

## 6.4: Charge Carriers in Semiconductors - Statics and Kinetics

Now let me demonstrate the application of the concepts discussed in the last section to understanding the basic kinetic properties of semiconductors and a few key semiconductor structures – which are the basis of most modern electronic and optoelectronic devices, and hence of all our IT civilization. For that, I will need to take a detour to discuss their equilibrium properties first.

I will use an approximate but reasonable picture in which the energy of the electron subsystem in a solid may be partitioned into the sum of effective energies  $\varepsilon$  of independent electrons. Quantum mechanics says<sup>32</sup> that in such periodic structures as crystals, the stationary state energy  $\varepsilon$  of a particle interacting with the atomic lattice follows one of periodic functions  $\varepsilon_n(\mathbf{q})$  of the *quasimomentum*  $\mathbf{q}$ , oscillating between two extreme values  $\varepsilon_{n|min}$  and  $\varepsilon_{n|max}$ . These *allowed energy bands* are separated by *bandgaps*, of widths  $\Delta_n \equiv \varepsilon_{n|min} - \varepsilon_{n-1|max}$ , with no allowed states inside them. Semiconductors and insulators (dielectrics) are defined as such crystals that in equilibrium at  $T = 0$ , all electron states in several energy bands (with the highest of them called the *valence band*) are completely filled,  $\langle N(\varepsilon_v) \rangle = 1$ , while those in the upper bands, starting from the lowest, *conduction band*, are completely empty,  $\langle N(\varepsilon_c) \rangle = 0$ .<sup>33</sup> Since the electrons follow the Fermi-Dirac statistics (2.8.5), this means that at  $T \rightarrow 0$ , the Fermi energy  $\varepsilon_F \equiv \mu(0)$  is located somewhere between the valence band's maximum  $\varepsilon_{v|max}$  (usually called simply  $\varepsilon_V$ ), and the conduction band's minimum  $\varepsilon_{c|min}$  (called  $\varepsilon_C$ ) – see Figure 6.4.1.

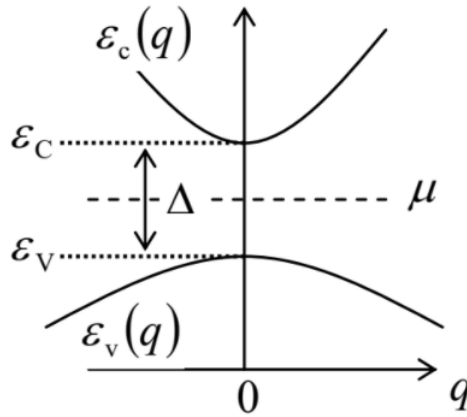


Figure 6.4.1: Calculating  $\mu$  in an intrinsic semiconductor.

$$\varepsilon = \begin{cases} \varepsilon_c + q^2/2m_c, & \text{for } \varepsilon \geq \varepsilon_c, \\ \varepsilon_v + q^2/2m_v, & \text{for } \varepsilon \leq \varepsilon_v, \end{cases} \quad \text{with } \varepsilon_c - \varepsilon_v \equiv \Delta. \quad (6.4.1)$$

The positive constants  $m_C$  and  $m_V$  are usually called the effective masses of, respectively, electrons and holes. (In a typical semiconductor,  $m_C$  is a few times smaller than the free electron mass  $m_e$ , while  $m_V$  is closer to  $m_e$ .)

Due to the similarity between the top line of Equation (6.4.1) and the dispersion law (3.1.3) of free particles, we may re-use Equation (3.2.11), with the appropriate particle mass  $m$ , the degeneracy factor  $g$ , and the energy origin, to calculate the full spatial density of populated states (in semiconductor physics, called *electrons* in the narrow sense of the word):

$$n \equiv \frac{N_c}{V} = \int_{\varepsilon_C}^{\infty} \langle N(\varepsilon) \rangle g_3(\varepsilon) d\varepsilon \equiv \frac{g_c m_c^{3/2}}{\sqrt{2\pi^2 \hbar^3}} \int_0^{\infty} \langle N(\tilde{\varepsilon} + \varepsilon_C) \rangle \tilde{\varepsilon}^{1/2} d\tilde{\varepsilon}, \quad (6.4.2)$$

where  $\tilde{\varepsilon} \equiv \varepsilon - \varepsilon_C \geq 0$ . Similarly, the density  $p$  of “no-electron” excitations (called *holes*) in the valence band is the number of *unfilled* states in the band, and hence may be calculated as

$$p \equiv \frac{N_h}{V} = \int_{-\infty}^{\varepsilon_v} [1 - \langle N(\varepsilon) \rangle] g_3(\varepsilon) d\varepsilon \equiv \frac{g_v m_v^{3/2}}{\sqrt{2\pi^2 \hbar^3}} \int_0^{\infty} [1 - \langle N(\varepsilon_v - \tilde{\varepsilon}) \rangle] \tilde{\varepsilon}^{1/2} d\tilde{\varepsilon}, \quad (6.4.3)$$

where in this case,  $\tilde{\varepsilon} \geq 0$  is defined as  $(\varepsilon_v - \varepsilon)$ . If the electrons and holes<sup>35</sup> are in the thermal and chemical equilibrium, the functions  $\langle N(\varepsilon) \rangle$  in these two relations should follow the Fermi-Dirac distribution (2.8.5) with the same temperature  $T$  and the same chemical potential  $\mu$ . Moreover, in our current case of an undoped (*intrinsic*) semiconductor, these densities have to be equal,

