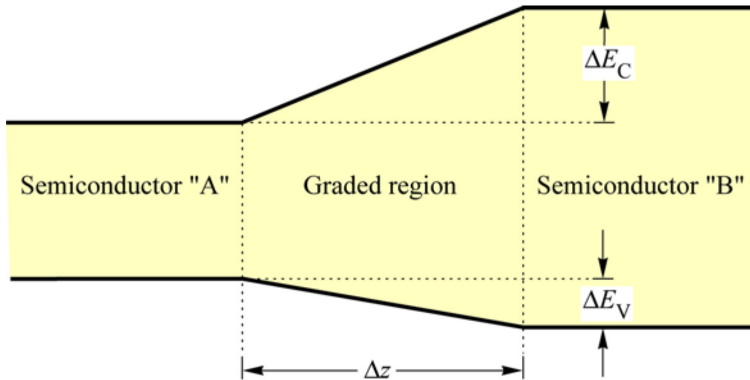

Fig. 17.8. Linearly graded region between a semiconductor "A" and "B" creating a quasi-electric field in the graded transition region.

The band diagram of a linearly graded semiconductor heterostructure is illustrated in ***Fig***. 17.8. The figure shows a narrow-gap semiconductor "A", a wide-gap semiconductor "B", and a linearly graded transition region "A_xB_{1-x}" with thickness Δz. It is assumed that ΔE_c, ΔE_V, and ΔE_g depend linearly on the composition x. Graded gap semiconductor structures were first considered by Kroemer (1957). He showed that the changes of the band edge energies with position can be understood as ***quasi-electric fields***. The quasi-electric field of the band diagram shown in ***Fig***. 17.8 is, in the conduction band, given by

$$|\mathcal{E}_C| = \Delta E_C / (e\,\Delta z) \quad (17.8)$$

In the valence band it is given by

$$|\mathcal{E}_V| = \Delta E_V / (e\,\Delta z) \quad (17.9)$$

Figure 17.8 reveals that the electric fields in the conduction band and in the valence band have opposite polarization. Therefore, electrons and holes are driven in the *same* direction (to the left-hand side of figure). This cannot be achieved by real electric fields in which electrons and holes are *always* driven in opposite directions. Due to this difference, Kroemer (1957) designated the fields occurring in graded semiconductor structures as ***quasi-electric fields***. Kroemer envisioned several different cases of graded gap structures and pointed out that in some graded gap structures, electrons and holes are pulled in the same direction, as discussed for the band diagram shown in ***Fig***. 17.8. In other graded gap structures, one of the bands could, *e. g*. the valence band, may be flat, while the other band could have a quasi-electric field. Kroemer also envisioned graded-gap heterobipolar transistors which enhance the minority carrier transport through the base.

Kroemers conjecture that the quasi-electric fields exert forces on free carriers was experimentally verified by Levine *et al*. (1982, 1983). The authors showed that the quasi-electric fields act on one carrier type just like regular electric fields of the same magnitude would. That

© E. F. Schubert

is, the drift velocity relation $v = \mu \xi$ (where ξ is the quasi-electric field) was confirmed by a modified Shockley-Haynes experiment (Levine *et al*., 1982).

There are many very interesting applications for graded gap structures including graded-base heterobipolar transistors (Kroemer, 1957; Miller *et al*. 1983; Hayes *et al*. 1983), the elimination of heterojunction band discontinuities (Schubert *et al*. 1992), and several graded gap structures for optoelectronic applications such as photodetectors. Graded-gap detectors have been reviewed by Capasso (1984, 1986).

Let us consider some experimental results of alloy semiconductors. The energy gap of the unstrained Si_xGe_{1-x} was analyzed as a function of composition by Weber and Alonso (1989) using low-temperature photoluminescence. Analytical expressions were given for the lowest energy gap as a function of the composition. For $x \le 0.85$, the X band is the lowest conduction band minimum. The energy gap of unstrained Si_xGe_{1-x} is given by

$$E_g(x) = \left(1.155 - 0.43\,x + 0.206\,x^2\right)\text{eV} \qquad \text{for} \qquad x \le 0.85 \tag{17.10a}$$

For $x > 0.85$, the L band is the lowest conduction band minimum. The energy gap is then given by

$$E_g(x) = \left(2.010 - 1.270\,x\right)\text{eV} \qquad \text{for} \qquad x > 0.85 \tag{17.10b}$$

For strained Si_xGe_{1-x} grown on Si substrates, Lang *et al*. (1985) showed that the degenerate valence band splits into two bands. The energy gap between the lowest conduction band and the highest valence band is then given by

$$E_g(x) = \left(1.155 - 0.65\,x + 0.22\,x^2\right)\text{eV} \qquad \text{for} \qquad x \le 0.70 \tag{17.10c}$$

This equation is an analytical expression of the low-temperature (90 K) photoluminescence data of Lang *et al*.

