Development Team

Principal Investigator	Prof. Subhasis Ghosh, School of Physical Sciences, Jawaharlal Nehru University, New Delhi
Paper Coordinator	
Content Writer	Prof. Subhasis Ghosh, School of Physical Sciences, Jawaharlal Nehru University, New Delhi
Content Reviewer	

Physics	**Physics at Nanoscale – IV** **Semiconductor Heterostructures**

Physics	**Physics at Nanoscale – IV** **Formation of heterostructure**

Description of Module	
Subject Name	Physics
Paper Name	Physics at Nanoscale – IV
Module Name/Title	Formation of heterostructure
Module Id	4.1

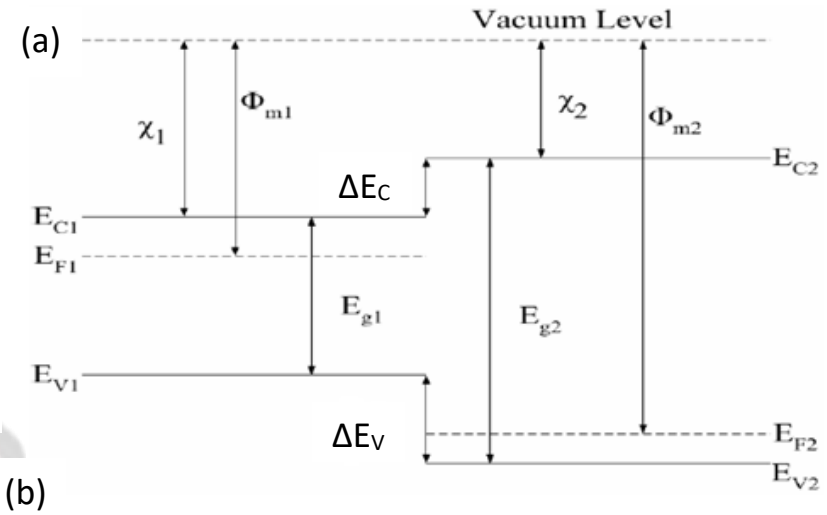
Table of Contents

4.1 Formation of heterostructure

4.1.1 What is semiconductor heterostructure ?

The bulk crystalline semiconductor exhibits perfect or nearly perfect translational symmetry i.e. the atoms within the crystal are regularly spaced throughout the entire bulk sample. However, bulk crystal can include point defects and extended defects, such as dislocation, grain boundaries, so that the perfect periodicity of the material is disrupted locally. But these local disruption in long range ordering does not disrupt global crystalline ordered structure. These defects can significantly affect the properties of the semiconducting material. The long range ordering can be also disrupted artificially in multilayered structures. Present day highly developed crystal growth technology, in particular different epitaxial growth techniques, has made it possible to grow ultrathin layers with atomic layer precision of different semiconductors stacked together, called *heterostructures*. A very thin layer of material can be grown on top of another material or can be sandwiched between layers grown with a different type of semiconductor material, whose lattice constant is either same or different. Since the constituent semiconductors within the heterostructure are of different types, many of their properties can be tuned differently resulting distinct properties which will be quite different than the constituent semiconductors. The most important properties which can be tuned are the material lattice constants, doping concentrations, electron-affinity, optical and electrical properties. The central feature of hetero-structure is that the band gaps of the constituent semiconductors are usually different and there will be discontinuities in both the conduction and valence band at the interface of two semiconductors. Most of the useful properties of hetero-structure depend on these discontinuities. Let us refer to the two semiconducting materials forming the heterostructure, with different energy gaps E_{g1} and E_{g2}, respectively. Hence, in the heterostructure, either both or at least one of the two, the conduction band or the valence band, must be discontinuous at the interface and the energy differences at the interface are called the conduction band and valence band discontinuities, respectively, as shown in Figure 1. Anderson's rule, also known as electron affinity rule is

generally employed for determining the energy band profile of the heterojunction between two semiconductor materials (Figure 1). According to this rule, to determine the energy band diagram, the vacuum levels of the two semiconductors on either side of the heterojunction should be aligned. Once the vacuum levels are aligned it is possible to use the electron affinity which is the energy difference between the lower edge of the conduction band and the vacuum level of the semiconductor and band gap values for each semiconductor to calculate the conduction band and valence band offsets. Each semiconductor has different electron affinity and band gap values. Once the relative positions of the conduction and valence bands for both



semiconductors are known, Anderson's rule allows the calculation of the band offsets of both the valence band (ΔE_v) and the conduction band (ΔE_c).