$$n = p \equiv n_i, \quad (6.4.4)$$

because if this *electroneutrality condition* was violated, the volume would acquire a non-zero electric charge density  $\rho = e(p - n)$ , which would result, in a bulk sample, in an extremely high electric field energy. From this condition, we get a system of two equations,

$$n_i = \frac{g_C m_c^{3/2}}{\sqrt{2\pi^2 \hbar^3}} \int_0^\infty \frac{\tilde{\varepsilon}^{1/2} d\tilde{\varepsilon}}{\exp\{(\tilde{\varepsilon} + \varepsilon_C - \mu)/T\} + 1} = \frac{g_V m_v^{3/2}}{\sqrt{2\pi^2 \hbar^3}} \int_0^\infty \frac{\tilde{\varepsilon}^{1/2} d\tilde{\varepsilon}}{\exp\{(\tilde{\varepsilon} - \varepsilon_V + \mu)/T\} + 1} \quad (6.4.5)$$

whose solution gives both the requested charge carrier density  $n_i$  and the Fermi level  $\mu$ .

For an arbitrary ratio  $\Delta/T$ , this solution may be found only numerically, but in most practical cases, this ratio is very large. (Again, for Si at room temperature,  $\Delta \approx 1.14$  eV, while  $T \approx 0.025$  eV.) In this case, we may use the same classical approximation as in Equation (3.2.16), to reduce Eqs. (6.4.2) and (6.4.3) to simple expressions

$$n = n_c \exp\left\{\frac{\mu - \varepsilon_C}{T}\right\}, \quad p = n_v \exp\left\{\frac{\varepsilon_V - \mu}{T}\right\}, \quad \text{for } T \ll \Delta, \quad (6.4.6)$$

where the temperature-dependent parameters

$$n_c \equiv \frac{g_C}{\hbar^3} \left(\frac{m_c T}{2\pi}\right)^{3/2} \quad \text{and} \quad n_v \equiv \frac{g_V}{\hbar^3} \left(\frac{m_v T}{2\pi}\right)^{3/2} \quad (6.4.7)$$

may be interpreted as the effective numbers of states (per unit volume) available for occupation in, respectively, the conduction and valence bands, in thermal equilibrium. For usual semiconductors (with  $g_C \sim g_V \sim 1$ , and  $m_C \sim m_V \sim m_e$ ), at room temperature, these numbers are of the order of  $3 \times 10^{25} m^{-3} \equiv 3 \times 10^{19} cm^{-3}$ . (Note that all results based on Eqs. (6.4.6) are only valid if both  $n$  and  $p$  are much lower than, respectively,  $n_C$  and  $n_V$ .)

With the substitution of Eqs. (6.4.6), the system of equations (6.4.4) allows a straightforward solution:

$$\mu = \frac{\varepsilon_V + \varepsilon_C}{2} + \frac{T}{2} \left( \ln \frac{g_V}{g_C} + 3 \ln \frac{m_v}{m_c} \right), \quad n_i = (n_C n_V)^{1/2} \exp\left\{-\frac{\Delta}{2T}\right\}. \quad (6.4.8)$$

Since in all practical materials the logarithms in the first of these expressions are never much larger than 1,<sup>36</sup> it shows that the Fermi level in intrinsic semiconductors never deviates substantially from the so called *midgap value*  $(\varepsilon_V + \varepsilon_C)/2$  – see the (schematic) Figure 6.4.1. In the result for  $n_i$ , the last (exponential) factor is very small, so that the equilibrium number of charge carriers is much lower than that of the atoms – for the most important case of silicon at room temperature,  $n_i \sim 10^{10} cm^{-3}$ . The exponential temperature dependence of  $n_i$  (and hence of the electric conductivity  $\sigma \propto n_i$ ) of intrinsic semiconductors is the basis of several applications, for example simple *germanium resistance thermometers*, efficient in the whole range from  $\sim 0.5$  K to  $\sim 100$  K. Another useful application of the same fact is the extraction of the bandgap of a semiconductor from the experimental measurement of the temperature dependence of  $\sigma \propto n_i$  – frequently, in just two well-separated temperature points.

However, most applications require a much higher concentration of carriers. It may be increased quite dramatically by planting into a semiconductor a relatively small number of slightly different atoms – either *donors* (e.g., phosphorus atoms for Si) or *acceptors* (e.g., boron atoms for Si). Let us analyze the first opportunity, called *n-doping*, using the same simple energy band model (6.4.1). If the donor atom is only slightly different from those in the crystal lattice, it may be easily ionized – giving an additional electron to the conduction band, and hence becoming a positive ion. This means that the effective ground state energy  $\varepsilon_D$  of the additional electrons is just slightly below the conduction band edge  $\varepsilon_C$  – see Figure 6.4.2a.<sup>37</sup>

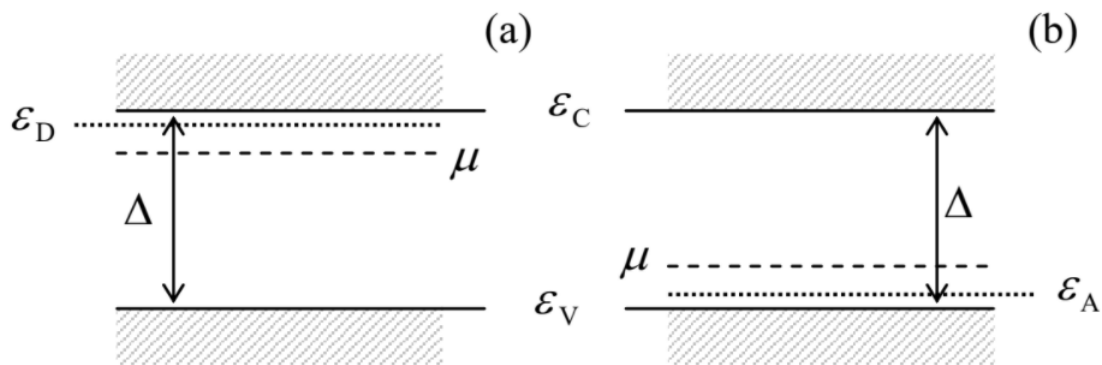


Figure 6.4.2: The Fermi levels  $\mu$  in (a)  $n$ -doped and (b)  $p$ -doped semiconductors. Hatching shows the ranges of unlocalized state energies.

