Diodes

Carsten Wulff, *2022-05-07*, v0.1.0

Abstract—I explain how diodes work.

Why

Diodes are a magical [1] semiconductor device that conduct current in one direction. It's one of the fundamental electronics components, and it's a good idea to understand how they work.

If you don't understand diodes, then you won't understand transistors, neither bipolar, or field effect transistors.

A useful feature of the diode is the expoential relationship between the forward current, and the voltage across the device.

To understand why a diode works it's necessary to understand the physics behind semiconductors.

This paper attempts to explain in the simplest possible terms how a diode works [2]

Intrinsic carrier concentration

The intrinsic carrier concentration of silicon, or how many free electrons and holes at a given temperature, is given by

$$n_i = \sqrt{N_c N_v} e^{-\frac{E_g}{2kT}}$$

where E_g is the bandgap energy (approx 1.12 eV), k is Boltsmanns constant, T is the temperature in Kelvin, N_c is the density of states in conduction band, and N_v is the density of states in the valence band.

The density of states are

$$N_c = 2 \left[\frac{2\pi kT m_n^*}{h^2} \right]^{3/2} \quad N_v = 2 \left[\frac{2\pi kT m_p^*}{h^2} \right]^{3/2}$$

where h is Planck's constant, m_n^* is the effective mass of holes, and m_p^* is the effective mass of holes. The effective mass of electrons and holes in silicon depend on direction of movement, strain of silicon, and I'm not entierly sure what is the correct number to use when computing density of states.

In [1] they claim the intrinsic carrier consentration is a constant, although they do mention n_i doubles every 11 degrees. In BSIM 4.8 [2] n_i is

---

[1] It doesn't stop being magic just because you know how it works. – Terry Pratchett, The Wee Free Men

[2] Simplify as mutch as possible, but no more. – Einstein

$$n_i = 1.45e10 \frac{TNOM}{300.15} \sqrt{\frac{T}{300.15}} \exp^{21.5565981 - \frac{E_g}{2kT}}$$

Comparing the three models in Fig. 1, we see the shape of BSIM and the full equation is almost the same, while the "doubling every 11 degrees" is just wrong.
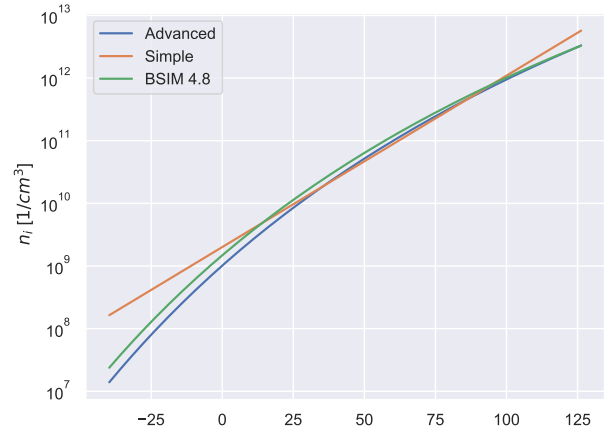

Fig. 1. Intrinsic carrier concentration versus temperature

At room temperature this intrinsic carrier consentation is about $n_i = 1 \times 10^{16} carriers/m^3$.

That may sound like a big number, however, if we calculate the electrons per um^3 it's $n_i = \frac{1 \times 10^{16}}{(1 \times 10^6)^3} carriers/\mu m^3 < 1$, so there are really not that many free carriers in intrinsic silicon.

Doping

We can change the property of silicon by introducing other elements. Phosphor has one more electron than silicon, Boron has one less electron. Injecting these elements into the silicon crystal lattice changes the number of free electron/holes.

These days, we usually dope with a high energy ion gun, while in the olden days, most doping was done by diffusion. You'd paint something containing Boron on the silicon, and then heat it in a furnance to "diffuse" the Boron atoms into the silicon.

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, which have negative charge, 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.

The doped material does not have a net charge, however, it's the same number of electrons and protons, so even though we dope silicon, it does remain neutral.

The 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 $n-, n, n+$.

The number of electrons and holes in a n-type material is

$$n_n = N_D \ , \ p_n = \frac{n_i^2}{N_D}$$

and in a p-type material

$$p_p = N_A \ , \ n_p = \frac{n_i^2}{N_A}$$

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.

PN JUNCTIONS

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. Within both materials there are free electrons, and free holes which move around constantly.

Now imagine we bring the two materials together, and we call where they meet the junction. Some of the electrons in the n-type will wander across the junction to the p-type material, and visa versa. On the opposite side of the junction they might find an opposite charge, and might get locked in place. They will become stuck.

After a while, the diffusion of charges across the junction 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.

The charge difference will create a field, and a built-in voltage will develop across the depletion region.

The magnitude of the built-in voltage can be computed from Fermi-Dirac distribution, stating that the average number of fermions in a single-particle state j is given by

$$n_j = \frac{1}{e^{(E_j - \mu)/kT} + 1}$$

where E_j is the energy of the single-particle state, μ is the chemical potential.

Assuming the exponential is much larger than 1, and taking the ratio number of free electrons on the n-side, and p-side, we get

$$\frac{n_n}{n_p} = \frac{e^{(E_{n_p} - \mu)/kT}}{e^{(E_{n_n} - \mu)/kT}} = e^{\frac{E_{n_p} - E_{n_n}}{kT}}$$

where $E_{n_p} - E_{n_n}$ is the energy difference between electrons on the p-side and n-side. This energy difference is equivalent to $q\Phi_0$ where Φ_0 is the built-in voltage. As a result, and inserting for n_p, we get

$$\frac{N_A N_D}{n_i^2} = e^{\frac{q\Phi_0}{kT}}$$

or

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

CURRENT

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 that 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.

