Topic 13: Diode I-V Characteristics

Preface:

This set of notes will analyze and derive the current-voltage expression for the ideal PN junction. By the end of these notes, you will know the assumptions which made the junction ideal, the typical IV characteristic, and several supplemental topics like Avalanche Breakdown and Quasi Fermi Levels.

Ideal IV Characteristic:

The current-voltage characteristic of any component is simply the plot of current over voltage. For a resistor, Ohm's law states $V = IR$. Rearranging the formula for current, $I = \frac{V}{R}$. Therefore, the IV characteristic is linear with a slope of $\frac{1}{R}$. Since many semiconductor devices have nonlinear current-voltage equations, it is incredibly convenient to provide the IV curve instead of a formula. We can apply the IV characteristic to inductors and capacitors as well, but this is not our focus.

Assumptions:

The nonidealities of the PN junction are beyond the scope of this course. Hence, we will make several assumptions to simplify the analysis.

1. The space-charge region has abrupt boundaries.
2. The region beyond the space charge is neutral.
3. Carrier density at the boundaries $-x_p$ and x_n are related by the junction's electrostatic potential.
4. Carrier injection is low.
5. Generation and recombination current components are negligible; zero.
6. Electron and hole currents are constant throughout the space-charge region; over distance.

Several of these assumptions seem arbitrary, so let's step through each one to address their necessity. The following list corresponds one-to-one with the list of assumptions above.

1. Abrupt boundaries lets us use the equations derived for an abrupt junction.
2. The neutral bulk forces the voltage across the device to be applied purely across the junction.
3. The electrostatic potential, ϕ, equals the built-in voltage at the boundaries. Assuming this relation allows the use of $V_{bi} = \frac{kT}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right)$.
4. High injection in the n-type region implies many holes are present. If the hole concentration becomes too large the region will behaves less an an n-type. The inverse is true for the p-type region with electrons.
5. Generation and recombination would unnecessarily complicate the mathematics.
6. Zero generation or recombination implies any current entering the SRC must leave the SRC. The current may change over time, but not distance.

Notation:

This set of notes will be math-intensive. Hence, we need to define some terminology to remove ambiguities. The subscript, o, is used to denote thermal equilibrium. If the o is not present the device is under an excitation, such as a bias voltage.

- p_{po}: The hole concentration in the p-type region at thermal equilibrium.

- p_{no}: The hold concentration in the n-type region at thermal equilibrium.
- n_{no}: The electron concentration in the n-type region at thermal equilibrium.
- n_{po}: The electron concentration in the p-type region at thermal equilibrium.
- p_p: The hole concentration in the p-type region under an excitation.
- p_n: The hole concentration in the n-type region under an excitation.
- n_n: The electron concentration in the n-type region under an excitation.
- n_p: The electron concentration in the p-type region under an excitation.
- Δn: The excess electron concentration within the device.
- Δp: The excess hole concentration within the device.

Setup:

We have a strong foundation in analyzing a semiconductor at thermal equilibrium. The mass-action law, $np = {n_i}^2$, becomes $p_{po}n_{po} = {n_i}^2$ or $n_{no}p_{no} = {n_i}^2$. It is incredibly easy to get lost with the subscripts, but remember that the subscripts indicate the region and thermal equilbrium. For example, n and p in the p-type become n_{po} and p_{po}, respectively. We know in the p and n-type bulk the concentrations are N_A and N_D respectively. Therefore, $p_{po} = N_A$ and $n_{no} = N_D$.

Since N_A and N_D are known, algebra may be used to determine the minority concentration in each bulk.

$$\begin{aligned}
n_{po} &= \frac{p_{po}}{{n_i}^2} = \frac{N_A}{{n_i}^2} && (1)\\
p_{no} &= \frac{n_{no}}{{n_i}^2} = \frac{N_D}{{n_i}^2} && (2)\\
V_{bi} &= \frac{kT}{q}\ln\left(\frac{p_{po}n_{no}}{{n_i}^2}\right) && (3)\\
&= \frac{kT}{q}\ln\left(\frac{N_A N_D}{{n_i}^2}\right) && (4)
\end{aligned}$$

