Notes for Lecture 16

Diode Equation+

The actual lecture 16 covered a significant part of LN 15 as well.

16.1 Current under Bias

In the previous lecture, we identified the physics of minority carrier injection (or depletion) just outside the depletion region.

In order to get the leading estimate of that, we assumed that the current is zero, as the first step of solving Eq. 15.4 (and an analogous equation $J_n = \mu_n enE + D_n e\partial n/\partial x$ for the electron) *within the depletion region*. Now, the question is how do we actually update J_p and J_n? For this, we look to the quasi-neutral region, and remind ourselves of point 1 of Section 15.1. The details will be described below. This is clearly the first iteration of the perturbation, and if more precision is necessary, then Eq. 15.4 and its analogous form for the electron given above can be integrated with the thus obtained J_p and J_n, to give the next order estimate of the minority carrier injection and depletion, and the iteration can continue on this way indefinitely, if desired. Here, we are interested in the leading order result only.

Note that the current in the quasi-neutral p region is due to the combination of the majority and the minority carriers. The current due to the minority carriers is easy to calculate, since it is due to diffusion only (point 2, page 2, LN 15). In contrast, the current due to the majority carrier is harder to calculate, since it involves both the diffusion current and the drift current. Furthermore, it involves a small E field as an unknown. As we will see, the majority current is actually not a necessary ingredient for obtaining the diode equation. Rather, from the solution, the majority current can be calculated as an outcome, and the small E field inside the quasi-neutral region can be calculated from it. For this reason, we need consider only minority carrier equations for the purpose of calculating current.

At $x = -x_p$, the current due to the minority current density is given by $J_n(x = -x_p) = eD_n \left.\frac{d\Delta n}{dx}\right|_{x=-x_p}$, and this is easily calculated from Eq. 15.14 as

$$J_n(x = -x_p) = \frac{eD_n \Delta n(x = -x_p)}{L_n} = \frac{eD_n n_i^2}{L_n N_A} \left[\exp(e\beta V_A) - 1 \right] \tag{16.1}$$

In fact, the full x dependence within the quasi-neutral p region is easy to calculate,

$$J_n(x) = J_n(x = -x_p) \exp\left(\frac{x + x_p}{L_n}\right) \tag{16.2}$$

Now, going to the quasi-neutral n region, we have

$$\Delta p(x) = \Delta p(x = x_n) \exp\left(-\frac{x - x_n}{L_p}\right) \qquad x \geq x_n \tag{16.3}$$

with $L_p = \sqrt{D_p \tau_p}$. From the law of junction,

$$\Delta p(x = x_n) = \frac{n_i^2}{N_D} \left[\exp(e\beta V_A) - 1 \right] \tag{16.4}$$

The minority current density at $x = x_n$ is given by

$$J_p(x = x_n) = \frac{eD_p \Delta n(x = x_n)}{L_p} = \frac{eD_p n_i^2}{L_p N_D} \left[\exp(e\beta V_A) - 1 \right] \tag{16.5}$$

And, the full dependence within the quasi-neutral n region is

$$J_p(x) = J_p(x = x_n) \exp\left(-\frac{x - x_n}{L_p}\right) \tag{16.6}$$

Lastly, note that all formula in this section is applicable when $V_A < 0$, as well. In that case, the minority carrier is not injected. Rather, because values of Δn and Δp above are negative, and the minority carriers are depleted.

16.2 Ideal Diode Equation

Clearly, in a steady state, the current through the depletion region is conserved in each channel *for an ideal diode* (page 7 of LN 15). In a more formal fashion, this follows from the continuity equation

$$\frac{\partial n}{\partial t} = \left.\frac{\partial n}{\partial t}\right|_{DD} + \left.\frac{\partial n}{\partial t}\right|_{RG} = \frac{1}{e}\frac{\partial J_n}{\partial x} + \left.\frac{\partial n}{\partial t}\right|_{RG} \tag{16.7}$$

Due to the assumption of the steady state, $\partial n/\partial t = 0$. Due to the assumption of the no net generation-recombination (ideal diode), the 2nd term on the RHS is zero. Thus, we conclude $\partial J_n/\partial x = 0$. Similarly, we also conclude that $\partial J_p/\partial x = 0$.

$$\frac{\partial J_p}{\partial x} = \frac{\partial J_n}{\partial x} = 0 \qquad\qquad -x_p \leq x \leq x_n \tag{16.8}$$

This means the total current density is simply obtained by adding $J_p(x = x_n)$ and $J_n(x = -x_p)$ of the previous section. By multiplying the result by A (the crossection of the device), we get the current I

$$I = I_0\left[\exp(e\beta V_A) - 1\right] \tag{16.9}$$

where

$$I_0 = eAn_i^2\left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D}\right) \tag{16.10}$$

This is the famous "ideal diode equation." Or, the "Shockley equation."

