

Notes for Lecture 15

Key Physics of Biased pn Junction

In the previous lecture, we went over the electrostatics of a pn junction, biased or not. In this lecture, we explore further what happens under bias, identifying key physics.

Let me introduce a new terminology and then make sure that everyone understands the on-going assumptions.

First, let us define a new terminology. The region outside the depletion region is called a “quasi-neutral” region. The neutrality is assured by the fast action of the majority carriers in this region. It is “quasi-neutral” since the charge neutrality is not strict in reality. It is strict only within the assumption of the depletion approximation. As we discussed in the last lecture, we assume that $E \approx 0$ in this region. Note that within the depletion approximation, E must be strictly zero, but, in reality, we can only claim that $E \approx 0$.

Second, note that we continue to assume the low level injection. This means that in the p -type quasi-neutral region, the modifications to p and n are much smaller than p . Similarly, in the n -type quasi-neutral region, the changes to p and n are much smaller than n . A precise statement regarding this assumption in the context of the bias driven modification of the carrier density will appear near the end of this note.

Third, we assume non-degenerate semiconductors.

Fourth, the p part is on the left and the n part is on the right, by convention.

Fifth, related to the first point, note that the assumption of no significant voltage drop in the quasi-neutral region means that we are implicitly assuming a small value for $|V_A|$. Why is this? It is because if $|V_A|$ is large, then a large current may flow (next

lecture). When a large current flows, a significant fraction of V_A will apply to the quasi-neutral region. This is because the quasi-neutral region has an (approximately) fixed resistance, while the resistance in the depletion region will decrease as the current increases. A large voltage drop in the quasi-neutral region requires a significant modification of our analysis here. This and other various large V_A physics phenomena will be discussed in the next lecture.

15.1 Fundamentals

Let us see how we can understand the current through a pn junction under a bias voltage, V_A . We continue to consider a step junction, and we assume a steady state, i.e. a current is flowing at a constant rate. Note that we assume that V_A is small. How small? The answer is “small enough so that other large V_A physics is not important.” There is already one way to characterize this definition. For instance, with a small V_A , it is a good approximation that the bias is applied entirely to the depletion region. Why so? Let us keep these basic facts in mind.

1. The current must be independent of the position x .
2. In the quasi-neutral regions, the minority carrier carries current only by diffusion. This is because E is small, and the minority carrier density is small. So the drift current is negligible (cf. Appendix 1). On the other hand, the majority carrier carries current by both diffusion and drift.
3. In the depletion region, both types of current, diffusion and drift, exist.
4. In a non-equilibrium state, there can be a skew of the electron distribution or the hole distribution, causing a current, while $n(x)$ or $p(x)$ does not experience any net birth rate (generation) or any net death (recombination) rate. From the electron point of view or the hole point of view, then each subsystem is a particle-conserving system. Let us call this a non-equilibrium mechanism (i). Then, we define a non-equilibrium mechanism (ii) as that which involves a generation or recombination of $n(x)$ or $p(x)$.
5. Far from the junction, the device acts like a uniform p type semiconductor or a uniform n type semiconductor, and the device should not know anything about the depletion region, as whatever action that goes on in the depletion region must be damped/screened out (by the minority diffusion length scale that we already learned in LN 13, as it turns out)¹. So, the physics at a far

¹Of course, this assumption fails if the quasi-neutral region is very small in a device – such a

end must be identical with a uniform semiconductor with a finite current. In such a situation, all that is happening is that the momentum space distribution of electrons is skewed, as is the momentum space distribution of holes. Other than this fact, the local state is much like in the equilibrium state, in that the number of electrons and the number of holes are just as in the equilibrium state. So, even though the Fermi level is not defined for non-equilibrium, the Fermi level could be operationally defined (more about this in the next lecture) as unchanged from the equilibrium Fermi level. Clearly, only the non-equilibrium mechanism (i) is at work here, and the non-equilibrium mechanism (ii) is not².

The way to think about the diffusion current and the drift current through the junction is as follows. We shall consider electrons only. The argument for holes is analogous. The diffusion will drive the electrons to the left, while the drift will drive the electrons to the right (if $|V_A| < V_{bi}$). First, consider the diffusion process. At the right edge of the junction there are many electrons. On average, then, these electrons are driven to the left by the diffusion process. However, as they diffuse, the energy must be conserved. Unless the energy of the electron is higher than $V_{bi} - V_A$, the electron will not be able to make it to the left end of the junction. Less energetic electrons will slow down and then turn backwards, driven by the electric field! Thus, the turned-around motion is mainly drift, while the initial motion is mainly diffusion. These two motions will cancel each other. So, the **net diffusion** of the electrons across the junction comes from the diffusion of those electrons with energy higher than $V_{bi} - V_A$. How about the drift? Having considered the net diffusion this way, the only remaining part of the drift process is the following. On the left end of the depletion region, there will be a small density of electrons, as they are minority there. These electrons, if they somehow end up within the depletion region, they will be “caught in the electric field” and accelerate towards the right end of the depletion region. This is the **net drift** of the electrons. How do they end up in within the depletion region? That is due to the diffusion.