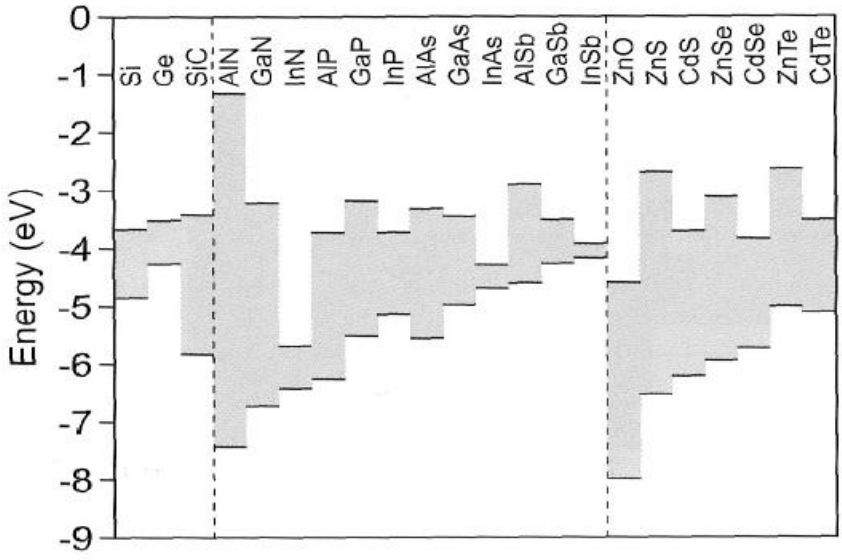


Figure 4.1 (a) *Parameters of semiconductors with different band gap of Eg1 and Eg2, electron affinity of χ_1 and χ_2, work function of Φ_{m1} and Φ_{m2} defining the band discontinuities in heterostructure. ΔE_C and ΔE_V are the band discontinuities at the interface between two semiconductors. (b) Band offsets between different semiconductors according to Anderson's common affinity rule.*

4.1.2 Different types of heterostructures

Band alignment at the semiconductor interfaces can be of three different types: straddling gap (type I), staggered gap (type II) or broken gap (type III), as shown in Figure 4.2.

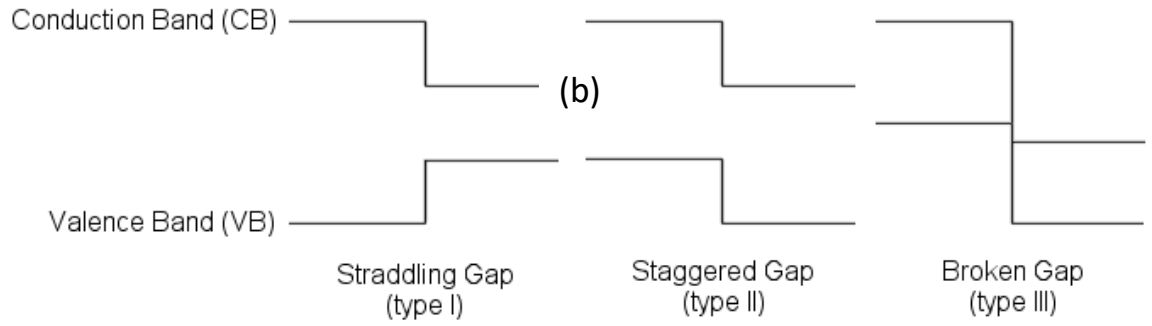


Figure 4.2 *The three types of heterostructures: (a) type I; (b) type II; (c) type III.*

The type I heterostructure is the most common. In type I heterostructure (Figure 4.2a), the sum of the conduction band and valence band edge discontinuities is equal to the energy gap difference of GaAs and AlGaAs and given by,

$$\Delta E_g = \Delta E_c + \Delta E_v \qquad (4.1)$$

The type II heterostructure (Figure 4.2b) is arranged such that the discontinuities at conduction and valence band have different signs. The bandgap discontinuity in this case is given as the difference between the conduction band and valence band edge discontinuities

and the most common type II heterostructure is the $Al_{1-x}In_xAs$-InP based heterostructure. In type III heterostructures (Figure 4.2c), the band alignment is such that the top of the conduction band in one semiconductor lies below the valence band maxima of the other semiconductor and the most common type III heterostructure is the GaSb-InAs based systems. Out of three types, type I heterostructure is most common based on GaAs- $Al_{1-x}Ga_xAs$ based systems. While the band gap of one component (GaAs) is fixed (1.4eV), but band gap in $Al_xGa_{1-x}As$ can be varied by the Al composition following the relation.

$$E_g = 1.424 + 1.247x \quad 0 < x < 0.45 \qquad (4.2)$$

And the close lattice matching between GaAs and $Al_{1-x}Ga_xAs$ makes this system most suitable for different devices in which strain free interface is required, such as high electron mobility transistor (HEMT). However, strain induced lattice mismatch can be useful for many applications in devices.