The interpretation of this result is actually simple, and has already been given in page 3 of LN 15. The drift current is unchanged relative to the equilibrium term, while the diffusion current has an exponential dependence on V_A. In the angled bracket above, $\exp(e\beta V_a)$ corresponds to the net diffusion current through the junction, while -1 corresponds to the net drift current through the junction. The difference is the net current.

16.3 Majority Current

Note that the current is conveniently calculated by considering only minority currents in the quasi-neutral region. However, since we know the total current, we also know the majority current. All we have to do is subtract the minority current from the total current. The majority current is comparable to the minority current near the depletion region, and becomes totally dominant far away from the junction. This is due to the limited minority carrier diffusion.

16.4 Summary of All Currents

Here is a summary of all currents, in terms of the current density. They are presented roughly in the order that we derived them in our discussions so far.

$$J_n(x) = J_{n,0}\, \exp\left(\frac{x+x_p}{L_n}\right) \tag{16.11}$$

$(x \le -x_p)$ minority current (Eq. 16.2)

$$J_n(x) = J_{n,0} = e n_i^2 \frac{D_n}{L_n N_A}\, [\exp(e\beta V_A) - 1] \tag{16.12}$$

$(-x_p \le x \le x_n)$ (Eqs. 16.1,16.8)

$$J_p(x) = J_{p,0}\, \exp\left(-\frac{x-x_n}{L_p}\right) \tag{16.13}$$

$(x \ge x_n)$ minority current (Eq. 16.6)

$$J_p(x) = J_{p,0} = e n_i^2 \frac{D_p}{L_p N_D}\, [\exp(e\beta V_A) - 1] \tag{16.14}$$

$(-x_p \le x \le x_n)$ (Eqs. 16.5,16.8)

$$J_p(x) + J_n(x) \equiv J_{tot} = J_{n,0} + J_{p,0} \tag{16.15}$$

$(-\infty < x < \infty)$ point 1 of page 2 of LN 15 (DC or near-DC bias)

$$J_p(x) = J_{tot} - J_n(x) = J_{p,0} + J_{n,0}\left[1 - \exp\left(\frac{x+x_p}{L_n}\right)\right] \tag{16.16}$$

$(x \le -x_p)$ majority current

$$J_n(x) = J_{tot} - J_p(x) = J_{n,0} + J_{p,0}\left[1 - \exp\left(-\frac{x-x_n}{L_p}\right)\right] \tag{16.17}$$

$(x \ge x_n)$ majority current

The current is carried by both electrons and holes except at far ends of the device. Within the quasi-neutral region, the minority current at the edge of the depletion region is converted to the majority current on the length scale of the minority carrier diffusion length, with the sum of the majority current and the minority current held constant.

16.5 Deviation from Ideal

Ideally, the reverse bias current saturates at $-I_0$. And the forward bias current current behaves as $\exp(e\beta V_A)$. However, this has only a limited validity.

At small forward bias, the current is much greater than predicted. At reverse bias, the current is also much greater, and it does not saturate.

These are due to recombination current (forward bias) and generation current (reverse bias). It is easy to understand why this must happen, since there has to be a net recombination (forward bias) and a net generation (reverse bias). The quantitative theory of it is rather involved. The discussion in Section T6.2.3 is excellent. Exercise T6.6 is worth noting. At small forward bias,

$$I = I_{01}\exp(e\beta V_A) + I_{02}\exp(e\beta V_A/2) \tag{16.18}$$

where $I_{02} \gg I_{01}$ and the 2nd term is due to recombination. At high bias, the first exponential function catches up, and becomes dominant.

