Lecture 20

The p-n Junction

The p-n junction is the interface between a p-doped and n-doped semiconductors (Figure 20.1a). The p-n junction is the fundamental building block of semiconductor electronic devices, due to its diode behavior. Similar to the metal-semiconductor interface we introduced in Lecture 19, the current of a p-n is very low under reverse bias ($V < 0$), while rapidly increasing under forward bias ($V > 0$), as shown in Figure 20.1b. However the mechanism is not exactly the same as in a metal-semiconductor junction. So what happens at the p-n junction interface? Intuitively speaking, when n-doped and p-doped semiconductors contact, the n-doped semiconductor has more electrons and the p-doped semiconductor has more holes. The imbalance of carriers cause a flow of electrons from the n-doped region to the p-doped region, and creates a charged region near the interface. In the following sections we'll introduce quantitative analysis of the band diagram and carrier transport of the p-n junction.

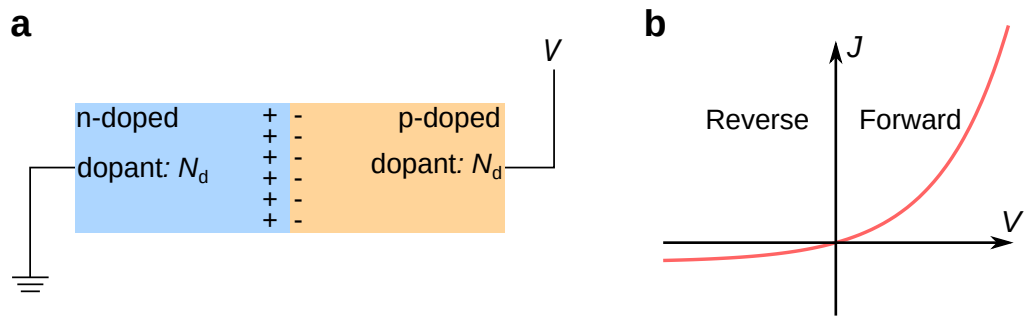

Figure 20.1: The p-n junction. (a) The device layout of a p-n junction, showing the charged interfacial layer. The voltage is applied on the p-doped side. (b) Current-bias behavior of a typical p-n junction.

20.1 Band alignment of p-n junction

Let's start with the band diagram of the p-n junction. Before contact, the Fermi level of the n-doped side is higher than the p-doped side. After contact, E_F is aligned. Since the p-n junction comes from the same semiconductor (Si) while the dopant concentration varies, the vacuum level E_{vac} is continuous at the semiconductor interface. As a result, band bending is observed for E_c, E_v and E_{vac}. An important quantity of the p-n junction is the mismatch between the Fermi levels in the different doped regions, the built-in potential ϕ_{bi}:

$$e\phi_{bi} = E_{F,n} - E_{F,p} = E_g - e\phi_p - e\phi_n \tag{20.1}$$

where $\phi_p = E_F - E_v$ and $\phi_n = E_c - E_F$ are the Fermi level offset in p-doped and n-doped regions, respectively. The electric potential ψ in the semiconductor is higher on the n-doped side and lower on the p-doped side, with a difference of ϕ_{bi}. The physical parameters of a p-n junction can be seen in Figure 20.2a.

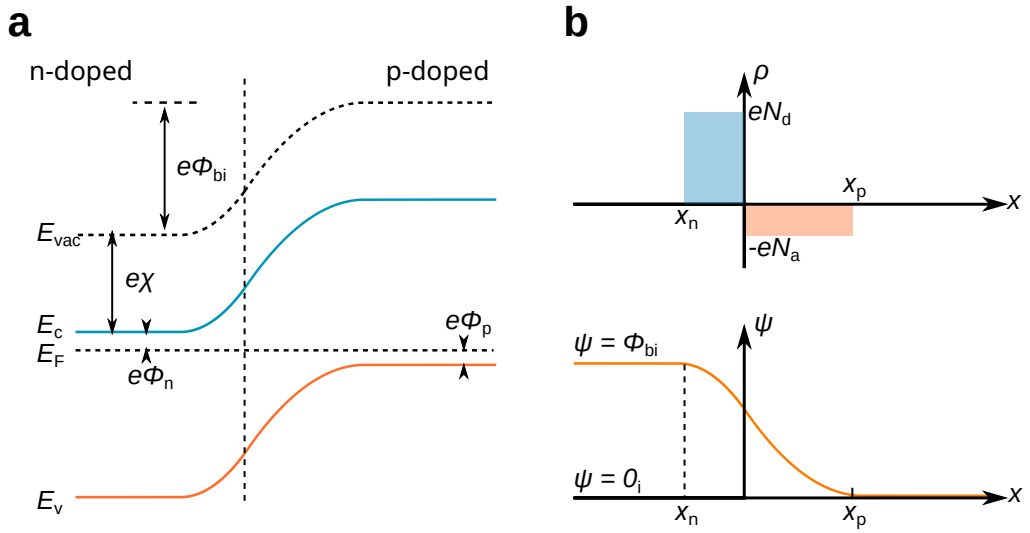

Figure 20.2: Band diagram of a p-n junction. (a) Band bending of a p-n junction, showing the total band offset as ϕ_{bi}, the built-in potential. (b) Charge density and potential profiles of a p-n junction under the abrupt junction approximation.

