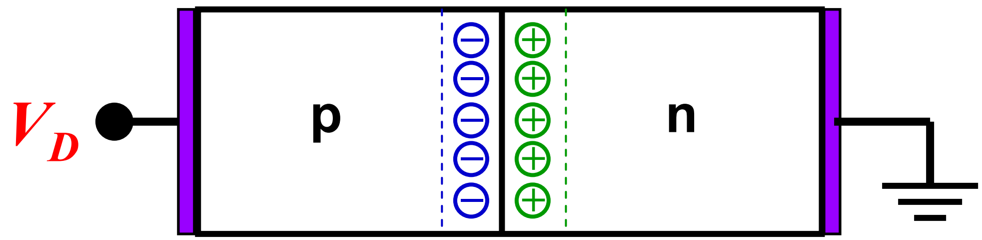
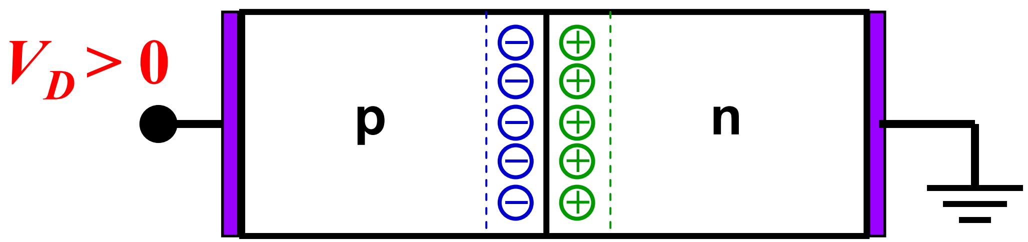
Effect of Applied Voltage



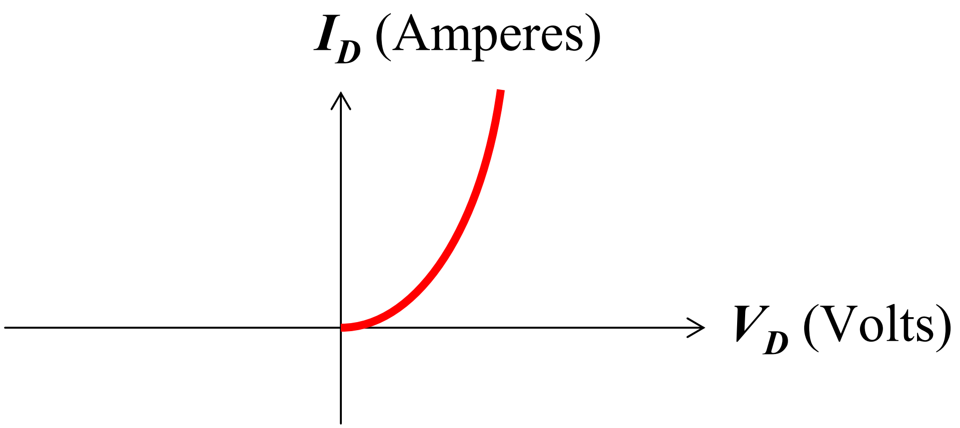
- The quasi-neutral p and n regions have low resistivity, whereas the depletion region has high resistivity. Thus, when an external voltage V_D is applied across the diode, almost all of this voltage is dropped across the depletion region. (Think of a voltage divider circuit.)
- If V_D > 0 (***forward bias***), the potential barrier to carrier diffusion is reduced by the applied voltage.
- If V_D < 0 (***reverse bias***), the potential barrier to carrier diffusion is increased by the applied voltage.

Forward Bias

- As V_D increases, the potential barrier to carrier diffusion across the junction decreases*, and current increases exponentially.