$$np = n_i^2. \quad (6.4.9)$$

However, for a doped semiconductor, the electroneutrality condition looks differently from Equation (6.4.4), because the total density of positive charges in a unit volume is not  $p$ , but rather  $(p + n_+)$ , where  $n_+$  is the density of positively-ionized (“activated”) donor atoms, so that the electroneutrality condition becomes

$$n = p + n_+. \quad (6.4.10)$$

If virtually all dopants are activated, as it is in most practical cases,<sup>39</sup> then we may take  $n_+ = n_D$ , where  $n_D$  is the total concentration of donor atoms, i.e. their number per unit volume, and Equation (6.4.10) becomes

$$n = p + n_D. \quad (6.4.11)$$

Plugging in the expression  $p = n_i^2/n$ , following from Equation (6.4.9), we get a simple quadratic equation for  $n$ , with the following physically acceptable (positive) solution:

$$n = \frac{n_D}{2} + \left( \frac{n_D^2}{4} + n_i^2 \right)^{1/2}. \quad (6.4.12)$$

This result shows that the doping affects  $n$  (and hence  $\mu = \varepsilon_C - T \ln(n_C/n)$  and  $p = n_i^2/n$ ) only if the dopant concentration  $n_D$  is comparable with, or higher than the intrinsic carrier density  $n_i$  given by Equation (6.4.8). For most applications,  $n_D$  is made *much* higher than  $n_i$ ; in this case Equation (6.4.12) yields

$$n \approx n_D \gg n_i, \quad p = \frac{n_i^2}{n} \approx \frac{n_i^2}{n_D} \ll n, \quad \mu \approx \mu_p \equiv \varepsilon_C - T \ln \frac{n_C}{n_D}. \quad (6.4.13)$$

Because of the reasons to be discussed very soon, modern electron devices require doping densities above  $10^{18} \text{ cm}^{-3}$ , so that the logarithm in Equation (6.4.13) is not much larger than 1. This means that the Fermi level rises from the midgap to a position only slightly below the conduction band edge  $\varepsilon_C$  – see Figure 6.4.2a

The opposite case of purely  $p$ -doping, with  $n_A$  acceptor atoms per unit volume, and a small activation (negative ionization) energy  $\varepsilon_A - \varepsilon_V \ll \Delta$ ,<sup>40</sup> may be considered absolutely similarly, using the electroneutrality condition in the form

$$n + n_- = p, \quad (6.4.14)$$

where  $n_-$  is the number of activated (and hence negatively charged) acceptors. For the relatively high concentration ( $n_i \ll n_A \ll n_V$ ), virtually all acceptors are activated, so that  $n_- \approx n_A$ , Equation (6.4.14) may be approximated as  $n + n_A = p$ , and the analysis gives the results dual to Equation (6.4.13):

$$p \approx n_A \gg n_i, \quad n = \frac{n_i^2}{p} \approx \frac{n_i^2}{n_A} \ll p, \quad \mu \approx \mu_n \equiv \varepsilon_V + T \ln \frac{n_V}{n_A}. \quad (6.4.15)$$

so that in this case, the Fermi level is just slightly above the valence band edge (Figure 6.4.2b), and the number of holes far exceeds that of electrons – again, in the narrow sense of the word. Let me leave the analysis of the simultaneous  $n$ - and  $p$ -doping

(which enables, in particular, so-called *compensated semiconductors* with the sign-variable difference  $n - p \approx n_D - n_A$ ) for the reader's exercise.

$$\frac{d^2\phi}{dx^2} = -\frac{\rho(x)}{\kappa\epsilon_0}. \quad (6.4.16)$$

Here  $\kappa$  is the dielectric constant of the semiconductor matrix – excluding the dopants and charge carriers, which in this approach are treated as explicit (“stand-alone”) charges, with the volumic density

$$\rho = e(pn_- - n). \quad (6.4.17)$$

(As a sanity check, Eqs. (6.4.16)-(6.4.17) show that if  $\mathcal{E} \equiv -d\phi/dx = 0$ , then  $\rho = 0$ , bringing us back to the electroneutrality condition (6.4.14), and hence the “flat” band-edge diagrams shown in Figs. 6.4.2b and 6.4.3a)

