Notes on

Heterostructure Fundamentals

EE-650Y

Fall 1995

Mark Lundstrom
School of Electrical and Computer Engineering
and the NSF MRSEC for
Technology-Enabling Heterostructures
Purdue University
West Lafayette, Indiana 47907

1. Introduction

The first step in thinking about how a semiconductor device operates usually consists of drawing its energy band diagram. If you can draw the energy band diagram, you can probably understand how the device will operate. For devices with uniform material composition you should know that:

(1) the electrostatic potential, $V(x)$, can be obtained by turning $E_C(x)$, $E_V(x)$, or $E_I(x)$ upside down (because $E_C(x)$ = constant - $qV(x)$ etc.)

(2) the electric field is simply 1/q times the slope of E_C, E_V, or $E_I(x)$.

(3) the charge density, $\rho(x)$ is $q/\kappa_s\varepsilon_o$ times the second derivative of E_C, E_V, or $E_I(x)$.

(See, for example, R. F. Pierret, *Semiconductor Fundamentals,* Vol. I of the Modular Series on Solid State Devices, Addison-Wesley, 1983.) For compositionally nonuniform semiconductors (so-called heterostructures), NONE of the above statements is true in general. The purpose of these notes is to explain the basic concepts needed to draw energy band diagrams for heterostructure devices.

Before we turn to heterojunctions, let's begin by re-examining the energy band diagram for a uniformly doped, compositionally uniform semiconductor as shown below.

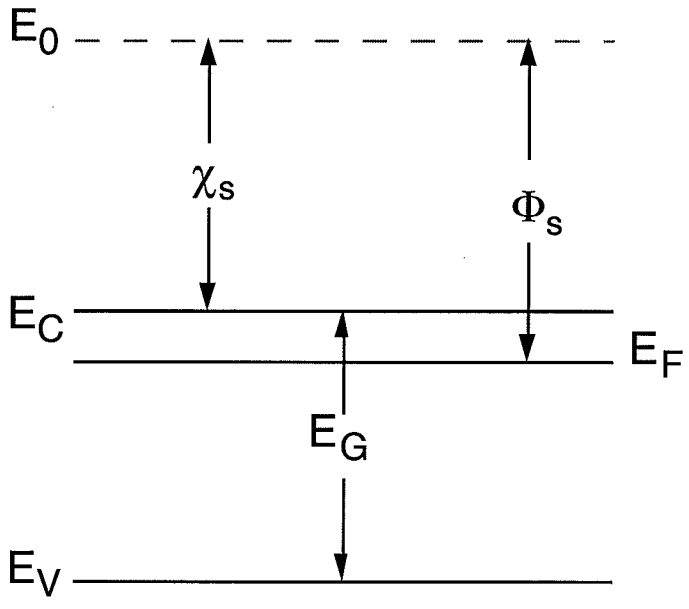


Fig. 1. Energy band diagram for a uniform semiconductor. The location of the conduction band is measured with respect to a reference energy level, E_0, the field-free vacuum level.

The positions of E_C and E_V are determined by the chemical bonding of the atoms and can be measured or calculated by solving Schroedinger's equation. These energy levels must be measured relative to some reference level. The reference I have chosen is E_0, the energy of a free electron just outside the neutral semiconductor (i.e. the vacuum level). The electron affinity, χ_s is the energy needed to remove from the semiconductor of an electron located at E_C and make it a free electron. The work function Φ_s is the distance between E_0 and the Fermi level, E_F. For this case we have

$$E_C = E_0 - \chi_S \tag{1}$$

$$E_V = E_C - E_G = E_0 - \chi_S - E_G \ . \tag{2}$$

For homostructures, the electron affinity and bandgap are position-independent, and there is no need to worry about the reference level, but for heterostructures, a reference level is essential. We have selected the field-free vacuum level as the reference, but the choice of a reference level is a subtle point that we shall return to in Sec. 9.

Now let's consider a slightly more complicated case - a semiconductor with an electric field. Note that electric fields are responsible for the chemical bonding of the atoms, but we have already accounted for these short range forces by solving Schroedinger's equation to find E_C and E_V. The field we are now considering is an additional, slowly varying field as produced, for example, by an MOS gate or by the nonuniform doping of a PN junction.

We know that the electric field produces a force on electrons

$$F_e = -qE(x) \tag{3}$$

and an equal but opposite force on holes

$$F_h = +qE(x). \tag{4}$$

The electron changes energy as it moves in the E-field by

$$\Delta E = -\int F_e dx = -qV(x) \ . \tag{5}$$

The result is that (1) and (2) must be modified to

$$E_C(x) = E_0 - \chi_S - qV(x) \tag{6}$$

and

$$E_V(x) = E_0 - \chi_S - E_G - qV(x) \tag{7}$$

and our energy band diagram will acquire a slope.

Figure 2 shows a familiar example. In this case, we begin with a charge neutral n-type semiconductor and another neutral p-type semiconductor. Before we place them in contact, their Fermi levels are separated by

$$E_{FN} - E_{FP} = E_G - \delta_N - \delta_P \ , \tag{8}$$

where

$$\delta_N = \left(E_C - E_{FN}\right) = kT \log(N_C / n_o) \tag{9}$$

and

$$\delta_P = \left(E_{FP} - E_V\right) = kT \log(N_V / p_o) \ . \tag{10}$$

Before they are in contact, there are separate Fermi levels for the n- and p-type semiconductors, but after they are in contact, the energy bands adjust to align the Fermi levels. The only way to move the bands is for an electrostatic potential, the *contact potential*, to develop. The contact, or *built-in potential* is determined by the separation in the Fermi levels of the separate semiconductors,

$$qV_{bi} = E_{FN} - E_{FP} = kT \log\left(\frac{N_A N_D}{n_i^2}\right) \ . \tag{11}$$

The last expression is ONLY VALID FOR HOMOJUNCTIONS, but the built-in potential is always given by the difference in the Fermi levels of the isolated semiconductors. If we reference these Fermi levels to the field-free vacuum level, then we obtain the general expression for the built-in potential as

$$qV_{bi} = \Phi_P - \Phi_N \tag{12}$$

where

$$\Phi_P = \chi_P + E_{GP} - \delta_P \tag{13a}$$

$$\Phi_N = \chi_N + \delta_N \tag{13b}$$

are the workfunctions of the p- and n-type semiconductors.

From Eqs. (12) and (13), we can express the built-in potential for a heterojunction as

$$qV_{bi} = (\chi_P - \chi_N) + E_{GP} - \delta_P - \delta_N, \tag{14a}$$

which can be re-expressed in several ways. For example, we can write

$$qV_{bi} = kT\log\left[\frac{N_A N_D}{n_{iN} n_{iP}}\right] + (\chi_P - \chi_N) + \frac{(E_{GP} - E_{GN})}{2} - \frac{kT}{2}\log\left[\frac{N_{VP} N_{CN}}{N_{CP} N_{VN}}\right], \tag{14b}$$

which generalizes (11) to heterojunctions. Note also that Eq. (12) applies to any heterojunction, not just to pn diodes if we replace the subscripts P and N by 1 and 2 to denote the two materials.

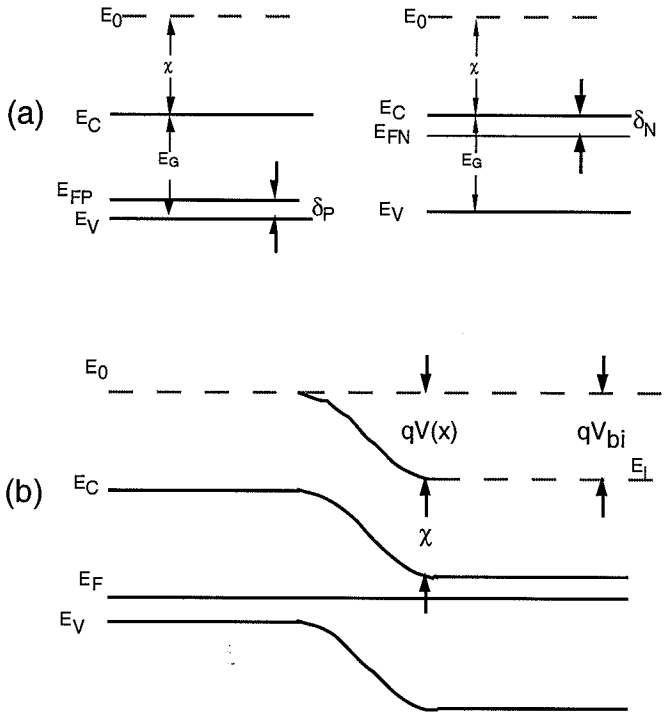


Fig. 2. Illustration of the formation of a pn junction. (a) the isolated semiconductors are characterized by separate Fermi levels, and (b) after placing them in contact, electrons flow from the region with the higher Fermi level to the region with the lower Fermi level. As a result of this flow, a space-charge layer builds up and a built-in potential results. In equilibrium, the final structure is characterized by a single Fermi level. Note that E_0 is lowered to $E_l(x)$, the local vacuum level.

To summarize, Eqs. (6) and (7) show that the energy of an electron at E_C (or E_V) is comprised of two different kinds of contributions: 1) chemical bonding as determined by χ_s (or $\chi_s + E_G$) and 2) $-qV(x)$ which results from the macroscopic electric field. It is easy to see that conditions (1), (2), and (3) on page 2 are satisfied by (6) and (7). So far, however, we have assumed χ_s and E_G to be constant. They won't be constant in a heterostructure, and Eqs. (1) - (3) will lose validity.

2. Energy Bands in Abrupt Heterojunctions

To draw energy band diagrams for devices with a position-dependent alloy composition (so-called *heterostructure devices*), it is essential to know more than simply how the band gap varies with position - we must also know how the bands line up at compositional junctions. Figure 3 shows the observed band alignments for $Al_{0.3}Ga_{0.7}As$/GaAs. For $Al_xGa_{1-x}As$, the offset in valence bands is observed to be about 40% of the difference in band gaps. For this material pair, the conduction and valence bands of the smaller bandgap semiconductor lie completely within the bandgap of the wider bandgap semiconductor. Heterojunction pairs which line up as illustrated in Fig. 3 are known as *type I heterojunctions*. Heterojunction pairs of III-V compounds in which either the group III or the group IV element differs, form type I heterojunctions. Examples include AlAs/GaAs and GaP/GaAs heterojunctions.

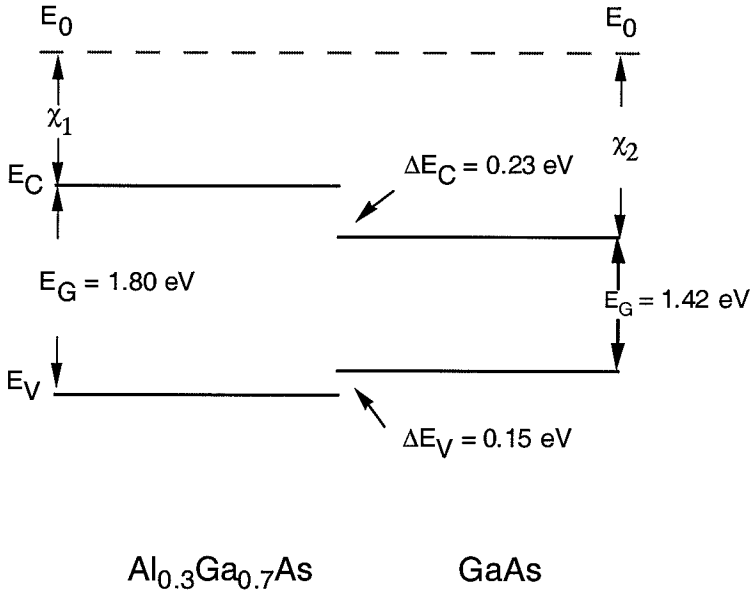


Fig. 3 Observed band alignments for $Al_xGa_{1-x}As$ for x = 0.3 and x = 0.0. Heterojunction pairs that line up like this are known as type I heterojunctions.

From Fig. 3, we observe that

$$\Delta E_C = \chi_2 - \chi_1 \tag{15}$$

where the subscript 2 refers to GaAs and subscript 1 to $Al_{0.3}Ga_{0.7}As$. Equation (15) is known as the *Electron Affinity Rule*; we shall have more to say about it in Sec. 9. For these kinds of heterojunctions, it is also apparent that

$$\Delta E_G = \Delta E_C + \Delta E_V \ . \tag{16}$$

Figure 4 shows another possibility, known as a *type III heterojunction* (some authors, however, refer to this as a type II misaligned heterojunction). III-V heterojunction pairs in which both the group III and group V elements differ (e.g. GaSb/InAs) form type III heterojunctions. In this case, the conduction band of one semiconductor lies below the valence band of the other. Transport is complicated by the fact that the electron wavefunction changes from electron-like to hole-like as the electrons moves across the heterojunction. If the electron affinity difference is less, so that the conduction and valence bands of the smaller bandgap semiconductor straddle the valence band of the larger bandgap semiconductor, the interface is known as *type II (*some authors refer to this as a type II staggered heterojunction). Examples of type II heterojunctions include the $In_xGa_{1-x}As$ / $Ga_xSb_{1-x}As$ and $Al_xIn_{1-x}As$ / InP systems.