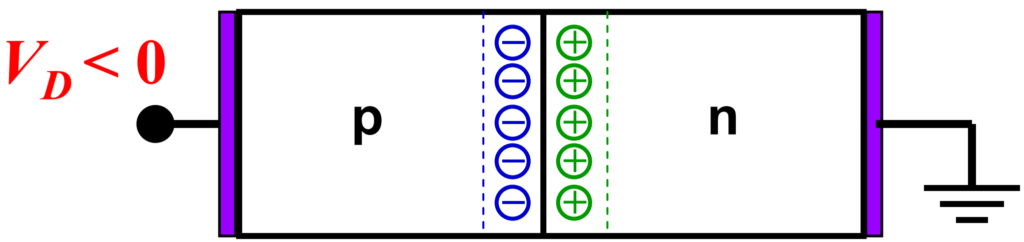
The carriers that diffuse across the junction become minority carriers in the quasi-neutral regions; they then recombine with majority carriers, "dying out" with distance.



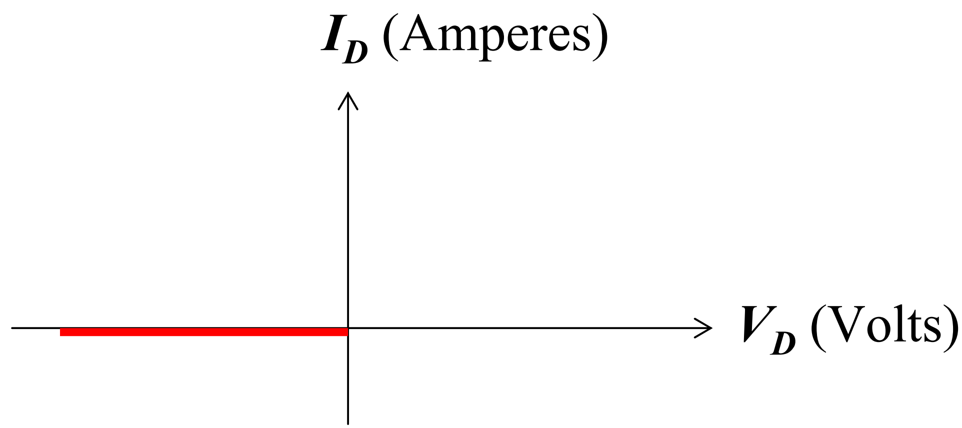
* Hence, the width of the depletion region decreases.

Reverse Bias

- As $|V_D|$ increases, the potential barrier to carrier diffusion across the junction increases*; thus, no carriers diffuse across the junction.



A very small amount of reverse current ($I_D < 0$) does flow, due to minority carriers diffusing from the quasi-neutral regions into the depletion region and drifting across the junction.



* Hence, the width of the depletion region increases.

1

To understand how the p-n junction behaves under biasing, we start with the unbiased junction. We use our usual convention as in the preceding lectures that the p-type region is on the left and the n-type region is on the right. See Figure 1. We use the terminology "bulk p-type region" to discuss the portion of the p-type region that is to the left of the depletion layer and the terminology "bulk n-type region" to discuss the portion of the n-type region that is to the right of the depletion layer.

We recall that at the edge of the depletion region in the bulk p-type region (and throughout the bulk p-type region) the majority carrier concentration (density of holes) was N_a, the minority carrier concentration (density of electrons) was $\frac{n_i^2}{N_a}$ and the potential was $-V_T \ln \frac{N_a}{n_i}$, where $V_T = \frac{kT}{q}$ denotes the thermal voltage.

Similarly, at the edge of the depletion region in the bulk n-type region (and throughout the bulk n-type region) the majority carrier concentration (density of electrons) was N_d, the minority carrier concentration (density of holes) was $\frac{n_i^2}{N_d}$ and the potential was $V_T \ln \frac{N_d}{n_i}$.

Finally, we recall that the built-in potential, i.e. the total rise in potential as we move from the bulk p-type region to the bulk n-type region, was

$$\phi_B = V_T \ln \frac{N_a N_d}{n_i^2}.$$

This entire potential rise occurs across the depletion region.

