Fundamentals of Semiconductor Devices Prof. Digbijoy N. Nath Centre for Nano Science and Engineering Indian Institute of Science, Bangalore

Lecture – 21 Breakdown of junction and C_V profiling

Welcome. So, what did we learn in the last class? We had ended the last class with the introduction of Breakdown. I told you that breakdown can happen in p-n junction diodes in reverse bias because primarily of two main mechanisms tunneling based mechanism breakdown and avalanche breakdown. What physically happens in breakdown is that suddenly the current will start increasing the reverse leakage current which is very low will suddenly start increasing at some large reverse voltage. So, something is happening, which is leading to that ok, I told you there are two reasons now.

So, today we shall discuss we will start the discussion in those two reasons why a device can breakdown. It is very important practically to understand breakdown and know what the limits are because you typically want to operate the device not near a breakdown; less than breakdown because if the device breaks down then you will lose the performance, right. So, you do not want the device to operate near breakdown. Also, the breakdown if the breakdown is for example, minus 20 V that sets an upper limit as to that extent to which you can use the device.

So, you cannot use the device at minus 25 V for examples. So, that is very important right. So, we'll come to the white board and will start understanding the two mechanisms one by one ok.

(Refer Slide Time: 01:37)



So, in a p-n junction breakdown you know could happen because of two reason one is tunneling, and one is avalanche, right. So, let us come to tunneling first; what does tunnel mean. At very high field at very high field reverse bias at very high reverse bias I can say which also gives you a very high field by the way a very high reverse bias also gives you very high field, right. So, how will the band diagram look like you know you have it will look like something like this at very reverse bias. So, your reverse bias is very large; depletion also is sort of large.

So, you know this is the whole quasi Fermi level E_{Fp} this is electro quasi Fermi level E_{Fn} . So, what happens is that if you have a very high doping of relatively high doping of pside and relatively high doping on the n-side then this depletion is generally narrow. Although you are applying a reverse bias and trying to increase it, but the depletion is narrow and what happens is that this becomes a thin barrier and the electrons that are there are many electrons in the valence band right the valence band is filled up with electron they will see empty states here.

These are empty states; these are empty states in conduction band. So, they can tunnel they can tunnel across the they can tunnel across the barrier. They can tunnel across the barrier and that because of the tunneling then this will happen at high field only. So, and that tunneling across the barrier will suddenly lead to a suddenly lead to a large current, suddenly lead to a large current and that is called tunneling current and that is called tunneling breakdown tunneling breakdown or it is called the Zener breakdown because Zener was the first person who proposed it or found it so, it's called Zener breakdown or tunneling breakdown.

So, high field these electrons will find a lot of empty space they will tunnel essentially, and this happens when your p and n side are sort of little highly doped. These are sort of highly doped.

(Refer Slide Time: 03:50)



So, it is like that right. So, it's there is a triangular barrier it must overcome. This is sort of the triangular barrier; this is sort of the triangular barrier. This is the barrier that is must overcome it is this barrier is the barrier which the electrons here will have to tunnel to reach to the other side. So, it's a triangular barrier and you there are equations to calculate what is the current that you will get here.

So, complicated looking equation you do not have to memorize it, but we just have to understand the physics and that is there you know the electrons can tunnel from one side to another side across this barrier you know this actually this looks like a triangular barrier. There can be different kinds of barrier this looks like a triangular barrier.

And I am assuming that you know what is tunneling from your high school physics. Tunneling is when you have a potential barrier; so, this is a potential barrier if electrons are here it can still tunnel from one state to another state through the barrier which is not possible in a real world. It depends how much can tunnel depends on the height lower the height more is the tunneling and depends on the thickness of the barrier, lower is the thickness thinner the barrier is more is the tunneling. So, that is why I said this barrier will be thinner if this barrier will be thinner when the doping is high. So, it has so, it can tunnel more. So, this is called tunneling.

Typically, tunneling occurs at a reverse bias breakdown only tunneling breakdown voltage is typically around 4 times the band gap. So, for example, if silicon is band gap is 1.1 eV then the tunneling based breakdown might typically occur at around 4.4 V it can vary little bit, but this is typically the around this value you can you might expect the tunneling based breakdown if it is a little relatively highly doped, because of the high field and because 4 times the band gap is little high you know because you are dropping 4 times the band gap essentially.