The energy gaps of ternary III-V alloy semiconductors have been compiled by Swaminathan and Macrander (1991). The data is summarized in Table 4. The energy gaps of quaternary III-V and for II-VI semiconductors will not be summarized here. The interested reader is referred to the literature (Pearsall, 1982; Landolt-Börnstein, 1987).

In graded semiconductor structures, the composition of the semiconductor is varied. This variation in chemical composition is not only accompanied by a change of the bandgap energy, but also by a change in the lattice constant. The change in lattice constant is, for all semiconductor alloys, governed by Vegard's law. Consider a semiconductor "A" with a lattice constant a_0^A and a semiconductor "B" with the lattice constant a_0^B. Then the lattice constant of the alloy A_xB_{1-x} is given by ***Vegard's law*** which states

$$\boxed{a_0^{AB} = a_0^A\,x + a_0^B\,(1-x)} \tag{17.11}$$

For most graded semiconductor structures, it is imperative that the lattice constant does not change as the composition of the alloy is varied. Such structures are called ***lattice-matched*** graded semiconductors. If semiconductors are not lattice matched, graded semiconductors. If semiconductors are not lattice matched, microscopic defects occur when the composition is varied. These defects degrade the quality, *e. g*. the radiative efficiency, of the semiconductor.

© E. F. Schubert

The relationship between the gap energy, the corresponding wavelength, and the lattice constant of group-IV and group III-V semiconductors is shown in ***Fig***. 17.9 (Tien, 1985). A similar plot is shown in ***Fig***. 17.10 for II-VI semiconductors (Feldman *et al*., 1992). The two plots allow one to select lattice-matched semiconductors with the desired bandgap energy. ***Figure*** 17.9 reveals that the lattice constant of the material system $Al_xGa_{1-x}As$ / GaAs does not change, as the composition x is varied. This advantageous properly allows one to easily grow lattice-matched graded structures with this material system.

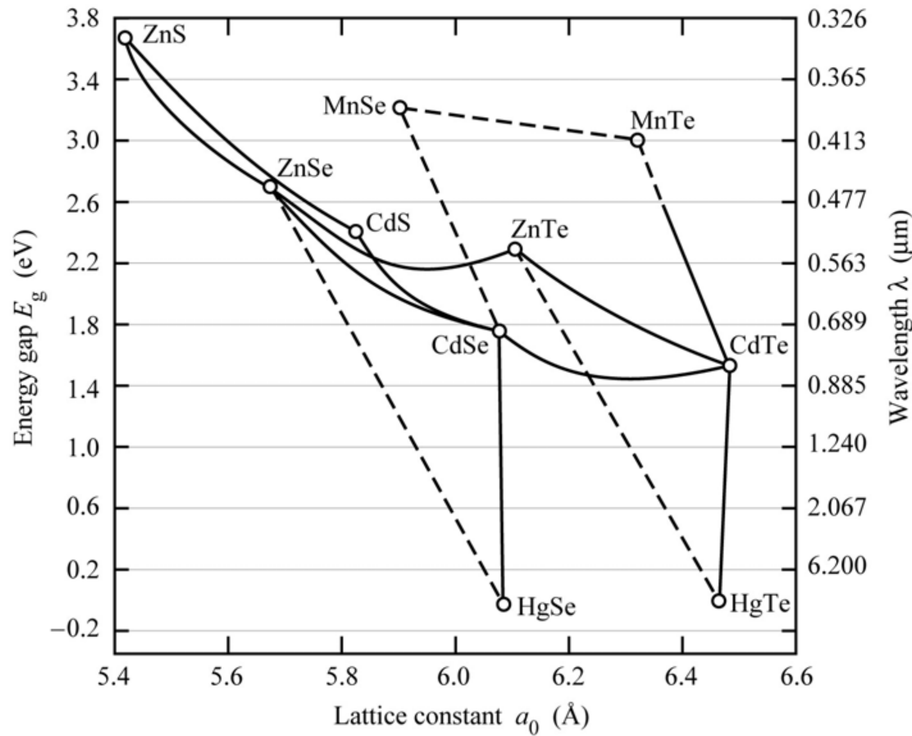

Fig. 17.11 Lattice constant versus energy gap at room temperature for various II–VI semiconductors and their alloys (after Feldman *et al*., 1992).