Now, observe that the ratio of the density of electrons at the edge of depletion region in the bulk p-type region to the density of electrons at the edge of the depletion region in the bulk n-type region is

$$\frac{n_i^2}{N_a N_d} = e^{-\frac{\phi_B}{V_T}}$$

Similarly, the ratio of the density of holes at the edge of the depletion region in the bulk n-type region to the density of holes at the edge of the depletion region in the bulk p-type region is also

$$\frac{n_i^2}{N_a N_d} = e^{-\frac{\phi_B}{V_T}}$$

The fact that these ratios are related to the difference in potential between these points in this exponential way is a general principle of statistical mechanics, called *Boltzmann's principle*. This principle will form the basis of our discussion of the p-n junction under biasing. Boltzmann's principle says that, in equilibrium, the ratio of the concentrations of particles at a location x to that at a location x_0 is proportional to the exponential of the negative of the difference between the per-particle potential energy at x to the per-particle potential energy at x_0, measured in units of thermal energy (in the case of charged particles in an electric potential, as we have here, the unit of thermal energy is kTq; we just happen to have expressed the ratio of energies as a ratio of potentials).

Our earlier analysis of the carrier concentrations across an unbiased p-n junction is consistent with Boltzmann's principle not just between the bulk regions on the two sides of the junction, as shown above, but also throughout the junction. Indeed, in Section 4 of Lecture 35 you will see the equation,

$$\phi(x) - \phi(x_0) = V_T \ln \frac{n(x)}{n(x_0)}$$

where x and x_0 are the locations of any two cross-sections of the junction and $n(x)$ and $n(x_0)$ are respectively the density of electrons at those two locations. This can be rearranged to read

$$n(x) = n(x_0) e^{\frac{\phi(x)-\phi(x_0)}{V_T}}$$

This is an instance of Boltzmann's principle (recall that electrons prefer locations with higher potential). Similarly, for instance using the mass action law, you can rewrite the same equation as

$$p(x) = p(x_0) e^{-\frac{\phi(x)-\phi(x_0)}{V_T}}$$

where $p(x)$ and $p(x_0)$ are respectively the density of holes at the two locations., which also respects Boltzmann's principle.

The Maxwell distribution on the velocities of particles confined to a box in thermal equilibrium at fixed temperature, which we discussed in Section 4 of Lecture 32 (and derived in Section 5, which was optional) is also consistent with Boltzmann's principle. [1]

2

Suppose now that the p-n junction is *forward biased* with a positive bias voltage $0 < V_D < \phi_B$. What this means is that potential on the p-side of the junction is raised. This means that the depletion region will narrow, because less of a potential rise is needed across the depletion region to get from the bulk p-type to the bulk n-type region. Indeed, the potential rise needed is exactly $\phi_B - V_D$ and occurs across the depletion region.

At the edge of the depletion region in the bulk n-type region the majority carrier concentration (density of electrons) is still roughly N_d (because N_d is so large, we can ignore any changes in the majority carrier concentration due to biasing). By applying Boltzmann's principle, we conclude that at the edge of the depletion region in the bulk p-type region, the minority carrier concentration (density of electrons) must be

$$n_p(-x_p) = N_d e^{-\frac{\phi_B - V_D}{V_T}} .$$

Here $-x_p$ denotes the edge of the depletion region in the bulk p-type region, and the notation $n_p(\cdot)$ reminds us that we are talking about electrons in the p-type region. See Figure 2.

Similarly, at the edge of the depletion region in the bulk p-type region the majority carrier concentration (density of holes) is still roughly N_a. By applying Boltzmann's principle, we conclude that at the edge of the depletion

[1] The consistency is most easily checked by considering each coordinate. In Section 5 we showed that the fraction of molecules whose velocity in the x-coordinate is v_x is given by $g(v_x) = (\frac{m}{2\pi kT})^{\frac{1}{2}} e^{-\frac{1}{2}\frac{mv_x^2}{kT}}$. The ratio of the fraction of particles having the x-component of the kinetic energy equal to E_x to those having the x-component of the kinetic energy equal to E_x^0 is therefore $e^{-\frac{E_x - E_x^0}{kT}}$, which is consistent with Boltzmann's principle.

region in the bulk n-type region, the minority carrier concentration (density of holes) must be

$$p_n(x_n) = N_a e^{-\frac{\phi_B - V_D}{V_T}}.$$

Here x_n denotes the edge of the depletion region in the bulk n-type region, and the notation $p_n(\cdot)$ reminds us that we are talking about holes in the n-type region. See Figure 2.

