Fundamentals of electronic materials, devices and fabrication

Dr. S. Parasuraman

Department of Metallurgical and Materials Engineering

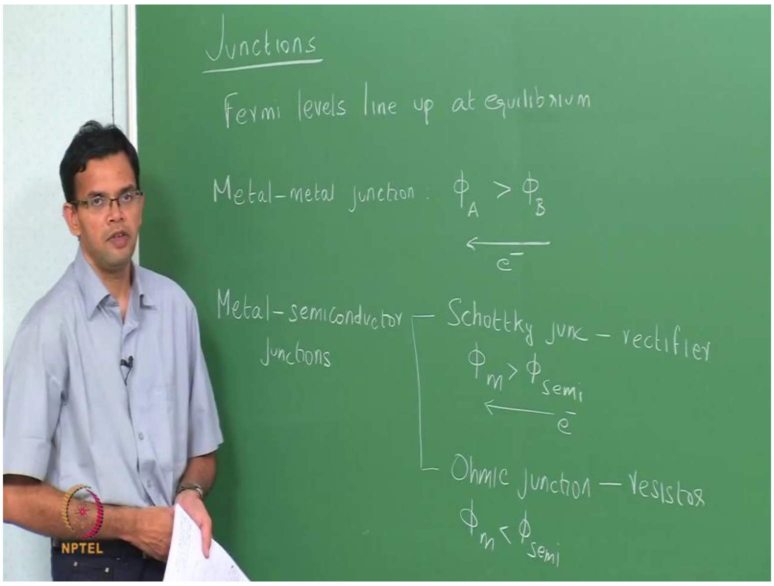
Indian Institute Technology, Madras

Lecture - 10

Pn junctions in equilibrium

Let us start with a brief review of last class. Last class we started looking at junctions.

(Refer Slide Time: 00:24)



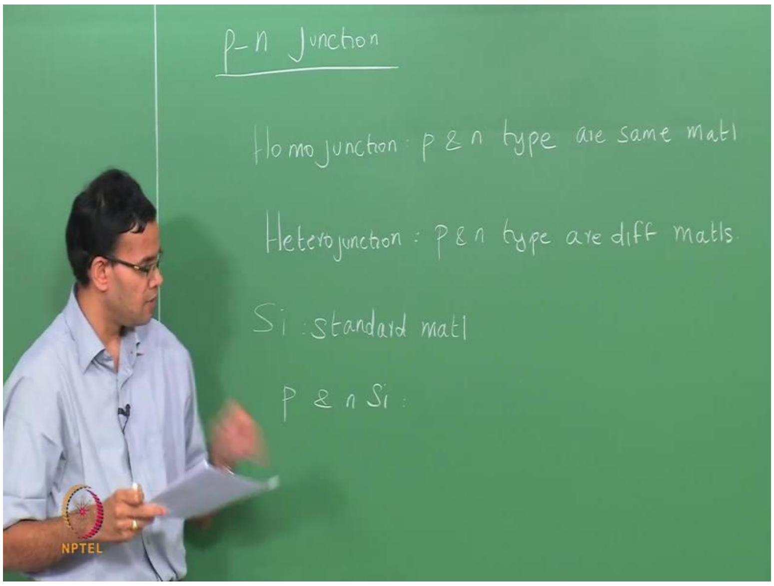
These junctions can be formed between 2 semiconductors or between a metal and a semiconductor or between 2 metals. We were not concerned about how we will form the junction; we just assume that we have an ideal junction or an ideal interface between the 2 materials with no defects. Whenever we have a junction and junction is at equilibrium, we said that the Fermi levels must line up. So, the Fermi levels line up at equilibrium, and Equilibrium is defined as when we have no external potential applied to the system. We first looked at a Metal-metal junction, so we have 2 metals A and B with work functions φ_A and φ_B. And, we also said that $\varphi_A > \varphi_B$; in such a case, we found out that electrons will move from B to A, and this will create a contact potential.

We then looked at metal semiconductor junctions. And, we saw that there were 2 types; the first one was called the Schottky Junction. So, this is a junction where the work function of the metal we call it φ_m is greater than the work function of a semiconductor.