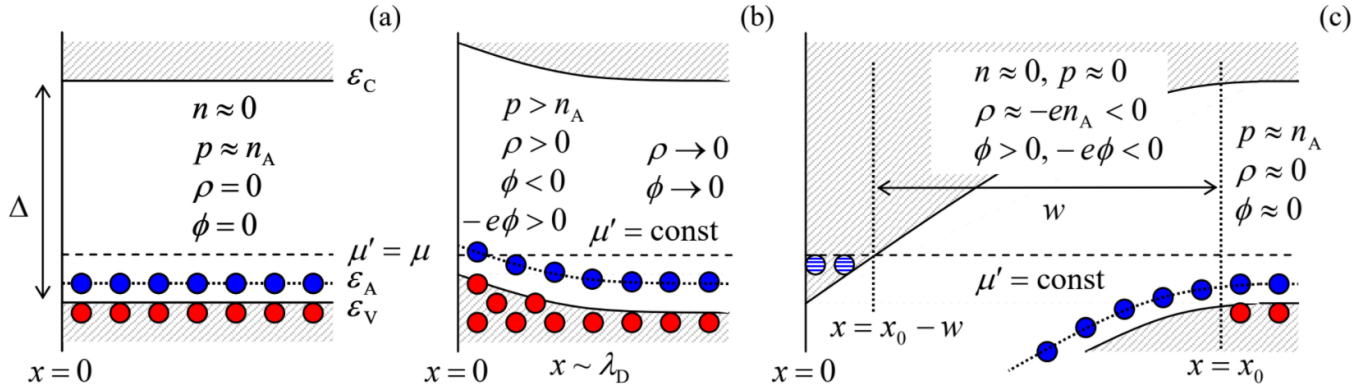


Figure 6.4.3: The band-edge diagrams of the electric field penetration into a uniform p-doped semiconductor: (a)  $\mathcal{E} = 0$ , (b)  $\mathcal{E} < 0$ , and (c)  $\mathcal{E} > \mathcal{E}_c > 0$ . Solid red points depict positive charges; solid blue points, negative charges; and hatched blue points, possible electrons in the inversion layer – all very schematically.

$$\frac{d\phi}{dx}(0) = -\mathcal{E}. \quad (6.4.18)$$

Note that the electrochemical potential  $\mu'$  (which, in accordance with the discussion in Sec. 3, replaces the chemical potential in presence of the electric field),<sup>46</sup> has to stay constant through the system in equilibrium, keeping the electric current equal to zero – see Equation (6.3.6). For arbitrary doping parameters, the system of equations (6.4.6) (with the replacements  $\epsilon_V \rightarrow \epsilon_V - e\phi$ , and  $\mu \rightarrow \mu'$ ) and (6.4.16)-(6.4.18), plus the relation between  $n_-$  and  $n_A$  (describing the acceptor activation), does not allow an analytical solution. However, as was discussed above, in the most practical cases  $n_A \gg n_i$ , we may use the approximate relations  $n_- \approx n_A$  and  $n \approx 0$  at virtually any values of  $\mu'$  within the locally shifted bandgap  $[\epsilon_V - e\phi(x), \epsilon_C - e\phi(x)]$ , so that the substitution of these relations, and the second of Eqs. (6.4.6), with the mentioned replacements, into Equation (6.4.17) yields

$$\rho \approx e n_V \exp\left\{\frac{\epsilon_V - e\phi - \mu'}{T}\right\} - e n_A \equiv e n_A \left[ \left( \frac{n_V}{n_A} \exp\left\{\frac{\epsilon_V - \mu'}{T}\right\} \right) \exp\left\{-\frac{e\phi}{T}\right\} - 1 \right]. \quad (6.4.19)$$

The  $x$ -independent electrochemical potential (a.k.a. Fermi level)  $\mu'$  in this relation should be equal to the value of the chemical potential  $\mu(x \rightarrow \infty)$  in the semiconductor's bulk, given by the last of Eqs. (6.4.15), which turns the expression in the parentheses into 1. With these substitutions, Equation (6.4.16) becomes

$$\frac{d^2\phi}{dx^2} = -\frac{e n_A}{\kappa\epsilon_0} \left[ \exp\left\{-\frac{e\phi}{T}\right\} - 1 \right], \quad \text{for } \epsilon_V - e\phi(x) < \mu' < \epsilon_C - e\phi(x). \quad (6.4.20)$$

This nonlinear differential equation may be solved analytically, but in order to avoid a distraction by this (rather bulky) solution, let me first consider the case when the electrostatic potential is sufficiently small – either because the external field is small, or because we focus on the distances sufficiently far from the surface – see Figure 6.4.3 again. In this case, in the Taylor expansion of the exponent in Equation (6.4.20), with respect to small  $\phi$ , we may keep only two leading terms, turning it into a linear equation:

$$\frac{d^2\phi}{dx^2} = -\frac{e^2 n_A}{\kappa\epsilon_0 T} \phi, \quad \text{i.e. } \frac{d^2\phi}{dx^2} = \frac{\phi}{\lambda_D^2}, \quad \text{where } \lambda_D \equiv \left( \frac{\kappa\epsilon_0 T}{e^2 n_A} \right)^{1/2}, \quad (6.4.21)$$

with the well-known exponential solution, satisfying also the boundary condition  $\phi \rightarrow 0$  at  $x \rightarrow \infty$ :

$$\phi = C \exp\left\{-\frac{x}{\lambda_D}\right\}, \quad \text{at } e|\phi| \ll T. \quad (6.4.22)$$

The constant  $\lambda_D$  given by the last of Eqs. (6.4.21) is called the *Debye screening length*. It may be rather substantial; for example, at  $T_K = 300$  K, even for the relatively high doping,  $n_A \approx 10^{18} \text{ cm}^{-3}$  typical for modern silicon ( $\kappa \approx 12$ ) integrated circuits, it is close to 4 nm – still much larger than the crystal lattice constant  $a \sim 0.3$  nm, so that the above analysis is indeed quantitatively valid. Note also that  $\lambda_D$  does not depend on the charge's sign; hence it should be no large surprise that repeating our analysis for an  $n$ -doped semiconductor, we may find out that Eqs. (6.4.21)-(6.4.22) are valid for that case as well, with the only replacement  $n_A \rightarrow n_D$ .