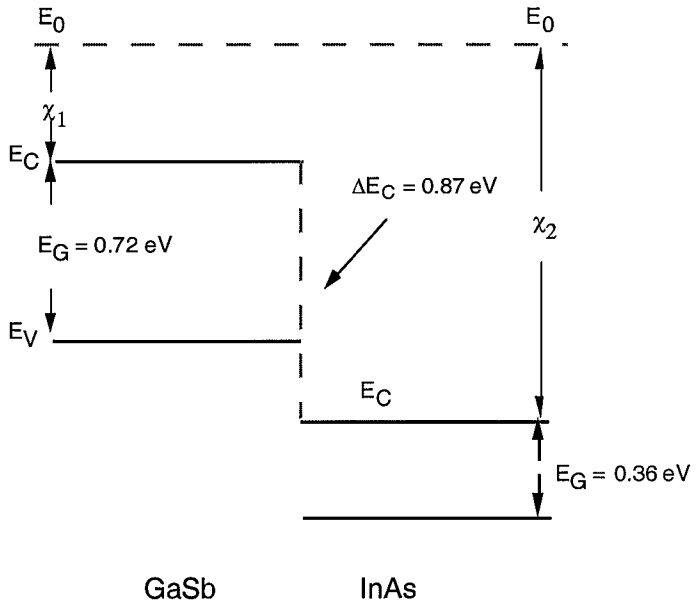


Fig. 4 Observed band alignments for GaSb and InAs. Heterojunction pairs that line up like this are known as Type III heterojunctions, or as Type II misaligned heterojunctions.

The band diagrams of Figs. 3 and 4 ignore electrostatic potentials due to re-arrangement of mobile carriers which occur near the compositional junction after the semiconductors are placed in contact. Energy band diagrams for N-p, P-n, and N-n $Al_{0.3}$ $Ga_{0.7}As$: GaAs heterojunctions are shown in Figs. 5 - 7. (By convention, the capital letter identifies the semiconductor with the larger bandgap.) The band diagram for the heterojunction is deduced conceptually just as it was for homojunctions, by bringing into contact two isolated, bulk semiconductors. When placed in contact, electrons move from the semiconductor with the higher Fermi-level to the other, and an electric field is produced to balance this transfer. The built-in potential is simply the difference in workfunctions of the isolated, bulk semiconductors as expressed by Eqs. (12) or (14).

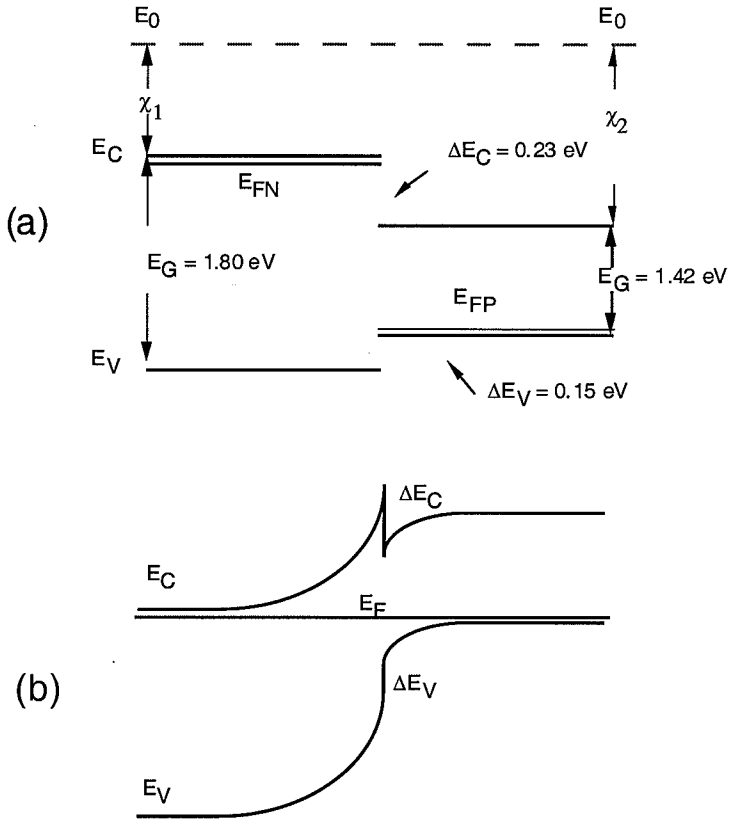


Fig. 5. $Al_{0.3}Ga_{0.7}As$: GaAs/ N-p heterojunction. (a) before contact. (b) after contact.

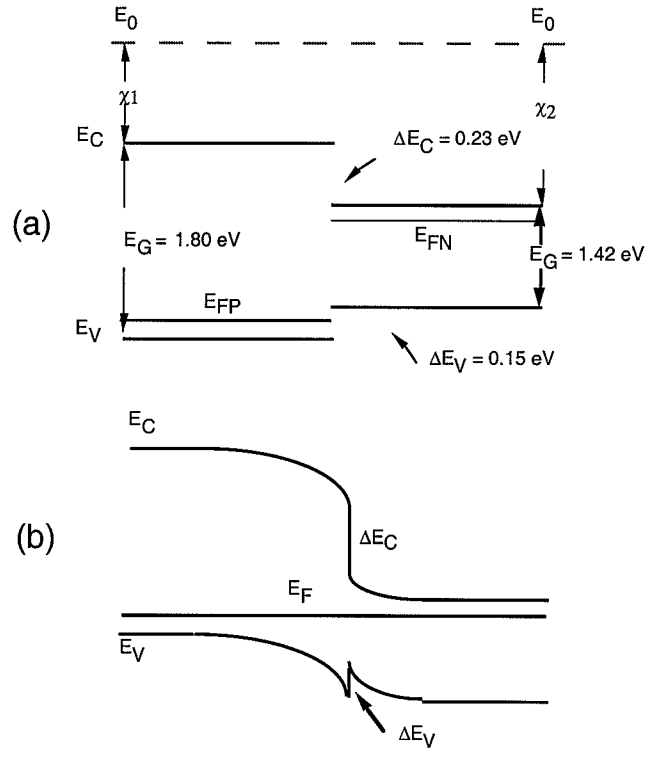


Fig. 6. $Al_{0.3}Ga_{0.7}As$: GaAs/ P-n heterojunction. (a) before contact. (b) after contact.

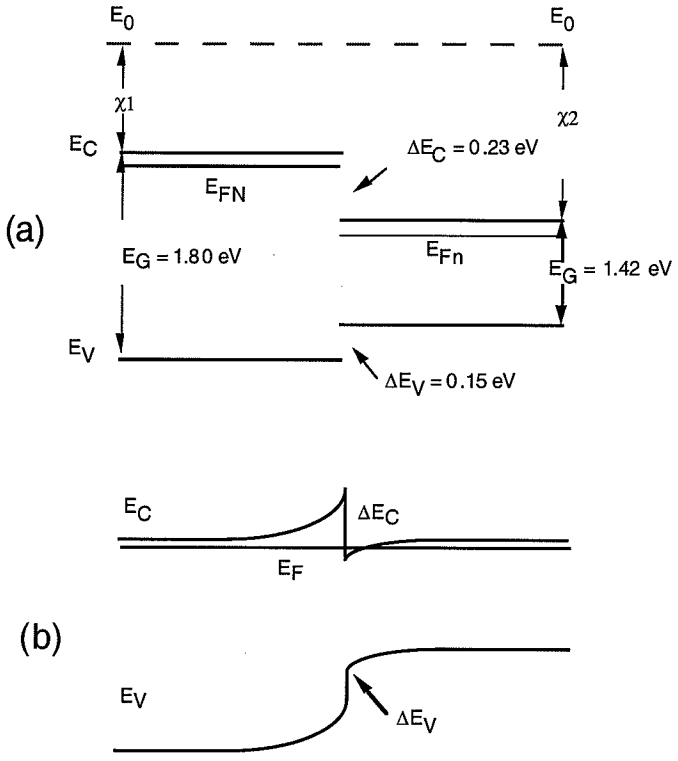


Fig. 7. $Al_{0.3}Ga_{0.7}As$: GaAs/ N-n heterojunction. (a) before contact. (b) after contact.

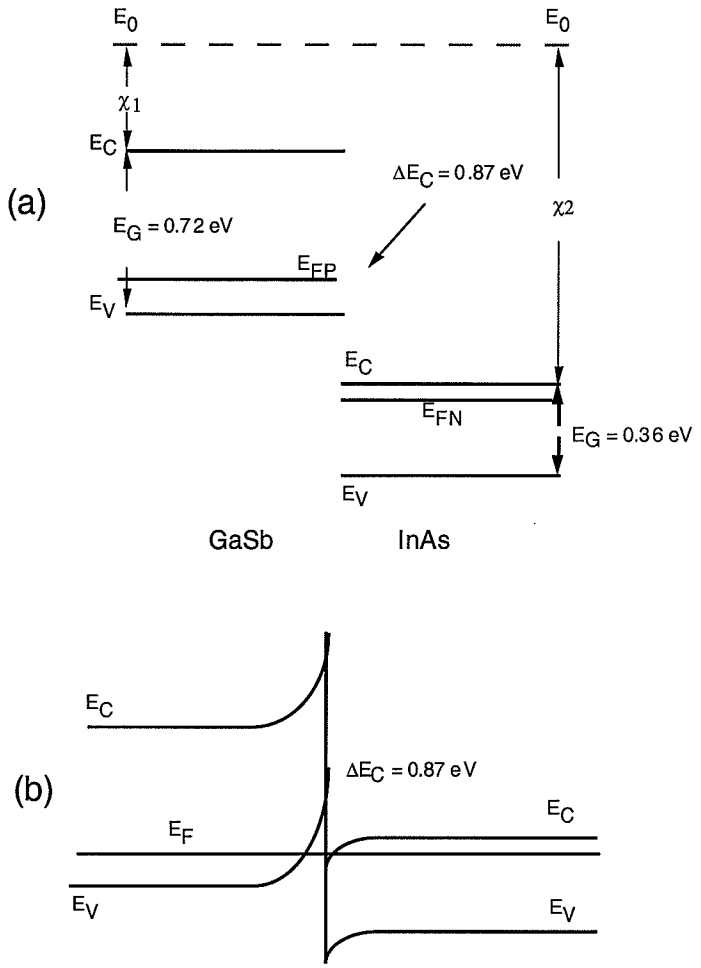


Fig. 8. GaSb : InAs/ P-n heterojunction. (a) before contact. (b) after contact.

The examples in Figs. 5, 6, and 7 were for type I heterojunctions. Figure 8 illustrates an example for a type III heterojunction. The same principles apply, electrons flow from the semiconductor with the higher Fermi level to the one with the lower Fermi level until a potential barrier to stop the flow is established. For the example shown in Fig. 8, we have an accumulation of electrons on the n-InAs side and an accumulation holes on the P-GaSb side. Because the conduction band on the InAs side of the interface lies lower than the valence band on the GaSb side, this interface is sometimes referred to as "semimetalic."

3. The Depletion Approximation

One of the first things taught in an introductory semiconductor course is the depletion approximation. The depletion approximation can also be applied to pn heterojunctions and often provides an adequate estimate for the electrostatic potential. Assume a pn heterojunction with the geometry illustrate in Fig. 9. (As far as the electrostatics are concerned, it doesn't really matter whether the n-type or p-type side has the larger bandgap.)

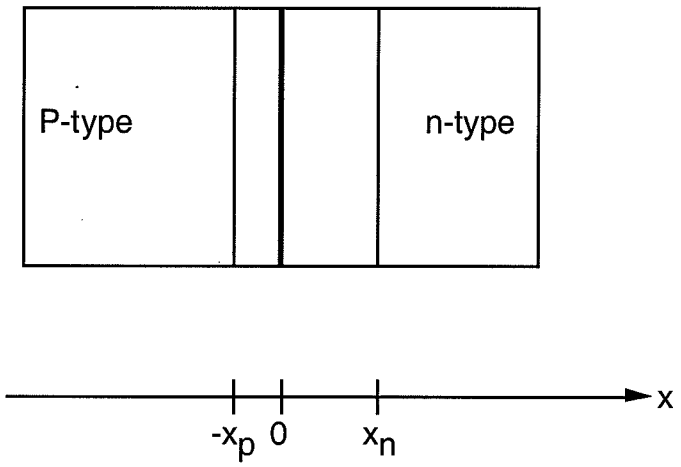


Fig. 9 Geometry of the example pn heterojunction.