In such a case, we found that electrons will travel from the semiconductor to the metal, and this will again create a depletion region in the semiconductor and a contact potential. We also looked at the Schottky junction under bias and found that it behaves as a rectifier. So, that in the case of forward bias, the junction will conduct, but when we apply a reverse bias there is a small saturation current, but the junction does not conduct.

The other type of junction that we saw was the Ohmic junction. In this case, the work function of the metal is smaller than that of the semiconductor, so that you have electrons flowing the other way and the Ohmic junction behaves as a resistor. This is where we left of from our last class. Today, we are going to look at a junction between 2 semiconductors. So, we are going to start with a p-n junction.

(Refer Slide Time: 03:58)



p-n Junction

Homojunction : p & n type are same matl

Heterojunction : p & n type are diff matls

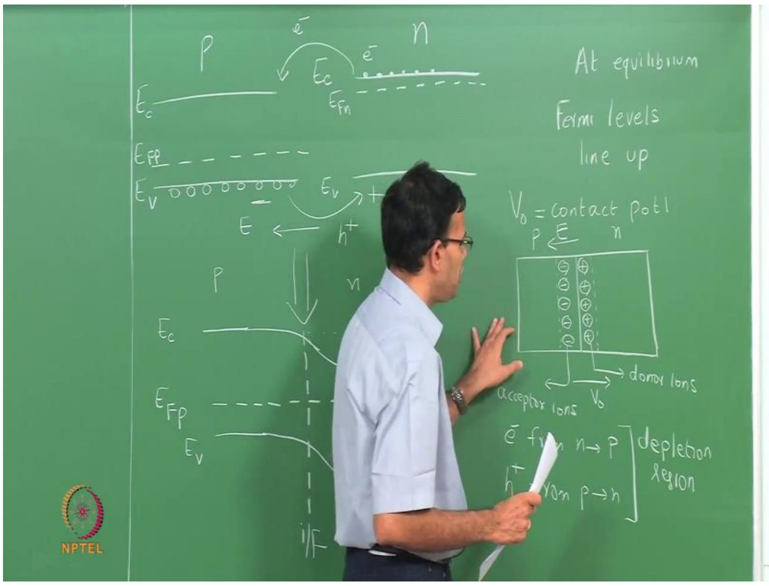
Si : standard matl

p & n Si :

As the name implies, a p-n junction is formed between a p-type semiconductor and an n-type semiconductor. Usually, both are formed from the same material, and such a kind of junction is called the Homo junction, these are where both your p and n type are from the same material. We can also have a junction found between different materials, in such a type of junction is called a Hetero junction. So, in this case p and n are different materials. We will mostly look at homo junctions, where they are the same material and we will develop concepts of the junction in the equilibrium and under bias and also calculations for the current. And towards the end, we will also look at the effect of having different materials. So, we will find that hetero junctions have some really interesting properties and later, when we talk about optical devices as well like, led or lasers, we will find that hetero junctions have some advantage there.

Homo junctions are easier to grow because you are essentially the same material, all you are doing is doping one side p-type and other side n-type. So, the interfaces are easier to form. In the case of hetero junctions, the materials have to be chosen carefully so that we have a good interface with no defects, so that posses some restrictions on the kinds of junctions they are formed. But for the analysis, we are going to do now; we are going to assume that we have an ideal junction that is there are no defects. Most of the time we will be talking about these p-n junctions with respect to silicon, so silicon is our standard material of choice, but we will also look at some examples of other materials. And when we come to hetero junctions, we will talk a lot about compound semiconductors. So, let us consider a junction formed between p and n silicon. So, let me start by drawing the band diagram for p and n type silicon. So, we will first put them far apart and then bring them together to form the junction.

(Refer Slide Time: 07:16)