Before contact, in the n-doped region, the free electron density $n \approx N_d$. Using our knowledge of the charge distribution in semiconductors (Equation 19.3), we get:

$$\phi_n = \frac{k_B T}{e} \ln\left(\frac{N_c}{N_d}\right) \tag{20.2}$$

and similarly in the p-doped region:

$$\phi_p = \frac{k_B T}{e} \ln\left(\frac{N_v}{N_a}\right) \tag{20.3}$$

Combine the definition of ϕ_n and ϕ_p with Equation 20.1, we can express the built-in potential as:

$$\phi_{bi} = \frac{k_B T}{e} \ln(\frac{N_a N_d}{n_i^2}) \tag{20.4}$$

where n_i is the intrinsic carrier concentration of silicon.

Next, we will work out the potential profile across the junction. For simplicity we can use the abrupt junction approximation as we used to solve the M-S interface. Consider the Poisson equation in the junction:

$$\frac{d^2\psi}{dx^2} = -\frac{\rho}{\varepsilon_0 \varepsilon_r} \tag{20.5}$$

The abrupt junction approximation is that the dopants are completely ionized in the depletion region, that the charge density $\rho = eN_d$ for $x_n < x < 0$ in the n-doped region, while $\rho = -eN_a$ for $0 < x < x_p$ in the p-doped region (Figure 20.2b). The boundary conditions are:

- $\psi(x = x_n) = \phi_{bi}$, in the n-doped region
- $\psi(x = x_p) = 0$, in the p-doped region
- $\frac{d\psi}{dx}(x = x_n) = \frac{d\psi}{dx}(x = x_p) = 0$

Similar to the procedure we adapted for the M-S junction, the solutions of the potential profiles in the n- and p-doped regions, are:

- n-doped region ($x_n < x < 0$)

$$\psi(x) = \phi_{bi} - \frac{eN_d}{2\varepsilon_0 \varepsilon_r}(x - x_n)^2 \tag{20.6}$$

- p-doped region ($0 < x < x_p$)

$$\psi(x) = \frac{eN_a}{2\varepsilon_0 \varepsilon_r}(x_p - x)^2 \tag{20.7}$$

At the interface, the continuity conditions for both the ψ and $d\psi/dx$ are continuous (due to same dielectric constant) lead to:

$$\phi_{bi} = \frac{eN_d}{2\varepsilon_0 \varepsilon_r} x_n^2 + \frac{eN_a}{2\varepsilon_0 \varepsilon_r} x_p^2 \tag{20.8}$$

and

$$N_d x_n + N_a x_p = 0 \tag{20.9}$$

Solving the above sets of equations gives the width of the depletion region:

$$x_p - x_n = W = \sqrt{\frac{2\varepsilon_0 \varepsilon_r \phi_{bi}}{e}(\frac{1}{N_a} + \frac{1}{N_d})} \tag{20.10}$$

with a doping density of $N_a = N_d = 10^{18}$ cm^{-3}, the depletion width W is at the order of 50 nm. With increasing N_a and N_d, the depletion width decreases due to $\phi_{bi} \sim \ln(N_a N_d)$, which is slower than $N_a^{-1} + N_d^{-1}$.

20.2 Carrier transport at the p-n junction

Now let's study the current-bias behavior of the p-n junction. We assume the bias at the n-doped side is zero in all cases. Within the reverse bias regime, $V < 0$, the Fermi level of the p-doped side, E_{Fp} is elevated, and *vice versa*. The potential drop across the junction $\phi_{\mathrm{b}} = \phi_{\mathrm{bi}} - V$, and the depletion width becomes:

$$W = \sqrt{\frac{2\varepsilon_0\varepsilon_{\mathrm{r}}(\phi_{\mathrm{bi}} - V)}{e}(\frac{1}{N_{\mathrm{a}}} + \frac{1}{N_{\mathrm{d}}})} \tag{20.11}$$

The comparison between the band diagrams under reverse and forward biases, can be seen in Figure 20.3. Similar to the analysis in M-S junction, the change of W as a function of applied bias V, can be measured using the capacitance of the system, that $1/C^2 \propto (\phi_{\mathrm{bi}} - V)$. By extrapolating the curve of $1/C^2$ versus $-V$ to the x-axis, we get the built-in potential ϕ_{bi}.