The key observation is that the net diffusion is strongly dependent on the bias voltage but the net drift is not. This is because the net drift simply requires that the minority carriers “happen to fall into the junction,” a process that is unchanged with or without the bias, while the net diffusion depends on the number of majority electron carriers at energy $V_{bi} - V_A$ above V_c , where this number depends exponentially on the voltage. Of course, there will be a weak dependence of both processes on the voltage as the electrostatics do depend on V_A with a power law ($x_p, x_n, E \propto \sqrt{V_{bi} - V_A}$

device is called a “narrow base” device, which is excluded in the current discussion. Such narrow base case will become important in the context of a transistor.

²The reverse is not true. Just because only the non-equilibrium mechanism (i) is at work does not mean that the values of $n(x)$ and $p(x)$ are equal to the equilibrium values.

as we saw in the last lecture), but those weak dependences are no matches for the exponential dependence of the net diffusion process, which is the dominant controlling physics here.

A summary is given now. Note that a *pn* junction device, when used in a circuit, is called a diode.

This is a qualitative summary of the diode action. The net diffusion across the depletion region, as defined above, is due to the majority carriers, and the net drift across the depletion region, as defined above, is due to the minority carriers. The former increases exponentially on forward bias, and decreases exponentially on reverse bias. In comparison, the net drift stays essentially the same with or without bias. The current in the reverse bias case is given dominated by the net drift current, which is quite small (nA \sim pA). The current on forward bias is much greater than the current on reverse bias. This is the basis of the diode property.

Fig. T6.1 is a good visual summary of the same physics.

It may be worthwhile to emphasize again that while the *net* diffusion current across the depletion region is carried by the majority carriers and the *net* drift current by the minority carriers, both types of carriers experience diffusion and drift everywhere, especially within the depletion region where the E field is the most significant.

15.2 Bias-Induced Minority Carrier Injection or Depletion

From the consideration of the previous section, it is clear that in the steady state, some increase in the electron density will be established on the left end of the depletion region (and some increase in the hole density on the right end of the depletion region) on forward bias.

The key physical effect of the transport of charge carriers through the depletion region is the minority carrier injection at the edges of the depletion region on forward bias. On reverse bias, the key physical effect is the minority carrier depletion in the same region.

Now, let us *calculate* the steady state value of $p_n(x_n)$ to figure out just *how much*

minority carrier injection or depletion occurs. We need some careful considerations for doing so.

We noted that Eqs. 14.12–14.14 are not applicable under bias, strictly speaking, but Eq. 14.15 is valid under bias, as is. The reason for the latter is, of course, the above point 5. Note that in equilibrium, we have, in fact,

$$V_{bi} = \frac{k_B T}{e} \ln \frac{n(x_r)}{n(x_l)} = \frac{k_B T}{e} \ln \frac{p(x_l)}{p(x_r)} \quad x_l \leq -x_p, x_r \geq x_n, V_A = 0 \quad (15.1)$$

Under bias, the point 5 means

$$V_{bi} = \frac{k_B T}{e} \ln \frac{n(\infty)}{n(-\infty)} = \frac{k_B T}{e} \ln \frac{p(-\infty)}{p(\infty)} \quad \text{for any } V_A \quad (15.2)$$

But, one might ask a question, how about at other values of x ? For instance, how would this change if one were to change ∞ to x_n and $-\infty$ to $-x_p$, corresponding to the edges of the depletion region? The answer is (cf. the activity sheet):

$$V_{bi} - V_A = \frac{k_B T}{e} \ln \frac{n(x_n)}{n(-x_p)} = \frac{k_B T}{e} \ln \frac{p(-x_p)}{p(x_n)} \quad \text{for any } V_A \quad (15.3)$$

Where does this answer come from? It relies on certain assumptions that make n and p mere “electrostatic quantities.” It is best understood if we treat the effect of V_A as a “perturbation” (see below for more discussion).

Let us consider p only, in detail. The analysis of n is completely analogous, and we will simply write down the solution for it below.