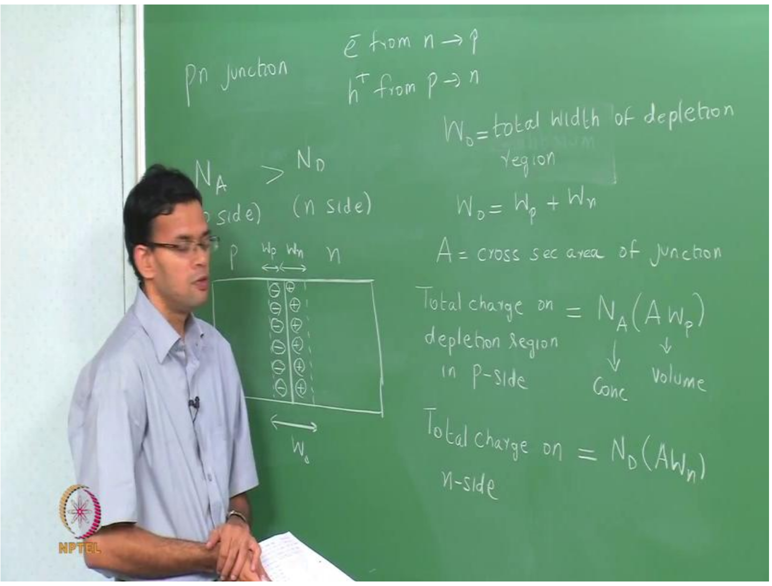
So, here is the band diagram for p silicon, we have a conduction band - E_c, a valence band E_v. So, E_c is the bottom of the conduction band, E_v is the top of the valence band. The material is p-type, so we know that the Fermi level lies closer to the valence band. So, let me put the Fermi level closer to the valence band, and call it E_{FP}. Now, I have n-type silicon; once again I have E_c , E_v and then E_{Fn}. So, the position of E_c and E_v does not change because both of them correspond to silicon, so the band gap is maintained. The only difference is where replace the Fermi level. So, we bring this p and n material together in order to form your p-n junction.

And we said the first rule is at equilibrium the Fermi levels must line up. Another way of thinking about it is that in the n-side, you have excess electrons in the conduction band, these electrons can move to the p-side; on the p-side, you have excess holes, these are in the valence band, and these holes can move to the n-side. Electrons moving here, holes moving the other way, because we have electrons moving we will have a net positive charged on the n-side because we have holes moving, we will have a net negative charge on the p-side. So, once again we will have a contact potential between the p-n junctions.

So, let me draw the junction in equilibrium. The first thing I will do is line up the Fermi level. Let me mark the interface region. So, this I will call I F, so I F represents the interface between the p and n. Far away from the junction the material behaves as a p or a n-type. So, far away from junction, I have a p, and I have an n. Now we said that electrons move from n to p, so that we have a net positive charge on the n-side, and a net negative charge on the p-side, which means there is an electrical field. So, the electric field goes from positive to negative. In last class, we also saw that whenever we have an electric field, we have band bending and the bands bend in the direction of the fields. So, the bands bend up in the direction of E. So, we have the bands bending on the n-side up, we have bands bending on the p-side down, so that is how the junction forms. You can do the same for valence band. So, this is my conduction band - E_c, this is my valence band – E_v, and I have a contact potential at the junction. So, V_o represents the contact potential.

So, we can draw a simplified picture of this. So, let me just draw a schematic. So, this is my p-side that is my n-side to form my p-n junction. Electrons move from the n to the p. So, when these electrons move, you are left with donor ions, since donor ions donate an electron and when these electrons move, they have a net positive charge. So, there is a net positive charge on the n-side, so this net positive charge refers to the donor ions. Similarly, there is a net negative charge on the p-side, because the electrons or the holes from p-side move to the n-side leaving behind your acceptor ions and the acceptor ions are negative, which means you have an electric field E and a built-in potential V_o. So, we have excess electrons from n to p, you have excess holes moving from p to n and when these meet they recombine and get destroyed. So, what you are left behind is a depletion region around the junction. So, let me redraw this diagram and mark the depletion region.

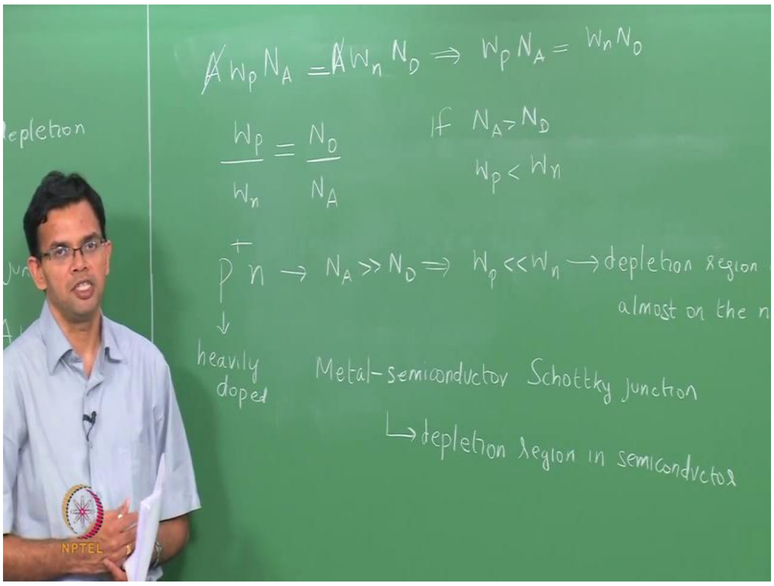
(Refer Slide Time: 14:06)