You should be able to sketch the charge density, ρ, the electric field E, and the potential V versus position for this diode using the same ideas employed for homojunctions. The results are displayed in Fig. 10. Our goal now is to calculate these quantities using the depletion approximation. We will simply extend the analysis for homojunctions to heterojunctions. (A good, basic reference for homojunctions is *The PN Junction Diode* by G. W. Neudeck (Vol. III of the Modular Series on Solid State Devices, Addison-Wesley, 1983).

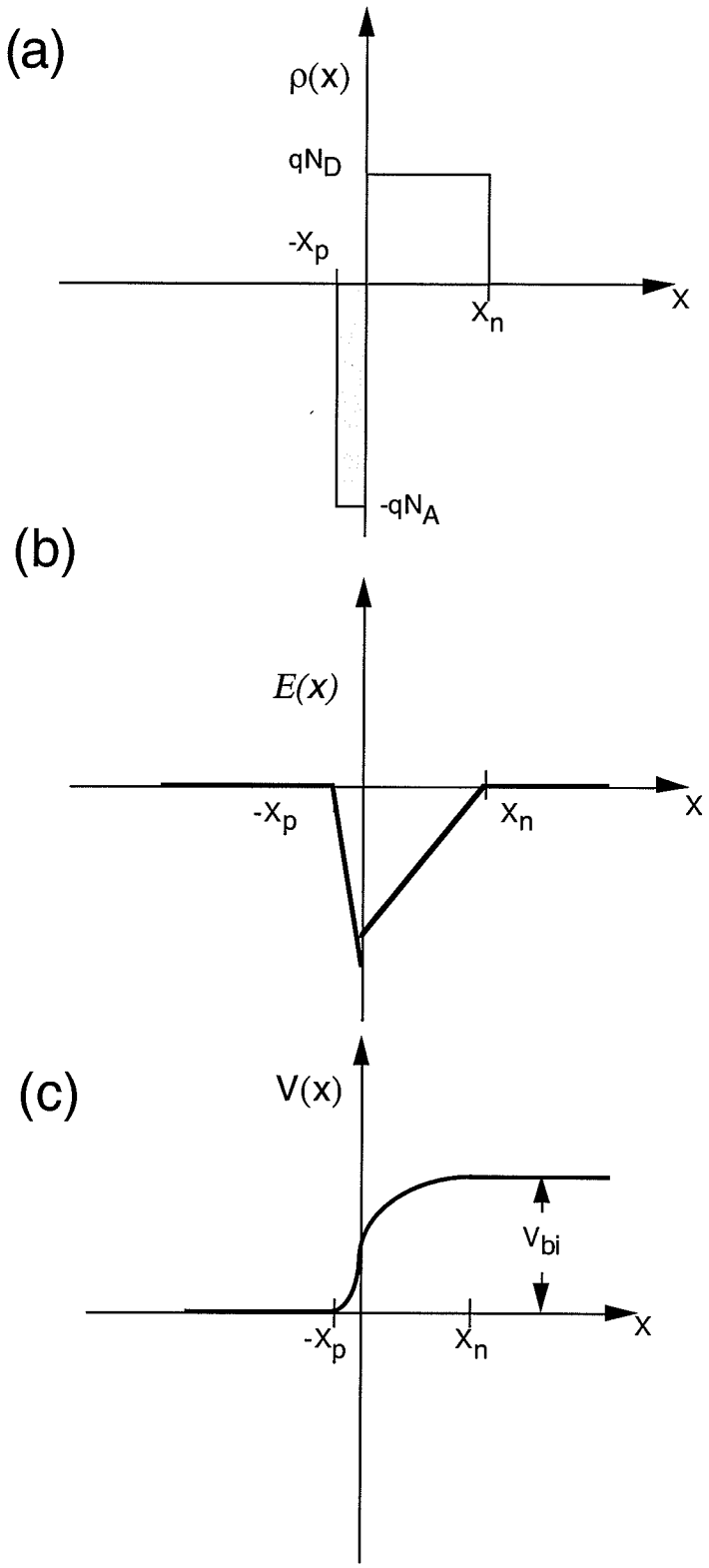


Fig. 10 Qualitative sketch of expected electrostatics for a pn heterojunction. (a) space-charge density vs. position, (b) electric field vs. position, and (c) electrostatic potential vs. position. Note the discontinuity in the electric field because the relative dielectric constant of the two sides may differ.

We begin with Poisson's Equation:

$$\frac{dE}{dx} = \frac{\rho}{\kappa\varepsilon_o} \tag{17}$$

where

$$\rho = q[p(x) - n(x) + N_D^+(x) - N_A^-(x)] \tag{18}$$

is the charge density. If we assume the depletion approximation for ρ, then

$$\begin{aligned} \rho(x) &= 0 && x < -x_p \\ \rho(x) &= -qN_A && -x_p < x < 0 \\ \rho(x) &= +qN_D && 0 < x < x_n \\ \rho(x) &= 0, && x_n < x \end{aligned}$$

and Poisson's equation. becomes

$$\frac{dE}{dx} = \frac{-qN_A}{\kappa_P\varepsilon_o} \qquad -x_p < x < 0 \tag{19}$$

$$\frac{dE}{dx} = \frac{+qN_D}{\kappa_N\varepsilon_o} \qquad 0 < x < x_n \tag{20}$$

and

$$E = 0. \qquad x < -x_p \text{ and } x > x_n$$

To obtain E (x) we integrate,

$$\int_0^{E(x)} dE = \frac{-qN_A}{\kappa_p\varepsilon_o} \int_{-x_p}^{x} dx$$

$$\int_{E(x)}^{0} dE = \frac{+qN_D}{\kappa_N\varepsilon_o} \int_{x}^{x_n} dx$$

and find

$$E(x) = \frac{-qN_A}{\kappa_P\varepsilon_o}(x + x_p) \qquad -x_p < x < 0 \tag{21a}$$

$$E(x) = \frac{-qN_D}{\kappa_N\varepsilon_o}(x_n - x)\ . \qquad 0 < x < x_n \tag{21b}$$

If there is no sheet charge at $x = 0$, Gauss' law states that

$$D(0^-) = \kappa_P E(0^-) = \kappa_N E(0^+) = D(0^+)$$

so

$$N_A x_p = N_D x_n \ , \tag{22}$$

which is simply a statement of charge conservation. To solve for $V(x)$ recall that

$$\frac{dV(x)}{dx} = -E(x)$$

defines $V(x)$ to within an arbitrary constant. Setting $V(-x_p) = 0$, we obtain from (21a)

$$\int_0^{V(x)} dV = \frac{qN_A}{\kappa_P \varepsilon_o} \int_{-x_p}^{x} (x + x_P) dx$$

or

$$V(x) = \frac{qN_A}{2\kappa_P \varepsilon_o} (x + x_p)^2 \ . \qquad -x_p < x < 0 \tag{23}$$

We also know that $V(x_n) = V_{bi} - V_A$. Integrating (21b) we find

$$V(x) = (V_{bi} - V_A) - \frac{qN_D}{2\kappa_N \varepsilon_o} (x_n - x)^2 \qquad 0 < x < x_n \tag{24}$$

Now to evaluate the depletion region widths, we invoke another boundary condition. We assume that no electrostatic dipoles exist at $x = 0$, so

$$V(0^-) = V(0^+),$$

which from (23) and (24) gives

$$\frac{qN_A x_p^2}{2\kappa_P \varepsilon_o} = (V_{bi} - V_A) - \frac{qN_D}{2\kappa_N \varepsilon_o} x_n^2 \ . \tag{25}$$

When (25) and (22) are solved, we find the depletion region widths as

$$x_p = \left[\frac{2\kappa_P \kappa_N \varepsilon_o N_D (V_{bi} - V_A)}{qN_A (N_A \kappa_P + N_D \kappa_N)} \right]^{1/2} \tag{26}$$

$$x_n = \left[\frac{2\kappa_P \kappa_n \varepsilon_o N_A (V_{bi} - V_A)}{qN_D(N_A\kappa_P + N_D\kappa_N)} \right]^{1/2} \tag{27}$$

and

$$W = x_p + x_n = \left[\frac{2\kappa_P \kappa_N \varepsilon_o (N_A + N_D)^2 (V_{bi} - V_A)}{q(N_A\kappa_P + N_D\kappa_N) N_A N_D} \right]^{1/2} . \tag{28}$$

It is also of interest to evaluate V_{JP} and V_{JN}, the potential drops on the P and N sides of the junction. These quantities are readily evaluated as

$$V_{JP} = V(0) = \frac{qN_A}{2\kappa_P \varepsilon_o} x_p^2 \qquad \text{from (23)}$$

and

$$V_{JN} = (V_{bi} - V_A) - V(0) = \frac{qN_D}{2\kappa_N \varepsilon_o} x_n^2 \quad . \qquad \text{from (24)}$$

Using (22) we find

$$\frac{V_{JP}}{V_{JN}} = \frac{N_D \kappa_N}{N_A \kappa_P} \tag{29}$$

Since, by definition

$$V_{bi} - V_A = V_{JP} + V_{JN},$$

We also have

$$V_{JP} = (V_{bi} - V_A) \left[\frac{\kappa_N N_D}{\kappa_N N_D + \kappa_P N_A} \right] \tag{30}$$

and

$$V_{JN} = (V_{bi} - V_A) \left[\frac{\kappa_P N_A}{\kappa_N N_D + \kappa_P N_A} \right] . \tag{31}$$

Using (30) and (31), we can express the depletion regions widths in terms of the potential drops on each side of the junction as

$$x_p = \left[\frac{2\kappa_P \varepsilon_o V_{JP}}{qN_A}\right]^{1/2} \tag{32}$$

and

$$x_n = \left[\frac{2\kappa_N \varepsilon_o V_{JN}}{qN_D}\right]^{1/2} \tag{33}$$

which are easier to remember than (26) and (27).

The key results derived above are collected and repeated below.

$$E(x) = \frac{-qN_A}{\kappa_P \varepsilon_o}(x + x_p) \qquad -x_p < x < 0$$

$$E(x) = \frac{-qN_D}{\kappa_N \varepsilon_o}(x_n - x) \qquad 0 < x < x_n$$

$$V(x) = \frac{qN_A}{2\kappa_P \varepsilon_o}(x + x_p)^2 \qquad -x_p < x < 0$$

$$V(x) = (V_{bi} - V_A) - \frac{qN_D}{2\kappa_N \varepsilon_o}(x_n - x)^2 \qquad 0 < x < x_n$$

$$V_{JP} = (V_{bi} - V_A)\left[\frac{\kappa_N N_D}{\kappa_N N_D + \kappa_P N_A}\right]$$

$$V_{JN} = (V_{bi} - V_A)\left[\frac{\kappa_P N_A}{\kappa_N N_D + \kappa_P N_A}\right]$$

$$\frac{V_{JP}}{V_{JN}} = \frac{N_D \kappa_N}{N_A \kappa_P}$$

$$x_p = \left[\frac{2\kappa_P \varepsilon_o V_{JP}}{qN_A}\right]^{1/2} \qquad x_n = \left[\frac{2\kappa_N \varepsilon_o V_{JN}}{qN_D}\right]^{1/2}$$

$$N_A x_p = N_D x_n \ , \qquad W = x_n + x_p$$

4. The Poisson-Boltzmann Equation

For device studies, a solution to Poisson's equation alone is often very useful. In equilibrium, it gives the equilibrium energy band diagram, and if we assume piece-wise constant quasi-Fermi levels (zero current), we can estimate the band diagram under bias. Often, however, we find that the depletion approximation cannot be assumed, so we need a general form of Poisson's equation which includes the mobile charge. Our purpose in this section is to formulate this equation, known as the *Poisson-Boltzmann equation.* The references in Sec. 10 discuss the solution of the Poisson-Boltzmann equation by analytical techniques under simple conditions or by numerical techniques under more general conditions.

