High Speed Devices and Circuits

Prof. K. N. Bhat

Department of Electrical Engineering

Indian Institute of Technology, Madras

Lecture 15

Ohmic Contacts on Semiconductors

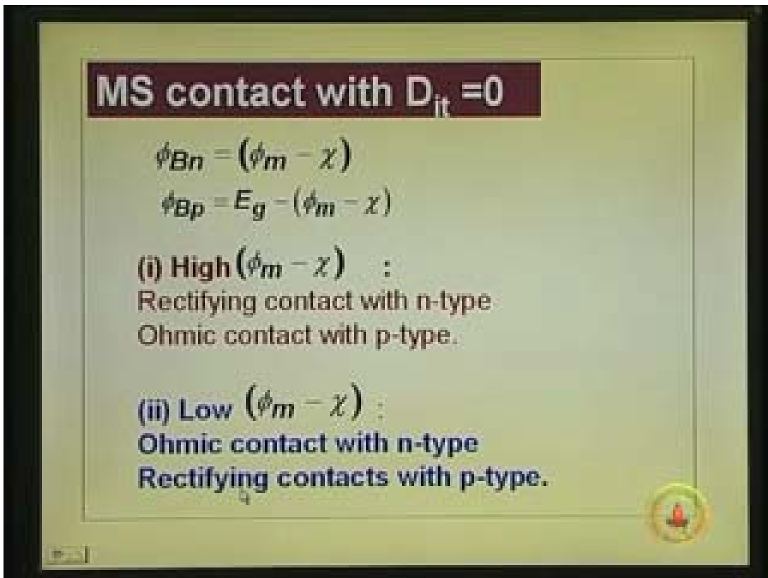
So we have been discussing about the metal semiconductor contacts. Today we will see in some detail, what difficulties one will have to make ohmic contacts onto high band gap semiconductors like gallium arsenide may be given wider band gap semiconductors.

(Refer Slide Time 01:10)

We will focus on gallium arsenide indium phosphide and of course silicon also we will just touch upon. These are common themes for forming the ohmic contact.

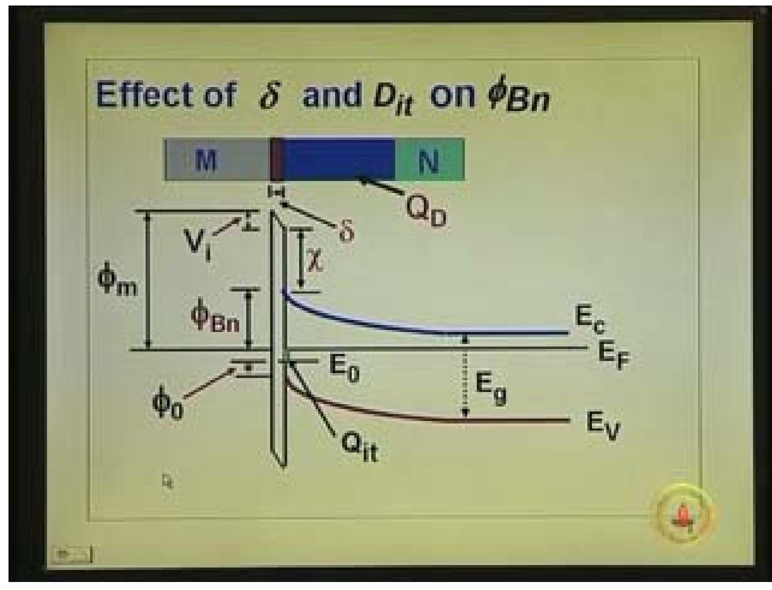
Now, what we saw in the ideal case was that, the barrier height phi_{Bn} is equal to phi_m minus chi.

(Refer Slide Time 01:44)