We say that minority carriers have been *injected* into the corresponding regions from the other side. Namely, electrons have been injected into the p-type region from the n-type region and holes have been injected into the n-type region from the p-type region.

What happens in the p-n junction under forward bias is that each of the minority carriers *diffuse* through the corresponding bulk region till they reach the metal gate at the corresponding end. Hole-electron pairs also get removed by recombination in the bulk regions because there is an imbalance in the mass action law because of the injection of minority carriers. The concentration of the minority carriers can be thought of as decreasing from the relatively high level at the edge of the depletion region back to the level they would normally have in the bulk region when we get to the metal contact at the edge of the region. See Figure 2. Specifically, the density of electrons in the bulk p-type region increases from $N_d e^{-\frac{\phi_B}{V_T}}$ at $-W_p$, i.e. at the metal contact to the left, to $N_d e^{-\frac{\phi_B - V_D}{V_T}}$ at $-x_p$. Similarly, the density of holes in the bulk n-type region decreases from $N_a e^{-\frac{\phi_B - V_D}{V_T}}$ at x_n to $N_a e^{-\frac{\phi_B}{V_T}}$ at W_n, i.e. at the metal contact on the right.

The precise description of the profile of the concentrations of the majority and minority carries in the bulk regions is quite involved. You will need to take more advanced courses to fully understand this. However, the discussion we have engaged in so far is already sufficient to understand the origin of the Shockley equation for the I/V characteristic of a diode in the forward bias region. The main point is that the equations describing the minority carrier concentration in the bulk regions (i.e. the equations that tell us how to compute $n_p(x)$ for $-W_p \leq x \leq -x_p$ and how to compute $p_n(x)$ for $x_n \leq x \leq W_n$ are *linear* equations). When the applied forward bias is V_D, we saw that $n_p(x)$ needs to decrease from $\frac{n_i^2}{N_a}e^{\frac{V_D}{V_T}}$ at $-x_p$ to $\frac{n_i^2}{N_a}$ at $-W_p$. Similarly, $p_n(x)$ needs to decrease from $\frac{n_i^2}{N_d}e^{\frac{V_D}{V_T}}$ at x_n to $\frac{n_i^2}{N_d}$ at W_n. The decrease required of $n_p(x)$ is proportional to $e^{\frac{V_D}{V_T}} - 1$, as is the decrease required of $p_n(x)$! Together with the linearity of the describing equations, this explains the form of Shockley's equation in the forward bias region.

We can be somewhat more precise in the case when the length of the bulk p-type and bulk n-type regions is sufficient small that we can approximate the profile of the minority carrier concentrations in the respective regions by straight lines (this called the case of the *short-base diode*). In this case the gradient of the minority carrier concentration in the bulk p-type region can be taken to be constant throughout the range $-W_p \leq x \leq -x_p$, equal to

$$(e^{\frac{V_D}{V_T}} - 1)\frac{n_i^2}{N_a}\frac{1}{W_p - x_p}.$$

Likewise, the gradient of the minority carrier concentration in the bulk n-type region can be taken to be constant throughout the range $x_n \leq x \leq W_n$, equal to

$$(e^{\frac{V_D}{V_T}} - 1)\frac{n_i^2}{N_d}\frac{1}{W_n - x_n}.$$

We further make the approximation that x_p and x_n are very close to zero. One can argue in this case (you will need to take more advanced courses to understand this) that the overall current is essentially the sum of the current due to diffusion of the electrons as minority carriers through the bulk p-type region and the current due to diffusion of holes as minority carriers through the bulk n-type region. We know that these can be computed from the gradient of the concentration profile. From this observation and the approximations we made, we get an explicit formula for the current through the diode as a function of the applied forward bias voltage V_D:

$$I_D = qn_i^2(\frac{D_p}{N_d W_n} + \frac{D_n}{N_a W_p})(e^{\frac{V_D}{V_T}} - 1)$$

where D_p denotes the diffusion coefficient for holes and D_n denotes the diffusion coefficient for electrons. Note that this way of writing the formula gives an explicit expression for I_s (what we called the saturation current) in the Shockley formula $I_D = I_s(e^{\frac{V_D}{V_T}} - 1)$. [2]