For homostructures, κ_S is constant, and Poisson's equation is

$$\frac{d^2V(x)}{dx^2} = \frac{-q}{\kappa_S \varepsilon_0}\left[p(x) - n(x) + DOP(x)\right] , \tag{34}$$

which relates the electrostatic potential to the mobile and fixed charge, $DOP(x) = N_D^+(x) - N_A^-(x)$. By relating $p(x)$ and $n(x)$ to $V(x)$, we find

$$\frac{d^2V}{dx^2} = \frac{-q}{\kappa_S \varepsilon_0}\left[n_i e^{-q(V-\phi_p)/k_BT} - n_i e^{q(V-\phi_n)k_BT} + DOP(x)\right] , \tag{35}$$

which is the Poisson-Boltzmann equation - a nonlinear equation for $V(x)$. In (35), ϕ_p and ϕ_n are the quasi-Fermi potentials,

$$\begin{aligned} \phi_p &= (E_F - F_p)/q \\ \phi_n &= (E_F - F_n)/q , \end{aligned} \tag{36}$$

which are zero in equilibrium and can often be assumed to be piece wise constant under bias. We now seek an equation analogous to (35) but valid for heterostructures.

In a non-degenerate heterostructure,

$$n(x) = n_i(x) e^{(F_n(x) - E_I(x))/k_BT} , \tag{37}$$

where

$$n_i(x) = \sqrt{N_C(x) N_V(x)}\, e^{-E_G(x)/2k_BT} \tag{38}$$

and

$$E_I(x) = E_{ref} - \chi(x) - \frac{E_G(x)}{2} + \frac{k_B T}{2} \log\left[\frac{N_V(x)}{N_C(x)}\right] - qV(x) \tag{39}$$

is the intrinsic level. By writing (37) as

$$n(x) = n_{ir}\ e^{(F_n - E_I + k_B T\ \log(n_I / n_{ir}))k_B T} \tag{40}$$

(where n_{ir} is $n_i(x)$ evaluated at a reference location) using (36), (38), and (39), we find

$$n(x) = n_{ir}\ e^{q(V+V_n-\phi_n)/k_B T} \tag{41}$$

where

$$qV_n \equiv \chi(x) - \chi_{ref} + k_B T \log(N_C(x) / N_{Cref}) \tag{42}$$

is the so-called band parameter. (To eliminate some constants in (41), we've made a specific choice for the reference potential for V.) A similar development for holes leads to

$$p(x) = n_{ir}\ e^{-q(V-V_p-\phi_p)/k_B T} \tag{43}$$

where

$$qV_p \equiv (\chi(x) - \chi_{ref}) - (E_G(x) - E_{Gref}) + k_B T \log(N_V(x) / N_{Vref}). \tag{44}$$

Equations (41) and (43) are just like the corresponding expressions for homostructures except that n_i is evaluated at the reference location and band parameters which measure the change in material composition are introduced. Using (41) and (43), we find the Poisson-Boltzmann equation for heterostructures as

$$\frac{d}{dx}\left(-\kappa_S \varepsilon_0 \frac{dV}{dx}\right) = q\left(n_{ir} e^{-(V-V_p-\phi_p)/kT} - n_{ir} e^{(V+V_n-\phi_n)/kT} + DOP(x)\right) \tag{45}$$

We've made several simplifying assumptions to derive (45), but they are readily removed. For example, (45) can be extended to treat multiple-valley, non-parabolic bands with Fermi-Dirac statistics (Lundstrom, 1982 and Gray, 1985).

The Poisson-Boltzmann equation for heterostructures can easily be solved by conventional numerical techniques. For one example, consult (Gray, 1985), which describes FISH1D, a computer program to solve Poisson's equation for one-dimensional heterostructures.

5. Energy Bands in Graded Heterojunctions

Modern crystal growth technology makes it possible to grow heterojunctions that are nearly abrupt on an atomic scale, but for some devices it is preferable to intentionally grade the composition. The next step, then, is to learn how to sketch energy band diagrams when the material composition varies smoothly. We could do this by solving the Poisson-Boltzmann equation for $V(x)$ given an arbitrary material composition as specified by the band parameters, $V_n(x)$ *and* $V_p(x)$. From the resulting $V(x)$ and Eqs. (6) and (7), we could find the energy band diagram. Our objective here, however, is to sketch energy band diagrams for graded heterostructures without resorting to computers.

To begin, let's re-draw our energy band diagram for the N-p junction using a different procedure, as outlined in Fig. 11. First, note that we can easily sketch the expected electrostatic potential for the junction shown in Fig. 11a; the result is displayed in Fig. 11b. Next, let's sketch an assumed electron affinity profile, $\chi(x)$ as also shown in Fig. 11b. According to Eqs. (6) and (7), the energy band diagram is simply a constant minus the electrostatic potential and minus the electron affinity. By turning the electrostatic potential and electron affinity profiles upside down and adding the two, we get the conduction band profile shown in Fig. 11c. We then find the valence band profile by subtracting the position dependent bandgap from $E_C(x)$. The final result shown in Fig. 11c is just the energy band diagram we sketched in Fig. 5, but the procedure followed here is readily generalized for compositionally graded heterojunctions.

The procedure consists of separating the problem into two parts: 1) we first sketch the electrostatic potential profile, which we can often estimate, then 2) add the compositional profile, which is determined from the growth process. The procedure becomes useful when the compositional profile is graded rather than abrupt. Figure 12 shows how we construct the energy band diagram for an N-p heterojunction in which the compositional transition is graded rather than abrupt. (We assume that the compositional grading is contained within the depletion region and does not extend into the quasi-neutral semiconductor.) The compositional grading does not change the electrostatics, which are determined by the depletion approximation, so the only difference is the electron affinity profile that we add to $V(x)$. As shown in Fig. 12, the compositional grading smoothes out the conduction band spike, and the energy band diagram appears much like the one expected for an abrupt N-p heterojunction with $\Delta E_C = 0$. We say that the compositional grading

has "removed" the band spike and, indeed, compositional grading is often used to remove band spikes that are detrimental to devices.

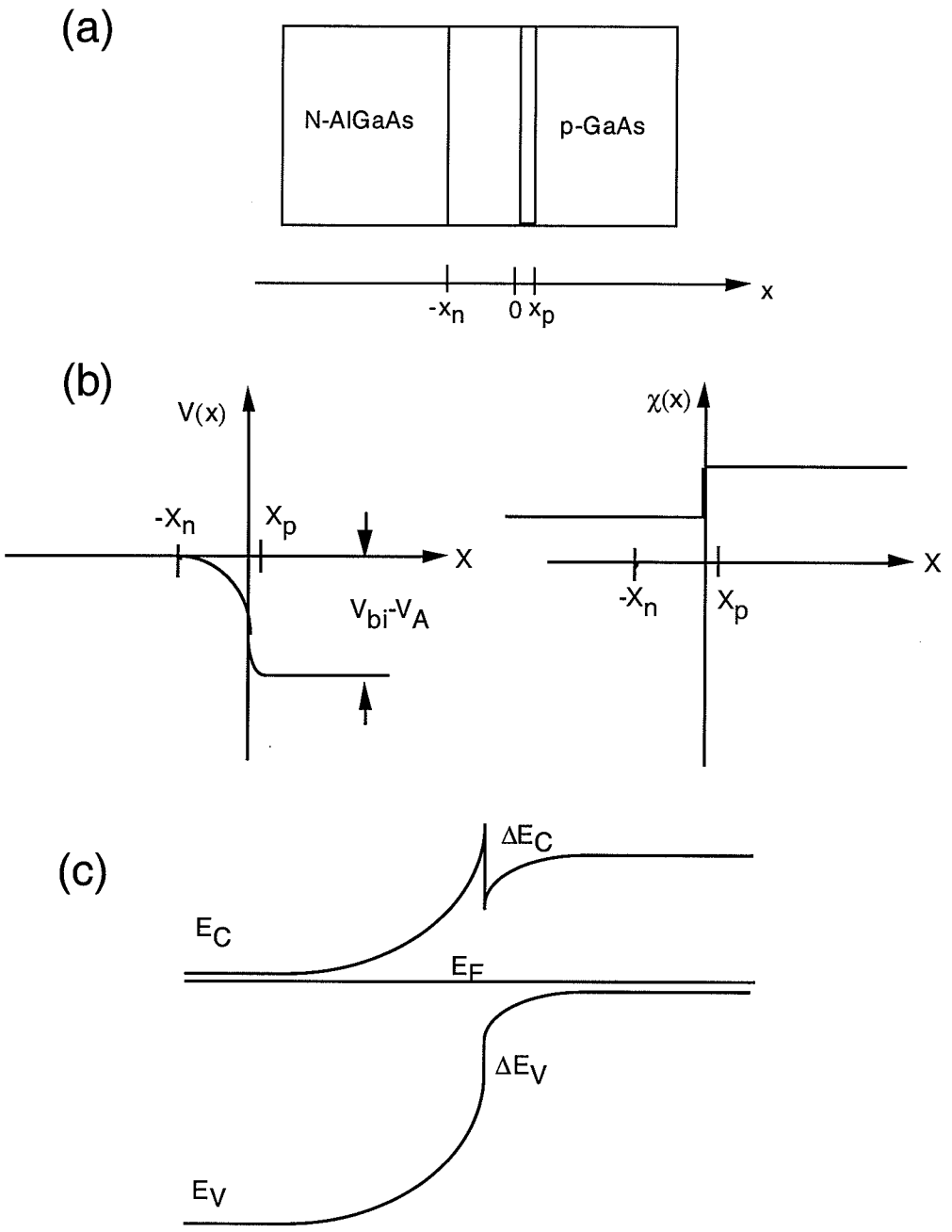


Fig. 11. The energy band diagram for an N-p abrupt heterojunction. In this case, we begin with the junction structure in (a) from which we can deduce how the charge will separate. From the charge profile, we can sketch the potential profile as shown in (b). Given an assumed electron affinity profile as shown also in (b), we can deduce the conduction band profile by flipping the potential and affinity profiles upside down and adding them. The final result shown in (c) is the result we obtained in Fig. 5.

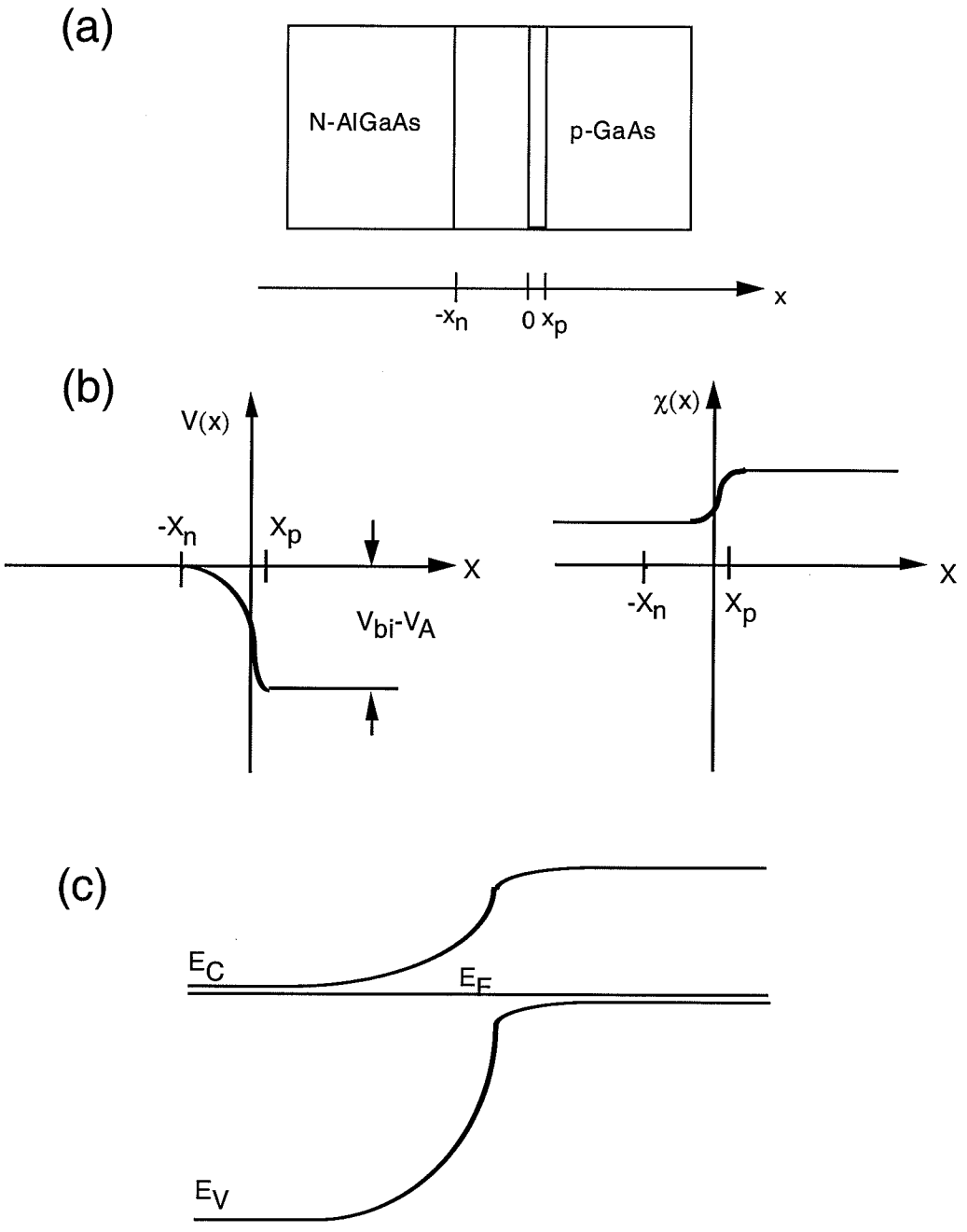


Fig. 12. The energy band diagram for an N-p graded heterojunction. The problem is identical to the one sketched in Fig. 11, except that the composition is graded about $x = 0$.