If we apply a voltage in the forward direction, opposite to the field, the current will be

$$I_D = I_S(e^{\frac{V_D}{V_T}} - 1)$$

where V_D is the voltage across the diode, $V_T = \frac{kT}{q}$ and

$$I_S = qAn_i^2 \left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D} \right)$$

where A is the area of the diode, D_n, D_p is the diffusion constant of electrons and holes and L_n, L_p is the diffusion length of electrons and holes.

FORWARD VOLTAGE TEMPERATURE DEPENDENCE

We can rearrange I_D equation to get

$$V_D = V_T \ln \left(\frac{I_D}{I_S} \right)$$

and at first glance, it appears like V_D has a positive temperature coefficient. That is, however, wrong.

First rewrite

$$V_D = V_T \ln I_D - V_T \ln I_S$$

$$\ln I_S = 2 \ln n_i + \ln qA + \ln \left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D} \right)$$

Assume that diffusion constants, and diffusion lengths are independent of temperature. That leaves n_i that varies with temperature.

$$n_i = 2 \left(\frac{4\pi^2 k^2 T^2 m_n^* m_p^*}{h^4} \right)^{3/4} e^{\frac{-E_g}{2kT}}$$

$$\ln n_i = \ln 2 + \frac{3}{4} \ln \frac{4\pi^2 k^2 m_n^* m_p^*}{h^4} + \frac{3}{2} \ln T - \frac{E_g}{2kT}$$

with $E_G = qV_G$ and inserting back into equation for V_D

$$V_D = \frac{kT}{q}(\ell - 3 \ln T) + V_G$$

Where ℓ is temperature independent, and given by

$$\ell = \ln \frac{I_D}{2Aq} - \ln \left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D} \right) - \frac{3}{4} \ln \frac{4\pi^2 k^2 m_n^* m_p^*}{h^4}$$

From equations above we can see that at 0 K, we expect the diode voltage to be equal to the bandgap of silicon. Diodes don't work at 0K though.

Although it's not trivial to see that the diode voltage has a negative temperature coefficient, if you do compute it as in vd.py, then you'll see it decreases.

The slope of the diode voltage can be seen to depend on the area, the current, doping, diffusion contstant, diffusion length and the effctive masses.

Fig. 2 shows the V_D and the deviation of V_D from a straight line. The non-linear component of V_D is only a few mV. If we could combine V_D with a voltage that increased with temperature, then we could get a stable voltage across temperature to within a few mV.

Bandgap references

Assume we have a circuit like Fig. 3 Here we have two diodes, biased at different current densities. The voltage on the left diode V_{D1} is equal to the sum of the voltage on the right diode V_{D2} and voltage accross the resistor R_1. The current in the two diodes are the same due to the current mirror. A such, we have that

$$I_S e^{\frac{qV_{D1}}{kT}} = N I_S e^{\frac{qV_{D2}}{kT}}$$

Taking logartihm of both sides, and rearranging, we see that
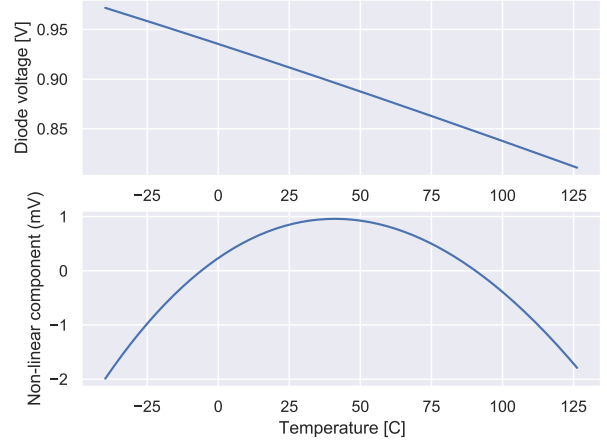

Fig. 2. Diode forward voltage as a function of temperature

$$V_{D1} - V_{D2} = \frac{kT}{q} \ln N$$

Or that the difference between two diode voltages biased at different current densities is proportional to absolute temperature.

In the circuit above, this ΔV_D is across the resistor R_1, as such, the $I_D = \Delta V_D / R_1$. We have a current that is proportional to temperature.

If we copied the current, and sent it into a series combination of a resistor R_2 and a diode, we could scale the R_2 value to give us the exactly right slope to compensate for the negative slope of the V_D voltage.

The voltage across the resistor and diode would be constant over temperature, with the small exception of the non-linear component of V_D.

References

[1]   T. C. Carusone, D. Johns, and K. Martin, *Analog integrated circuit design.* Wiley, 2011 [Online]. Available: https://books.google.no/books?id=1OIJZzLvVhcC

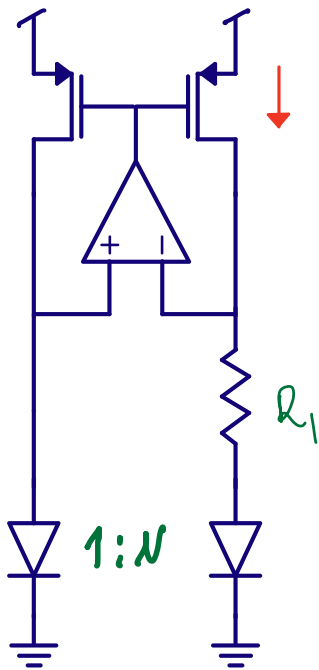[2]   Berkeley, "Berkeley short-channel IGFET model." [Online]. Available: http://bsim.berkeley.edu/models/bsim4/

Fig. 3. Circuit to generate a current proportional to kT