The key to the understanding the behavior of the heterostructure is the energy-band profile which provides position dependence of energy of the conduction and valence band. The position-dependent band-edge energies can be obtained from the Schroindger equation

$$ih\frac{\partial\psi}{\partial t} = -\frac{h^2}{2m^*}\frac{\partial^2\psi}{\partial x^2} + \left[E_n - U_{V,C}(x)\right] \qquad (4.3)$$

where ψ is the envelope function, m^* is the effective mass, E_n is the energy at the edge of the nth band, and $U_{V,C}(x)$ is the potential at the interface of heterojunction. The band structure is incorporated in the material-dependent parameters En and m^*. $U_{V,C}(x)$ can be represented as

$$U_{V,C}(x) = E_{V,C}(x) - qV(x) \qquad (4.4)$$

$U_C(x)$ and $U_V(x)$ are the effective potential for the conduction and valence bands, respectively and $E_C(x)$ and $E_V(x)$ are position dependent conduction and valence band edges, respectively. In a heterostructure the dependence of U_C and U_V on x are due to the combined effects of the electrostatic potential V(x) and the energy-band discontinuities or shifts due to the heterostructure. As a zeorth approximation, it can be assumed that heterostructure is charge neutral, so that V can be assumed as constant and can be neglected. In such cases, the behaviour of E_C and E_V at the interface can be determined. Though, there are several

theoretical proposals of universal band alignments(one such proposal is known as virtual crystal approximation, discussed later in this chapter), but it has been found that there is no a priori relation between the band-edge energies of the two semiconductors forming a heterojunction,. The quantities used to describe the band alignment are defined in Figure 4.1. The one quantity which is known with great certainty is the total band gap discontinuity,

$$\Delta E_G = E_G{}^{(B)} - E_G{}^{(A)} \qquad (4.5)$$

Where $E_G^{(A)}$ and $E_G^{(B)}$ are the energy gaps of materials A and B, respectively. The total discontinuity ΔE_G is distributed between the valence and conduction band discontinuities defined as

$$\Delta E_V{}^{(AB)} = E_V{}^{(A)} - E_V{}^{(B)} \qquad (4.6a)$$

$$\Delta E_C{}^{(AB)} = E_C{}^{(A)} - E_C{}^{(B)} \qquad (4.6b)$$

And the individual discontinuities add up to result the total discontinuity,

$$\Delta E_G = \Delta E_V + \Delta E_C \qquad (4.7)$$

4.1.3 III-V Semiconductor and their alloy for heterostructures

How the discontinuities are distributed between the valence and conduction bands is the major question to be answered by theory and experiment.

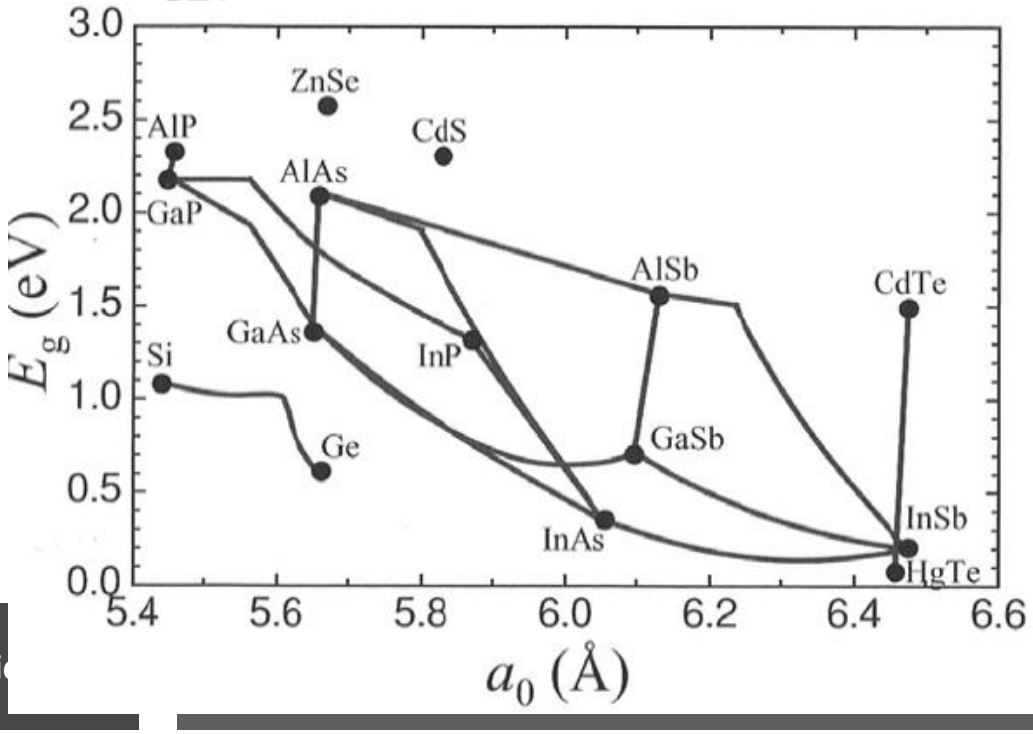
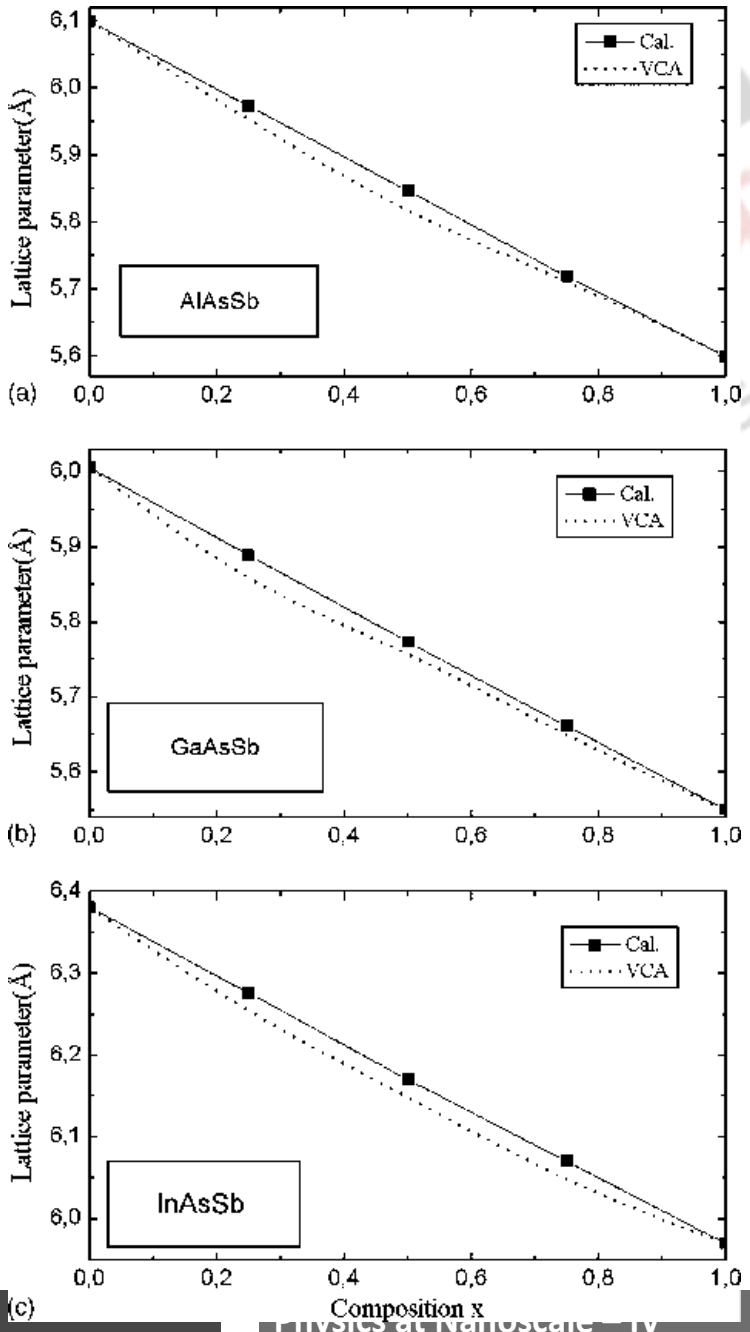