Design Example: Optimal Grading Profile for an Npn HBT

Let's consider a problem, specifying a compositional profile to remove the conduction band spike in an AlGaAs/GaAs/GaAs Npn heterojunction bipolar transistor. For this device, the base doping density greatly exceeds the emitter doping density, $N_{AB} >> N_{DE}$, so the electrostatic potential drops mostly across the N-AlGaAs emitter. For $N_A >> N_D$, Eq. (24) reduces to

$$V(x) = (V_{bi} - V_A)\left[1 - (1 + x / x_n)^2\right] \qquad -x_n < x < 0 \tag{46}$$

where

$$x_n = \left[\frac{2\kappa_n \varepsilon_0 (V_{bi} - V_A)}{qN_D}\right]^{1/2}. \tag{47}$$

The compositional profile is given by

$$\chi(x) = \chi_{GaAs} - \Delta\chi(x) \tag{48}$$

where the maximum value of $\Delta\chi$ is limited to ΔE_C.

The key idea is to recall that the electrostatic potential will decrease $E_C(x)$, as we move into the N-AlGaAs region, but $\chi(x)$ decreases, which tends to increase $E_C(x)$ in opposition to the potential. To prevent a spike from occurring, we should balance the two effects,

$$\Delta\chi(x) = q(V_{bi} - V_A)\left[1 - (1 + x / x_n)^2\right] \tag{49}$$

Since the maximum value of $\Delta\chi$ is limited to ΔE_C, we can find the extent of the graded region by solving

$$\Delta E_C = q(V_{bi} - V_A)\left[1 - (1 + x_g / x_n)^2\right] \tag{50}$$

for the grading distance, x_g, to find

$$x_g = x_n\left[1 - \sqrt{1 - \Delta E_C / q(V_{bi} - V_A)}\right]. \tag{51}$$

Finally, dividing (49) by (50), we find

$$\frac{\Delta\chi(x)}{\Delta E_C} = \frac{1-\left(1+x/x_n\right)^2}{1-\left(1+x_g/x_n\right)^2}. \tag{52}$$

So, given a ΔE_C that must be removed by grading, we first evaluate the grading distance from (51), then grade the composition quadratically over this distance according to (52). Using typical numbers for an N-$Al_{0.3}Ga_{0.7}As$ / -p-GaAs heterojunction ($V_{bi} \approx 1.25$ V and $\Delta E_C \approx 0.25$ V), we find from Eq. (51) that $x_g \approx 0.1\ x_n$ so conduction band spikes are readily removed by compositional grading over a short distance.

Compositional Grading in Quasi-neutral Regions

The procedure sketched in Fig. 12 for constructing band diagrams of graded heterojunctions relies on a key assumption, that compositional grading does not influence the electrostatic potential. As a result, we can first deduce the electrostatic potential from the depletion approximation, then add the compositional profile. This assumption works well for pn heterojunctions because the potential is determined by charge in a depletion region which is set by the doping density. For other situations, however, it may not be possible to decouple grading from the electrostatic potential. In general, numerical techniques like those discussed in Sec. 4 might be necessary, but there is a simple, important case that we should discuss.

Figure 13 considers the case of a uniformly doped p-type semiconductor with a linear compositional grading. Figure 13 (a) shows the energy band diagram for a similar graded semiconductor, but this one is intrinsic rather than p-type. The correct equilibrium energy band diagram for the uniformly doped p-type structure is shown in Fig. 13 (b). We can establish that this is the correct band diagram from a simple argument. If the doping is uniform, then the resulting hole density will also be very nearly uniform because any separation of the mobile charge from the dopants will uncover charge which will set up an electric field that opposes further separation. Since the Fermi level is constant in equilibrium and because the separation of the valence band and Fermi level determines the hole density, we conclude that the valence band is very nearly constant. (We are ignoring the small slope in E_V caused by the variation of the effective density of states, N_V.) For this case, the electrostatics force the valence band to be nearly constant, and the full variation of the bandgap, $E_G(0) - E_G(W)$, is felt by the conduction band. The key point is that the majority carrier band is flat and the minority carrier band acquires a slope of dE_G/dx.

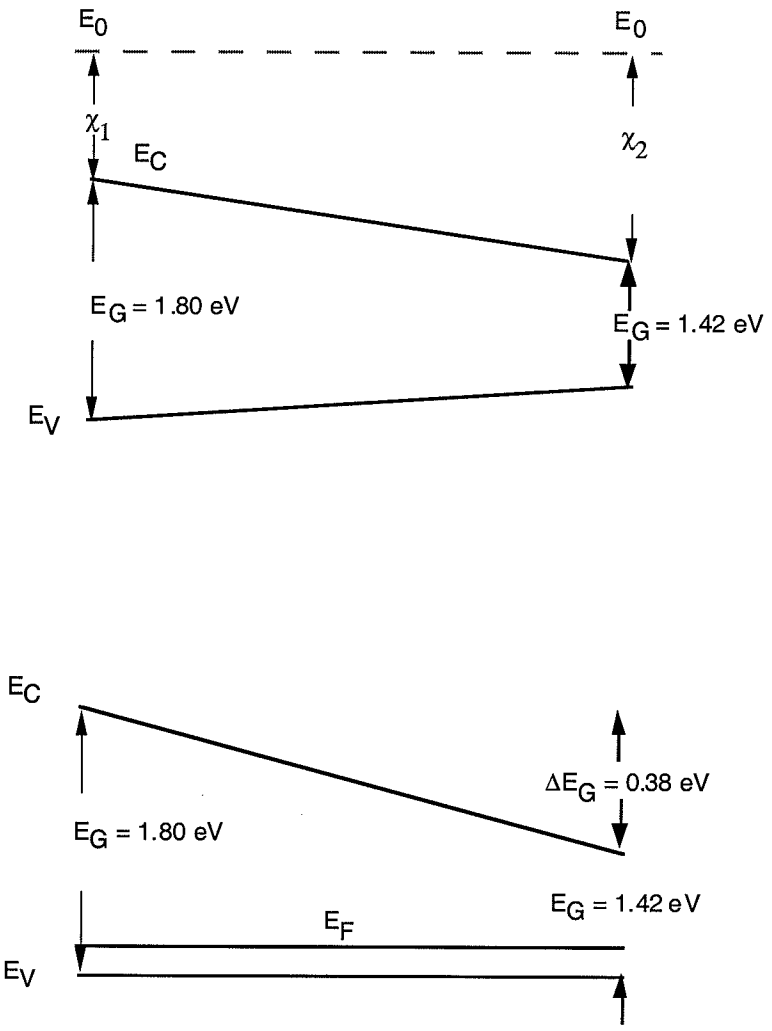


Fig. 13. The energy band diagram for a compositionally graded, uniformly doped p-type semiconductor. (a) the band diagram for the intrinsic semiconductor, and (b) the band diagram for a quasi-neutral, p-type semiconductor.

6. General Picture of a Compositionally Graded Semiconductor

In the most general case, both *V(x)* and the composition of the semiconductor vary with position Since the composition is nonuniform, the crystal periodicity is broken, and one may question the whole concept of energy bands. Consult (Gora, 1967) and van Vliet (1994) for a discussion of this point. Suffice is to say that if the composition varies slowly, we may take the band structure at any point to be the band structure of the corresponding bulk semiconductor with the composition at that point.

Because the composition is nonuniform, E_C and E_V (and therefore χ and E_G) will also be nonuniform. As a result, (6) and (7) become

$$E_C(x) = E_o - \chi_S(x) - qV(x) \tag{53}$$

$$E_V(x) = E_o - \chi_S(x) - qV(x) - E_G(x). \tag{54}$$

An energy band for this case might look like Fig. 14

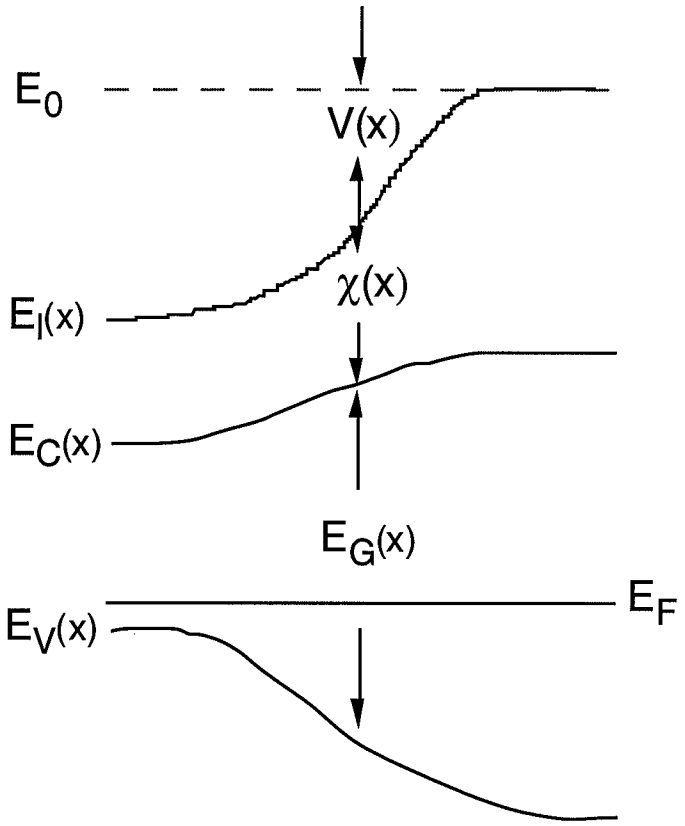


Fig. 14. An energy band diagram for a compositionally graded semiconductor.

Figure 14 shows a semiconductor with band-bending which is due to both an electric field and to compositional variations. The slope of the conduction band gives the force on an electron, but it is impossible to deduce the electric field from the energy band diagram.

Let's consider the force acting on an electron in the conduction band,

$$F_e = \frac{-dE_C}{dx} = q\frac{dV(x)}{dx} + \frac{d\chi_S}{dx} \tag{55}$$

and on a hole in the valence band

$$F_h = \frac{+dE_V}{dx} = q\frac{dV(x)}{dx} - \frac{d(\chi_S + E_G)}{dx}. \tag{56}$$

The force on electrons is not equal in magnitude and opposite in direction to the force on holes, as we would expect for forces due to electric fields. The electric field is only one component of the force on a carrier. Since we are used to thinking of electric fields producing forces on carriers, we can define *quasi-electric fields* for electrons by

$$F_e = -qE(x) - qE_{QN}(x) \tag{57}$$

and for holes by

$$F_h = +qE(x) + qE_{QP}(x). \tag{58}$$

With these definitions we have

$$E_{QN} = -\frac{1}{q}\frac{d\chi_S}{dx} \tag{59}$$

and

$$E_{QP} = -\frac{1}{q}\frac{d}{dx}(\chi_S + E_G). \tag{60}$$

Notice that the quasi E-field for electrons can differ both in magnitude and direction from the quasi E-field for holes. These quasi-E fields give the device designer an additional degree of freedom since they can be controlled by the nonuniform composition (Kroemer, 1957). Neither conditions (1), (2) nor (3) on page 2 apply to the E-band diagram of Fig. 13. Notice also that E_C and E_V are not constrained to be parallel in a heterostructure.

Example: Removing Band Spike in Npn AlGaAs/GaAs/GaAs HBTs:

We have already discussed how to grade the emitter base junction of an HBT optimally. The concept of quasi-electric fields provide a simple, quick estimate for the required grading. Assume that Figure 11 illustrates the emitter base junction of a heterojunction bipolar transistor. For such devices, the p-type base is heavily doped and the band bending occurs on the N-AlGaAs side of the heterojunction. The height of the band spike is then ΔE_C above the conduction band in the p-GaAs, and we seek a simple estimate for the compositional profile needed to eliminate the band spike.

