Lecture 10: pn junctions in equilibrium

Contents

1 Introduction

Electronic devices are formed from interfaces formed from different materials (metals, semiconductors, insulators). These can be devices with a single interface/junction (diode) or with 2 interfaces (transistors). The properties of the device depends on the properties developed due to the formation of the junction. For a metal and semiconductor there can be two types of junctions, with correspondingly different electrical behavior. We can have a Schottky junction, which acts as a rectifier or an Ohmic junction, which acts as a resistor. A *pn* junction is one formed between two semiconductors with different types of majority carriers (holes or electrons). The semiconductors can be made of the same material, *homojunction*, or different materials, *heterojunction*. It is also possible to form junctions between two semiconductors with the same majority carrier type but at different concentrations. Such junctions are called *iso type junctions*.

2 Contact potential

Consider a junction between p type and n type Si. The band diagrams of the two semiconductors, where apart, are shown in figure 2. When the two materials are brought together, the Fermi levels line up at equilibrium. There are excess electrons on the n-side and excess holes on the p-side. The electrons from the n side move to the p side and combine with the holes, while similarly, holes move from p to n side and combine with the electrons. Thus, there is a region around the interface where electrons and holes are annihilated. This is called the **depletion region** or the **space charge layer**. Because of the movement of majority carriers a net positive charge develops on the n-side (due to the positively charged donor ions) and a negative charge develops on the p-side (due to negatively charged acceptor ions). Thus, an electric field is formed going from the n-side to the p-side and there is a built-in potential, from the p to the n side. This potential is called the *junction potential* or *contact potential*. The presence of the electric field leads to band bending, with the bands bending up, moving from n-type to p-type semiconductor. The band diagram when the pn junction is formed (at equilibrium) is shown in figure 3. The space charge layer (SCL) extends on both the n and p side and the contact potential (eV_0) is equal to the difference between the workfunctions of the p and n side ($\phi_p - \phi_n$).

Figure 2: Energy band diagram of p and n type Si before the junction is formed. The location of the Fermi level is different for the two semiconductors.

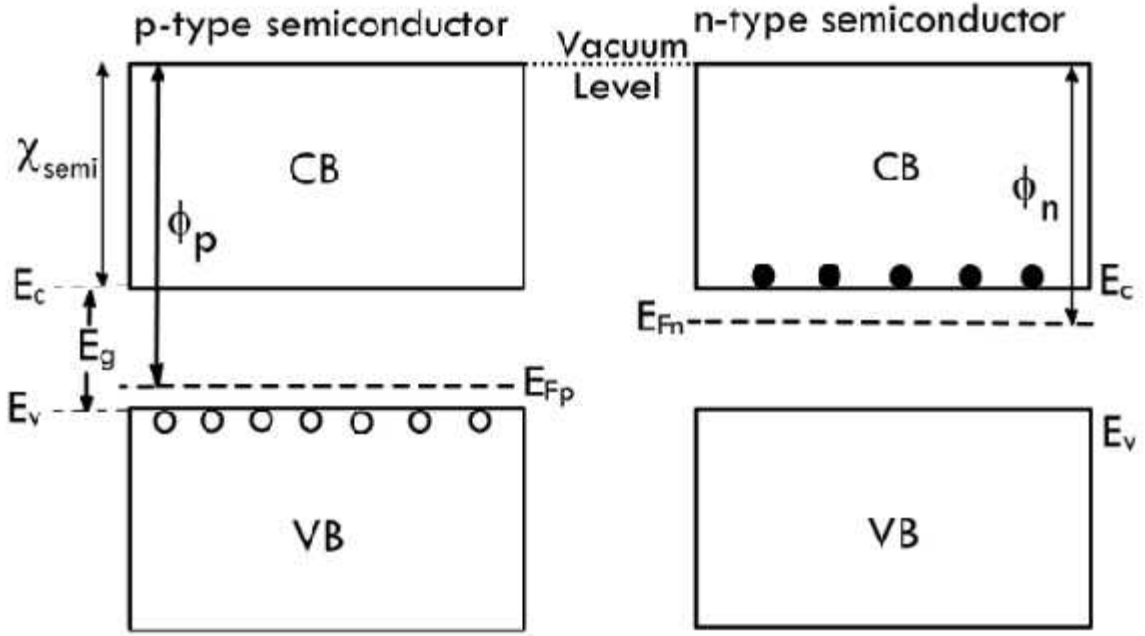
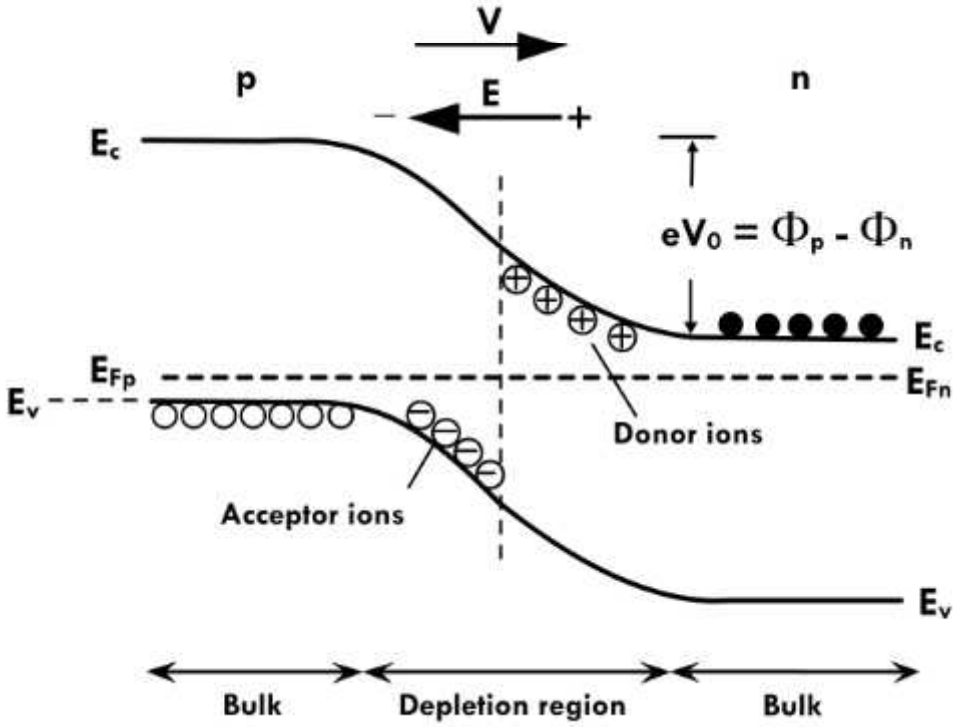