(Refer Slide Time: 05:38)



So, secondly, the avalanche breakdown comes, and avalanche breakdown is even more important actually. Avalanche breakdown is more common breakdown and many of the devices suffer avalanche breakdown only. So, what happens is that and for this avalanche breakdown you do not need to have a high p-n doping you even do not need to have any doping sometimes. You do not even need to have a thin barrier it's a different kind of breakdown mechanism.

So, what is happening is that high field, at high field, in a reverse bias voltage what happens? In a reverse bias it is a p-type, right. So, any minority electron that comes here it will be immediately swept away here, right. So, at very high reverse bias at very high V_R which means the very high reverse bias this field also increases know. So, the electron will be accelerated very fast this electron that is coming will be accelerated very fast or very high you know accelerated very fast and it will acquire high kinetic energy.

This electron because is accelerated fast by the field will acquire very high kinetic energy. So, what will happen is that the electron because by the virtue of it is kinetic energy it will take it will knock out an electron from an atom and the neutral atom, it might knock out an electron from a neutral atom and it might create an electron hole pair. If it knocks an electron out, it might create an electron hole pair.

So, an electron that was coming it strikes one atom it creates an electron hole pair because it was too energetic; it was too energetic because it has a lot of kinetic energy that is because it was accelerated fast by the very high field. And this electron that is formed here will now get accelerated the second electron that this formed is now get accelerated and might knock out another electron hole pair and create one electron hole pair. So, this electron itself might create one electron hole pair.

Similarly, the hole that was formed also might accelerate and that hole also might also create one electron hole pair because that hole will be now highly energetic. So, that hole will also knock out an electron or a hole from a semiconductor and it will create an electron hole pair. Of course, electron must travel more distance here the hole has to return less distance here.

So, this is like a this is a electron initiated process is an electron initiated process, sometimes they can be hole initiated process also. You know if we talk about the holes this is n-type the hole coming from here it will knock out and create an electron hole pair that electron will again accelerate knock out, another it will knock out. It will keep going.

So, it may be hole initiated, this might be electron-initiated process, but you know it creates one electron will create electron hole pair, one electron again will create the secondary this will create then this will create two. So, it will multiply, it will multiply like an avalanche it will multiply like an avalanche, right. It will multiply like an

avalanche and suddenly and a very rapidly in a very smaller time frame this will lead to a huge number of carriers that will keep flowing this side because they are knocking and creating 2, 2, 2, 2, 2 factors of 2, right.

So, this avalanche will multiply like an avalanche rapid increase in the electrons and this leads to a sudden charge surge in the reverse bias current in the reverse bias current it will lead to a sudden charge. So, this is called your avalanche breakdown ok. This is called your avalanche breakdown and there is a critical field at which will happen typically it will happen at a reverse bias voltage greater than 6 times the band gap. So, silicon band gap is 1.1 eV. So, the avalanche breakdown will happen at a voltage or reverse bias voltage of more than 6.6 electron volt at least ok.

So, you know the avalanche is basically this rapid multiplication of carriers because they gain kinetic energy and keep producing the electron hole pairs in a very rapid succession.



(Refer Slide Time: 09:18)

And so, associated with this there are two terms; one term is called α_n (alpha), one is called α_p . These are called the ionization coefficient; impact ionization coefficient of electrons and holes respectively. This is electron, this is hole, impact ionization coefficient what is impact ionization? This term impact ionization is used actually to describe this process when the electron gain sufficient kinetic energy and it knocks out the carriers and produces one electron hole pair that is that rapid creation, we call it avalanche creation that is happening technically it is called impact ionization ok.

What is happening is that electron by virtue of higher kinetic energy is impacting you know, its creating an impact on another atom and creating a electron hole pair. So, this ionizing that electron hole pair so, we call it impact ionization; we call it impact ionization. An impact ionization energy you know this impact we have an impact ionization for electrons, we have an impact ionization for hole and this way the impact and this is a coefficient where a constant.