Figure 4.3 *Lattice constant versus bandgap for common cubic III-V semiconductor alloys and Si-based alloys*

Semiconductor heterostructures are generally fabricated from materials that can be grown upon each other epitaxially with relatively low defect densities at interface. The most preferred heterojunction systems are comprised of semiconductor materials that are relatively closely lattice matched. Figure 4.3 shows bandgaps versus lattice constants of common cubic III–V and Si-based materials. The GaAs is lattice matched with AlGaAs is the first heterojunction systems to be developed and till today remains to be most popular heterojunction for basic and applied physics applications. The alloying of two or more semiconductor materials, for instance Si and Ge or GaAs and AlAs, is a powerful tool in order to control the band structure. Let us consider an alloy consisting of two components: A with a fraction x, and B, with a fraction 1-x. If A and B have similar crystalline lattices, one can expect that the alloy A_xB_{1-x} has a lattice constant a_c given by $a_c = a_A\, x + a_B\,(1-x)$ which is known as Vegard's law, also known as *Virtual-crystal approximation*(VCA), according to which certain parameters of the alloy, such as band gap, can be characterized as a function of the



fraction *x*. Figure 4.4 shows the variation of lattice parameters in three III-V alloy semiconductor systems. Prediction according to VCA has been compared with experimental values. It is interesting to note that the experimental data can be explained adequately by VCA.

Figure 4.4 *Lattice constant of AlAsSb (a), GaAsSb (b) and InAsSb (c) alloys vs. composition (solid square) compared with the virtual crystal approximation (dot line).*