If the applied field  $E$  is weak, Equation (6.4.22) is valid in the whole sample, and the constant  $C$  in it may be readily calculated using the boundary condition (6.4.18), giving

$$|\phi|_{x=0} \equiv C = \lambda_D \mathcal{E} \equiv \left(\frac{\kappa \varepsilon_0 T}{e^2 n_A}\right)^{1/2} \mathcal{E}. \quad (6.4.23)$$

This formula allows us to express the condition of validity of the linear approximation leading to Equation (6.4.22),  $e|\phi| \ll T$ , in terms of the applied field:

$$|\mathcal{E}| \ll \mathcal{E}_{max}, \quad \text{with } \mathcal{E}_{max} \equiv \frac{T}{e \lambda_D} \equiv \left(\frac{T n_A}{\kappa \varepsilon_0}\right)^{1/2}; \quad (6.4.24)$$

in the above example,  $\mathcal{E}_{max} \sim 60$  kV/cm. On the lab scale, such field is not low at all (it is twice higher than the threshold of electric breakdown in the air at ambient conditions), but may be sustained by many solid-state materials that are much less prone to the breakdown.<sup>47</sup> This is why we should be interested in what happens if the applied field is higher than this value.

$$\lambda_{ef}(0) \sim \lambda_{TF} \equiv \left[\frac{\kappa \varepsilon_0}{e^2 g_3(\varepsilon_F)}\right]^{1/2}. \quad (6.4.25)$$

The effects taking place at the opposite polarity of the field,  $\mathcal{E} > 0$ , are much more interesting – and more useful for applications. Indeed, in this case, the band bending down leads to an exponential decrease of  $\rho(x)$  as soon as the valence band edge  $\varepsilon_V - e\phi(x)$  drops down by just a few  $T$  below its unperturbed value  $\varepsilon_V$ . If the applied field is large enough,  $E > E_{max}$  (as it is in the situation shown in Figure 6.4.3d), it forms, on the left of such point  $x_0$  the so-called *depletion layer*, of a certain width  $w$ . Within this layer, not only the electron density  $n$ , but the hole density  $p$  as well, are negligible, so that the only substantial contribution to the charge density  $\rho$  is given by the fully ionized acceptors:  $\rho \approx -en_- \approx -en_A$ , and Equation (6.4.20) becomes very simple:

$$\frac{d^2 \phi}{dx^2} = \frac{en_A}{\kappa \varepsilon_0} = \text{const}, \quad \text{for } x_0 - w < x < x_0. \quad (6.4.26)$$

Let us use this equation to calculate the largest possible width  $w$  of the depletion layer, and the critical value,  $\mathcal{E}_c$ , of the applied field necessary for this. (By definition, at  $\mathcal{E} = \mathcal{E}_c$ , the left boundary of the layer, where  $\varepsilon_V - e\phi(x) = \varepsilon_C$ , i.e.  $e\phi(x) = \varepsilon_V - \varepsilon_A \equiv \Delta$ , just touches the semiconductor surface:  $x_0 - w = 0$ , i.e.  $x_0 = w$ . (Figure 6.4.3c shows the case when  $\mathcal{E}$  is slightly larger than  $\mathcal{E}_c$ .) For this, Equation (6.4.26) has to be solved with the following boundary conditions:

$$\phi(0) = \frac{\Delta}{e}, \quad \frac{d\phi}{dx}(0) = -\mathcal{E}_c, \quad \phi(w) = 0, \quad \frac{d\phi}{dx}(w) = 0. \quad (6.4.27)$$

Note that the first of these conditions is strictly valid only if  $T \ll \Delta$ , i.e. at the assumption we have made from the very beginning, while the last two conditions are asymptotically correct only if  $\lambda_D \ll w$  – the assumption we should not forget to check after the solution.

After all the undergraduate experience with projective motion problems, the reader certainly knows by heart that the solution of Equation (6.4.26) is a quadratic parabola, so that let me immediately write its final form satisfying the boundary conditions (6.4.27):

$$\phi(x) = \frac{en_A}{\kappa \varepsilon_0} \frac{(w-x)^2}{2}, \quad \text{with } w = \left(\frac{2\kappa \varepsilon_0 \Delta}{e^2 n_A}\right)^{1/2}, \quad \text{at } \mathcal{E}_c = \frac{2\Delta}{e \varepsilon_0 w}. \quad (6.4.28)$$

Comparing the result for  $w$  with Equation (6.4.21), we see that if our basic condition  $T \ll \Delta$  is fulfilled, then  $\lambda D \ll w$ , confirming the qualitative validity of the whole solution (6.4.28). For the same particular parameters as in the example before ( $n_A \approx 10^{18} \text{ cm}^{-3}$ ,  $\kappa \approx 10$ ), and  $\Delta \approx 1 \text{ eV}$ , Eqs. (6.4.28) give  $w \approx 40 \text{ nm}$  and  $\mathcal{E}_c \approx 600 \text{ kV/cm}$  – still a practicable field. (As Figure 6.4.3c shows, to create it, we need a gate voltage only slightly larger than  $\Delta/e$ , i.e. close to 1 V for typical semiconductors.)

Figure 6.4.3c also shows that if the applied field exceeds this critical value, near the surface of the semiconductor the conduction band edge drops below the Fermi level. This is the so-called *inversion layer*, in which electrons with energies below  $\mu'$  form a highly conductive degenerate Fermi gas. However, typical rates of electron tunneling from the bulk through the depletion layer are very low, so that after the inversion layer has been created (say, by the gate voltage application), it may be only populated from another source – hence the hatched blue points in Figure 6.4.3c. This is exactly the fact used in the workhorse device of semiconductor integrated circuits – the *field-effect transistor* (FET) – see Figure 6.4.4.