This impact ionization coefficient is defined as the change of electron concentration for example, it is electron the change of electron concentration, so, Δ_n (delta). What is the change of electron concentration with respect to position, because you see you know this is an electron and it rapidly it is creating one electron hole pair again is coming electron hole pair its creating and going know. So, what is the change of the excess carriers you are had generating because of this impact ionization with respect to the position this is position, right. This whole thing divided by the total electron at any point. This is your impact ionization coefficient.

And the way they are multiplying the carriers are creating and multiplying know they are creating and multiplying. There is a multiplication factor. There is a multiplication factor defined as M ok. This multiplication factor it will tells you how many carriers are getting generated and what is the rate at which they are generating and so on and so forth? So, multiplication factor you know gives you an idea how rapidly they are multiplying and so on.

So, multiplication factor is defined as

$$M = \frac{1}{1 - \int_W \alpha_n dx}$$

What is W? The depletion over which it is because only the depletion will have a higher field, there is no field here. So, those no points considering here and the moment this integration goes to 1 your M tends to infinity and that is when your avalanche breakdown takes place.

now, one electron that is injected can virtually lead to an infinite number of electron hole pairs by the time it comes out here, that is what it means. So, when M tends to infinity it's called avalanche breakdown and that means, that one electron that are injecting can virtually lead to an infinite number of carriers that are there and that happens at a certain voltage, right.

(Refer Slide Time: 12:07)



And, so, there is a field if there is critical field a critical field for each material and a critical field is the electric field at which this avalanche will typically this avalanche breakdown will happen. This critical electric field is called say you know F critical I will call it electric critical field, and this depends very weakly on the doping concentration of the material, it is very weak dependent. So, I can say that this is independent of the voltage the electric field. For example, in silicon this is probably close to around 400 kV/cm. So, at the field of 400 kV/cm there is electric field your silicon might start giving you know avalanche breakdown this is called critical field.

And then there is a voltage; the breakdown voltage will depend on the depletion width of course; it will depend on the depletion width. So, the breakdown voltage for example, in a p-n junction if you recall if you apply the if you this is x for example, this is 0, this is W_n , this is minus W_p . Your field profile was given by something like that, right. This is the maximum field which is F_{max} we call it that that F_{max} the maximum value that is

 $F_{max} = (qN_DW_n)/\epsilon_0 \epsilon_s$

The same thing as $(qN_AW_p)/\epsilon_0 \epsilon_s$ for a material to work at the limit of breakdown the maximum this can be is actually F critical which is defined in a material each material has it is own critical field.

And the voltage that will block is the area under this triangle. The area under a triangle is $\frac{1}{2} *W* F_{CR} = V_{BR}, F_{max}$ which is your F critical. This is your V_{BR} break down. So, if you know the breakdown, if you know the critical field, if you know the depletion width here then you know what at what voltage breakdown it will happen?

So, for example, silicon has a critical field of around 400 kV /cm and suppose the depletion with that you have in a semiconductor, you know the depletion with that you have in a semiconductor here. Suppose the depletion width is given as a 1 micron which is 1000 nanometer or 1 micron the same thing as 10^{-4} cm.

So, then the maximum breakdown voltage that you can block is $\frac{1}{2} \times 10^{-4}$ cm $\times 400$ kV/cm. So, centimeter-centimeter gone, this is 10^{3} , this is 10^{-4} . So, this is 10^{-1} , this 1 goes here 20/2. So, this is 20 V.

So, the breakdown voltage is 20 V. You can block 20 V which have depletion region of 1 micron or 1000 nanometer. If your silicon critical field is 400 kV per centimeter ok, that is the breakdown voltage you can essentially block. But, again of course, if your material has a higher band gap; for example, gallium nitride has a band gap of 3.4 eV. So, the critical field of gallium nitride is 3 MV/cm, that is a much larger voltage than silicon. So, you can drop much larger you can block much larger voltage.

So, then the for the same depletion region 1 micron your breakdown voltage that you can block in gallium nitride p-n junction it the same depletion will be half into the depletion which is 1000 nanometer into 3 MV/cm is you know 3000 kV/cm which is roughly 9 times more than this. So, if silicon you can block 20 V here right with gallium nitride you can block around 180 V. So, which means you can use a gallium nitride p-n junction to do much higher voltage blocking.