Figure 3: Energy band diagram of the pn junction in equilibrium. There is a positive charge on the n side and negative charge on the p side. Bands bend up moving from n to p.



3 Depletion region

Consider a pn junction at equilibrium, as shown in figure 3. The total depletion region width is w_o, with widths w_p and w_n on the p and n side respectively.

$$w_0 = w_p + w_n \tag{1}$$

The depletion region forms because holes from the p side and electrons from the n side diffuse and get annihilated. Let N_A and N_D be the doping concentrations in the p and n regions. The overall pn junction should be electrically neutral since it is in equilibrium and isolated. So the net positive charge on the n side (due to the positive donors) should be balanced by the net negative charge on the p side (due to the negative acceptors).

Let the cross-sectional area of the junction be A. Hence, for charge balance

Charge density $n = q/V$

→ $q_{wp} = q_{wn}$

$$N_A A w_p = N_D A w_n \tag{2}$$

Rearranging this gives the ratio of the depletion widths on the p and n side in terms of the dopant concentration.

$$\boxed{\frac{w_p}{w_n} = \frac{N_D}{N_A}} \tag{3}$$

Thus, the depletion width is *inversely proportional* to the dopant concentration. It is thinner in the region with higher dopant concentration. So, if $N_A > N_D$ then $w_p < w_n$. If one dopant concentration is much higher, i.e. $N_A \gg N_D$ then $w_p \ll w_n$, or the depletion region lies *almost entirely* in the n side. This kind of a junction is called a p^+n junction. We can also have a pn^+ junction, with a heavily doped n side. A metal-semiconductor Schottky junction can be considered an extreme case of a heavily doped junction where the depletion region lies entirely in the semiconductor side. This is because the metal has a much higher electron concentration than the semiconductor.

4 Junction parameters

For a *pn* junction in equilibrium the important junction parameters are the contact potential and the total depletion width. This is related to the variation of charge density and electric field within the junction. The parameters for a *pn* junction are summarized in figure 4.