So, we are forming a p-n junction. So, in the case of a p-n junction, we said that we have excess electrons on the n-side move to p. We can think of this in terms of diffusion, where diffusion usually occurs from a higher concentration gradient to a lower gradient. So, the higher concentration of electrons is on the n-side, and this moves into p. Similarly, we have holes which move from to p-side to the n-side, and when these meet they recombine and form a depletion region. So, let us consider a p-n junction where the acceptor concentration N_A - this is on the p-side is greater than the donor concentration N_D and this is on the n-side. So, let me redraw my p-n junction, this is the p that is the n; there is a depletion region that is formed because of the diffusion of carriers, we call this W_o, so W_o is the total width of the depletion region. So, this forms at the interface of the p-n junction and it extends to both the p and the n side. So, we have some width on the p-side, let me call W_p; some width on the n-side W_n, so that W_o is just the sum of the depletion region in the p side and the depletion region on the n-side.

So, we said that in the case of a p-n junction, we have a net charge on the n-side, because the electrons are gone leaving behind the positively charged donors; and you have a net negative charge on the p-side, because the holes are gone leaving behind the acceptor ions. So, if A is the cross section of the junctions, so A is the cross-sectional area of the junction, we must have a balance of charge between the p and the n-side. So, the total charge on the depletion region in the p-side is nothing but the concentration of acceptor ions times the volume; the volume is A x W_p. So, this is the concentration, this is the volume. We can similarly calculate the total charge on the n-side, which is N_D, which is the concentration of donor ions times the volume AW_n. In order to maintain the neutrality of charge, this total positive charge must be equal to the total negative charge. So, let me equate these 2 and write the expression.

(Refer Slide Time: 18:48)

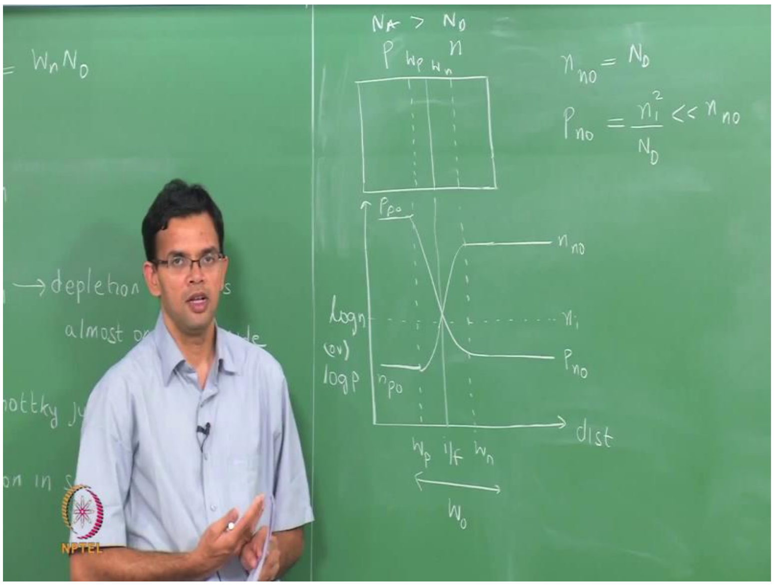


So, if you want to maintain charge neutrality, you have $A\ W_p\ N_A$, which is the total negative charge on the p-side must be AW_nN_D; the cross section area is the same, so that can be removed. So, what we are left with is $W_p\ N_A = W_n\ N_D$. On other way of writing this is at $\frac{W_p}{W_n} = \frac{N_D}{N_A}$. So, the ratio of depletion region widths on the p and n-side is inversely proportional to the concentration of the dopants whether they are the donors or acceptors. So, if $N_A > N_D$, so N_A is greater, which means $W_p < W_n$. So, the depletion region is larger on the n-side then on the p-side. They are certain p-n junctions that are formed between a heavily doped p^+ region and an n region. So, p^+ refers to a heavily doped region.