4.2 Modulation doping

4.2.1. What is modulation doping ?

Modulation doping introduces remote doping concept in semiconductor and offers an important advantage in band gap engineering. By this method, the free carrier concentration within a semiconductor layer can be increased significantly without the introduction of dopant impurities. Generally conventional doping in semiconductors is employed to increase free carrier concentration and conductivity, however doping is also responsible for enhanced ionized impurity scattering which reduces the carrier mobility. This problem can be overcome using modulation doping by which the free carriers can be spatially separated from the dopants. The spatial separation of the dopants and free carriers reduces the deleterious effect of ionized impurity scattering. Hence, the free carrier concentration and conductivity can be increased significantly without reducing the carrier mobility. In modulation doping, the semiconductor with higher band gap is generally doped. In case of GaAs-AlGaAs system, AlGaAs layer is intentionally doped n-type and GaAs layer is unintentionally doped. In equilibrium the Fermi level must align throughout the heterostructure structure. As it is explained in Figure 4.5, when the two semiconductors AlGaAs and GaAs are initially apart, the Fermi level lies closer to the conduction band edge in the AlGaAs than in the GaAs since the AlGaAs is doped n-type. When the heterostructure is formed, as shown in Figure 4.5, Fermi level can only be aligned if electrons are transferred from the AlGaAs layer into the GaAs. This results in increase in the electron concentration of GaAs without doping! As the electrons leave AlGaAs, the ionized donor atoms within the AlGaAs result in a net positive charge, which are balanced by the net negative charge due to the electrons transferred in the GaAs layer from AlGaAs layer. As the transferred electrons in GaAs are spatially separated from ionized donor atoms in the AlGaAs, effect of Coulomb interaction reduces between them resulting the suppression of ionized impurity scattering of the transferred electrons. This leads to higher electron mobility. If an un-doped AlGaAs spacer layer is introduced between the doped AlGaAs and un-doped GaAs layers, to increase the spatial separation of the

electrons from the ionized donors, further reduction of the ionized impurity scattering is possible resulting very high mobility (>$10^7 cm^2/Vsec$) in GaAs-AlGaAs system.