Eq. 3 is helpful but it is in terms of variables we already know. We need some way to derive expressions for the minority concentration in each region. Coincidentally, we may plug in either Eq. 1 or 2 into ${n_i}^2$ and see what happens.

$$\begin{aligned}
V_{bi} &= \frac{kT}{q}\ln\left(\frac{p_{po}n_{no}}{p_{po}n_{po}}\right)\\
V_{bi} &= \frac{kT}{q}\ln\left(\frac{n_{no}}{n_{po}}\right)\\
\Rightarrow \frac{n_{no}}{n_{po}} &= e^{qV_{bi}/kT}\\
n_{po} &= n_{no}e^{-qV_{bi}/kT} && (5)\\
n_p &= n_n e^{-q\,(V_{bi}\,-\,V_{bias})/kT} && (6)\\
p_{no} &= p_{po}e^{-qV_{bi}/kT} && (7)\\
p_n &= p_p e^{-q\,(V_{bi}\,-\,V_{bias})/kT} && (8)
\end{aligned}$$

Where Eq. 6 and 8 have accounted for a bias voltage. The bias voltage is still subtracted from V_{bi} since the factored negative is from algebra. Note the subscript o is no longer present under a bias. The derivation has not been included for Eq. 7, but it is nearly identical. If you plug Eq. 1 into Eq. 3 only a few steps of algebra are needed. At this moment, we can appreciate why low-level injection is necessary.

Under an excitation, Eq. 6 and 8, will depend on how the thermal equilibrium concentration is affected by the bias voltage. We can say the bias will introduce a number of carriers, Δn, and Δp.

$$\begin{aligned} n_n &= n_{no} + \Delta n \approx n_{no} \\ p_p &= p_{po} + \Delta p \approx ppo \end{aligned}$$

If we allowed for high-level injection, large Δn and Δp would occur. The majority concentration of each region could change significantly. As the injection increases, the typing of each region becomes ambiguous, and an intrinsic-like material is produced. Therefore, low-level injection allows the majority concentration to be independent of the excitation strength. Implying that the majority concentration under an excitation is equal to the thermal equilibrium value.

Continuing, let's expand Eq. 6 and 8 and see if any simplifications can be made. We start by replacing n_n and p_p with n_{no} and p_{po}, due to low-level injection.

$$\begin{aligned} p_n &= p_{po}\left(e^{qV_{bi}/kT} \cdot e^{qV_{bias}/kT}\right) \\ &= p_{no}e^{qV_{bias}/kT} \text{ at } x = x_n \qquad (9) \\ n_p &= n_{no}\left(e^{qV_{bi}/kT} \cdot e^{qV_{bias}/kT}\right) \\ &= n_{po}e^{qV_{bias}/kT} \text{ at } x = -x_p \qquad (10) \end{aligned}$$

Interestingly, the minority carrier concentration in the bulk under thermal equilibrium or excitation is dependent on the concentration across the junction. We need to ask ourselves if this makes sense. Under a strong forward bias, $V_{bias} > V_{bi}$ the exponential term will be fairly large; increasing the minority concentration. Inversely, a strong reverse bias will place the exponential term in the denominator. Recall that reverse bias implies $V_{bias} < 0$, so we must account for the negative. Significantly large bias will trend the exponential toward negative infinity, driving the minority concentration to zero.

These observations behave identically to our intuition of the junction. A large forward bias removes the depletion region and allows diffusion to occur. The majority carrier will diffuse to become the minority in the opposite region. The large influx of carriers increases the minority concentration of each region. Inversely, a large reverse bias implies a large electric field across the space charge region. Any carrier which stumbles into the space charge region will be quickly thrown back to the majority region. However, the carrier will accelerate and hold a non-zero momentum. Therefore, as the carrier reaches its majority region it does not immediately stop. Implying the minority concentration at the edge of the SCR is zero; provided sufficient reverse bias.

It now seems we have a set of boundary conditions for various biases. These conditions will be helpful in further analysis. The last comment before continuing the derivation is that equations 9 and 10 apply to the boundaries of the SCR. Why? We have assumed negligible generation and recombination in the SRC but not in the bulk. As a minority carrier travels throughout the p or n-type bulk, recombination is likely. Hence, the concentration deep into the bulk must be less than the concentration at the SCR boundary. Hence, we claim these two equations apply at the boundary since they will only decrease over distance.