Recall that the electrostatic potential pulls the conduction band down in Fig. 11, and the compositional change from GaAs to AlGaAs pulls the conduction band up, producing the band spike. We can think of these two effects in terms of electric fields. At the junction, the electric field is *E(0)* and if we linearly grade the composition from GaAs to $Al_{0.3}Ga_{0.7}As$ over a distance of x_g, then we have a quasi-electric field for electrons of $\Delta E_C/x_g$ which opposes the electric field. To remove the band spike, we want the quasi-electric field to be smaller than the actual electric field, so

$$\frac{\Delta E_c / q}{x_g} < \left|E(0^-)\right|$$

which gives the required grading distance as

$$x_g > \frac{\Delta E_C / q}{\left|E(0^-)\right|} \quad . \tag{61}$$

(We are assuming that the electric field is approximately constant over the compositionally graded region.) Equation (61) gives a rough estimate for the grading distance. Of course the resulting band profile will not be as smooth as the optimal grading profile discussed in Sec. 5, in which the quadratic compositional profile precisely matched the quadratic potential profile.

7. Drift-Diffusion Equations for Heterostructures

Our goal in this section is to derive the hole and electron current equations, J_p and J_n, for a semiconductor with nonuniform composition.

To begin, recall:

$$J_p = p\mu_p \frac{dF_p}{dx} \tag{62a}$$

and

$$J_n = n\mu_n \frac{dF_n}{dx}, \tag{62b}$$

where F_p and F_n are the quasi-Fermi levels. If the semiconductor is nondegenerate, the carrier densities are related to the quasi-Fermi levels by

$$p = N_v(x)e^{(E_v - F_p)/kT} \tag{63a}$$

and

$$n = N_c(x)e^{(F_n - E_C)/kT}. \tag{63b}$$

From (63a) and (63b), we obtain:

$$F_p = E_V - kT\log\left(\frac{p}{N_v}\right) \tag{64}$$

and

$$F_n = E_C - kT\log\left(\frac{n}{N_c}\right). \tag{65}$$

Consequently

$$\frac{dF_p}{dx} = \frac{dE_V}{dx} - kT\left[\frac{1}{p}\frac{dp}{dx} - \frac{1}{N_c}\frac{dN_v}{dx}\right] \tag{66}$$

$$\frac{dF_n}{dx} = \frac{dE_C}{dx} - kT\left[\frac{1}{p}\frac{dn}{dx} - \frac{1}{N_c}\frac{dN_c}{dx}\right] \tag{67}$$

Using the above in (62a) and (62b) we get

$$J_p = p\mu_p\left[\frac{dE_V}{dx} + \frac{kT}{N_v}\frac{dN_v}{dx}\right] - kT\mu_p\frac{dp}{dx} \tag{68}$$

$$J_n = n\mu_n\left[\frac{dE_C}{dx} + \frac{kT}{N_c}\frac{dN_c}{dx}\right] - kT\mu_n\frac{dn}{dx} \tag{69}$$

These expressions look like drift-diffusion equations - especially if we use the Einstein relation

$$\frac{D_p}{\mu_p} = \frac{D_n}{\mu_n} = \frac{kT}{q} \tag{70}$$

To complete the derivation, we must express dE_C/dx and dE_V/dx in terms of material parameters and the electrostatic potential, $V(x)$. From our band model for a nonuniform semiconductor, we have

$$E_C(x) = E_o - \chi(x) - qV(x) \tag{71}$$

and

$$E_V(x) = E_C(x) - E_G(x) = E_o - \chi(x) - E_G(x) - qV(x), \tag{72}$$

where E_0 is a (constant) reference energy level. If we differentiate these, and use (68) and (69),we get

$$J_p = -pq\mu_p\left[\frac{d}{dx}\left\{V + \frac{\chi(x)}{q} + \frac{E_G(x)}{q}\right\} - \frac{kT}{q}\frac{1}{N_v}\frac{dN_v}{dx}\right] - qD_p\frac{dp}{dx} \tag{73}$$

$$J_n = -nq\mu_n\left[\frac{d}{dx}\left\{V + \frac{\chi(x)}{q}\right\} + \frac{kT}{q}\frac{1}{N_c}\frac{dN_c}{dx}\right] + qD_n\frac{dn}{dx} \tag{74}$$

In examining Eqs. (73) and (74), we see that compositional variations introduce additional terms into the drift-diffusion equations. The drift current (the part in { } brackets) includes both electric and quasi-electric fields. We also see a conventional diffusion current, but there is another term involving the gradient of the effective densities of states. It is not clear whether to write this *density-of-states effect* (Marshak, 1978) as a drift current (because it is proportional to the carrier concentration) or as a diffusion current (because it involves a gradient). This term arises because carriers tend to diffuse in the direction of increasing density of states, because there are more states available for the random walk. In practice, however, the biggest effect is drift in the quasi-electric fields, and the density of states effect is rather small.

Finally, the current equations can be made more notationally compact by defining two band parameters, V_p *and* V_n, as

$$qV_p(x) = -[\chi(x) + E_G(x) - kT\log N_v(x)] + const \tag{75}$$

and

$$qV_n(x) = \chi(x) + kT\log N_c(x) + const \tag{76}$$

With these definitions, we finally obtain

$$\boxed{J_p = -pq\mu_p\frac{d}{dx}(V - V_p) - qD_p\frac{dp}{dx}} \tag{77}$$

$$\boxed{J_n = -nq\mu_n\frac{d}{dx}(V + V_n) + qD_n\frac{dn}{dx}} \tag{78}$$

which look exactly like the conventional current equations except for the two new terms, V_p and V_n. If we choose the constants in Eqs. (75) and (76) to make the band parameters zero in a reference material, then we get Eqs. (42) and (44), the same band parameters used in the Poisson-Boltzmann equation.

Example: Application to a linearly graded, quasi-neutral heterostructure

As an example of applying the current equations, (73) and (74), let's consider the graded heterostructure of Fig. 13. If electrons are injected from the left, then under low injection conditions, there will be a strong force producing an electron current, but the hole current will be small. Assuming that $J_p \approx 0$, we can use Eq. (73) to find

$$J_p = -pq\mu_p \frac{d}{dx}\left\{V + \frac{\chi(x)}{q} + \frac{E_G(x)}{q}\right\} \approx 0 \tag{79}$$

(we have assumed that the hole concentration is uniform and that the density-of-states does not vary). From Eq. (79) we can find the electric field as

$$E = -\frac{dV}{dx} = \frac{d}{dx}\left\{\frac{\chi(x)}{q} + \frac{E_G(x)}{q}\right\} \tag{80}$$

which is the electric field set up to flatten the valence band and keep the hole concentration constant. If we insert this electric field into Eq. (74), we find the minority electron current as

$$J_n = nq\mu_n \frac{d(E_G / q)}{dx} \; . \tag{81}$$

So the minority carrier electrons drift in a total field (the actual field plus the quasi-electric field) which is minus the gradient of the bandgap variation. This is exactly the conclusion we reached from the energy band diagram we sketched in Sec. 5. The approach here, however, is easily generalized to treat spatial variations of the effective densities of states and of the majority hole density.

8. Heavy Doping Effects and Heterostructures

In a heavily doped semiconductor, carrier-carrier interactions reduce the energy gap and carrier-dopant interactions distort the bands by introducing band tails (Abram, 1993). The resulting position-dependent, perturbed band structure influences carrier transport and has a strong effect on the electrical performance of devices (del Alamo 1987, Slotboom, 1976). Modeling transport in such structures is much like modeling transport in any heterostructure, but the details of the perturbed band structure are still not well-understood, so a specialized terminology has developed

(Lundstrom, 1981, van Vliet, 1993). The specialized terminology is reviewed in this section and is related to the conventional treatment of transport in heterostructures described in Sec. 7.

Band structure of Heavily Doped Semiconductors

Figure 15 illustrates the band structure perturbations induced by heavy impurity doping (del Alamo, 1987). When the semiconductor is lightly doped, the conduction and valence bands are parabolic with sharply defined band edges. As the doping density increases, the localized impurity states broaden into an impurity band and the fluctuating potential associated with the dopants located on randomly situated lattice sites produces tails in the conduction and valence bands. For very heavy impurity doping, the impurity band merges with the conduction band. Under such conditions, electron-hole correlation and exchange effects dominate and these many body effects produce a more or less rigid shrinkage of the band gap. The electronic properties of heavily doped semiconductors are profoundly influenced by these perturbations in the band structure.

Fig. 15 Illustration of how heavy impurity doping affects the band structure of an n-type semiconductor (from del Alamo, 1987).

The $n_o p_o$ Product

In heavily doped semiconductors, the perturbed band structure alters $n_o p_o$ which must be evaluated from

$$n_o = \int_{E_{C0}'}^{E_{top}} \rho_{C'}(E - E_{C0}')f(E - E_F)dE\,, \tag{82a}$$

and

$$p_o = \int_{E_{bot}}^{E_{V0}'} \rho_{V'}(E_{V0}' - E)[1 - f(E - E_F)dE\,, \tag{82b}$$

where ρ_C' is the perturbed conduction band density-of-states, f is the Fermi function, and E_{C0}' is the conduction band mobility edge (a corresponding set of definitions for the valence band apply to the quantities with a subscript, V). For lightly-doped semiconductors, the densities-of-states are parabolic and the product of Eqs. (82a) and (82b) reduce to

$$n_o p_o = n_{io}^2 = N_C N_V\, e^{-E_G/k_B T}\,, \tag{83}$$

where N_C and N_V are the effective densities-of-states for the conduction and valence bands. For heavily-doped semiconductors, however, the result is much different. The doping-induced perturbation of the energy bands can be viewed at the simplest level as an effective narrowing of the energy gap from which we conclude that $n_o p_o$ will increase with doping density. Both detailed many body calculations and measurements confirm that $n_o p_o$ does increase in heavily doped p-GaAs (Abram, 1993; Harmon, 1994).

The $n_o p_o$ product is an important factor in device design and analysis, and for such purposes it is convenient to express the product in the simple form (Lundstrom, 1981),

$$n_o p_o \equiv n_{ie}^2 \equiv n_{io}^2\, e^{\Delta_G^0/k_B T}\,. \tag{84}$$

The effective intrinsic carrier concentration n_{ie} is greater than n_{io}, which refers to a lightly doped semiconductor. It is important to note that Eq. (84) is a defining equation for Δ_G^o, the equilibrium *effective* (or *apparent* or *electrical*) energy gap shrinkage, which is not a physical energy gap shrinkage but is, rather, related to the band structure by equating Eq. (84) to the product of Eqs. (82a) and (82b). If we adopt a simple parabolic band model for the band structure of a heavily doped semiconductor, then the effective gap shrinkage is

$$\Delta_G(x) = \Delta E_G + k_B T \log\left(\frac{N_V(x)N_C(x)}{N_{Vref}N_{Cref}}\right) + k_B T \log\left(\frac{F_{1/2}(\eta_V)F_{1/2}(\eta_C)}{e^{\eta_V}\quad e^{\eta_C}}\right) \tag{85}$$

where $\eta_V = (E_{V0} - E_F)/k_BT$ and $\eta_C = (E_F - E_{C0})/k_BT$, and $F_{1/2}$ is the Fermi-Dirac integral of order one-half. For a degenerate p-type semiconductor, $F_{1/2}(\eta_V)/e^{\eta_V} < 1$ and $F_{1/2}(\eta_C)/e^{\eta_C} \cong 1$, so the effect of majority carrier degeneracy is to widen the effective energy gap, which is analogous to the well-known, Burstein-Moss shift in the optical gap. Equation (85) is valid only when the energy bands are parabolic, but it does illustrate how band gap shrinkage and Fermi-Dirac statistics influence the $n_o p_o$ product. Under nondegenerate conditions and if the densities of states are not perturbed, then the effective gap shrinkage reduces to the actual bandgap shrinkage.

It must be emphasized that a variety of effective or apparent energy gap shrinkages have been defined and that these nonphysical, defined quantities should not be confused (Lundstrom, 1981; del Alamo, 1987; Marshak, 1987). This distinction is especially important when comparing the results of electrical measurements of devices to optical absorption or photoluminescence measurements. The effective gap shrinkage (or equivalently, the effective intrinsic carrier concentration) is the important quantity for modeling bipolar devices because it determines the minority carrier concentration and, as we demonstrate below, the minority carrier current in a heavily doped semiconductor.