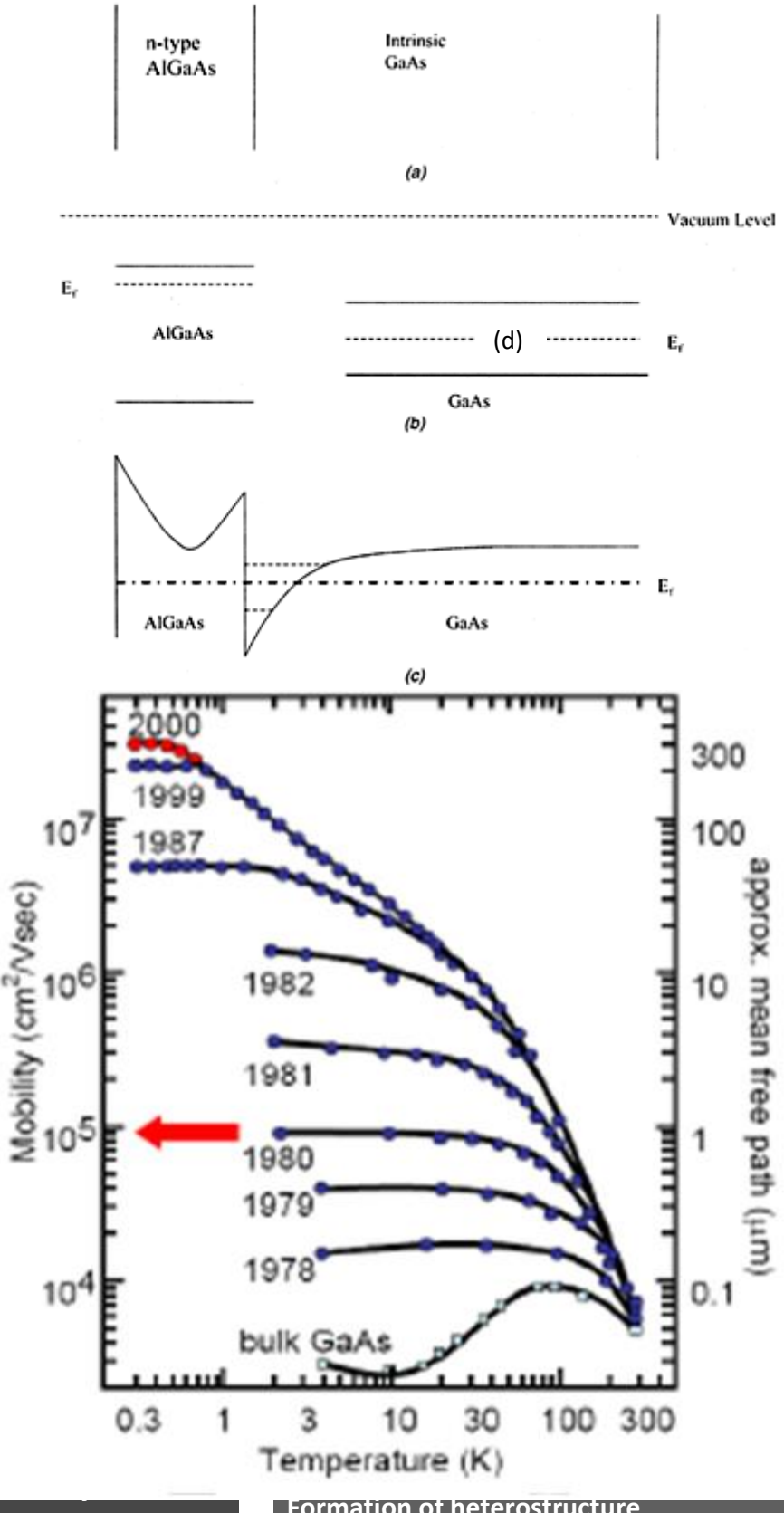


Figure 4.5 (a) *GaAs-AlGaAs modulation-doped heterostructure; (b) energy band diagrams of the AlGaAs and GaAs layers when they are apart; (c) once the heterostructure is formed between GaAs and AlGaAs, resulting energy band diagram in equilibrium. The two dashed horizontal lines at the heterointerface in the GaAs layer represent energy subbands arising from spatial quantization effects in triangular quantum well. (d) Temperature dependence of electron mobility in GaAs-AlGaAs heterostructure. How mobility has been increased in GaAs-GaAlAs heterostructure over the years. It is also clear that absence of scattering in modulation doping lead to dramatic increase in electron mean free path.*

4.2.2 Subbands in single heterotructure

As seen in Figure 4.5, the conduction band edge in the GaAs layer is strongly bent near the hetero interface due to transfer of electrons from AlGaAs to GaAs. The sharp bending of the conduction band edge and the presence of the conduction band edge discontinuity forms a triangular potential well within the GaAs layer. It can be shown that the band bending is sufficiently strong so that the spatial dimensions of the potential well are comparable to the de Broglie wavelength of electrons. As discussed before, spatial quantization effects occur resulting discrete energy bands, called *sub bands*.

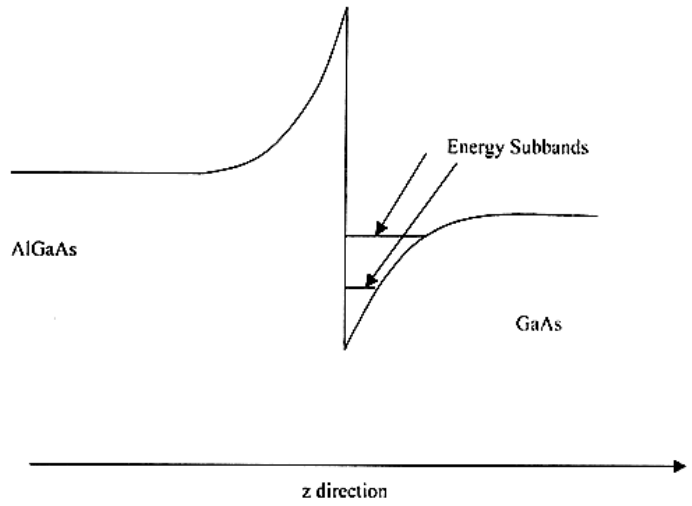


Figure 4.6 *Band profile of heterostructure based on n-doped AlGaAs and undoped GaAs. Electrons in triangular quantum well showing two energy sub bands.*

Figure 4.6 shows the conduction band profile of band edges at the interface of n-type AlGaAs and undoped GaAs. The direction of quantization is perpendicular to the hetero interface, i.e. along the z direction. However, in the directions parallel to the interface i.e. x and y direction, the energy of electrons is not quantized, since the motion of the electrons is not restricted by the band bending. Essentially the electrons behave as free particles in the x-y plane. Using a parabolic band model approximation, the electron energies within the potential well in the conduction band can be given by

$$E = \frac{\hbar^2 k_x^2}{2m} + \frac{\hbar^2 k_y^2}{2m} + E_n \qquad (4.8)$$

Where k_x and k_y represent the x and y coordinates of the electron k-vector, respectively and E_n represents the energies due to the spatial quantization in the z direction. The corresponding wave function for an electron in the potential well is

$$\psi(r,z) = \chi(z)e^{ikr} \qquad (4.9)$$

Where k is the two-dimensional wave vector, r the two-dimensional spatial vector consisting of the x and y coordinates, and $\chi(z)$ the sub band wave-function.

The general solution for the eigenenergies of the finite potential well formed in the conduction band requires self-consistent solution of the Schrodinger and Poisson equations, which are discussed briefly in the next section. The eigen energies can be roughly approximated using the solution for an infinite triangular potential well as

$$E_i = \left(\frac{1^2}{2m}\right)^{1/3}\left(\frac{3}{2}\pi qF\right)^{2/3}\left(i+\frac{3}{4}\right)^{2/3} \qquad (4.10)$$

Where F is the electric field strength corresponding to the slope of the energy band and i is an integer representing the band index. Using Eqs. 4.10 and 4.8, the energy of an electron in the conduction band well can be roughly approximated.