So, for high power electronic devices or a device that need a lot of voltage to be blocked you can do you can use gallium nitride with the same device dimension compared to silicon. That is why for emerging power electronics like where you need to do all this power conversion blocking high voltage like electric vehicles and other things that are emerging gallium nitride is a more suitable material because it can block larger voltage than for example, silicon.

So, power electronics where you need this power switching and large voltage blocking capability gallium nitride devices are proven are now getting established as superior to silicon because this is a wide band gap material. The other wide band gap material also like silicon carbide and so on. So, those are established to be much better at blocking voltage. So, this comes from this breakdown kind of a thing. So, the breakdown of a material defines you know the limit to the breakdown voltage you can operate the device at and that sets a limit to the technology also.

Like, if you look at the electric vehicle you need to charge the vehicle you need to charge the vehicle and the on-board charger probably will need 3 kW power to charge it. So, it will have some 600 V for example, bias that you must block when you switch. So, you know it depends on the material, depends on the p-n junction property where are you are not you are able to block that 600 voltage if not then you to go for a very slow charging over a long time. So, those are the things that are important in real world devices, you know how p-n junction and how it how the blocking of a p-n junction is very critical to that.

So, this is about the breakdown that we have studied and finally, you know remember that the breakdown is of course, larger for larger band gap material because the critical field typically is higher for a higher band gap material. (Refer Slide Time: 17:20)



The final thing that we remember is the capacitance voltage what does capacitance voltage mean? So, you have a p-n junction as I told keep telling there is a charge here know, there is a negative charge here, there is a positive charge here, but this is a depleted region. This is a depleted region there are no carriers there and this p-type has holes.

So, this is highly conducting for example, this is also electrons so, this is highly conducting. So, it looks like these are very highly conducting you know layers that are sandwiching or that are wrapping around an insulating layer because this is a there are charges here in this layer, but it is insulating. Because there are no free careers, there no mobile carriers in this region. So, this is a depleted region.

So, essentially its like you have a conducting region you have a conducting region and in between you have a depleted region, like an insulator. So, this is a capacitor. This is a capacitor and this capacitor has a capacitance given by area which is the area of you know the cross-sectional area times $\varepsilon_0\varepsilon_s/W$, (depletion width) W. Of course, I want the express normalized capacitance which means I will express the take the capacitance here area here and I will take it as a normalized capacitance in the units of F/cm².

Typically, the area the unit of capacitance is farad and normalizing at the area. So, it is area you know capacitance per centimeter square, this is your formula. But you can derive this also in a more complicated way and you will still end up getting this. A capacitance then this is by the way call you know this is called the depletion capacitance because its corresponding to the depletion region this called depletion capacitance or its also called junction capacitance, it's also called junction capacitance.

The reason is because this is corresponding to the junction here, this also corresponds to the depletion region here, so, there is charge here, there is charge here, there is a depleted region here. So, it's a depleted capacitance is the depletion region capacitance you can think of it is like a there is a highly conducting metal, this a metal actually it is non metal, but there is a metal there is a metal and in between there is a dielectric, this is depleted dielectric. So, essentially that is like similar kind of a thing. So, it is a depletion capacitance.

And when you use the diode in different circuits and other things this depletion capacitance because there is a capacitance it can lead to you know RC delay for example, there is always a resister associated with it note that a resistance and capacitance can give you a delay make you circuit slower. So, this a very important things and you can derive this value of this capacitance by going the more traditional route.



(Refer Slide Time: 19:50)

For example, you know this is your p-type, n-type for example, there is a depletion region here I call it W. So, essentially the capacitance is defined as the small change in the depletion charge. What is the charge stored here? which is d you know which the

small change in the charge is stored there I will call it del Q delta D the depletion charge divided by the small change in the applied voltage V a.

$$C = \Delta Q / \Delta D,$$
$$\Delta Q_D = \frac{q N_D W_n}{\varepsilon_0 \varepsilon_S}$$
$$Q_D = \frac{q N_D (N_A) W}{\varepsilon_0 \varepsilon_S (N_D + N_A)}$$

So, if I apply a voltage ok, a small change in the voltage, the small change in the voltage whether it is forward or reverse, but typically we talk about reverse bias only. But even you know typically talk about in the reverse bias only. So, a small change in the voltage that you are applying delta V_a can lead to a small change in the charge that you are storing the ratio is called the capacitor.