The above observations and comments are easily summed into a carrier concentration plot. The plot includes the majority and minority carrier concentration over the length of the device. Say we have a PN junction where $N_A > N_D$, the carrier concentration plot is provided on the following page.

Note that in Fig. 1 only the thermal equilibrium values have been plotted. We stated under the reverse bias that the concentration at the boundary is zero, however, this will be included shortly. Note that $p_{po} > n_{no}$ since $N_A > N_D$. Similarly, $p_{no} > n_{po}$ from the mass-action law and $N_A > N_D$. I encourage you to verify that this is a true statement. This figure indicates that regardless of the bias, deep into the bulk the function must curve to n_{po} and p_{no}. The red and blue lines have been included to provide a visualization of how the carriers diffuse across the junction. The physics of how they traverse the junction is not our focus now but is still included. The complete thermal equilibrium plot is provided below.

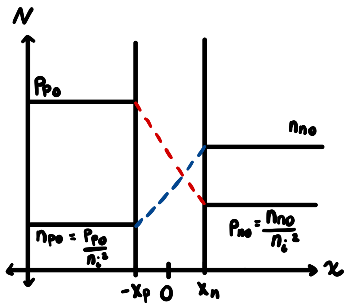

Figure 1: PN junction thermal equilibrium carrier concentration plot.

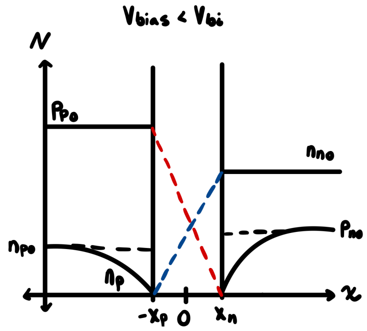

Figure 2: PN junction reverse biased concentration plot.

As stated earlier, under a reverse bias, the minority concentration curves must bend downward toward zero. The equations above show the bending is exponential. However, the sketched figure is not exactly to scale. Toward the bulk, $x \lll -x_p$ and $x \ggg x_n$ the curves converge to n_{po} and p_{no}, respectively. For the forward bias, we expect the curve to reflect upward.

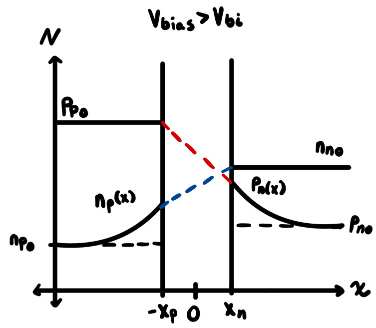

Figure 3: PN junction forward biased concentration plot.

Again, note that the curvature is not to scale. The height of the $p_n(x)$ curve should not be that large since it implies high-level injection. However, I wanted to add emphasis toward the curvature by drawing it large. Regardless, these plots allow us to visualize how the concentration should vary over time. Remember that our analysis has not found the functions $n_p(x)$ and $p_n(x)$, we are only theorizing based on boundary conditions how the curve must look.

Continuity Equation

With the minority carrier curves, we have exhausted all our avenues. To continue our pursuit of the IV characteristic, we need to utilize the continuity equation. For reference,

$$\frac{\partial n}{\partial t} = \mu_n\left(n\frac{d\mathscr{E}}{dx} + \frac{dn}{dx}\mathscr{E}\right) + D_n\frac{d^2n}{dx^2} + (G_n - R_n) \tag{11}$$

$$\frac{\partial p}{\partial t} = -\mu_p (p\frac{d\mathscr{E}}{dx} + \frac{dp}{dx}\mathscr{E}) + D_p\frac{d^2p}{dx^2} + (G_p - R_p) \tag{12}$$

N-type Bulk:

Let's start by examining the n-type bulk; $x \ggg x_n$. We may ignore the G_n term and the electric field terms. Assuming charge neutrality in the bulk implies the electric field throughout the bulk is negligible, or zero. We assume generation in the bulk is negligible but no assumption has been made about recombination. The remaining terms include diffusion and recombination.