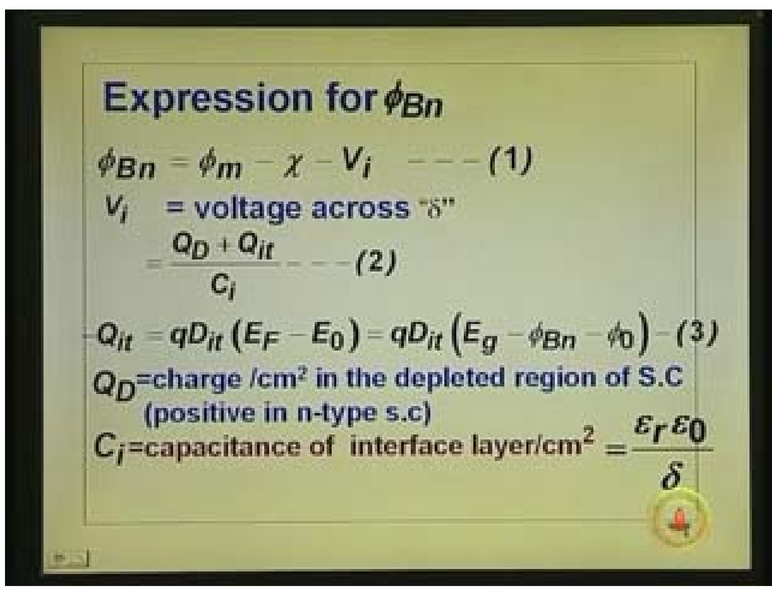
So once you fix the metal, the barrier height is fixed. Once phi_m minus chi is high you can get rectifying contact on n-type substrate, that is what we saw and you can get ohmic contacts if phi_m minus chi is high. In the same way, if phi_m minus chi is low, phi_{Bn} is low and you get ohmic contact onto n-type semiconductor and rectifying contact on to p-type. It looks very simple but things get complicated because of this which you have discussed last time.

(Refer Slide Time 02:21)



That is, a thin layer of material may be dielectric material and high density of states here, in the interface, does not give barrier height equal to phi_m minus chi but there is additional drop coming up here. Now, that additional drop I have shown it actually in this polarity. It could be this way or the other way because if it is this way, the phi_m the barrier height will be less than phi_m minus chi, depending upon what charge is present here, that is, depending upon the relative position on the Fermi level to the E_0 will have plus charge or minus charge. If you have minus charge, it will have some polarity. If you have plus charge, it has some polarity. So what I am trying to point out is this could either increase the phi_{Bn} from phi_m minus chi or decrease. We have gone through the whole thing and ultimately arrived at these three equations.

(Refer Slide Time 03:14)



We can see phi_{Bn} is phi_m minus chi plus or minus an additional term; if it is with the polarity that you have shown there it is minus; whereas if it is the other way it will be plus. So with taking those things into account, what we saw was, the phi_{Bn} gets a new form altogether.

(Refer Slide Time 03:33)

From equations (1),(2) and (3)

$$\phi_{Bn} = \phi_m - \chi + \frac{qD_{it}}{C_i}\left(E_g - \phi_{Bn} - \phi_0\right) - \frac{Q_D}{C_i}$$

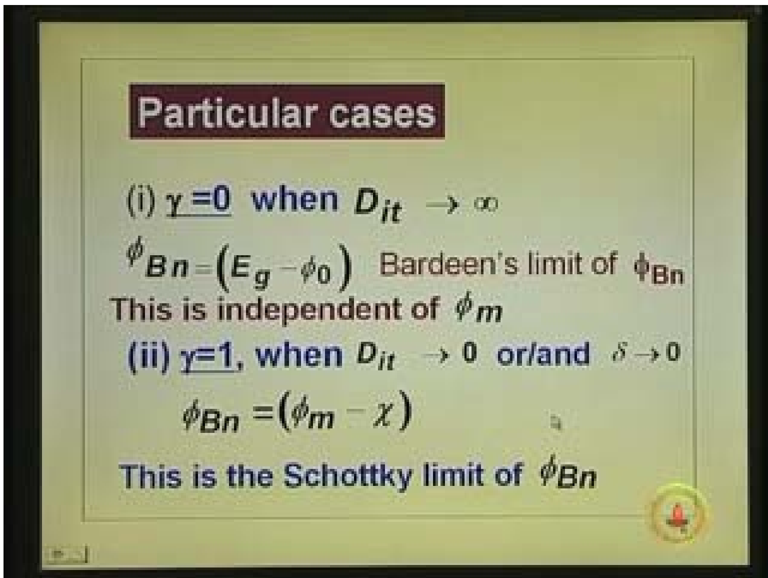
Rearranging,

$$\phi_{Bn} = \gamma(\phi_m - \chi) + (1-\gamma)\left(E_g - \phi_0\right) - \frac{\gamma Q_D}{C_i}$$

$$\gamma = \frac{1}{1 + \frac{qD_{it}}{C_i}} = \frac{1}{1 + \frac{qD_{it}\delta}{\varepsilon_r \varepsilon_0}}$$