See if you increase this suppose you are applying a voltage say minus 3 volt you are increasing a small amount from minus 3 V to minus 3.1 V what will happen is that this depletion that was there will slightly increase slightly increase with reverse bias it slightly increases know. So, the moment is slightly increasing the Δ some extra Δ has come. So, what is the change in the charge stored now because it has slightly increased, there is a slow more slightly more increase in the charge that is storing what is that ratio with respect to the voltage that you are changing, right.

So, what is the actual charge that you are storing? The charge actually that you are storing if you look at each of this block you know qN_DW_n which is the depletion the n side divide by ϵ_S . I can write this as $qN_D/\epsilon_0\epsilon_S$ and the W_n can be written as $(N_A/(N_A+N_D))$ *W; the W is the total depletion if you recall. This is the charge this your charge; the charge that you are storing.

And if you recall this W; this W actually can be written as $2\epsilon_0\epsilon_S/q$ because there is a reverse bias that you are applying here $(V_{bi}+V_R) *(1/N_A+1/N_D)$, right. So, this is depending on that right.

$$W = \sqrt{\frac{2\varepsilon_0\varepsilon_S}{q}} (V_{bi} + V_R)(\frac{1}{(N_A} + \frac{1}{N_D})$$

(Refer Slide Time: 22:15)



So, essentially what I can do is that I can say that your charge that you are having in the depletion is qN_D. So, I look here is $(qN_DN_A/\epsilon_0\epsilon_S) *W_n$ I mean that is because it is $(N_A + N_D) *2$ epsilon by square s $(V_{bi} + V_R) * (1/N_A + 1/N_D)$.

$$Q_D = \frac{qN_D(N_A)}{\varepsilon_0\varepsilon_S(N_D + N_A)} \sqrt{\frac{2\varepsilon_0\varepsilon_S}{q}(V_{bi} + V_R)(\frac{1}{(N_A} + \frac{1}{N_D}))}$$

Now, a capacitance is defined as the derivative of this Q with respect to V_R , I mean that is the applied voltage right V_R or V_A . So, essentially this expression needs to be differentiated with respect to V_R . If you do that you will end up getting $\varepsilon_0\varepsilon_S/W$ you can do the math, this is F/cm². So, that is why your capacitance is a parallel plate capacitor only. And which more reverse bias if you apply more reverse bias voltage your W will increase, right. If your W increases your capacitance will decrease. So, it more and more reverse bias your capacitance also decreases.

Now, you can of course, choose to write the expression of W in a detail from like I had written here, right. There is detailed form you can choose the write and then what will happen is that if you do a little bit of more mathematics around that if you put this W as the detail value and if you take $1/C^2$, right. If you just take as $1/C^2$ square the value will

come out to be something like if you do all the math $2/q \epsilon_0 \epsilon_s$ by doping say I am talking about one sided abrupt junction.

$$C = \frac{dQ_D}{dV_R} = \frac{\varepsilon_0 \varepsilon_S}{W} F/cm^2$$
$$\frac{1}{C^2} = \frac{2}{q\varepsilon_0 \varepsilon_S N_D} (V_{bi} - V_R)$$

Which means I am having p plus very highly doped p-type and very lightly doped n-type or light relatively lightly doped n-type then I told you that depletion is predominantly on the n-side only the depletion on the p side is negligible and we can neglect that. So, the depletion is predominantly on the n-side only it is called one sided abrupt junction one sided abrupt junction and the reason is called one sided abrupt is called because the depletion is only extending to one side which is the n side, the lighter side. The heavier may there is almost zero depletion, very small depletion so, we neglect that.

So, if we take a one-sided depletion like and this is supposing N_D doping and n-side this is very high doping we do not care. So, in this kind of a one-sided depletion region if you do this capacitance voltage put the W accordingly then you will get the expression like this into the built in voltage of the p-n junction. If you remember every p-n junction has a built-in voltage, right, that is a built-in voltage here also plus V minus V applied voltage this is the relation you will get.