4.1 Concentration gradient

To calculate the various junction parameters it is important to define some concentration terms. On the p-side, let N_A be the concentration of the acceptors. If n_i is the intrinsic carrier concentration, then the electron and hole concentrations in the p side are defined as

$$\boxed{\begin{aligned} p_{p0} &= N_A \\ n_{p0} &= \frac{n_i^2}{N_A} \end{aligned}} \tag{4}$$

Similarly, on the n-side, with donor concentration of N_D, the electron and hole concentrations are

$$\boxed{\begin{aligned} n_{n0} &= N_D \\ p_{n0} &= \frac{n_i^2}{N_D} \end{aligned}} \tag{5}$$

The change in concentration of electrons and holes as we move across the device is shown in figure 4(b).

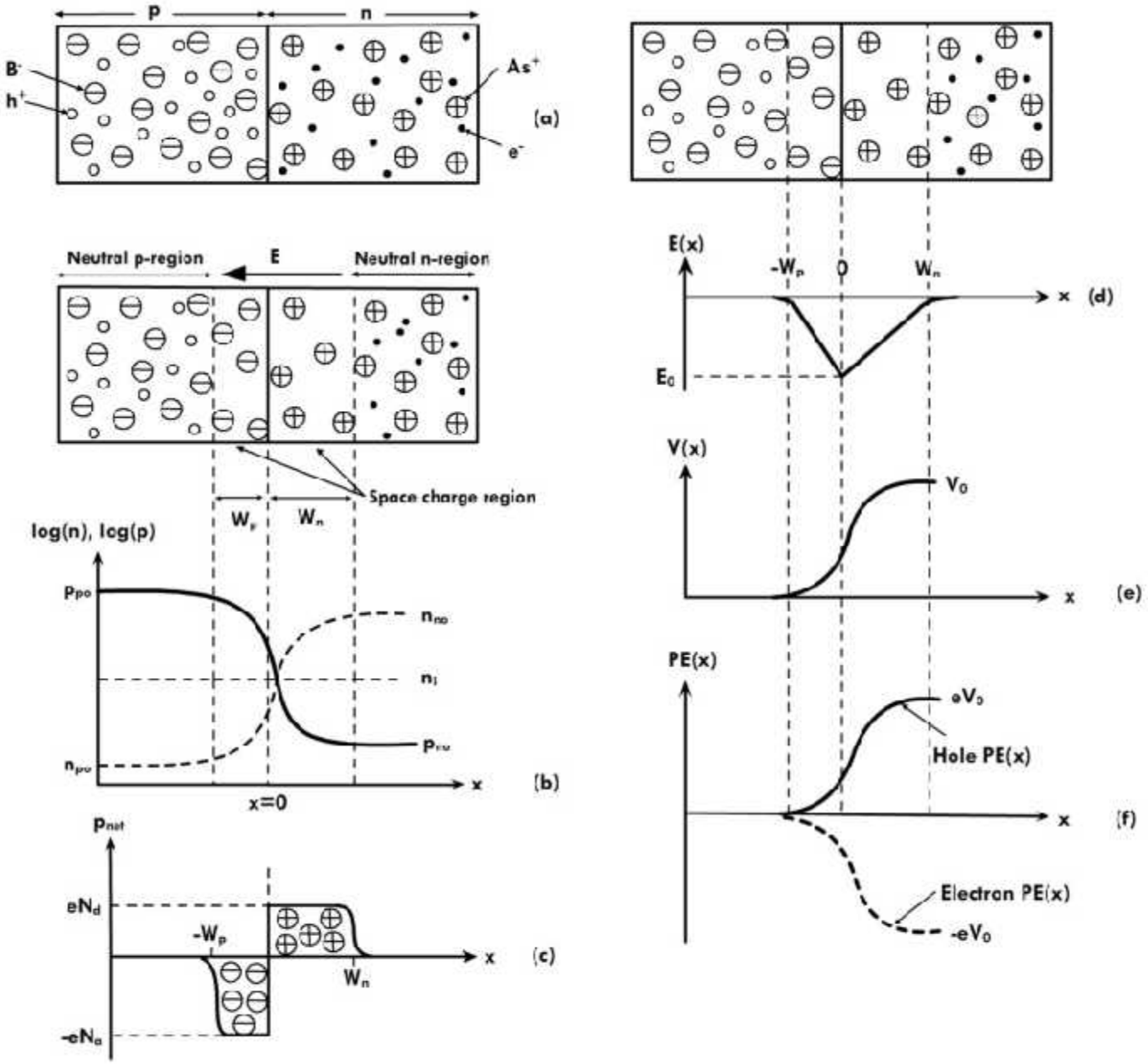


Figure 4: Properties of an abrupt pn junction. (a) Schematic of the pn junction, showing the depletion layer. (b) Carrier concentration across the junction. (c) Charge density variation. (d) Electric eld,(e) Potential, and (f) Potential energy in the depletion region.