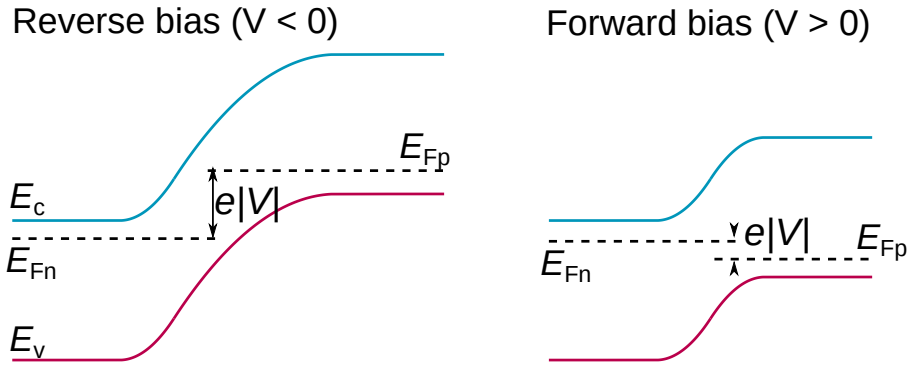

Figure 20.3: Band diagram of the p-n junction under reverse (left) and forward (right) bias. The depletion layer thickness increases with reverse bias while decrease with forward bias.

Now let's consider the carrier transport across the p-n junction. From the Drude model, the current in a homogeneous semiconductor proportional to the carrier density, therefore the current in the n-doped region is dominated by electrons, and p-doped region is dominated by holes. However within the depletion region, merely no free carriers exists as we know from the abrupt junction approximation. What happens near the boundary of the depletion region, is recombination of e^--h^+ pairs. when $V \neq 0$, the semiconductor is not at thermal equilibrium due to the carrier injection. For instance, when $V > 0$ (forward bias), more holes flow from the p-doped side to the n-doped region, creating region of excess minority carrier (hole) inside the n-doped side. Similarly, excess electrons also exist in the p-doped side. The minority carriers are actually conducting current in the p-n junction, which we'll see in the derivations that follow. To distinguish, we name the majority carriers n_{n} and p_{p}, as well as the minority carriers n_{p} and p_{n}.

In the p-doped region and far from the interface, the minority carrier (electron) density follows:

$$n_{p0} = N_{\mathrm{c}} \exp(-\frac{E_{\mathrm{g}} - e\phi_{\mathrm{p}}}{k_{\mathrm{B}}T}) = n_{\mathrm{n0}} \exp(-\frac{e\phi_{\mathrm{bi}}}{k_{\mathrm{B}}T}) \tag{20.12}$$

By applying a bias V, the minority carrier density at the boundary of the depletion region becomes: $n_{\mathrm{p}}(x_{\mathrm{p}}) = n_{\mathrm{p0}}\exp(\frac{eV}{k_{\mathrm{B}}T}) = n_{\mathrm{n0}}\exp\left[-\frac{e(\phi_{\mathrm{bi}}-V)}{k_{\mathrm{B}}T}\right]$ On the other hand, $p_{\mathrm{n}}(x_{\mathrm{n}}) = p_{\mathrm{n0}}\exp(\frac{eV}{k_{\mathrm{B}}T})$. Take minority carrier in the n-doped region for example, the concentration of p_{n} remains constant over time, and there is no photocarrier generation, the continuity equation is then:

$$u_{\mathrm{p}}k_{\mathrm{B}}T\frac{\mathrm{d}^2p_{\mathrm{n}}}{\mathrm{d}x^2} - \frac{p'_{\mathrm{n}}}{\tau_{\mathrm{p}}} = 0 \tag{20.13}$$

with boundary conditions:

$$\begin{aligned} p_{\mathrm{n}}(x_{\mathrm{n}}) &= p_{\mathrm{n0}}\exp(\frac{eV}{k_{\mathrm{B}}T}) \\ p_{\mathrm{n}}(-\infty) &= p_{\mathrm{n0}} \end{aligned} \tag{20.14}$$

Solve the differential equations gives:

$$p_{\mathrm{n}}(x) = p_{\mathrm{n0}} + \underbrace{p_{\mathrm{n0}}\left[\exp(\frac{eV}{k_{\mathrm{B}}T}) - 1\right]\exp(-\frac{x-x_{\mathrm{n}}}{\sqrt{u_{\mathrm{p}}k_{\mathrm{B}}T\tau_{\mathrm{p}}}})}_{p'_{\mathrm{n}}} \tag{20.15}$$