In this particular case, N_A is usually much larger than N_D. So, using this above charged neutrality expression you have W_p is much smaller than W_n, so that the depletion region is almost entirely on the n-side. An extreme example of this is in the case of your metal semiconductor Schottky junction, we saw this in last class where we form the junction between the metal and the semiconductor. And, we had electrons moving from the semiconductor to the metal, so that you had a depletion region, and the depletion region was almost entirely on the semiconductor. And this is again because, if you try look at an expression similar to here, the charge density in the case of a metal is much higher than that of a semiconductor. So, in order to maintain charge neutrality, you have electrons coming not only from the surface of the semiconductor, but also from the bulk creating a depletion region, and this depletion region lies entirely in the semiconductor side.

So, the next thing is to do is to calculate this built-in potential that comes between the p that forms when you have a p-n junction. So, we want to calculate the contact potential that forms in a p-n junction. To do that let me first draw how the carrier concentration that is the hole concentration changes and the electron concentration changes as we go from one end to the other.

(Refer Slide Time: 23:15)

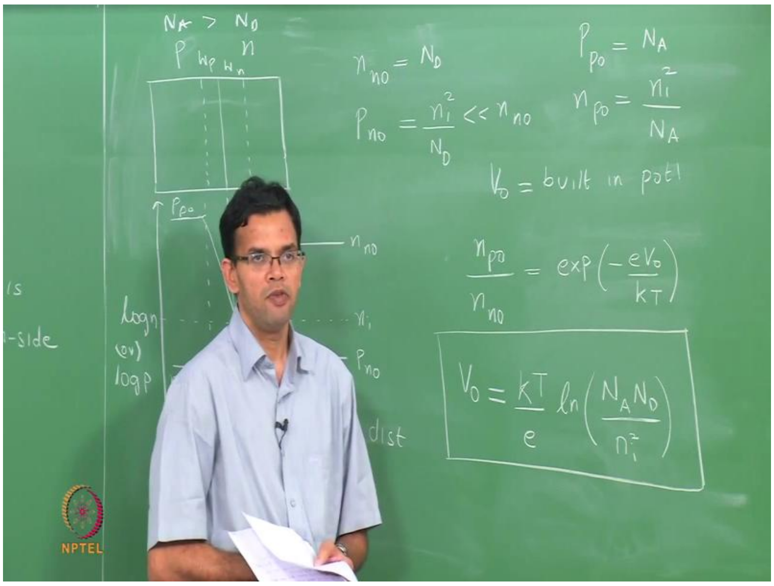


So, let me just redraw my p-n junction here. This is the interface. So, the p-side and the n-side, we have N_A and N_D as the concentration of acceptors and the concentration of donors. And we have $N_A > N_D$. We just saw that this means that the depletion width is larger on the n-side. So, there is a depletion region on the p-side, and there is a larger depletion width on the n-side. So, let me call this W_p and then W_n. So, if we plot a log of how the concentration of electrons or holes changes as a function of distance. So, I will plot log(n) or log(p) as a function of distance. So, let me again mark my interface this is W_p; this is W_n. So, this represents the depletion region. Let me also mark n_i, n_i is the intrinsic carrier concentration.

In the case of electrons, and we are in the n-side, the concentration of electrons n in the n-side let me call it n_{no} is nothing but N_D. The concentration of holes in the n-side, we can use the law of mass action is nothing but $\frac{n_i^2}{N_D}$. And usually this is much less than n_{no}. Similarly, the concentration of holes in the p-side is just N_A, and the concentration of electrons in the p-side is just $\frac{n_i^2}{N_A}$. So, these are all just notations, but we have done these calculations before when we looked at extrinsic semiconductors. If you go ahead and plot this, this is n_{no}; it is equal to N_D. At the depletion region, the concentration begins to drop, because we said we have a depletion region, because electrons move from the n to the p-side, so the concentration drops. And then finally, in the p-side the concentration becomes equal to n_{po}. We can do the same for the holes and we said that the hole concentration is higher than the electron concentration we said N_A is more than N_D.