phi$_{Bn}$ is equal to gamma into phi$_m$ minus chi plus 1 minus gamma into E_g minus phi$_0$ minus this quantity. We saw this last time, now gamma is related to the interface state density D_{it}. If D_{it} is 0, that term goes off and you have gamma equal to 1 then gamma equal to 1; if gamma equal to 1, this term goes off and we will have this small term coming up. So mostly, phi$_{Bn}$ will be equal to phi$_m$ minus chi when gamma equal to 1 that is, when D_{it} equal to 0 and also with delta equal to 0, so these both the terms will go off ideal, that is the ideal situation, but now when D_{it} is very large, suppose this tends to infinity then the entire thing becomes equal to 0 which would mean that the first term is 0; the second term will be gamma is 0; E_g minus phi$_0$ and this also goes off because gamma is 0. So we will have only E_g minus phi$_0$, what we are telling is, if the D_{it} is very very large, gamma equals 0 and phi$_{Bn}$ is E_g minus phi$_0$.

(Refer Slide Time 04:51)

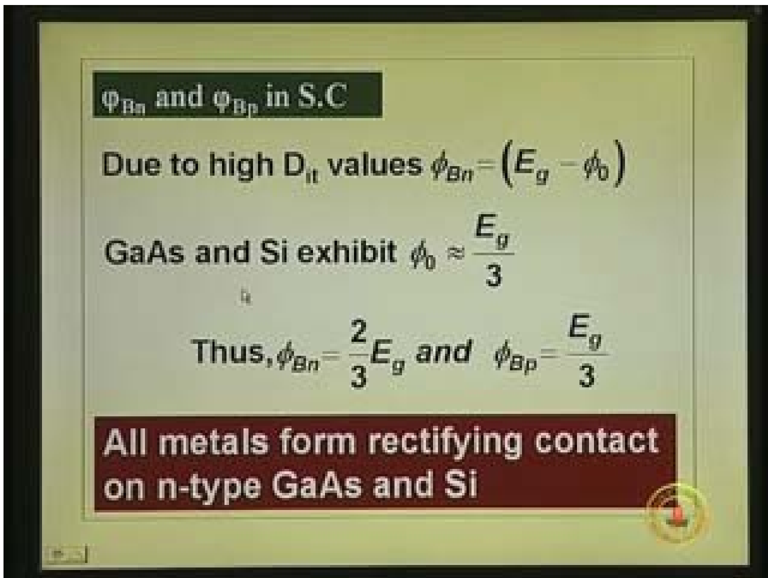


So it depends upon the difference between E_g and phi_0. The phi_{Bn} now depends upon what the value of phi_0 is- phi_0 is actually the energy gap between the neutral level $E_0$and E_g. So that depends upon a particular material and the surface condition of material and actually the material for a gallium arsenide and silicon it so turns out that phi_0 is about one third of E_g, if it is one third of E_g then this whole thing will become two thirds of E_g we will take a look at it further.

So if gamma is 1 when D_{it} equal to 0, we have to just point it out, we will have phi_{Bn} is equal to phi_m minus chi, which you will never get, experimentally we have found that we cannot get it. It turns this closer to that because of high density of states.

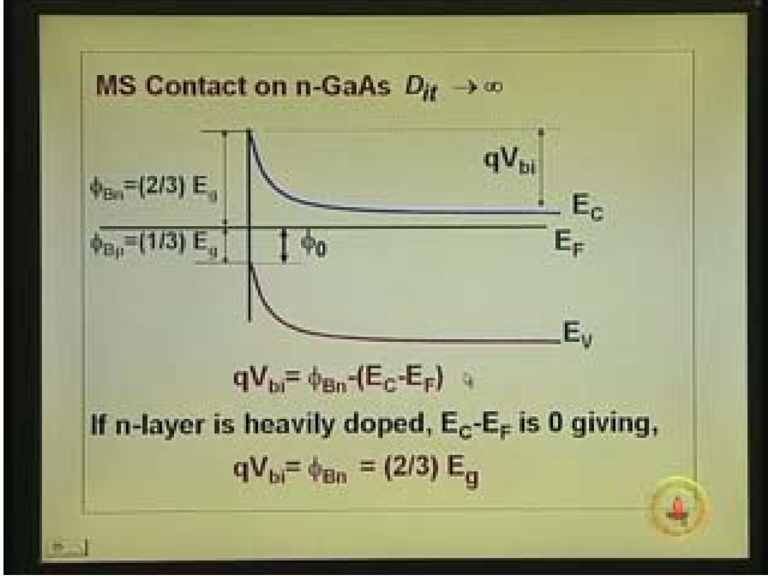
So due to these now I comment on that, it is due to high interface state density phi_{Bn} is equal to E_g minus phi_0, we have seen it just now, because the gamma equal to 0. And for gallium arsenide and silicon phi_0 is E_g by 3.