(Refer Slide Time: 25:18)



So, let us see again you will get a relation like $1/C^2 = 2/q\epsilon_0\epsilon_s$ the doping on the n-side; pside doping does not matter because is very high now by the built in voltage minus the applied voltage. So, if you actually plot $1/C^2$ now, versus the voltage that you applying then you will get a curve like this. It has a negative slope; the negative slope is given by this minus is there know this quantity.

So, the slope is given by minus 2 the slope is negative of course, $-2/q\epsilon_0\epsilon_SN_D$. So, if you plot 1 by so, if you measure capacitance and you take $1/C^2$ when you plot with respect to the applied voltage then you are going to get minus 2 by this. From this slope you can find out the doping on the n-side. So, it's a very powerful technique to find out the doping and you see this point if you extrapolate it will cut somewhere here know this point is where the y-axis is 0 which means $1/C^2$ is 0. If $1/C^2$ is 0, then V_{bi} equal to V_R which means this point is where your built-in voltage, from this equation.

So, the extrapolated value on this axis will give you the built-in voltage. So, you can calculate the built-in voltage also from the capacitance voltage profile. So, it's a very powerful tool. It is a very powerful tool and if you take up you know is the same thing if you take derivative of this of course, d by d $V_R 1/C^2$ then you will end up getting - $2/q\epsilon_0\epsilon_S N_D$. If you take derivative of this you will get the slope and that slope is this that is the same thing essentially, I told you, right.

$$\frac{d}{dV_R} \left(\frac{1}{C^2} \right) = \frac{-2}{q\varepsilon_0 \varepsilon_S N_D}$$

So, capacitance voltage profiling is a very powerful tool or technique to find out the doping as well as the built-in potential and a p-n junction behaves like a capacitor. So, you can take a p-n junction, you can apply reverse bias for example, you can apply reverse bias and you can sweep the voltage in a measurement tool. You can sweep the reverse voltage and you can monitor capacitance, so, you can get a C versus you know V_R plot.

Now, you take this values that you are getting in a the excel sheet, you reverse it 1/C, then you square it and then that will be your you know that is $1/C^2$ versus V_R plot and that gives you the idea of you know the doping and other things.

(Refer Slide Time: 27:41)



Now, if you recall your capacitance is $\varepsilon_0\varepsilon_S/W$ and W depends on square root of $V_{bi} + V_R$ into many other terms are there, right. So, essentially your capacitance depends inversely as $1/(V_{bi} + V_R)$, there voltage that you are applying if your reverse voltage is large then you can compare to built in then you can even neglect that. So, this will be essentially V_R .

$$C = \frac{\varepsilon_0 \varepsilon_S}{\sqrt{V_{bi} + V_R}}; C \propto \frac{1}{\sqrt{V_R}}$$

So, essentially your if you are plotting the capacitance here and you applying the V_R , then this relation this relation will go like this. This relation is basically 1 by square root of V_R ok. This is like that it will go ok. It will go like that I think it probably will not go like it will go the slope is positive know. So, it will go like that V_R is if the V_R keeps increasing then the capacitance becomes lower and lower. So, it be like that that is what will happen actually.

So, this is all about capacitance voltage profiling of C V of a p-n junction and this is an important concept that you need to keep in mind because in future when you discuss about many other things like BJTs or even MOSFETs, these capacitance voltage profiling and breakdown will become very useful and very handy in discussing those devices ok.

So, we will wrap up the lecture here today. We have now practically and virtually finished up p-n junction, whatever we needed to learn for the course we have discussed. There are many other things about p-n junction like diffusion capacitance and small signal analysis which we are not covering in this course, those are becoming more advanced those are sort of little bit more advanced topics. So, I have given you a flavor of how breakdown voltage relates to the p-n junction behavior, how different band gap materials can have different breakdown, different power electronic applications need higher breakdown voltage, for example.

I told you about the capacitance force profiling that the p-n junction is like a capacitor there is a junction capacitance and using that junction capacitance we have been able to extract the doping profile the built in voltage and so on. This a very useful things that we are going to need as we discuss future devices. So, we will wrap up the lecture here.

What shall we discuss in the next class? From the next class we will start metal semiconductor junction. If you put a metal on a semiconductor, what is the junction and what are the properties and before that we should ask ourselves why do we need to put a metal, what is the necessity of putting a metal on a semiconductor. So, those things we will study from the next class ok.

Thank you.