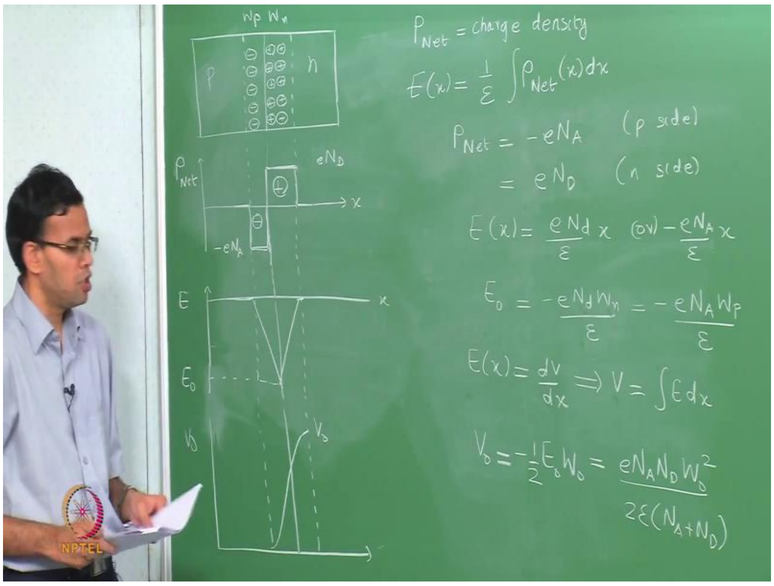
So, let me just draw these axis a bit up, so that this is p_{po}. Once again when we reach the depletion region this number is going to drop, because you have holes moving from the p-side to the n-side and then far away from the junction, we have the concentration p_{no}. So, this graph shows you how the electron and the hole concentration change, as we move from the n to the p or from p to n. So, this difference is related to the built-in potential because what the potential does is a prevents further motion of electrons from n to p or holes from p to n. So, in this way, it is similar to what happens when we have a Schottky junction. So, there also we had a built-in potential that prevents further motion of electrons.

(Refer Slide Time: 28:07)



So, V_o is the built in potential, it is related to the concentration of electrons on the n and the p-side. So, n_{po} which is the concentration of electrons in the p-side by n_{no} which is the concentration of electrons on the n-side is equal to $exp\left[\frac{-eV_o}{kT}\right]$. So, once again V_o represents the barrier that the electrons have to overcome in order to go from the n to the p-side. We can substitute these values for n_o, n_{po} and n_{no}. And rearrange this expression to give you your contact potential V_o is nothing but $\frac{kT}{e}\ln\frac{N_A N_D}{n_i^2}$. We get this expression by taking natural lawn on both sides and then substituting for these values and rearrange it. So, the contact potential in the case of a p-n junction depends upon the concentration of the acceptors and the concentration of the donors and also the intrinsic carrier concentration. We can also calculate the width of the depletion region that forms when we have a p-n junction.

(Refer Slide Time: 30:03)



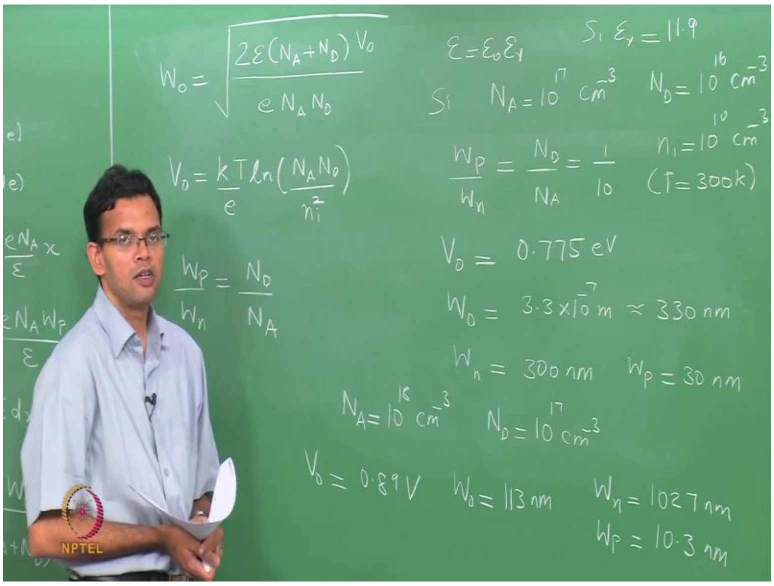
Let me just redraw the junction again. So, we have a p-n junction with $N_A > N_D$ so that we have some W_p which is smaller than W_n. So, we said that in the depletion region, the excess electrons and holes recombine and get annihilated, so that you have a net negative charge on the p-side and a net positive charge on the n-side. I am putting 2 positive charges to indicate that we have a wider depletion region on the n-side. If you plot the charge density as a function of distance, so ρ; ρ represents the charge density as a function of distance. We can usually say that the depletion region is devoid of carriers, so that the charge density is just a delta function. So, let me plot the interface; this is the n-side so, on the n-side the charge density is just given by the concentration of the donor ions. So, this is - e N_D or sorry this is $+N_D$. Because you have a positive charge on the p-side, the charge density is given by the acceptor ions which have a negative charge, so that this is $-e\ N_A$, and the total charge has to be zero, which means the positive charge on the n-side has to balance the negative charge on the p-side. So, the areas under these 2 graphs are the same.