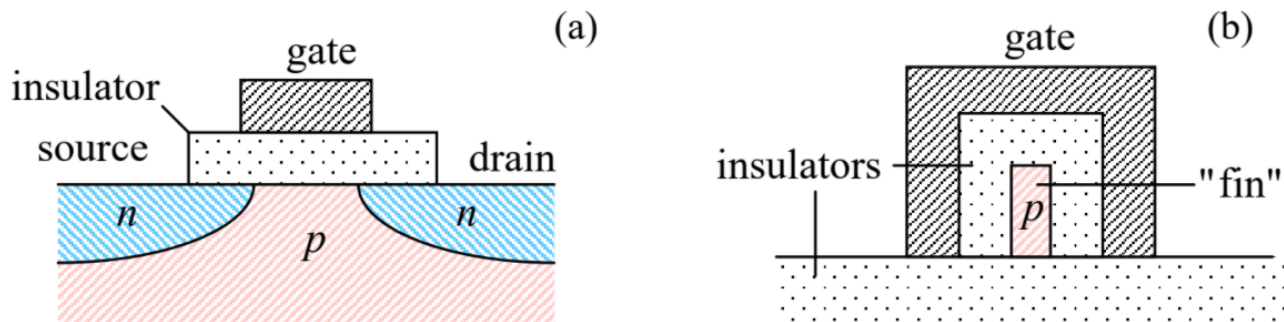


Figure 6.4.4: Two main species of the  $n$ -FET: (a) the *bulk FET*, and (b) the *FinFET*. While on panel (a), the current flow from the source to the drain is parallel to the plane of the drawing, on panel (b) it is normal to the plane, with the  $n$ -doped source and drain contacting the thin “fin” from two sides off this plane.

In the “bulk” variety of this structure (Figure 6.4.4a), a gate electrode overlaps a gap between two similar highly- $n$ -doped regions near the surface, called *source* and *drain*, formed by  $n$ -doping inside a  $p$  doped semiconductor. It is more or less obvious (and will be shown in a moment) that in the absence of gate voltage, the electrons cannot pass through the  $p$ -doped region, so that virtually no current flows between the source and the drain, even if a modest voltage is applied between these electrodes. However, if the gate voltage is positive and large enough to induce the electric field  $\mathcal{E} > \mathcal{E}_c$  at the surface of the  $p$ -doped semiconductor, it creates the inversion layer as shown in Figure 6.4.3c, and the electron current between the source and drain electrodes may readily flow through this *surface channel*. (Very unfortunately, in this course I would not have time/space for a detailed analysis of transport properties of this keystone electron device, and have to refer the reader to special literature.<sup>49</sup>)

Figure 6.4.4a makes it obvious that another major (and virtually unavoidable) structure of semiconductor integrated circuits is the famous  $p - n$  *junction* – an interface between  $p$ - and  $n$ -doped regions. Let us analyze its simple model, in which the interface is in the plane  $x = 0$ , and the doping profiles  $n_D(x)$  and  $n_A(x)$  are step-like, making an abrupt jump at the interface:

$$n_A(x) = \begin{cases} n_A = \text{const} & \text{at } x < 0, \\ 0, & \text{at } x > 0, \end{cases} \quad n_D(x) = \begin{cases} 0 & \text{at } x < 0, \\ n_D = \text{const} & \text{at } x > 0. \end{cases} \quad (6.4.29)$$

(This model is very reasonable for modern integrated circuits, where the doping is performed by *implantation*, using high-energy ion beams.)

To start with, let us assume that no voltage is applied between the  $p$ - and  $n$ -regions, so that the system may be in thermodynamic equilibrium. In the equilibrium, the Fermi level  $\mu'$  should be flat through the structure, and at  $x \rightarrow -\infty$  and  $x \rightarrow +\infty$ , where  $\phi \rightarrow 0$ , the level structure has to approach the positions shown, respectively, on panels (a) and (b) of Figure 6.4.2. In addition, the distribution of the electric potential  $\phi(x)$ , shifting the level structure vertically by  $-e\phi(x)$ , has to be continuous to avoid unphysical infinite electric fields. With that, we inevitably arrive at the band-edge diagram that is (schematically) shown in Figure 6.4.5.



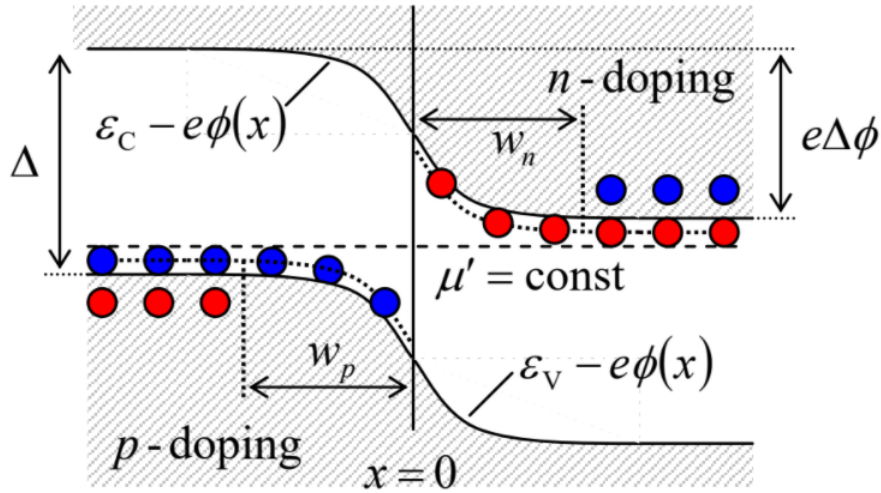


Figure 6.4.5: The band-edge diagram of a  $p-n$  junction in thermodynamic equilibrium ( $T = \text{const}$ ,  $\mu' = \text{const}$ ). The notation is the same as in Figs. 6.4.2 and 6.4.3.