3

We turn next to the case of the *p*-*n* junction diode under reverse bias. In this case the potential in the *p*-type region is lowered. This will lead to an increase in the potential rise that needs to occur across the depletion later, to $\phi_B - V_D$ (where now the applied bias V_D is negative). Using Boltzmann's principle once again, we conclude as before that the minority carrier concentration (concentration of electrons) at the edge of the depletion layer in the bulk *p*-type region (i.e. at $-x_p$ in Figure 3) is $N_d e^{-\frac{\phi_B - V_D}{V_T}}$, which is in fact much smaller than $N_d e^{-\frac{\phi_B}{V_T}} = \frac{n_i^2}{N_a}$, because now $V_D < 0$. Similarly, the minority carrier concentration (concentration of holes) at the edge of the depletion layer in the bulk *n*-type region (i.e. at x_n in Figure 3) is $N_a e^{-\frac{\phi_B - V_D}{V_T}}$, which is in fact much smaller than $N_a e^{-\frac{\phi_B}{V_T}} = \frac{n_i^2}{N_d}$, because $V_D < 0$. The concentration of electrons in the bulk *p*-type region, $n_p(x)$ for $-W_p \leq x \leq -x_p$, will now decrease from $\frac{n_i^2}{N_a}$ at $-W_p$ to $\frac{n_i^2}{N_a} e^{\frac{V_D}{V_T}}$ at $-x_p$, while the concentration of holes in the bulk *n*-type region, $p_n(x)$ for $x_n \leq x \leq W_n$, will now increase from $\frac{n_i^2}{N_d} e^{\frac{V_D}{V_T}}$ at x_n to $\frac{n_i^2}{N_d}$ at W_n, as shown in Figure 3. In the case of the short-base diode, as illustrated in Figure 3, the profile of the minority carrier concentrations in the respective bulk regions can be taken to be linear. The gradient of the concentration of electrons in the bulk *p*-type can be approximated by approximating the concentration at the edge of the depletion layer as zero, and so roughly equals

$$\frac{n_i^2}{N_a} \frac{1}{W_p - x_p}.$$

Similarly, the gradient of the concentration of holes in the bulk *n*-type can be approximated by approximating the concentration at the edge of the de-

[2] The Shockley equation, (10.1) of the text, also had a term called the *emission coefficient*. This is also called the *ideality factor*. It can be viewed as a model-fitting device to allow for nonidealities in the diode. We have "derived" the equation assuming an ideal diode, thus getting an emission factor of 1.

pletion layer as zero, and so roughly equals

$$\frac{n_i^2}{N_d}\frac{1}{W_n - x_n}.$$

We may further approximate x_n and x_p as roughly zero. The current is once again the sum of that carried by the electrons as minority carriers through the bulk p-type region and the holes as minority carriers through the bulk n-type region. We get a formula for the current through the diode in reverse bias as:

$$I_D = qn_i^2(\frac{D_p}{N_d W_n} + \frac{D_n}{N_a W_p}).$$

Note that this equals I_s as defined in the preceding section. This completes our "derivation" of Shockley's equation.

4

The discussion above needs to be modified at high reverse bias voltages. There are two distinct phenomena by which this occurs. One is *Zener breakdown*, which occurs in the case of narrow depletion layers (i.e. when the bulk regions are highly doped). In this case the electric field across the depletion region in reverse bias can become sufficiently large (at reverse bias voltages of a few volts) that electrons are torn out the covalent bonds in the bulk p-type region and swept into the bulk n-type region (recall that electrons want to move in a direction opposite to that of the electric field). This is the breakdown phenomenon underlying Zener diodes. Another phenomenon is *avalanche breakdown*. This happens in diodes with a relatively thicker depletion region (corresponding to more moderate doping densities in the bulk regions). Electrons accelerated in the electric field of the depletion region then get enough energy to knock other electrons out of their bonds (a process called impact ionization) which then creates a chain reaction (like an avalanche), so the diode breaks down.