The dominance of minority carriers for the transport of a p-n junction, can be seen in Figure 20.4. Since outside the depletion region $x < x_{\mathrm{n}}$, the potential profile is flat, the current in

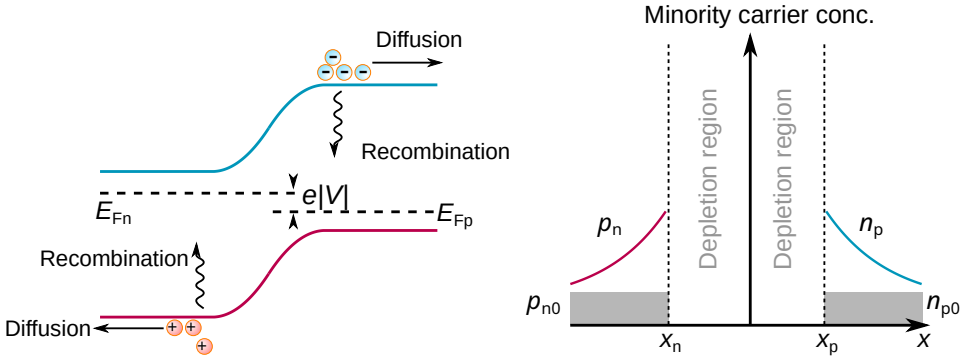

Figure 20.4: Dominance of minority carriers in p-n junction. Left: diffusion and recombination of minority carriers outside the depletion regions. Right:the concentration of minority carriers using the abrupt junction approximation.

the n-doped is solely due to the diffusion of excess minority carrier, that

$$J_{\mathrm{p}} = -eu_{\mathrm{p}}k_{\mathrm{B}}T\frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}x} = e\sqrt{\frac{u_{\mathrm{p}}k_{\mathrm{B}}T}{\tau_{\mathrm{p}}}}p_{\mathrm{n0}}\left[\exp(\frac{eV}{k_{\mathrm{B}}T}) - 1\right]\exp(-\frac{x-x_{\mathrm{n}}}{\sqrt{u_{\mathrm{p}}k_{\mathrm{B}}T\tau_{\mathrm{p}}}}) \tag{20.16}$$

Following the same procedure, the current density in the p-doped side, is:

$$J_{\mathrm{n}} = e\sqrt{\frac{u_{\mathrm{n}}k_{\mathrm{B}}T}{\tau_{\mathrm{n}}}}n_{\mathrm{p0}}\left[\exp(\frac{eV}{k_{\mathrm{B}}T}) - 1\right]\exp(-\frac{x_{\mathrm{p}}-x}{\sqrt{u_{\mathrm{n}}k_{\mathrm{B}}T\tau_{\mathrm{n}}}}) \tag{20.17}$$

The quantity $L_r = \sqrt{uk_BT\tau}$ is the characteristic length of combination in the doped semiconductors. Usually $L_r \gg W$, which means the recombination inside the depletion region is negligible, and we can therefore neglect the last exponential part in Equations 20.16 and 20.17. Therefore , we approximate the current density of both carriers as:

$$\begin{aligned} J_p &\approx e\sqrt{\frac{u_p k_B T}{\tau_p}} p_{n0}\left[\exp(\frac{eV}{k_B T}) - 1\right] \\ J_n &\approx e\sqrt{\frac{u_n k_B T}{\tau_n}} n_{p0}\left[\exp(\frac{eV}{k_B T}) - 1\right] \end{aligned} \tag{20.18}$$

and the total current density:

$$J_{tot} = e\left[\sqrt{\frac{u_p k_B T}{\tau_p}} p_{n0} + \sqrt{\frac{u_n k_B T}{\tau_n}} n_{p0}\right]\left[\exp(\frac{eV}{k_B T} - 1)\right] = J_0\left[\exp(\frac{eV}{k_B T} - 1)\right] \tag{20.19}$$

which has a very similar form with the current density in a Schottky M-S junction 19.13. However, we should not that in a Schottky junction, the majority of carrier dominates the current, while in a p-n junction it is the minority carrier. The threshold current J_0 in a Schottky junction is influenced by the Schottky barrier height, the difference between the electron affinity, while in a p-n junction, it is mainly the built-in potential (difference between Fermi level) that controls J_0. Another difference is that the flow of both carriers in a p-n junction makes it also suitable to separate photo-generated e-h pairs after illumination, the key process in a solar cell, while this cannot be achieved in a Schottky barrier due to the unipolar transport behavior.