$$\frac{\partial p_n}{\partial t} = D_p\frac{\partial^2 p}{\partial x^2} - R_p$$

$$D_P\frac{d^2 p_n}{dx^2} = R_p \tag{13}$$

The $\frac{\partial p}{\partial t}$ term was set to zero because we care about the concentrations at steady state. If we wanted to examine the transient behavior we would need to solve this partial differential equation, however, steady state is more than enough for our purposes. Plugging in the expression for R_p from our generation and recombination discussion,

$$D_P\frac{d^2 p_n}{dx^2} - \frac{p_n - p_{no}}{\tau_p} = 0$$

$$\frac{d^2 p_n}{dx^2} - \frac{p_n}{D_p\tau_p} = -\frac{p_n o}{\tau_p} \tag{14}$$

Hopefully, the form of Eq. 14 is familiar. The equation is a second-order, ordinary differential equation with constant coefficients. This question is not incredibly difficult to solve, the general form of which is,

$$p_n(x) = Ae^{(x\ -\ x_n)/L_p} + Be^{-(x\ -\ x_n)/L_p}\ ,\ \text{for}\ x \geq x_n \tag{15}$$

Where $L_p = \sqrt{D_p\tau_p}$. Recall that the $(x - x_n)$ shifts the solution to start at $x = x_n$. From this point we can find the homogeneous and particular solution to derive the complete $p_n(x)$ equation. However, we should make a quick observation. Is Eq. 15 bounded? As $x \to \infty$ the term multiplied by A will approach infinity. To keep the solution bounded, it is a requirement that $A = 0$. Again, we care about large x because carriers will diffuse in the direction of increasing x.

$$p_{n,homogeneous}(x) = Be^{-(x\ -\ x_n)/L_p}\ ,\ \text{for}\ x \geq x_n \tag{16}$$

The equation above is the homogeneous solution, it does not account for the nonhomogeneous $-\frac{p_{no}}{\tau_p}$ term. p_{no}, τ_p, and D_p are all constants, implying the particular solution must equal a constant, denoted C. Implying,

$$p_n(x) = C + Be^{-(x\ -\ x_n)/L_p}$$

As x becomes large, the B term tends to zero. Implying, $p_n(x \to \infty) = C = p_{no}$. C must equal p_{no} because it is known that given enough space all minority carriers will recombine. With C known, the last term to find is B. The concentration at $x = x_n$ was found at the beginning of these notes. Plugging in, we find:

$$p_n(x_n) = p_{no}e^{qV_{bias}/kT} = C + B$$

$$p_{no}e^{qV_{bias}/kT} = p_{no} + B$$

$$\Rightarrow B = p_{no}\left(e^{qV_{bias}/kT} - 1\right)$$

Substituting the values for B and C, the final minority carrier curve is found.

$$p_n = p_{no} + p_{no}\left(e^{qV_{bias}/kT} - 1\right)e^{-\frac{(x - x_n)}{L_p}} \tag{17}$$

P-type Bulk:

The derivation for the p-type bulk is similar to the n-type. Instead of solving for p_n we solve for n_p. The full derivation will not be included, but the important steps to aid those deriving it themselves. Beginning with the continuity equation. Beginning with the continuity equation,

$$\frac{\partial n_p}{\partial t} = \mu_n(n_p\frac{d\mathscr{E}}{dx} + \frac{dn_p}{dx}\mathscr{E}) + D_n\frac{d^2n_p}{dx^2} + (G_n - R_n)$$

$$0 = D_n\frac{d^2n_p}{dt^2} - R_n$$

$$0 = D_n\frac{d^2n_p}{dt^2} - \frac{n_p - n_{po}}{\tau_n}$$

$$n_p(x) = Ae^{(x\,+\,x_p)/L_n} + Be^{-(x\,+\,x_p)/L_p} + C \text{ , for } x \leq -x_p$$

$$n_p(x \to -\infty) = \infty \Rightarrow B = 0$$

$$n_p(x) = Ae^{(x\,+\,x_p)/L_n} + C \text{ , for } x \leq -x_p$$

$$n_p(x \to -\infty) = C = n_{po}$$

$$n_p(-x_p) = A + C = n_{po}e^{qV_{bias}/kT}$$

Therefore,

$$n_p(x) = n_{po} + n_{po}\left(e^{qV_{bias}/kT} - 1\right)e^{(x\,+\,x_p)/L_n} \tag{18}$$