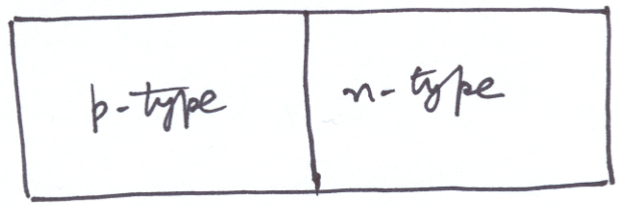


Figure 1

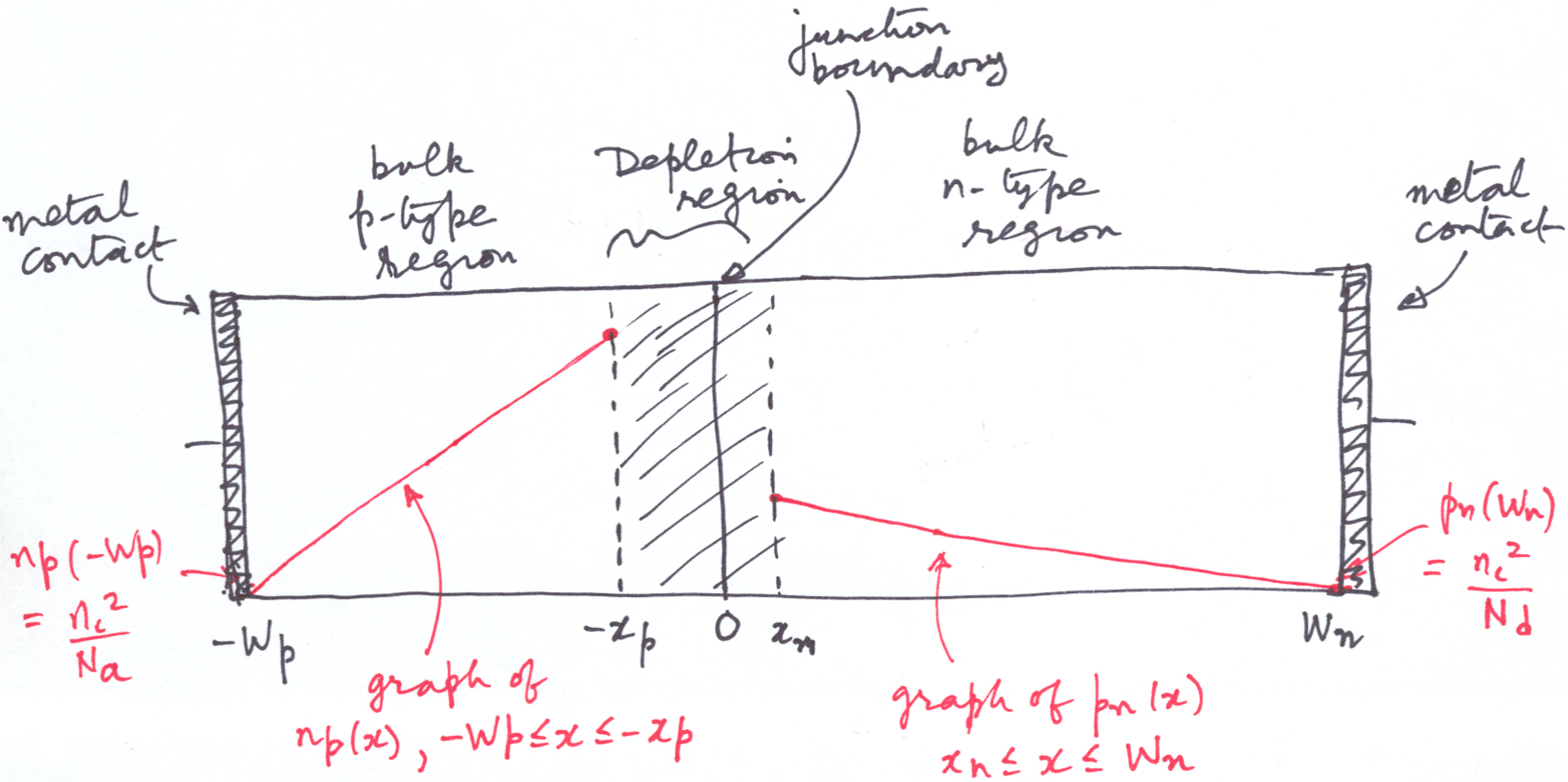


Figure 2: p-n junction under forward bias. The short-base case is shown. The profiles are curved in general