At a large bias value, the $I - V$ curve "slopes over" for the forward bias due to the series resistance effect of the quasi-neutral region. This is because the actual bias applied to the junction is now reduced

$$V_A \to V_A - IR_S \tag{16.19}$$

The resistance can be measured by measuring, for a given I, $V_A(actual) - V_A(theory - without - R_S)$. A semi-log plot ($\ln I$ vs. V_A) is convenient in this case. Also, at a large bias, the low level injection assumption will break down. In this case, the bias voltage dependence crosses over to $\exp(e\beta V_A/2)$ again. This is called the "high current $e\beta/2$ region" because in a semi-log plot ($\ln I$ vs. V_A), it will show the slope $e\beta/2$, instead of $e\beta$.

Finally, for reverse bias, one has Zener tunneling breakdown or avalanche breakdown.

16.6 Quasi-Fermi Level

The law of junction can be summarized well with the concept of the quasi-Fermi level.

In a non-equilibrium case, the concept of Fermi level is lost, but what if one were to "fit" the known carrier density to an artificial Fermi level? More specifically, one defines

$$n \equiv n_i \exp(\beta(F_N - E_i)) \tag{16.20}$$
$$p \equiv p_i \exp(\beta(E_i - F_P)) \tag{16.21}$$

F_N and F_P defined in this way are called the "quasi-Fermi level."

For a material in a non-equilibrium state, F_N and F_P differ in general, while as the material is brought into the equilibrium, F_N and F_P will converge to each other.

One good use of the quasi-Fermi level is that

$$J_p = \mu_p p \nabla F_P \tag{16.22}$$
$$J_n = \mu_n n \nabla F_N \tag{16.23}$$

Namely, the gradient of the quasi-Fermi level divided by e is an effective electric field. The derivation is the following (for the electron only – the hole case left for your exercise):

$$J_n = \mu_n neE + D_n e \nabla n \tag{16.24}$$
$$= \mu_n neE + \frac{\mu_n}{\beta} \nabla n \qquad \text{Einstein relation} \tag{16.25}$$
$$= \mu_n neE + \frac{\mu_n}{\beta} n_i \exp(\beta(F_N - E_i)) \cdot \beta \cdot (\nabla F_N - \nabla E_i) \qquad \text{Eq. 16.20} \tag{16.26}$$
$$= \mu_n neE + \mu_n n(\nabla F_N - eE) \qquad \nabla E_i = eE \text{ (LN 10)} \tag{16.27}$$
$$= \mu_n n \nabla F_N \tag{16.28}$$

Intuitively, F_N or F_P can be thought of as an "effective chemical potential" in the presence of the electric field. The particle will flow from the high chemical potential to the low chemical potential. For an electron, the current flows in the opposite direction to that flow, and so this is why the current flows in the direction of ∇F_N. For a hole, the current does flow in the same direction as the negative gradient of its chemical potential. Then, why is the current proportional to ∇F_P, not $-\nabla F_P$? It is because these energies are (as always) expressed from the electron point of view. The "quasi-Fermi level" for the hole is actually $-F_P$.

But, beware. For minority carriers, the gradient of the quasi-Fermi level is a good indicator for the current. However, for non-minority carriers, the gradient of the quasi-Fermi level can be very small in a figure, but can give rise to a comparable value to that due to minority carriers! This is because n or p is a large number for them!

In general, $E_F = F_N = F_P$ in equilibrium. However, the converse is not true: just because $F_N = F_P$ does not mean that the system is in equilibrium.

In the ideal diode treatment of ours, what are the values of F_N and F_P? First, note that F_N is constant in the quasi-neutral n region, and so is F_P in the quasi-neutral p region. These are not E_F, as E_F is not defined in non-equilibrium. All we can say is that these numbers will be identical to the E_F value that each region would have if each region were separately in equilibrium. Second, at $x \pm \infty$, $F_P = F_N$. This does not mean that the system is in equilibrium, since the current is flowing. Third, as the depletion region is reached, F_P (for n part) and F_N (for p part) will bend. At $x = -x_p$ or $x = x_n$, $V_A = F_N - F_P$, and in fact for all $-x_p \leq x \leq x_n$ by the law of the junction. Figure T6.4 summarizes these behaviors.

It is important to note that F_N or F_P being constant in such a figure does *not* mean that ∇F_N or ∇F_P is identically zero, the main reason being that our solutions are only perturbatively correct. Therefore, flat F_N or F_P does *not* mean that the corresponding current is zero (see Section 16.4) as discussed in the "But, beware" paragraph above.