4.2.3 Band Profile and Carrier Distribution in Heterostructure

For any application of hetero structure, it is required to understand the energy-band profile as function of position, i.e. position dependent band-edge energies U_V and U_C which will incl ude $E_{V,C}(x)$ and the electrostatic potential V(x). The charge distribution at the interface ρ(x) is responsible for electrostatic potential V(x). Carrier densities and Fermi level across the hetero structure have to be calculated self consistently by solving Schrodinger's equation and Poisson's equation. Assuming parabolic bands quasi-equilibrium electron and hole densities, n(x) and p(x), respectively, can be given by

$$p(x) = N_V(x)\mathcal{F}_{1/2}\left\{[E_V(x) - qV(x) - E_F(x)]/_{kT}\right\} \qquad (4.11a)$$

$$n(x) = N_C(x)\mathcal{F}_{1/2}\left\{[E_F(x) + qV(x) - E_C(x)]/_{kT}\right\} \qquad (4.11b)$$

Where $\mathcal{F}_{1/2}$ is the Fermi-Dirac integral of order $\frac{1}{2}$,

$$\mathcal{F}_{1/2}(\eta) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{\xi^{3/2}\, d\xi}{1+e^{\xi-\eta}} \qquad (4.12)$$

and $N_c(x)$ and $N_v(x)$ are the effective densities of stats in the conduction and valence bands, respectively and given by

$$N_C(x) = 2\left[\frac{2\pi m_C^*(x)kT}{h^2}\right]^{3/2} \qquad (4.13a)$$

$$N_V(x) = 2\left[\frac{2\pi m_V^*(x)kT}{h^2}\right]^{3/2} \qquad (4.13b)$$

The net charge density is contributed by the mobile carrier densities n(x) and p(x), and the ionized donor and acceptor impurity densities N_D^+ and N_A^- , respectively. N_D^+ and N_A^-, will also depend on effective potential at heterointerface and can be given by

$$N_D^+(\mathrm{x}) = \frac{\mathrm{N_D}}{1+\mathrm{g_D}\exp\left\{[E_F(x)+qV(x)-E_D(x)]/kT\right\}} \quad (4.14a)$$

$$N_A^-(\mathrm{x}) = \frac{\mathrm{N_A}}{1+\mathrm{g_A}\exp\left\{[E_A(x)-qV(x)-E_F(x)]/kT\right\}} \quad (4.14b)$$

Where g_D *and* g_A are the degeneracy factors of the donors and acceptors, respectively, and the impurity state energies E_D and E_A are defined with respect to the same energy scale as $E_{V,C}$. The total charge density is then

$$\rho(x) = q\left[p(x) - n(x) + N_D^+(\mathrm{x}) + N_A^-(\mathrm{x})\right] \quad (4.15)$$

With the aid of appropriate boundary conditions, V(x) can be obtained by solving following Poisson's equation

$$\frac{d}{dx}\varepsilon(x)\frac{dV}{dx} = \rho(x) \qquad (4.16)$$

Where ε is the dielectric constant of the participating semiconductors for which ε may be quite different and hence, depends on position. The screening equation for a hetero structure is obtained by combining all of the equations in this section into (4.16) which is a nonlinear differential equation for V(x), as the material parameters are fixed by the design of the hetero structure. The boundary conditions to be applied to this screening equation follows from the condition that each semiconductor material must be charge neutral far from the hetero junction. Let the boundary points be x_l and x_r . These can be taken to be $\pm\infty$ if one is solving

for the potential analytically, but if numerical techniques are used x_l and x_r should be finite but deep enough into the bulk semiconductor that charge neutrally may be assumed. One then determines $V(x_l)$ and $V(x_r)$ simply by solving

$$\rho(x_l) = 0 \qquad (4.17a)$$

$$\rho(x_r) = 0 \qquad (4.17b)$$

It is assumed that the Fermi energy which is different in different regions of the heterostructure is set by the bias applied across the hetero structure and the entire energy band structure, floats up or down until charge neutrally in the bulk is achieved. Thus the origin of the scale of V is set by the combined choice of the energy scale for the band structure energies E_V and E_C. The Fermi energies on each side of the junction E_F^+ are determined by the externally applied voltages at the respectively contacts. It is most convenient to define the Fermi energy so that

$$E_F{}^+ = -qV_a \qquad (4.18)$$

Where V_a is the applied voltage. If the carrier densities are neither degenerate nor closely compensated, the Fermi Function in (4.17a and 4.17b) may be approximated by exponentials and one may directly solve for $V(x_{r,r})$ to obtain the more familiar expressions:

$$V\left(x_{l,r}\right) = \begin{cases} \left\{E_C(x_{l,r}) - E_{Fl,r} + kTln\left[N_D(x_{l,r})/N_C(x_{l,r})\right]\right\}/q & N-\text{type} \\ \left\{E_V(x_{l,r}) - E_{Fl,r} - kTln\left[N_A(x_{l,r})/N_V(x_{l,r})\right]\right\}/q & P-\text{type} \end{cases} \qquad (4.19)$$