Within the depletion region,

$$J_p = \mu_p e p E - D_p e \frac{\partial p}{\partial x} \quad (15.4)$$

Note that we already know the exact form and value of E within the depletion approximation (last lecture). Thus, this equation can be considered as a differential equation for p with J_p as a sole input. Treating J_p as a small value (i.e. as proportional to a “perturbation” parameter ε), then J_p can be taken to be zero at first, which then will give the first order correction for p (see Appendix 1 and <http://griffin.ucsc.edu/teaching/10Q4-105/A02-Perturbation.pdf> for more details about the perturbation theory)³ As we shall see in the next lecture, treating J_p as a perturbation parameter is equivalent to taking the minority carrier injection

³Understanding B.2 and B.2.1 of the latter document will be enough to appreciate the current point. Here is a very quick bare-bone summary of them. At this point, I am not really trying to teach the perturbation theory. However, I realize that I cannot avoid teaching the essence of this important theory. So, here it goes, in the most bare-bone form. Suppose you were asked to solve

as a perturbation parameter in comparison to the majority carrier density, a perfectly reasonable thing to do in light of our assumption of low carrier injection.

Thus, in this perturbation theory, we can put $J_p = 0$ in the above equation, to obtain the leading order solution, for n and p ! This reduces the problem for p to that for the equilibrium except that E is now changed due to V_A . We have

$$p(x) = p(-x_p) \exp \left[\beta e \int_{-x_p}^x E(x') dx' \right] \quad -x_p \leq x \leq x_n \quad (15.5)$$

The result for $n(x)$ is, using an analogous argument,

$$n(x) = n(-x_p) \exp \left[-\beta e \int_{-x_p}^x E(x') dx' \right] \quad -x_p \leq x \leq x_n \quad (15.6)$$

Thus,

$$n(x)p(x) = n(-x_p)p(-x_p) \quad -x_p \leq x \leq x_n \quad (15.7)$$

Noting that $\int_{-x_p}^x E(x') dx' = -V_{bi} + V_A$ (LN 14), we see that

$$p(x_n) = p(-x_p) \exp [-e\beta(V_{bi} - V_A)] \quad (15.8)$$

$$n(x_n) = n(-x_p) \exp [e\beta(V_{bi} - V_A)] \quad (15.9)$$

proving Eqs. 15.3, as we set out to do. Good news – we can derive more stuff! Using $n(x_n) \approx n_0(x_n)$ and $p(-x_p) \approx p_0(-x_p)$, where the subscript 0 means the equilibrium value, again due to the low level injection assumption, and noting that $n_0(-x_p) = n_0(x_n) \exp(-e\beta V_{bi})$ and $p_0(x_n) = p_0(-x_p) \exp(-e\beta V_{bi})$, we can turn these two equations into

$x - 1 = \sin x$, where ε is a small dimensionless parameter, i.e. a perturbation parameter. It is not possible to solve this equation exactly. Within the perturbation theory, it is noted that there must be a solution near $x = 1$ and that solution can be systematically obtained. The initial guess $x = 1$ called the zero-th order solution, as it is correct up to $O(1)$. The first order solution, correct up to $O(\varepsilon)$, can be obtained as follows. First, take the zero-th order solution and plug it into the right hand side, and *only* into the right hand side: $x - 1 = \varepsilon \sin 1$. This is an easily solvable equation, which gives the first order solution, $x = 1 + \varepsilon \sin 1$. This calculation can be systematically carried out up to any order of ε , producing increasingly accurate result as the calculation goes to higher order (and becomes quite involved)! The correspondence of this simple example and the above equation of ours is the following. Unknown variable to solve for: $x \Leftrightarrow p$. Zero-th order term: $x - 1 \Leftrightarrow \mu_p e p E - D_p e \frac{\partial p}{\partial x}$. Perturbation term: $\varepsilon \sin x \Leftrightarrow J_p$. The perturbation parameter: $\varepsilon \Leftrightarrow n_i^2 [\exp(\beta e V_A) - 1] / (N_D N_A)$, the ratio of the minority injection/depletion per volume to the majority carrier density. Note that J_p is not a simple function of p ; it is rather a complicated functional of p . J_p is actually determined by solving equations in the quasi-neutral region (next lecture).

$$p(x_n) = p_0(x_n) \exp(e\beta V_A) \quad (15.10)$$

$$n(-x_p) = n_0(-x_p) \exp(e\beta V_A) \quad (15.11)$$

Note that for a positive V_A , $p(x_n) > p_0(x_n)$ and $n(-x_p) > n_0(-x_p)$. Thus we have **minority carrier injection** by the bias voltage! For a negative V_A , we have the minority carrier depletion instead, since $p(x_n) < p_0(x_n)$ and $n(-x_p) < n_0(-x_p)$.