The diagram shows that the contact of differently doped semiconductors gives rise to a built-in electric potential difference  $\Delta\phi$ , equal to the difference of their values of  $\mu$  in the absence of the contact – see Eqs. (6.4.13) and (6.4.15):

$$e\Delta\phi \equiv e\phi(+\infty) - e\phi(-\infty) = \mu_n - \mu_p = \Delta - T \ln \frac{n_C n_V}{n_D n_A}, \quad (6.4.30)$$

which is usually just slightly smaller than the bandgap.<sup>50</sup> (Qualitatively, this is the same contact potential difference that was discussed, for the case of metals, in Sec. 3 – see Figure 6.3.1.) The arising internal electrostatic field  $\mathcal{E} = -d\phi/dx$  induces, in both semiconductors, depletion layers similar to that induced by an external field (Figure 6.4.3c). Their widths  $w_p$  and  $w_n$  may also be calculated similarly, by solving the following boundary problem of electrostatics, mostly similar to that given by Eqs. (6.4.26)–(6.4.27):

$$\frac{d^2\phi}{dx^2} = \frac{e}{\kappa\epsilon_0} \times \begin{cases} n_A, & \text{for } -w_p < x < 0, \\ (-n_D), & \text{for } 0 < x < +w_n, \end{cases} \quad (6.4.31)$$

$$\phi(w_n) = \phi(-w_p) + \Delta\phi, \quad \frac{d\phi}{dx}(w_n) = \frac{d\phi}{dx}(-w_p) = 0, \quad \phi(-0) = \phi(+0), \quad \frac{d\phi}{dx}(-0) = \frac{d\phi}{dx}(+0), \quad (6.4.32)$$

also exact only in the limit  $\tau \ll \Delta$ ,  $n_i \ll n_D, n_A$ . Its (easy) solution gives the result similar to Equation (6.4.28):

$$\phi = \text{const} + \begin{cases} en_A(w_p + x)^2 / 2\kappa\epsilon_0, & \text{for } -w_p < x < 0, \\ \Delta\phi - en_D(w_n - x)^2 / 2\kappa\epsilon_0, & \text{for } 0 < x < +w_n, \end{cases} \quad (6.4.33)$$

with expressions for  $w_p$  and  $w_n$  giving the following formula for the full depletion layer width:

$$w \equiv w_p + w_n = \left( \frac{2\kappa\epsilon_0\Delta\phi}{en_{ef}} \right)^{1/2}, \quad \text{with } n_{ef} \equiv \frac{n_A n_D}{n_A + n_D}, \text{ i.e. } \frac{1}{n_{ef}} = \frac{1}{n_A} + \frac{1}{n_D}. \quad (6.4.34)$$

This expression is similar to that given by Equation (6.4.28), so that for typical highly doped semiconductors ( $n_{ef} \sim 10^{18} \text{ cm}^{-3}$ ) it gives for  $w$  a similar estimate of a few tens nm.<sup>51</sup> Returning to Figure 6.4.4a, we see that this scale imposes an essential limit on the reduction of bulk FETs (whose scaling down is at the heart of the well-known *Moore's law*),<sup>52</sup> explaining why such high doping is necessary. In the early 2010s, the problems with implementing even higher doping, plus issues with dissipated power management, have motivated the transition of advanced silicon integrated circuit technology from the bulk FETs to the *FinFET* (also called “double-gate”, or “tri-gate”, or “wrap-around-gate”) variety of these devices, schematically shown in Figure 6.4.4b, despite their essentially 3D structure and hence a more complex fabrication technology. In the FinFETs, the role of  $p-n$  junctions is reduced, but these structures remain an important feature of semiconductor integrated circuits.

Now let us have a look at the  $p-n$  junction in equilibrium from the point of view of Equation (6.3.19). In the simple model we are considering now (in particular, at  $T \ll \Delta$ ), this equation is applicable separately to the electron and hole subsystems, because in this model the gases of these charge carriers are classical in all parts of the system, and the *generation-recombination* processes<sup>53</sup>

coupling these subsystems have relatively small rates – see below. Hence, for the electron subsystem, we may rewrite Equation (6.3.19) as

$$j_n = n\mu_m q \mathcal{E} - D_n \frac{\partial n}{\partial x}, \quad (6.4.35)$$

where  $q = -e$ . Let us discuss how each term of the right-hand of this equality depends on the system's parameters. Because of the  $n$ -doping at  $x > 0$ , there are many more electrons in this part of the system. According to the Boltzmann distribution (6.4.6), some number of them,