4.2 Electric field

The depletion width typically extends on both the n and p side with relative widths inversely proportional to the dopant concentration, see equation 3. To calculate the total depletion width we need to consider first the distribution of electric field within the depletion region. This is related to the net charge density, given by **Gauss law**,

Divergence form of Gauss Law → $$\nabla . E = \frac{\rho_{net}}{\varepsilon} \quad (6)$$

This is a general 3D equation that relates the electric field, E, to the total charge density, ρ_{net}, and ε is the permittivity of the semiconductor ($\varepsilon = \varepsilon_0 \varepsilon_r$, where ε_0 is the permittivity of free space and ε_r is the relative permittivity of the semiconductor). For an one-dimensional interface this equation reduces to

$$\frac{dE}{dx} = \frac{\rho_{net}}{\varepsilon} \quad (7)$$

Thus the field can be obtained by integrating the total charge density in the pn junction.

$$E(x) = \frac{1}{\epsilon} \int \rho_{net}(x) dx \quad (8)$$

Beyond the depletion region charges are balanced so that the integration has to be performed only in the depletion region. There are different models for considering the charge distribution in the depletion region, but the simplest is the *abrupt charge distribution*.

$$\begin{aligned} \rho_{net} &= -eN_A \; for \; -w_p < x < 0 \\ \rho_{net} &= eN_D \; for \; 0 < x < w_n \end{aligned} \tag{9}$$

Substituting this in equation 8 and integrating gives the electric field distribution within the junction. The boundary condition is that the field must go to zero at the boundaries i.e. $x = -x_p$ and $x = x_n$.

$$E(x) = \frac{1}{\epsilon}\int_{-w_p}^{x} \rho_{net}(x)dx \rightarrow E(x) = -\frac{eN_A}{\varepsilon}(x + w_p) \; for \; -w_p < x < 0$$

$$E(x) = \frac{1}{\epsilon}\int_{x}^{w_n} \rho_{net}(x)dx \rightarrow E(x) = \frac{eN_D}{\varepsilon}(x - w_n) \; for \; -0 < x < w_n \tag{10}$$

This is plotted in figure 4(d). At $x = 0$ the electric fields must match and hence putting this in equation 10 gives

→ Maximum E-field $$E_0 = \frac{-eN_A w_p}{\varepsilon} = \frac{-eN_D w_n}{\varepsilon} \tag{11}$$

This is another way of deriving the charge balance equation, seen earlier in equation 3.

4.3 Electrical potential

The electric field is related to the potential by

$$E(x) = -\frac{dV}{dx} \tag{12}$$

The expression for electric field from equation 10 can be substituted and integrated for the two regions. The boundary conditions are set such that $V = 0$ at $x = -w_p$ and $V = V_0$ at $x = x_n$, where V_0 is the total potential i.e. the *contact* or *built-in potential.*

$$\begin{aligned} V(x) &= \frac{eN_A}{2\varepsilon}(x + w_p)^2 \; for \; -w_p < x < 0 \\ V(x) &= V_0 - \frac{eN_D}{2\varepsilon}(x - w_n)^2 \; for \; 0 < x < w_n \end{aligned} \tag{13}$$

Given that the potential function should be continuous the two expressions in equation 13 should be equal at $x = 0$. This gives

$$\frac{eN_A}{2\varepsilon} w_p^2 = V_0 - \frac{eN_D}{2\varepsilon} w_n^2 \tag{14}$$

This can be rearranged as

$$V_0 = \frac{e}{2\varepsilon}(N_A w_p^2 + N_D w_n^2) \tag{15}$$

This can be rearranged as

$$V_0 = \frac{e}{2\varepsilon}\left(N_A w_p^2 + N_D w_n^2\right) \tag{15}$$

Using equation 3 this can be rewritten in terms of w_p only as

$$V_0 = \frac{e}{2\varepsilon} w_p^2 N_A \left(1 + \frac{N_A}{N_D}\right) \tag{16}$$

The total depletion width, w_0 is given by equation 1 and the ratio of w_p and w_n is given by equation 3. Using this w_p can be written in terms of w_0 as follows

$$\begin{aligned} w_p + w_n &= w_0 \\ w_p N_A &= w_n N_D \\ w_p &= \frac{w_0}{1 + \frac{N_A}{N_D}} \end{aligned} \tag{17}$$