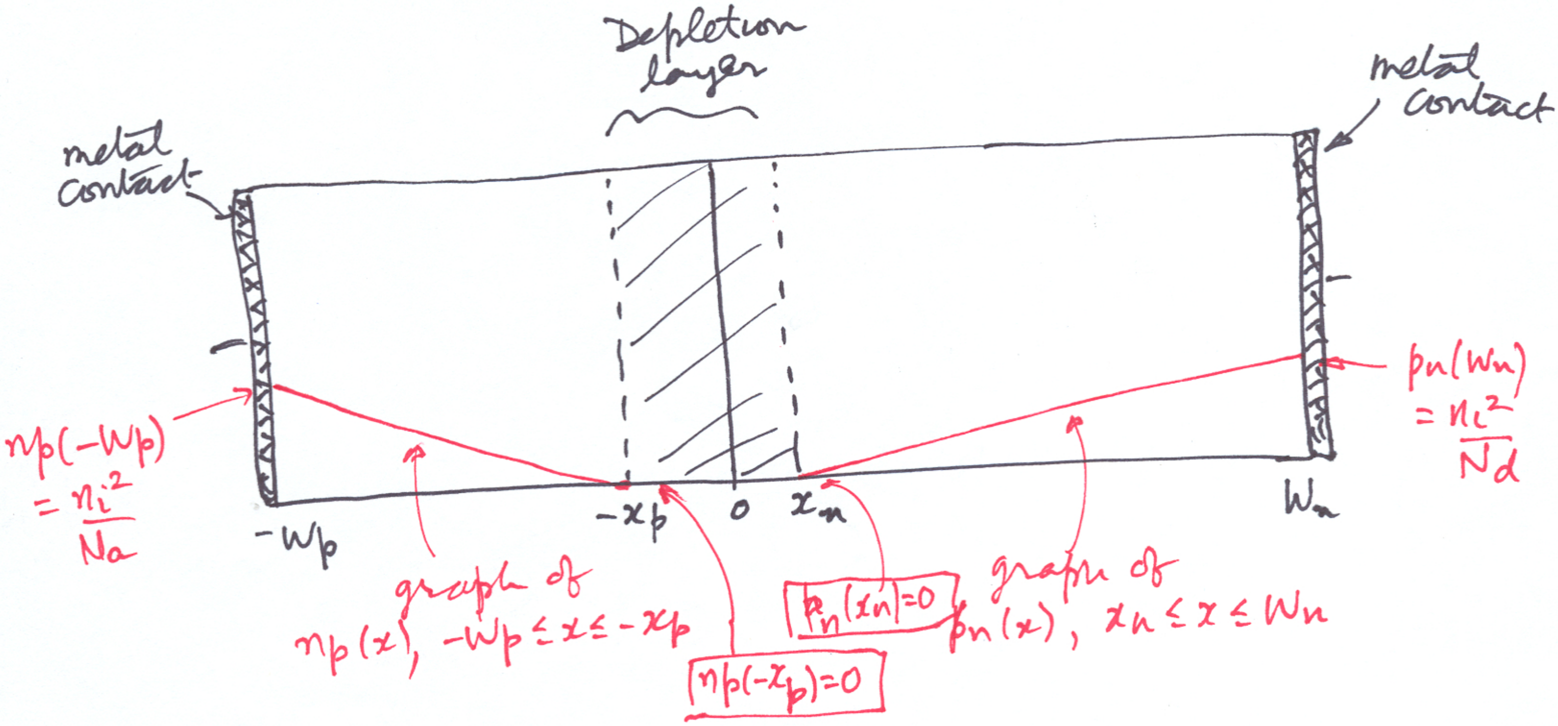


Figure 3

p-n junction under reverse bias.
The short base case is shown
Even though the notation for the depletion layer boundaries is the same as in Figure 2, the actual values x_n and x_p differ.