$$n_{>} \propto \exp\left\{-\frac{e\Delta\phi}{T}\right\}, \quad (6.4.36)$$

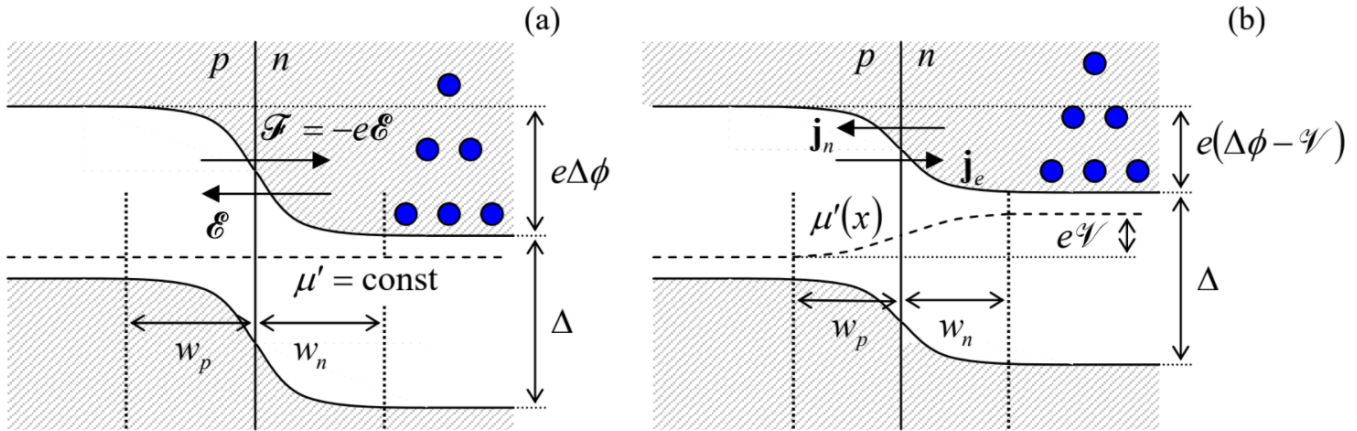


Figure 6.4.6: Electrons in the conduction band of a  $p-n$  junction at: (a)  $\mathcal{V} = 0$ , and (b)  $\mathcal{V} > 0$ . For clarity, other charges (of the holes and all ionized dopant atoms) are not shown.

$$e\Delta\phi \rightarrow e\Delta\phi + \Delta\mu' \equiv e\Delta\phi + q\mathcal{V} \equiv e(\Delta\phi - \mathcal{V}). \quad (6.4.37)$$

This change results in an exponential change of the number of electrons able to diffuse into the  $p$ -side of the junction – cf. Equation (6.4.36):

$$n_{>}(\mathcal{V}) \approx n_{>}(0) \exp\left\{\frac{e\mathcal{V}}{T}\right\}, \quad (6.4.38)$$

and hence in a proportional change of the diffusion flow  $j_n$  of electrons from the  $n$ -side to the  $p$ -side of the system, i.e. of the oppositely directed density of the electron current  $j_e = -ej_n$  – see Figure 6.4.6b

On the other hand, the drift counter-flow of electrons is not altered too much by the applied voltage: though it does change the electrostatic field  $\mathcal{E} = -\nabla\phi$  inside the depletion layer, and also the depletion layer width,<sup>57</sup> these changes are incremental, not exponential. As the result, the net density of the current carried by electrons may be approximately expressed as

$$j_e(\mathcal{V}) = j_{diffusion} - j_{drift} \approx j_e(0) \exp\left\{\frac{e\mathcal{V}}{T}\right\} - \text{const.} \quad (6.4.39)$$

As was discussed above, at  $\mathcal{V} = 0$ , the net current has to vanish, so that the constant in Equation (6.4.39) has to equal  $j_e(0)$ , and we may rewrite this equality as

$$j_e(\mathcal{V}) = j_e(0) \left( \exp\left\{\frac{e\mathcal{V}}{T}\right\} - 1 \right). \quad (6.4.40)$$

$$j(\mathcal{V}) \equiv j_e(\mathcal{V}) + j_h(\mathcal{V}) = j(0) \left( \exp\left\{\frac{e\mathcal{V}}{T}\right\} - 1 \right), \text{ with } j(0) \equiv j_e(0) + j_h(0), \quad (6.4.41)$$

describing the main  $p-n$  junction's property as an *electric diode* – a two-terminal device passing the current more “readily” in one direction (from the  $p$ - to the  $n$ -terminal) than in the opposite one.<sup>59</sup> Besides numerous practical applications in electrical and electronic engineering, such diodes have very interesting statistical properties, in particular performing very non-trivial



transformations of the spectra of deterministic and random signals. Very unfortunately, I would not have time for their discussion and have to refer the interested reader to the special literature.<sup>60</sup>

Still, before proceeding to our next (and last!) topic, let me give for the reader reference, without proof, the expression for the scaling factor  $j(0)$  in Equation (6.4.41), which follows from a simple, but broadly used model of the recombination process:

$$j(0) = en_i^2 \left( \frac{D_e}{l_e n_A} + \frac{D_h}{l_h n_D} \right). \quad (6.4.42)$$

Here  $l_e$  and  $l_h$  are the characteristic lengths of diffusion of electrons and holes before their recombination, which may be expressed by Equation (5.6.8),  $l_e = (2D_e\tau_e)^{1/2}$  and  $l_h = (2D_h\tau_h)^{1/2}$ , with  $\tau_e$  and  $\tau_h$  being the characteristic times of recombination of the so-called *minority carriers* – of electrons in the  $p$ -doped part, and of holes in the  $n$ -doped part of the structure. Since the recombination is an inelastic process, its times are typically rather long – of the order of  $10^{-7}$  s, i.e. much longer than the typical times of elastic scattering of the same carriers, that define their diffusion coefficients – see Equation (6.3.16 – 6.3.18).

---

This page titled [6.4: Charge Carriers in Semiconductors - Statics and Kinetics](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Konstantin K. Likharev](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.