Current Equations:

With equations for $n_p(x)$ and $p_n(x)$, they may be substituted into the diffusion current equations.

$$J_p = -qD_p\frac{dp}{dx} = \frac{qD_p p_{no}}{L_p}\left(e^{qV_{bias}/kT} - 1\right)e^{(x\,+\,x_p)/L_p} \tag{19}$$

$$J_n = qD_n\frac{dn}{dx} = \frac{qD_n n_{po}}{L_n}\left(e^{qV_{bias}/kT} - 1\right)e^{-(x\,-\,x_n)/L_n} \tag{20}$$

We can evaluate the two equations above at any x within the respective bulk. It is convenient to evaluate the two at x_n and x_p respectively.

$$J_p(-x_p) = \frac{qD_p p_{no}}{L_p}\left(e^{qV_{bias}/kT} - 1\right) \tag{21}$$

$$J_n(x_n) = \frac{qD_n n_{po}}{L_n}\left(e^{qV_{bias}/kT} - 1\right) \tag{22}$$

Our assumption of the current being constant through the space charge region implies the total current through the device is the sum of $J_p(-x_p)$ and $J_n(x_n)$, substituting the respective values,

$$J_{total} = \left(\frac{qD_p p_{no}}{L_p} + \frac{qD_n n_{po}}{L_n}\right)\left(e^{qV_{bias}/kT} - 1\right) \tag{23}$$

$$I_{total} = A\left(\frac{qD_p p_{no}}{L_p} + \frac{qD_n n_{po}}{L_n}\right)\left(e^{qV_{bias}/kT} - 1\right) \tag{24}$$

$$\text{Let } I_s = A\left(\frac{qD_p p_{no}}{L_p} + \frac{qD_n n_{po}}{L_n}\right) \tag{25}$$

$$\Rightarrow I = I_s\left(e^{qV_{bias}/kT} - 1\right) \tag{26}$$

The final equation should be iconic to those who have learned diode behavior in Electronics 1. We have derived the standard current equation of a diode from scratch. Let's take a moment to discuss the past few steps and marvel at the fact that such a standard equation in electronics took many pages to derive. Eq. 23 and 24 are interchangeable depending if you want the net current or current density through the device. Instead of writing the long list of constants, we abstract the term into the current density. Since we know the argument of the exponential is unitless Eq. 25 must have units of amps. It is incredibly intriguing that the saturation current depends only on the material properties of the junction. We can tune these various parameters to design some desired device characteristics.

Example:

To reiterate the incredibly small magnitude of I_s, let's use some practical values. Say we have a device with:

$N_A = 5 \cdot 10^{16}$ cm^{-3}

$N_D = 10^{16}$ cm^{-3}

$n_i = 1.5 \cdot 10^{10}$ cm^{-3}

$D_n = 21 \frac{\text{cm}^2}{s}$

$D_p = 10 \frac{\text{cm}^2}{s}$

$\tau_p = \tau_n = 0.5 \mu s$

$A = 2 \cdot 10^{-4} \text{cm}^2$

All we need to do is plug these various values into Eq. 23,

$$n_{po} = \frac{{n_i}^2}{N_D} = \frac{2.25 \cdot 10^{20} \text{cm}^{-6}}{10^{16} \text{cm}^{-3}} = 22,500 \text{cm}^{-3}$$
$$p_{no} = \frac{{n_i}^2}{N_A} = \frac{2.25 \cdot 10^{20} \text{cm}^{-6}}{5 \cdot 10^{16} \text{cm}^{-3}} = 4,500 \text{cm}^{-3}$$
$$L_n = \sqrt{21 \frac{\text{cm}^2}{s} * 0.5 \cdot 10^{-7} s} = 10.2 \mu\text{m}$$
$$L_p = \sqrt{10 \frac{\text{cm}^2}{s} * 0.5 \cdot 10^{-7} s} = 7.07 \mu\text{m}$$
$$I_s = 1.8 \cdot 10^{-15} A$$

An interesting observation is the incredibly small concentrations at thermal equilibrium. The inverse relationship of $n_{no}p_{no}$ and $n_{po}p_{po}$ implies that increasing the doping concentration will decrease the minority concentration. A small distribution of minority carriers will not contribute to a large reverse saturation current. Therefore, it seems that we want to dope each region as large as possible. However, there are practical issues with this, one of which is called breakdown. Breakdown limits the range of voltages that can be applied to the device without permanently damaging the device. Feel free to search avalanche and Zener breakdown for more information.