Minority Carrier Transport

Equations (77) and (78) show that the drift-diffusion equations for heterostructures are just like the conventional equations but with additional terms that account for the quasi-electric fields and position-dependent effective mass. It is interesting to observe that

$$n_i^2(x) = n_i^2(ref) \times e^{q(V_n(x)+V_p(x))/k_BT}, \tag{86}$$

which shows that the quantity $V_n + V_p$ acts like an effective energy gap shrinkage given by

$$\Delta_G(x) = qV_n + qV_p = \Delta E_G + k_BT \log\left(\frac{N_V(x)N_C(x)}{N_{Vref}N_{Cref}}\right). \tag{87}$$

Equation (87) relates the band parameters introduced for heterostructures to the effective bandgap shrinkage used to described heavily doped semiconductors. Equation (87) is a simplified form of Eq. (85) for nondegenerate semiconductors.

For heavily doped semiconductors, it is convenient to re-cast the drift-diffusion equations in terms of the effective energy gap shrinkage. A little algebra shows that (Lundstrom, 1981)

$$J_n = -nq\mu_n \frac{d}{dx}\{V + \gamma\Delta_G/q\} + k_BT\mu_n \frac{dn}{dx}, \tag{88a}$$

and

$$J_{px} = -pq\mu_p \frac{d}{dx}\{V - (1-\gamma)\Delta_G / q\} - k_B T\mu_p \frac{dp}{dx} \tag{88b}$$

where

$$\gamma \equiv \frac{V_n}{V_n + V_p} \tag{89}$$

is the *effective asymmetry factor*, a measure of the change in band edge and band structure associated with the conduction band.

The form of Eqs. (88) suggests that they are restricted to non-degenerate semiconductors, but they are easily generalized. When carriers are degenerate, $k_B T\mu \neq qD$, so the final term in the current equations should not be interpreted as a diffusion current. Instead, the influence of carrier degeneracy on the diffusion coefficient, D, can be included in the parameters, γ and Δ_G, if they are generalized for Fermi-Dirac statistics (Lundstrom, 1981).

According to Eqs. (88), the perturbed band structure introduces two additional parameters into the current equations. The first, Δ_G, is precisely the effective gap shrinkage defined in equilibrium by Eq. (84), and the second, the effective asymmetry factor, γ, is a measure of how much of the perturbation is associated with the conduction band. Equation (84) defines the effective gap shrinkage in equilibrium. Under low injection conditions, it is common to assume that the effective gap shrinkage is equal to its value in equilibrium, but high densities of injected carriers may perturb the band structure.

Equations (88a) and (88b) are equivalent to those commonly used for analyzing semiconductor heterostructures (Sutherland, 1977; Lundstrom, 1983) but are expressed in the form most useful when the compositional variation is due to heavy doping effects (Lundstrom, 1981). Detailed expressions for Δ_G and γ are available (Lundstrom, 1981), but it may be helpful to note that if the semiconductor is non degenerate and the band structure perturbation is a simple, rigid shift of the bands without any change in band shape, then Δ_G is the actual energy gap shrinkage and γ is the fraction of the shrinkage associated with a perturbation in the conduction band edge.

Minority Carrier Transport in Quasi-Neutral Regions

Many minority carrier devices are controlled by diffusion of carriers across a quasi-neutral region in low level injection. Consider a quasi-neutral p-type region in low level injection. The equilibrium field is found from Eq. (88b) by setting $\vec{J}_p = 0$ which is then inserted in Eq. (88a) to find

$$\vec{J}_n = -nq\mu_n\left[-(k_BT/q)\frac{\nabla p_o}{p_o} + \nabla(\Delta_G^o/q)\right] + k_BT\mu_n\nabla n, \tag{90}$$

where a subscript or superscript, "o," denotes the equilibrium value of a quantity. Equation (90) should accurately describe low injection electron transport in p^+-GaAs because under such conditions, the electrostatic potential, the effective gap shrinkage, and the majority carrier hole concentration are all very near their equilibrium values. The important point is that to describe low injection electron transport in quasi-neutral p^+-GaAs, the only information required about the perturbed band structure is the quantity, Δ_G - the effective gap shrinkage, which explains its widespread use in device modeling. Conversely, electrical measurements of minority carrier currents only provide information on Δ_G; they are insensitive to γ. When the simplifying assumptions of minority carrier diffusion in quasi-neutral regions under low level injection are not met, then a detailed understanding of the perturbed band structure is necessary. For example, just to compute the built-in potential of a p-n junction requires knowledge of both Δ_G^o and γ^o (Lundstrom, 1981).

9. Band Offsets

We have saved the most difficult part for last. We now know how to draw band diagrams, compute electrostatic potentials, and deduce current flows, but everything depends on how the bands line up. Given a pair of materials, we need a way to determine the band offsets. In this section, I'll discuss a few theories and approaches and will try to give a pragmatic approach to the question. A problem that all approaches face, is that while one might predict a band offset to 0.1 eV or so, they frequently influence devices through exp[-ΔE_C /kT] terms, so very accurate values are often needed.

The Electron Affinity Rule:

In the approach we have been taking, the conduction band offset was given by the difference in the electron affinities of the two materials in the heterojunction (recall Eq. (15)). This approach is also known as the Anderson Rule, after R.L. Anderson, one of the early heterojunction researchers (Anderson, 1962). In practice, however, the electron affinity rule has not proven to be a reliable way to deduce band offsets.

There are several reasons for the difficulties with the electron affinity rule. First, recall that electron affinities are typically about 4 eV and that band discontinuities are frequently one-tenth of this. Subtracting two large numbers with their experimental errors to get a small result is bound to produce large errors. In addition, electron affinities depend on surface charges and dipoles, but the

band offsets should depend on properties of the two bulk semiconductors, surfaces should not have to enter into the problem. Finally, note that when two semiconductors are brought into contact, there is a charge transfer between the bond at the interface which sets up a dipole. There is no reason to believe that the dipole at a semiconductor-semiconductor interface will be the same as the dipole at the semiconductor vacuum interface. For these reasons, researchers sought better ways to deduce band offsets.

The Common Anion Rule:

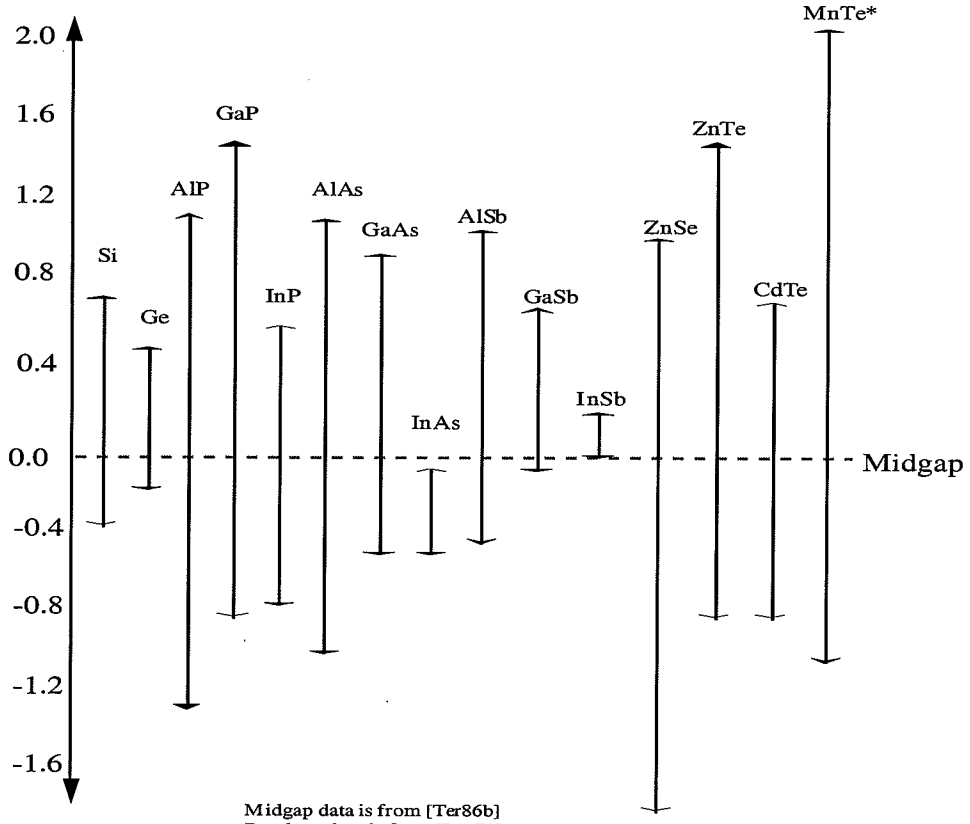
When one deduces the band structure of a semiconductor by a tight binding approach, one finds that to zeroth order, the valence band is made up of the anion wavefunction. For a heterojunction like AlAs/GaAs (or AlGaAs/GaAs) for which As is the common anion, the valence bands for each of the materials should be similar. The expectation would be that the valence band discontinuity would be small and most of the discontinuity would be taken up by the conduction band. For a long time, people thought that the conduction band discontinuity in AlGaAs/GaAs was 85% of the bandgap discontinuity, and the common anion rule was cited as the theoretical explanation. Later measurements, however, revealed that only about 60% of the bandgap discontinuity went to the conduction band, and people began to look more critically at the common anion argument. The valence band is determined in the first approximation by the anion, but the first order corrections involve the cation and are large enough to render the common anion rule unreliable. (Remember that band discontinuities tend to be small, so small shifts in the valence band energies can have a large effect on the resulting discontinuity.) The common anion rule is now used qualitatively. When there is a common anion, people expect the conduction band discontinuity to be larger than the valence band discontinuity, but how much is hard to say.

Tersoff's Quantum Dipole Theory:

Band discontinuities must be related to the microscopic properties of the heterointerface and there have been several different theoretical approaches to the problem. One of the most successful theories is due to Tersoff (Tersoff, 1984). In Tersoff's view, a band discontinuity arises from a quantum dipole set up when bulk states from one semiconductor tunnel a few Angstroms into gap states located within the bandgap of the other. These "interface induced gap states" are analogous to the "metal induced gap states" that form at a metal-semiconductor interface. The band discontinuity adjusts itself to force the dipole to zero. The view is that the gap states are mostly conduction band in character near the top of the bandgap (i.e. negatively charged when filled, or acceptor-like) and valence band in character near the bottom of the bandgap (neutral when filled, or donor-like). Somewhere in between is a neutral point, where the gap states take equal contributions from the conduction and valence bands. Tersoff argues that the semiconductors align

to match these effective midgap points. By calculating the effective midgap points for several semiconductors and evaluating band discontinuities by subtracting them, Tersoff obtained remarkably good agreement with experimentally measured discontinuities. Figure 16 shows the absolute location of the bands for several semiconductors using Tersoff's midgap energy as the reference.

An interesting feature of Tersoff's theory is that it also provides an explanation for the observed Schottky barrier heights of several metal-semiconductor pairs. In Tersoff's theory, the effective midgap point of the semiconductor aligns with the Fermi level of the metal. His theory is equally successful at explaining the observed barrier heights of metal-semiconductor junctions. While Tersoff's theory has gained a wide acceptance among theorists and experimentalists, other theories have also been proposed to explain the close connection between Schottky barrier heights and heterojunction band offsets (e.g. Freeouf, 1986).



[Ter86b] J. Tersoff, Phys. Rev. Lett. 56, 2755 (1986).
[Pan71] J.I. Pankove, Optical processes in semiconductors, Dover Publications, Inc., NY (1971).
[Han91] J. Han et al., J. Cryst. Growth 111, 767 (1991).

Fig. 16 The conduction and valence band positions for several common semiconductors as deduced from Tersoff's theory. (Courtesy of Jung Han, Purdue University)

Deducing Band Offsets from Schottky Barrier Heights

The correlation between measured Schottky barrier heights and heterojunction band offsets had been noticed well before Tersoff developed his theory. This correlation provides a practical approach to estimating band offsets. Figure 17 illustrates how it works. We show two metal-semiconductor junction, with a common metal. According to Tersoff, the effective midgap point of each semiconductor aligns to the Fermi level of the metal. Therefore, we take the Fermi level as our long sought reference level. From the figure, we see that

$$E_C = \phi_{bn} \tag{91}$$

and

$$E_V = E_G - \phi_{bn}. \tag{92}$$

Extensive tables of measured Schottky barrier heights are available (e.g. see S.M. Sze, Physics of Semiconductor Devices, John Wiley and Son, Inc. 1981, page 291), we can construct a table of the energies of the conduction and valence bands of various semiconductors. When we form a heterojunction, the band offset is just the difference between the two energies. Alternatively, we can view the process of forming a heterojunction as reducing the thickness of the metal in Fig. 18 until the two semiconductors meet. It is clear from Fig. 17 that

$$\Delta E_C = \phi_{bn}(AlAs) - \phi_{bn}(GaAs) \ . \tag{93}$$

The valence band discontinuity is then deduced from the known bandgap difference and ΔE_C. (In practice, the procedure is usually reversed, one deduces the valence band discontinuity first from the Schottky barrier heights for p-type semiconductors. This procedure is more reliable because the valence band maximum is always at $\boldsymbol{k} = 0$.)