Substituting this in equation 16

$$\boxed{V_0 = \frac{e}{2\varepsilon} w_0^2 \left(\frac{N_A N_D}{N_A + N_D}\right)} \tag{18}$$

Rearranging, w_0 can then be written in terms of V_0 as

$$w_0 = \sqrt{\frac{2\varepsilon V_0}{e}\left(\frac{N_A + N_D}{N_A N_D}\right)} \tag{19}$$

The distribution of the electrical potential is shown in figure 4(e). The potential energy is obtained by multiplying the potential by $-e$ and is plotted in figure 4(f). These plots are for the pn junction with abrupt interfaces and a constant charge distribution. If the charge distribution changes then the distribution of the field and potential also changes.

4.4 Junction potential

To derive the expression for the junction potential, we must realize that the potential prevents further motion of carriers once equilibrium is reached. Thus, the electron (or hole) concentrations on either side of the junction can be written in terms of the contact potential (V_0) as

$$\frac{n_{p0}}{n_{n0}} = \frac{p_{n0}}{p_{p0}} = \exp\left(-\frac{eV_0}{k_B T}\right) \tag{20}$$

$$\begin{aligned} n_{n0} &= N_D \\ n_{p0} &= \frac{n_i^2}{N_A} \end{aligned} \tag{21}$$

Equation 21 can be substituted in 20 and rearranged to give the contact potential

$$\boxed{V_0 = \frac{k_B T}{e} \ln\left(\frac{N_A N_D}{n_i^2}\right)} \tag{22}$$

The contact potential is directly related to the dopant concentrations. Higher the value of N_A and N_D greater is the contact potential. Once the contact potential is calculated this can be used to calculate the total depletion width and the individual widths in the p and n side.

5 Si pn junction at equilibrium (Example)

Consider a Si based pn junction with N_A of 2×10^{16} cm^{-3} and N_D of 10^{16} cm^{-3}. The widths of the depletion region are inversely proportional to the dopant concentrations, so that using equation 3, the ratio of w_p to w_n is

0.5. The depletion region is thicker on the n side than the p side by a factor of 2. Since the material is Si, at room temperature the intrinsic carrier concentration is equal to $10^{10}\ cm^{-3}$. So that the contact potential can be calculated using equation 22 and gives a value of 0.73 V. The absolute value of the depletion width, w_0, can be calculated by using 19. For Si, the relative permittivity, ϵ_r, is 11.9. This gives a depletion width of 381 nm. This depletion width extends to both the p and n sides. Using the ratios of the individual depletion widths, the values of w_p and w_n are 127 nm and 254 nm.The distribution of electric field and potential within the junction can be calculated using equations 10 and 13.

Consider the effect of the change in donor and acceptor concentrations to $2.5 \times 10^{16}\ cm^{-3}$ and $8 \times 10^{16}\ cm^{-3}$. The contact potential still remains the same but the ratio of the depletion width changes with the width on the n side now more than thrice the width on the p side. The total depletion width is now higher at 399 nm with w_p of 302 nm and w_n of 97 nm. The field and potential distributions also change. The plots for both these cases are summarized in figure 5. Thus, typical depletion widths are of the order of hundreds of nm. This would play a role in device miniaturization where typical device dimensions are of the order of tens of nm.