IV Characteristic Plot:

Combining all the information so far, we can finally draw the IV curve of a PN junction.

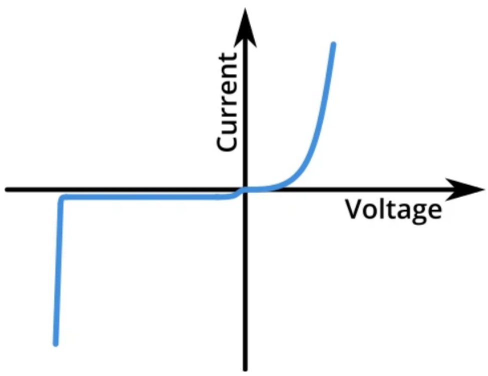

Figure 4: PN junction IV characteristic curve. ([2])

The image provided by Oscilla was included to show, with accuracy, the reverse saturation current. It can be difficult to have the saturation current to be close to zero but not zero. For now, ignore the dip in the far left side of the figure. This point corresponds to breakdown and the device becomes permanently damaged.

When we have a strong reverse bias, the built-in field adds, or is strengthened, by the field produces by the voltage source.

Energy Band Diagram:

The last part of our analysis relates this wealth of analysis to the energy band diagram. The energy band diagram under thermal equilibrium was introduced in the device introduction notes. However, NanoHub has an interesting simulation that will let us visualize the diagram under various biasing voltages. At thermal equilibrium, the following plot was generated.

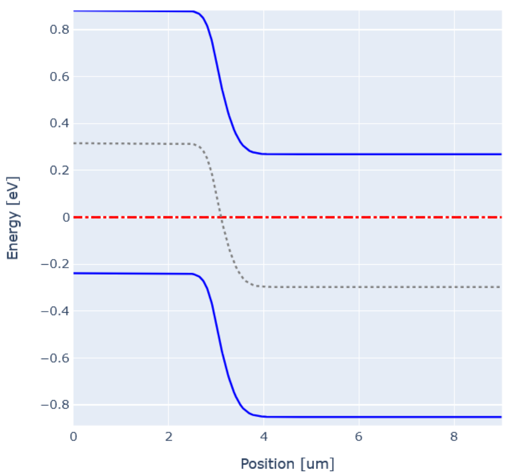

Figure 5: PN junction energy band diagram at thermal equilibrium.

Note the two blue lines correspond to the conduction and valence bands. The grey line represents the intrinsic level, and the red the Fermi level. Remember in our analysis the Fermi level is the reference. Hence, the level is centered at 0 eV. The left side of the figure is the p-type since E_F is near E_v. This means the n-type must be to the right. Now we may ask what would happen if we increased the reverse bias? The slope in the depletion region is related to the electric field $\mathscr{E} = \frac{dE_i}{dx}$. Implying that increasing the electric field will increase the slope, forcing the n-type region to move downward.

Thinking of regions moving upward or downward is ambiguous. Another way to visualize the shifting of bands is that adding electrons will shift a band upward. If we reverse bias the junction, the positive terminal is attached to the n-type and the negative terminal to the p-type. The metal will inject electrons into the p-type, causing a vertical shift. With a 1V reverse bias, the following curve was generated.

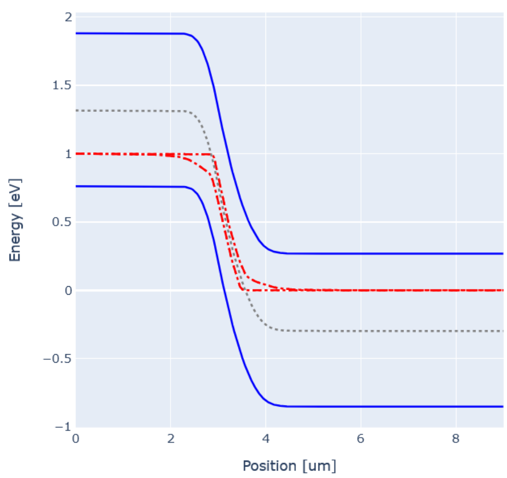

Figure 6: PN junction reverse bias energy band diagram. ([1])

Note that the Fermi level is no longer constant, which is true under an excitation. We see a discontinuity between the n and p-type bulk. The p-type bulk is held at 1 eV whereas the n-type is at 0 eV. The p-type Fermi level has shifted because of the 1 V reverse bias. We will discuss how the two Fermi levels are stitched together shortly.