We can relate the charge density to the electric field; the equation is electric field E as a function of x is just $\frac{1}{\varepsilon}$, where ε is the permittivity of the material integral ρ_{net} as a function of x d x. So, we just said that the charged density is a delta function and it is a constant. So, we saw that the $\rho_{net} = -e\ N_A$, this is the p-side and it is equal to $e\ N_D$ is the n-side. We can substitute for this here and then integrate over the entire width of the depletion region to get the electric field E. So, the expression for the field $E = \frac{e\,N_D}{\varepsilon}\,x$ or $\frac{-e\,N_D}{\varepsilon}\,x$. We can define an $E_o = -e\ N_D\ W_n\ \varepsilon$, which is equal to $-e\ N_A\ W_p$. And these 2 are the same, because we know that to maintain charge neutrality $N_D\ W_n = N_A\ W_p$. So, we can plot the electric field E as a function of distance. So, I will use the same plot here electric field-E the function of distance, the electric field is essentially negative.

So, let me just redraw it only on the negative side that is my interface; this is the p-side, this is the n-side. Outside the depletion region the electric field is 0, but within the depletion region E is a linear function of distance and the maximum value is E_o. So, the maximum value of E is E_o. E is also related to the potential by the expression $\frac{E\,dv}{dx}$, which means the potential V is $\int E\ d\ x$. We can substitute this expression and do the integration. The total potential which is the contact potential which we are interested in goes from the p-side to the n-side and V_o is nothing but $-0.5\ E_o$ and W_o, where W_o is the entire width of the depletion region that is equal to $\frac{eN_A N_A {W_o}^2}{2\varepsilon(N_A+N_D)}$.

We can plot V_o as a function of x. If you look at this expression V is the $\int E\ d\ x$ E is a linear function in x. So, the integral of a linear function is a parabolic function. So, if we plot V_o/x that is my interface that is the n-side, p-side potential goes from 0 up to the contact potential V_o and the expression for V_o is given here. Let me rearrange this to get the total width of the depletion region in terms of the contact potential.

(Refer Slide Time: 38:06)



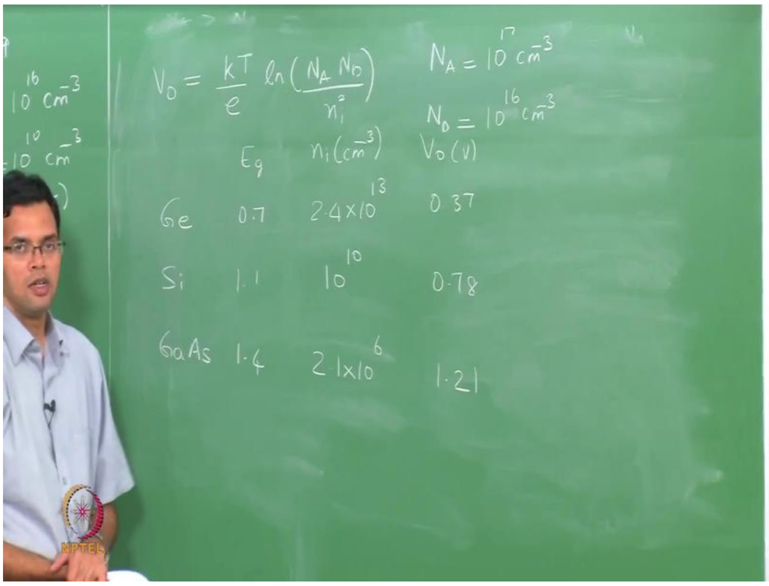
If we rewrite this expression, we will get W_o is nothing but $\sqrt{\frac{2\varepsilon(N_A+N_D)V_o}{eN_AN_D}}$. ε which is the permittivity of the material is nothing but epsilon ε_r nothing but ε_o which is the permittivity of free space times ε_r, which is the relative permittivity of the material. In the case of silicon, ε_r we saw this earlier as a value of 11.9. So, we have an expression for the contact potential $\frac{kT}{e}\ln\frac{N_AN_D}{n_i^2}$. You have an expression for the total width of the depletion region, and we also saw that the individual widths are inversely proportional to the concentrations. So, let us plug in some numbers to get a sense of what these values are.