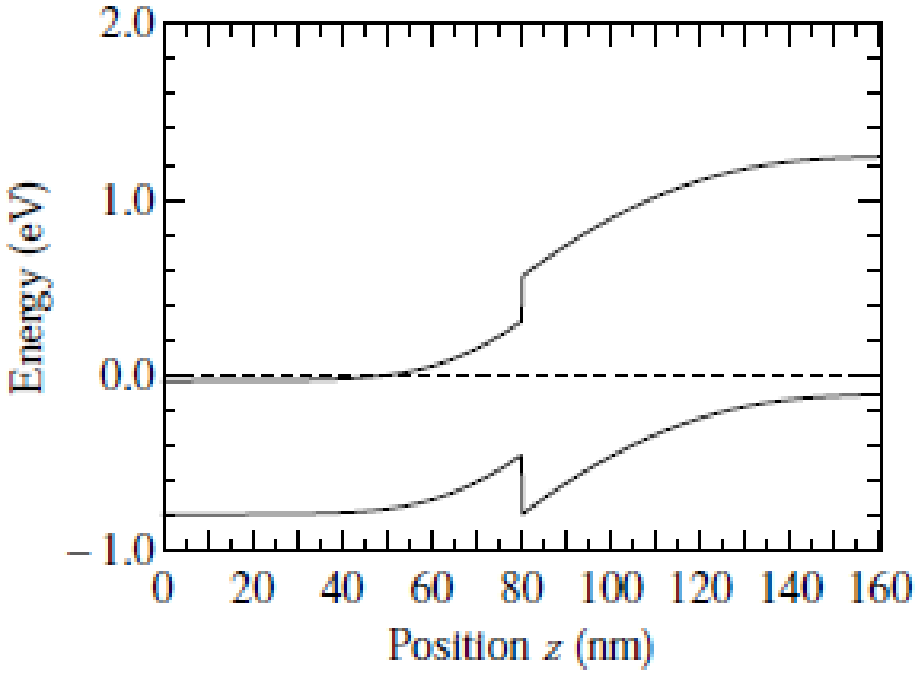


Figure 4.7 *Self-consistent band profile of an anisotype (n-p) type-I heterojunction in equilibrium, based on InGaAs-InP.*

In case of hetero structure, the diffusion voltage is the difference across the hetero junction $V_d = |V(x_r) - V(x_l)|$. The screening equation consisting of Poisson's equation (4.16) combined with the charge density expression (4.15) and subject to the boundary values obtained by solving (4.17) is a non-linear differential equation for the electrostatic potential V(x). It has to solved numerically for each specific case. However, an approximate solution to the problem is to make a finite difference approximation to the equation, reducing it to a set of simultaneous nonlinear algebraic equations, and solve these using Newton's method. In an isotype hetero junction, both sides of the junction doped with same conductivity type (n-n or p-p). In case of opposite conductivity types (p-n or n-p), it is an an-isotype junction. Figure 4.7 shows the band profile of an an-isotype type-I junction in equilibrium and looks similar to p-n homo junction. Figure 8 shows the iso-type type-I junction which resembles that of a Schottky junction.

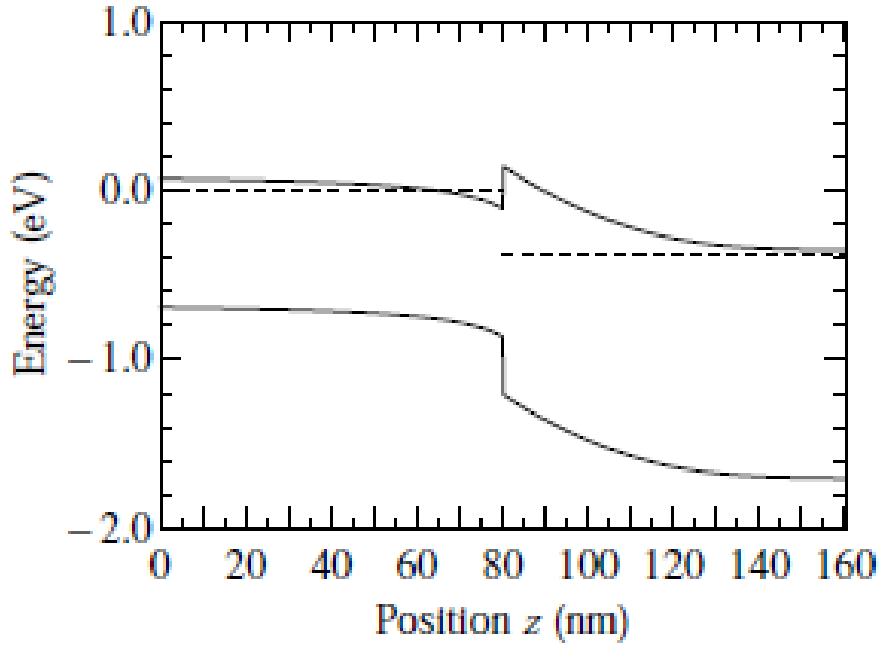


Figure 4.8 *Self-consistent band profile of an isotype (n-n) type-I hetero-junction in equilibrium, based on InGaAs-InP.*