Avalanche Breakdown:

Imagine if we dropped an electron within the depletion region. The large slope metaphorically accelerates the electron down the hill. The large field will induce significant momentum into a carrier. The carrier can then collide with other covalent bonds, and with sufficient energy, the bond will break. The generated electron-hole pair is swept away by the electric field and the process continues. This process is called Avalanche breakdown and results in the large negative current for negative voltages in the IV characteristics. The voltage where breakdown occurs is called V_B and is described by,

$$V_B = \frac{\mathscr{E}_{critical} w_{breakdown}}{2} \tag{27}$$

Where the critical electric field is the field strength in breakdown and the breakdown width is the width of the depletion region under breakdown. Typical values for the critical electric field are in the order of hundreds of kV per cm. This may seem incredibly large but 1 V applied to a 1 μ m device produces a field of 10 kV over 1 cm. We can find the depletion region from our standard PN junction equations. The algebra will simplify to a quadratic equation and assumptions can be made if the junction is n+ or p+. However, the point is that breakdown voltage is in the realm of tens of volts. However, it is heavily dependent on the doping of the device, large doping produces small breakdown voltages.

If we reverse the potential and apply approximately 0.4 V of forward bias we generate the following diagram.

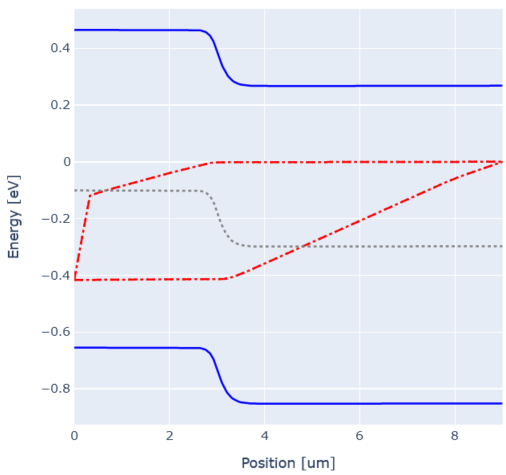

Figure 7: PN junction energy band diagram under small excitation.

I encourage you to ignore the drastic slope in the Fermi levels at first. Notice instead notice the height between the bulk conduction band. The slope of the depletion region is less than the prior simulations, implying the built-in field is weakening.

Quasi Fermi Level:

As for the weird polygon representing the Fermi level, these are called the **Quasi Fermi Levels**. The quasi Fermi levels were introduced in the Generation and Recombination, but are also useful for general excitations like voltage sources. There is a quasi level for the p-type, E_{Fp} and the n-type E_{Fn}. They are similar to the Fermi level but occur only under an excitation. The quasi levels attempt to extend the property of $\frac{dE_F}{dx} = 0$ by bending the Fermi level linearly. Beyond this, quasi levels are beyond the scope of this course, and we opt to draw nothing for the Fermi level in the depletion region under excitation.

Conclusion:

This set of notes wraps the several-part "saga" of the PN junction. Now that we have a firm understanding intuitively of how the junction behaves, and a set of expressions deriving the behavior we can analyze more interesting devices. The next device we will look at is the Bipolar-Junction Transistor or BJT for short. This set of notes was quite analytic, so please take the moment to justify the several assumptions made and the derivations. I cannot iterate how important it is to have a firm foundation with the PN junction because every subsequent topic builds upon it.

References

[1] Klimeck Daniel Mejia & Gerhard. *PN Junction Lab.* NanoHub https://nanohub.org/resources/pnjunctionlab. 2021.

[2] Ossila. *I-V Curves: A Guide to Measurement.* URL: https://www.ossila.com/pages/iv-curves-measurement.