Multiplying the above two equations by $n(x_n) \approx n_0(x_n)$ and $p(x_p) \approx p_0(x_p)$, respectively, we see that $p(x_n)n(x_n) = n(x_p)p(x_p) = n_0p_0 \exp(e\beta V_A) = n_i^2 \exp(e\beta V_A)$, where in the last step, we used the law of mass action, $n_i^2 = n_i p_i = n_0(x)p_0(x)$, which is valid at *any* x value within the device (due to the local equilibrium assumption for any thin slice of the device). Given the fact that we already proved that the product $n(x)p(x) = \text{constant}$ within the depletion region (Eq. 15.7), it is not surprising at all that $n(x_n)p(x_n) = n(-x_p)p(-x_p)$. Now, we know what that constant value is. The result is the so-called the **law of the junction**:

$$n(x)p(x) = n_i^2 \exp(e\beta V_A) \quad -x_p \leq x \leq x_n \quad (15.12)$$

While I made it sound, in class, that this law was not as good as it sounds to be, now I have a different opinion, having proved it using the perturbation theory: it is a sound principle justified by the perturbation theory⁴

Let me mention one important thing at this point. The treatment here is the theory of an *ideal pn junction*, by which it is meant that the net generation rate (for reverse bias since np is lower than the equilibrium value) or the net recombination rate (for forward bias since np is higher than the equilibrium value) in the depletion region is ignored. Indeed, for $n(x)p(x)$ to remain a constant across the depletion region, there cannot be any net generation or recombination process. This means that within the depletion region, we are still considering the non-equilibrium mechanism of type (i) only (page 2). This is a necessary condition for the law of the junction. In the next

⁴The book's take on the nature of this law is somewhat different from my take here. As far as I can see, all references on this subject have only vague remarks, if at all, to justify the law, at best, and the textbook is justified in rightfully questioning such remarks. However, as I present here, there is a systematic perturbation theory at work, while it may be a bit complicated.

lecture, we will discuss what happens if we do consider the net generation rate or the net recombination rate⁵.

However, even for an ideal pn junction, there is no way to ignore the net recombination (for forward bias) or the net generation (for reverse bias) that occurs in the quasi-neutral region, as is clear from Eqs. 15.2 and 15.3. So, just outside the depletion region, both non-equilibrium mechanisms (i) and (ii) (page 2) are at work.

15.3 Minority Carrier Diffusion

Consider the minority diffusion equation in the quasi-neutral p region.

$$0 = D_n \frac{\partial^2 \Delta n}{\partial x^2} - \frac{\Delta n}{\tau_n} \quad x \leq -x_p \quad (15.13)$$

As we already examined previously (LN 13), the solution of this equation is

$$\Delta n(x) = \Delta n(x = -x_p) \exp\left(\frac{x + x_p}{L_n}\right) \quad x \leq -x_p \quad (15.14)$$

where the diffusion length $L_n = \sqrt{D_n \tau_n}$. Here, we applied the boundary condition that $\Delta n(x = -\infty) = 0$ (point 5 of page 2). From the law of the junction,

$$\Delta n(x = -x_p) = \frac{n_i^2}{N_A} [\exp(e\beta V_A) - 1] \quad (15.15)$$

Note that this charge fluctuation is neutralized quickly by majority carriers (LN 13 and HW 7.3): $\Delta p(x) = \Delta n(x)$. This means that just as the minority carrier has an exponential distribution such as the above, the majority carrier will have an exactly the opposite variation in the density so that $p - n + N_D - N_A = 0$ for $x \leq -x_p$.

Similarly,

$$\Delta p(x) = \Delta p(x = x_n) \exp\left(\frac{x - x_n}{L_p}\right) \quad x \geq x_n \quad (15.16)$$

⁵ These processes ignored here actually have a very large effect, and so one may ask whether the current theory must be viewed as questionable. Yes and no. We would need to interpret the results very carefully when we compare the results of this theory with the experiment (some parameter such as the saturation current will have to be interpreted with very large correction, as well as the saturation behavior in the reverse bias case), but there is no doubt that this theory provides us with understanding of the fundamental physical processes of a pn junction, a good starting point to which other realistic effects can be added for a more realistic modeling of experiments.

where $L_p = \sqrt{D_p \tau_p}$ and

$$\Delta p(x = x_n) = \frac{n_i^2}{N_D} [\exp(e\beta V_A) - 1] \quad (15.17)$$

So, either the minority carrier injection (for forward bias) or the minority carrier depletion (for reverse bias) is damped out within the length scale of the minority diffusion length.

The small parameter of the current theory is identified as $|\Delta p(x_n)/n(x_n)|$ or $|\Delta n(-x_p)/p(-x_p)|$, both of which is approximately equal to $n_i^2 [\exp(\beta e V_A) - 1] / (N_D N_A)$. This quantity is dependent only on the intrinsic parameters of the crystal and the external condition (V_A), and thus qualifies excellently for, and plays the role of, the perturbation parameter of the theory of an ideal pn junction.