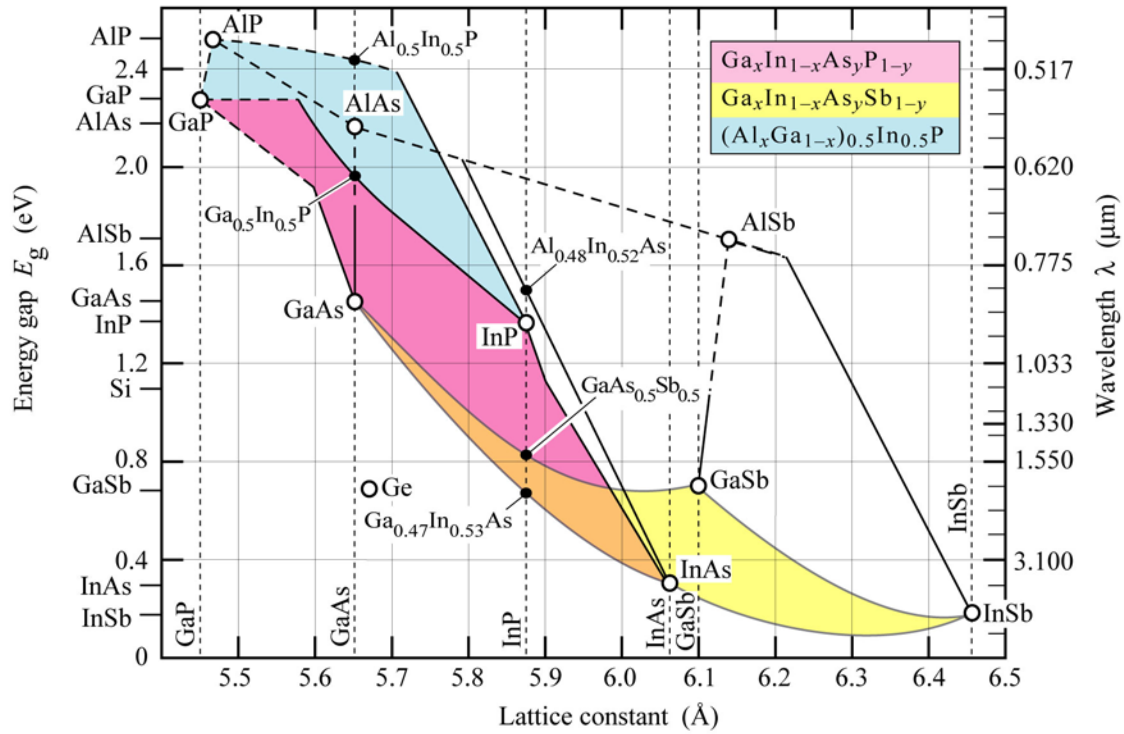

Fig. 17.9 Lattice constant versus energy gap at room temperature for various III–V semiconductors and their alloys (after Tien, 1985).

Alloy	Direct Energy Gap E_Γ (eV)	Indirect Energy Gap E_X (eV)	E_L (eV)
$Al_xIn_{1-x}P$	$1.34 + 2.23x$	$2.24 + 0.18x$	
$Al_xGa_{1-x}As$	$1.424 + 1.247x$ ($x < 0.45$) $1.424 + 1.087x + 0.438x^2$	$1.905 + 0.10x + 0.16x^2$	$1.705 + 0.695x$
$Al_xIn_{1-x}As$	$0.36 + 2.35x + 0.24x^2$	$1.8 + 0.4x$	
$Al_xGa_{1-x}Sb$	$0.73 + 1.10x + 0.47x^2$	$1.05 + 0.56x$	
$Al_xIn_{1-x}Sb$	$0.172 + 1.621x + 0.43x^2$		
$Ga_xIn_{1-x}P$	$1.34 + 0.511x + 0.604x^2$ ($0.49 < x < 0.55$)		
$Ga_xIn_{1-x}As$	$0.356 + 0.7x + 0.4x^2$		
$Ga_xIn_{1-x}Sb$	$0.172 + 0.165\ x + 0.413x^2$		
GaP_xAs_{1-x}	$1.424 + 1.172x + 0.186x^2$		
$GaAs_xSb_{1-x}$	$0.73 - 0.5x + 1.2x^2$		
InP_xAs_{1-x}	$0.356 + 0.675x + 0.32x^2$		
$InAs_xSb_{1-x}$	$0.18 - 0.41x + 0.58x^2$		

Table 17.2: Compositional dependence of the gap energy in ternary III-V semiconductors at room temperature.

© E. F. Schubert

17.4 Semiconductor heterostructures

- Lattice matching required for low defect density
- This is particularly important for minority carrier devices
- This is not so important for majority carrier devices
- Ideal: Heterostructures are formed by semiconductors with the same crystal structure and the same lattice constant: An example is $Al_xGa_{1-x}As$ on GaAs
- Often: Mismatched structures result in misfit dislocations defects which act as recombination centers. An example is GaN on sapphire
- Diagrams of energy gap-versus-lattice-constant for of different semiconductors

© E. F. Schubert