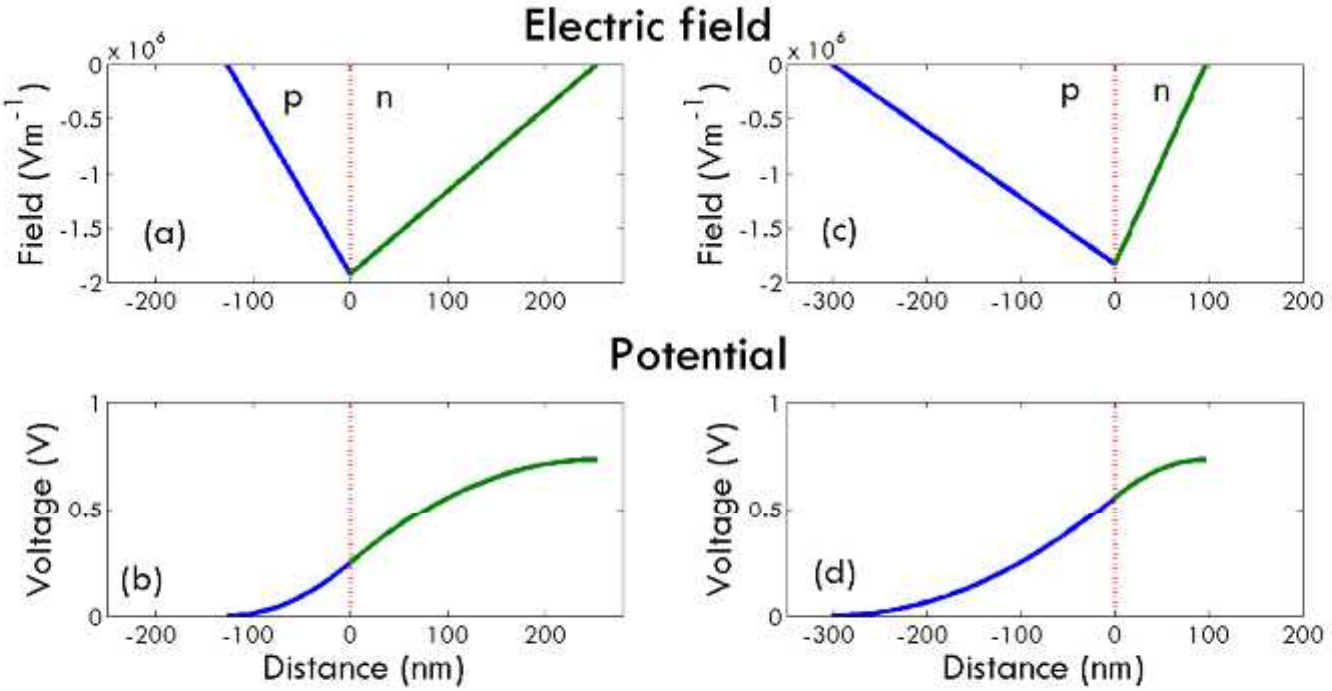


Figure 5: (a) - (d) Electric field and potential distribution within a Si pn junction depletion layer. The field is calculated using equation 10 and potential using equation 13. The plots are generated in MATLAB with carrier concentrations given in the text.The dotted lines represent the abrupt junction between the p and n sides.

6 Junction potential vs. Fermi level position

Consider the band diagram of the pn junction before the junction is formed, as shown in figure 2. For simplicity the acceptor concentration can be taken to be $10^{17}\ cm^{-3}$ and the donor concentration to be $10^{16}\ cm^{-3}$. Before the junction is formed, the Fermi levels in the extrinsic semiconductors can be calculated individually by considering the shift from the intrinsic Fermi level. This is given by

$$\begin{aligned} E_{Fn} - E_{Fi} &= k_B T \ln(\frac{N_D}{n_i}) = 0.36\ eV \\ E_{Fp} - E_{Fi} &= -k_B T \ln(\frac{N_A}{n_i}) = -0.37\ eV \end{aligned} \tag{23}$$

For the n type semiconductor the Fermi level is located above the intrinsic level while for the p type semiconductor it is located below the intrinsic level. When the junction is formed, as shown in the energy band diagram of 3, the Fermi levels line up. For this to happen, either the p Fermi level can be considered to have shifted up or the n level can be considered to have shifted down or both. The magnitude of the shift is given by the difference in the work functions, which is proportional the Fermi level positions. So the shift is equal to the difference in Fermi level positions and is given by

$$\text{Shift} = E_{Fn} - E_{Fp} = 0.36 - (-0.37) = 0.73\ eV \tag{24}$$

7 Contact potential for various semiconductors

The equation for the contact potential depends on the intrinsic carrier concentration, n_i, as given by equation 22.

$$V_0 = \frac{k_B T}{e} \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{25}$$

For different materials, the intrinsic carrier concentrations, n_i, are different since the band gaps (E_g) are different. These are related by

$$n_i = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2 k_B T}\right) \tag{26}$$

Thus, a higher value of n_i, due to smaller band gap, lower the contact potential, for the same dopant concentration. Contact potentials for 3 semiconductors, Ge, Si, and GaAs, with N_A of $10^{17}\ cm^{-3}$ and N_D of $10^{16}\ cm^{-3}$ are tabulated in table 1. With increasing band gap, the contact potential also increases.

Table 1: Contact potential for different semiconductors

Material	E_g	n_i (cm^{-3})	V_0 (V)
Ge	0.7	2.4×10^{13}	0.37
Si	1.1	10^{10}	0.78
GaAs	1.4	2.1×10^{6}	1.21