So, we will take the example of silicon with acceptor concentration $N_A = 10^{17}$ cm^{-3} and donor concentration is 10^{16} cm^{-3}. So, $\frac{W_p}{W_n}$ is nothing but $\frac{N_D}{N_A}$ which is $\frac{1}{10}$. So, the width of the depletion region in the p-side is 10 times smaller than that of the n-side and that is because the concentration of acceptors is 10 times more than the concentration of donors. In the case of silicon, we know that the intrinsic carrier concentration is 10^{10}, and we are doing these calculations at room temperature, so T is 300 Kelvin. So, we can calculate the potential, we will just use this expression where we will plug in the values. If you do that V_o is 0.78 electron volts.

You can also calculate the width of the depletion region W_o. Once again it is a straight case of using this equation and plugging in the numbers. If you do that, you get a value of W naught to be around 3.3×10^{-7} m or approximately 330 nm. So, the total width of the depletion region is slightly a less than 1 μ is around 0.3 μm. You can also calculate the individual widths for using the ratio. If you do that, we will get W_n is around 100 sorry, if you do that we get W_n to be around 300 nanometers and W_p to be 30, so that this ratio of $\frac{1}{10}$ is maintained. The width of the depletion region is inversely proportional to the concentration of the carriers. So, if you have a higher value of N_A or N_D then the total width will be lower.

If you want to rework this with N_A increased 10×10^{18} cm^{-3}. Similarly, N_D is 10^{17}. So, you have increased both N_A and N_A 10 times, but the ratio is still the same. We can redo these calculations. Let me just write it down V_o is also higher, because $V_o = N_A N_D$. The potential V_o is around 0.89 volts. The total width is lower W_o, it is around 113 nanometers. And once again the width on the n-side is 102.7 and the width on the p-side is 10.3. So, we can reduce the width of the depletion region by increasing the carrier concentration on the p and the n-side. So, this is the p-n junction, as far as silicon is concerned. Now, let us just look at what happens if we change the material. What we want to know is how the contact potential changes when we change the material?

(Refer Slide Time: 44:32)



Now, V_o is related to $\frac{kT}{e} \ln \frac{N_A N_D}{n_i^2}$. For sake of comparison, I am going to keep $N_A = 10^{17}$, and $N_D = 10^{16}$, but V_o also depends upon n_i which is the intrinsic carrier concentration and that depends upon the band gap. So, the higher the value of n_i which happens if you have a lower band gap then smaller is the contact potential. So, if you look at 3 materials germanium, silicon, gallium arsenide, so we are forming p-n junctions between 2 germanium p and n, same with silicon, same with gallium arsenide. The band gap values at room temperature 0.7, 1.1, and 1.4. n_i is different, we have done this calculations before for this intrinsic materials 2.4×10^{13}, 1×10^{10}, 2.1×10^{6}. So, these are the values of n_i and if you plug in this equation V_o in volts is 0.37, 0.78 and 1.21. So, the contact potential increases as the value of n_i goes down, and this happens when you have a larger band gap.

So, today we have looked at a p-n junction that is in equilibrium, that is, there is no external potential applied. We saw that we have electrons moving from the n-side to the p-side; holes moving the other way and this creates a depletion region.

In the next class, we are going to look at the I-V characteristics of a p-n junction and what happens when we apply a bias to this junction.