Schottky barrier heights are typically several tenths of an electron volt and band discontinuities are typically a few tenths of an eV, so the subtraction in Eq. (93) is not prone to excessive errors. Deducing band discontinuities from measured Schottky barrier heights has proven to be easy and fairly reliable in practice. Figure 18 displays the absolute locations of the conduction and valence bands for several common III-V semiconductors as deduced from Schottky barrier heights (Tiwari, 1992).

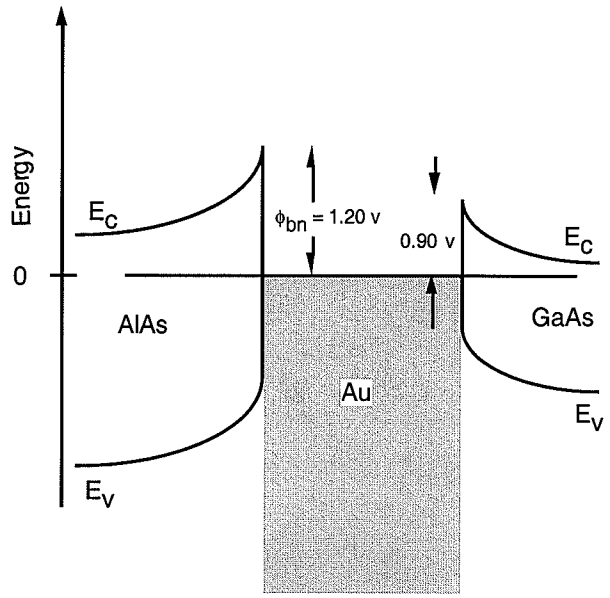


Fig. 17 Illustration of how one deduces band offsets from measured Schottky barrier heights.

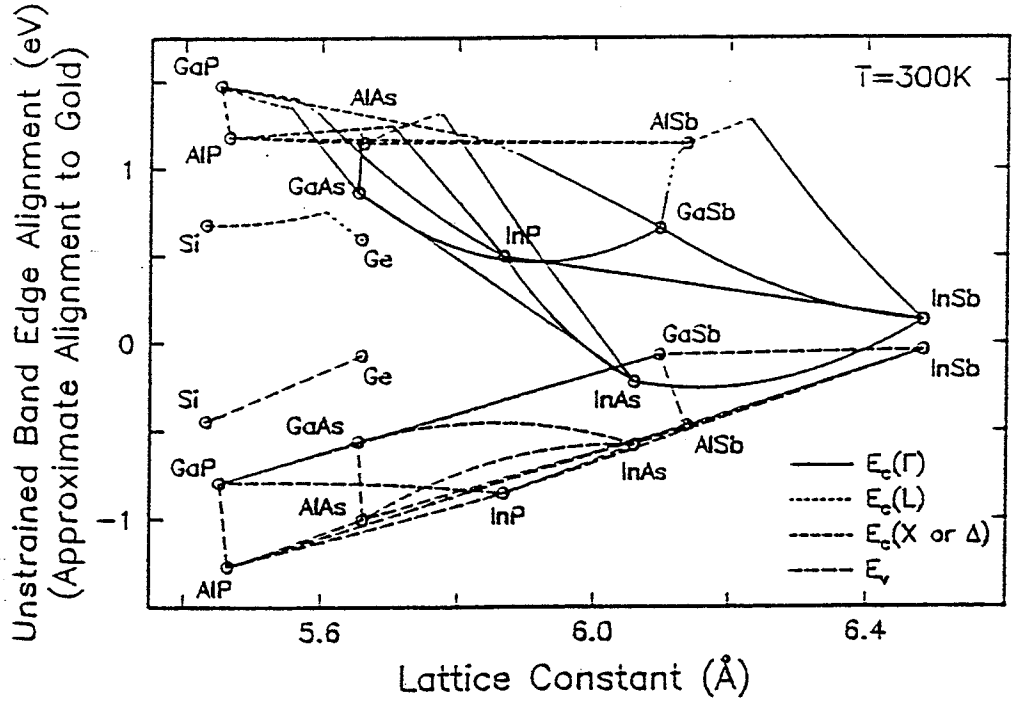


Fig. 18 The conduction and valence band energies plotted as a function of lattice constant of semiconductors. The circles indicate the band edges of binary semiconductors and the lines show the band edges of the ternary alloys. The two endpoints of each ternary line are the binary constituents of that ternary. Discontinuities between two lattice matched or nearly lattice matched semiconductor alloys may be found from the difference in energy between their band-edge energies. The zero energy point represents the approximate gold Schottky barrier position in the band gap of any given alloy. (From Tiwari, 1992)

10. Suggested Reading

Abram, R.A. (1993) in Minority Carriers in Compound Semiconductors: Physics and Applications, Eds. R.A. Ahrenkiel and M.S. Lundstrom, *Semiconductors and Semimetals*, **39** p. 259-317.

Anderson, R.L. (1962), "Experiments on Ge-GaAs heterojunctions," *Solid-State Electron.*, **5**, p. 341.

Cheung, D.T., Chiang, S.Y., and Pearson, G.L. (1975), "A Simplified Model for Graded Gap Heterojunctions," *Solid-State Electron.*, **18**, pp. 263-266.

del Alamo J.A., and Swanson, R.M. (1987), "Modeling of Minority Carrier Transport in Heavily Doped Silicon Emitters," *Solid-State Electron.*, **30**, 1127.

Freeouf, J.L. and Woodall, J.M. (1986), *Surf. Sci.*, **168**, p. 518.

Gora, T. and Williams, F. (1967), "Electronic States of Homogeneous and Inhomogeneous Mixed Semiconductors," II-VI Semiconducting Compounds, ed. by D.G. Thomas, Benjamin, New York, 639.

Gray J.L. and Lundstrom, M.S. (1985), "Solution to Poisson's Equation with Application to C-V Analysis of Heterojunction Capacitors," *IEEE Trans. Electron Dev.*, **ED-32**, pp. 2102-2109.

Harmon, E.S., Melloch, M.R., and Lundstrom, M.S. (1994), "Effective Bandgap Shrinkage in GaAs," *Appl. Phys. Lett.*, **64**, pp. 502-504.

Harrison, W.A. (1977), "Elementary Theory of Heterojunctions," *J. Vac. Sci. and Technol.*, **14,** pp. 1016.

Harrison, W.A. (1985), "Theory of Band Line-ups," *J. Vac. Sci. Technol. B.*, **3**, pp. 1231-1238.

Hayes, R., Cappaso, F., Malik, R.J., Gossard, A.C., and Wiegmann, W.(1983), "Optimum Emitter Grading for the Heterojunction Bipolar Transistor," *Appl. Phys. Lett.*, **43**, pp. 949-951.

H. Kroemer, (1957), "Quasi-Electric and Quasi-Magnetic Fields in Nonuniform Semiconductors," *RCA Review,* pp. 332-342.

Lundstrom, M.S., Schwartz, R.J., and Gray, J.L. (1981), "Transport Equations for the Analysis of Heavily Doped Semiconductor Devices," *Solid-State Electron.*, **24**, 413.

Lundstrom M.S. and Schuelke, R.J. (1982), "Modeling Semiconductor Heterojunctions in Equilibrium," *Solid-State Electron.*, **25**, pp. 683-692.

Lundstrom M.S. and Schuelke, R.J. (1983), "Numerical Simulation of Heterostructure Semiconductor Devices," *IEEE Trans. Electron Dev.*, **ED-30**, pp. 1151-1159.

Lundstrom, Mark (1990), Fundamentals of Carrier Transport, Vol. X in the Modular Series on Solid State Devices, Addison-Wesley, Reading, Mass.

Lundstrom, M.S. (1993), "Minority Carrier Transport in III-V Semiconductors," in Minority Carriers in Compound Semiconductors; Physics and Applications, Eds. R.A. Ahrenkiel and M.S. Lundstrom, *Semiconductors and Semimetals*, **39** p. 193-258.

Marshak, A.H. and van Vliet, K.M (1978a), ``Electrical Currents in Solids with Position-Dependent Band Structure, "*Solid-State Electron.*," **21,** 417.

Marshak, A.H. and van Vliet, K.M (1978b), "Carrier Densities and Emitter Efficiency in Degenerate Materials With Position-Dependent Band Structure," *Solid-State Electron.*, **21**,

Slotboom, J.W. and de Graaff, H.C. (1976), "Measurements of Bandgap Narrowing in Si Bipolar Transistors," *Solid-State Electron.*, **29**, 857.

Lundstrom, Mark (1990), Fundamentals of Carrier Transport, Vol. X in the Modular Series on Solid State Devices, Addison-Wesley, Reading, Mass.

Tersoff, J. (1984) "Theory of Semiconductor Heterojunctions: The Role of Quantum Dipoles," *Phys. Rev. B*, **30**, pp. 4874-4877.

Tiwari, S. and Frank, D.J. (1992) "Empirical Fit to Band Discontinuities and Barrier Heights in III-V Alloy Systems," *Appl. Phys. Lett*, **60**, pp. 630-632.

van Vliet , C.M. (1993), "Bandgap Narrowing and Emitter Efficiency in Heavily Doped Emitter Structures Revisited," *IEEE Trans. Electron Dev.*, **40**, p. 1140-1147.

Van Vliet, C. M., Nainaparampil, J. J., and Marshak, Alan H. (1994), "Ehrenfest derivation of the mean forces acting in materials with non-uniform band structure: A canonical approach," Solid-*State Electron.*, **37**.

11. Exercises

To test your understanding of the concepts discussed in these notes, try working the following problems.

1) Using the observed band offsets for AlGaAs/GaAs as illustrated in Fig. 1, sketch an energy band diagram for a doubly intrinsic, $Al_{0.3}Ga_{0.7}As$ / GaAs heterojunction.

2) Sketch the band diagram for an $Al_{0.3}Ga_{0.7}As$ / GaAs Pp heterojunction.

3) Sketch the energy band diagram for an $Al_{0.3}Ga_{0.7}As$ / GaAs / $Al_{0.3}Ga_{0.7}As$ NpN heterojunction.

4) Compute the built-in potential for an $Al_{0.3}Ga_{0.7}As$ / GaAs Np heterojunction assuming that $N_D = 2.0 \times 10^{17}$ and $N_A = 1.0 \times 10^{19}$ cm^{-3}.

5) For the $Al_{0.3}Ga_{0.7}As$ / GaAs heterojunction in problem 4, use the depletion approximation to compute, $E(0^-)$, $E(0^+)$, x_n, x_p, V_{JP}, and V_{JN}. Assume that $\kappa = 10.8$ for $Al_{0.3}Ga_{0.7}As$ and 12.1 for GaAs.

6) Repeat the problem sketch in Fig. 13, but this time assume n-type doping.

7) For heavily doped semiconductors, another common practice is to express the current equations in terms of the position dependent effective intrinsic carrier concentration. What assumption is necessary to do so for Eqs. (77a) and (77b)? What assumptions are necessary to do the same in Eq. (79)?

8) Assuming the AlGaAs/GaAs Np heterojunction of problems 4 and 5, deduce the grading width to remove the band spike under equilibrium conditions. Assume linear compositional grading and use Eq. (49). Construct an energy band diagram for the graded junction under equilibrium and under a forward bias of 1.2 volts.

9) Assuming the AlGaAs/GaAs Np heterojunction of problems 4 and 5 to deduce the grading width to remove the band spike under equilibrium conditions. Assume quadratic compositional grading. Construct an energy band diagram for the graded junction under equilibrium and under a forward bias of 1.2 volts.

10) Derive a minority carrier transport equation like Eq. (70) but allow the hole concentration and the effective densities of states to vary with position. Compare your result to Eq. (79).

Additional information for Exercises 2–5 on page T3-42.

(2) Assume $E_F - E_v = 0.1$eV for both the $Al_{0.3}Ga_{0.7}As$ and GaAs components.

(3) Assume abrupt heterojunctions, $E_F = E_v$ for the GaAs base, and E_F slightly below E_c for both the AlGaAs emitter and collector.

(4) Note that both the $Al_{0.3}Ga_{0.7}As$ and GaAs are weakly degenerately doped. Refer to Table 4.2 and the $N_{C,V}$ relationship in the nearby text of Volume VI in determining N_V and N_C. Employ the electron density of states effective mass for $Al_{0.3}Ga_{0.7}As$ as deduced from Table 1.1 on page T4-1. Make use of Fig. 4.15 in Volume VI in deducing the approximate positions of the Fermi levels.

(5) Perform the calculations for $V_A = 0$.