Diodes

Carsten Wulff, *2021-07-08*, v0.1.0

***Abstract*—I explain how diodes work.**

I. Why

Diodes are a magical semiconductor device that mostly conducts current in one direction. This is a useful feature, and it is one way to convert from an AC voltage to a DC voltage using a https://en.wikipedia.org/wiki/Diode_bridge.

In integrated circuits we don't intentionally use them that much. They have a forward voltage of about 0.5 V, and for low voltage circuits they are not that useful. But there are a few instances where they are very useful.

All integrated circuits are plagued by electrostatic discharge (ESD), both during assembling printed circuit boards (PCB), and after, when people touch the PCB. The ESD events can push huge currents into our intregrated circuits – 2 kV ESD zap is approximately 1.3 A –, and normal transistor simply don't survive those currents. At the pins of the IC we often use diodes to carry the ESD current safely between the pins of the device, and the grounded pin.

Although we don't intentionally use them, they are an inherent feature of almost all MOSFETs. The drain and source regions of an NMOS are doped with a donor element, and the bulk is doped with acceptors. This formes diodes between drain/source and bulk. These parasitic diodes have an capacitance that loads all circuits, and must be taken into account.

Another useful feature of the diode is the expoential relationship betwen the forward current, and the voltage across the device. If you push a constant current into a diode, then small changes to the current does not change the diode voltage significantly. You could use this as a reference voltage, however, the forward current does change with the temperature, so it requires sligthly more than one diode and a current to make a reference voltage that is stable over temperature.

II. How

Silicon is a crystal where all electrons are used in covalent bonds between the silicon atoms. If we ignore temperature, then none of the electronics are free to move. The temperature, the vibrations of the atoms, do sometimes break the covalent bond, so there is a continuous generation of electron/hole pairs in pure silicon. At room temperature this intrinsic carrier consentration is about $n_i = 1.1 \times 10^{16} carriers/m^3$.

We can change the property of silicon by introducing other elements. Phosphor has one more electron/proton than silicon, Boron has one less electron/proton. Injecting these elements into the silicon crystal lattice changes the number of free electron/holes (those not used in covalent bonds), this is commonly referred to as doping. If we have an element with more electrons we call it a donor, and the donor concentration N_D. Since the crystal now has an abundance of electrons we call it n-type. If the element has less electrons we call it an acceptor, and the acceptor concentration N_A. Since the crystal now has an abundance of holes, we call it p-type. Usually these doping concentrations are larger than the intrinsic carrier concentration, from maybe 10^{21} to 10^{27} $carriers/m^3$. To separate between these concentrations we use $p-, p, p+$ or $n-, n, n+$. In most instances the doping concentration is so much higher that the intrinsic carrier concentration that we can safely assume that the number of electroncs, and number of holes are the same as N_D and N_A.

In a p-type material, although holes dominate, there will still be a minority of electrons moving around, which is given by

$$p_n = \frac{n_i^2}{N_D} \tag{1}$$

, and a similar equation for the hole concentration in an n-type. If you need to know why it's this equation, then you need to check a solid-state physics book. This is one of those equations that you just need to remember

In a p-type crystal there is a majority of holes, and a minority of electrons. Thus we name holes majority carriers, and electrons minority carriers. For n-type it's opposite.

III. What

Imagine an n-type material, and a p-type material, both are neutral in charge, because they have the same number of electrons and protons. The free carriers will move around the material constantly.

Now imagine we bring the two materials together. Some of the electrons in the n-type will wander over to the p-type material, and visa versa. Here they will find an opposite charge, and will get locked in place. They will become stuck. This creates a depletion region with immobile charges. Where as the two materials used to be neutrally charged, there will now be a build up of negative charge on the p-side, and positive charge on the n-side. There will also be a field created by the charge difference, and a built-in voltage will develop across the depletion region. The built in voltage depends on the carrier concentrations, and is given by Equation 2 , where k is Boltzmanns constant, q is the charge of an electron.

$$\Phi_0 = \frac{kT}{q} ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{2}$$

IV. Reverse bias

We can apply an external voltage to the pn diode. If we apply a field in the same direction as the built-in voltage, we call it reverse bias. Under reverse bias the current in the diode is small.

As mentioned before, we continuously have electron/hole pairs generated by the temperature. In addition, we can have electron/hole pairs generated by for example photons (photo diodes), or impact ionization (charges at high speed, like radiation). Those electron/hole pairs come into existence in the depletion region, or happen to wander into the depletion region before recombining will be swept across the depletion region due to the electric field. The electron will drift to the n-type, and holes will drift to the p-type. This drift creates a leakage current in the diode.

To estimate the leakage current we would need to know how many electron/hole pairs are generated per second, and how many reach the depletion region before recombining. Not at all a trivial calculation. However, we do expect that the leakage current also doubles per 11^oC, similar to the n_i.

The width of a depletion region in the n-type material can be approximated by Equation 3 , where l_1 is Equation 4 , where ε_0 is the permittivity of free space (8.854×10^{12} F/m), and $K_s = 11.8$ the relative permittivity of silicon. At $V_R = 0$, the depletion width is l_1. As we increase V_R the depletion region will grow, but it's not proportional to V_R. The depletion width has the same equation for p-type, just replace N_A with N_D.

$$x_n(V_R) = l_1\sqrt{1 + \frac{V_R}{\Phi_0}} \tag{3}$$

$$l_1 = \sqrt{\frac{2K_s\varepsilon_0}{q\Phi_0}\frac{N_A}{N_D(N_A + N_D)}} \tag{4}$$

Remember that $I = C\frac{dV}{dt}$, and $I = \frac{dQ}{dt}$ thus $C = \frac{dQ}{dV}$, so if we find the charge in the depletion region, then we can calculate the small signal capacitance. For the n-side the depletion region charge per unit area can be approximated by $Q = qN_Dx_n(V_R)$ so the capacitance per unit area is Equation 5 , where C_{j0} is Equation 6

$$C = \frac{C_{j0}}{\sqrt{1 + \frac{V_R}{\Phi_0}}} \tag{5}$$

$$C_{j0} = \frac{qN_Dl_1}{2} = \sqrt{\frac{qK_s\varepsilon_0}{2\Phi_0}\frac{N_AN_D}{N_A + N_D}} \tag{6}$$

V. Reverse bias

VI. What

VII. Conclusion

References

[1] Brittain, Bardeen, Shockley "Point-contact transistor", online:https://en.wikipedia.org/wiki/Transistor