(Refer Slide Time 05:52)



So if this is E_g by 3, the entire phi_{Bn} is equal to two thirds of E_g and phi_{Bp} correspondingly will be you can see that there. So this is the total E_g and this is phi_0. E_g phi_0 by E_g minus phi_0 is this quantity that is phi_{Bn} that is two thirds of E_g,

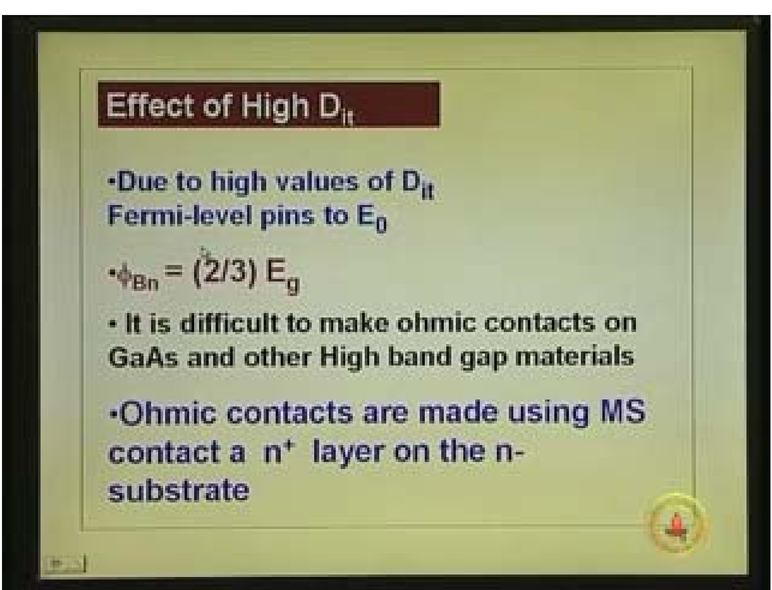
(Refer Slide Time 06:27)



And then automatically phi_{Bp} would be one third of E_g. Now we have seen that, if phi_{Bn} is large it turns out to be rectifying contact. Now here also, if you just see if this is two

thirds of E_g, or if this is $phi_{Bn,}$ the 'a' built-in potential here is equal to phi_{Bn} minus this quantity E_C minus E_F. After all, the total quantity minus this quantity, these two is equal to $V_{bi,}$ so that is that quantity. Now suppose, EF merges with E_C, phi_{Bn} becomes equal to q V_{bi} that incidental I am just pointing out. The main thing I want to point out is, you end up with phi_{Bn} which is large in n-type semiconductor and since it depends only on that E_g into two thirds, it is practically independent of metal. What is the consequence now? You have a hopeless situation, a very bad situation if you want to make ohmic contact on n-type substrate particularly with materials like silicon gallium arsenide etcetera or all the wide band gap materials because the barrier height turns out to be that two thirds of E_g. Now you can see that if it is one point four three electron volts, this will turn out to be about point nine five three, this it is too large to form a ohmic contact.

(Refer Slide Time 08:06)



Now pre-iterating what we said due to high values of D_{it}, Fermi level pins to $E_{0.}$ See what has happened now is, the Fermi level has gone and locked on to that E_0 level, when you say the high value of D_{it} makes this quantity equal to phi_0 what you said is that, EF has aligned with respect to $E_{0,}$ we will come back and read details of the physics of this little bit we take at the end of this lecture today.

But right now what we say is, the Fermi level will align with except in respect to $E_{0,}$ as a result this is $phi_{0,}$ as a result this is E_g by $phi_{0,}$ that is two third of E_g, we will come back to see what is the reason for this locking on to that.

In fact I think we will spend little time just later because right now my focus is on ohmic contact, so once phi_m is Fermi level aligns with $E_{0,}$ phi_{Bn} is actually equal to two thirds of E_g. So it is difficult to make ohmic contacts on gallium arsenide and other high band gap materials including silicon. Silicon is 1.1 and two thirds of it is that sometimes comes in like .75 which is not small, it is not good enough to make an ohmic contact. So ohmic contacts are therefore have to be made, this is actually the solution which has arrived at, for making ohmic contacts on n- type materials you have to make using metal semiconductor contact on n plus layer which is formed on n layer.

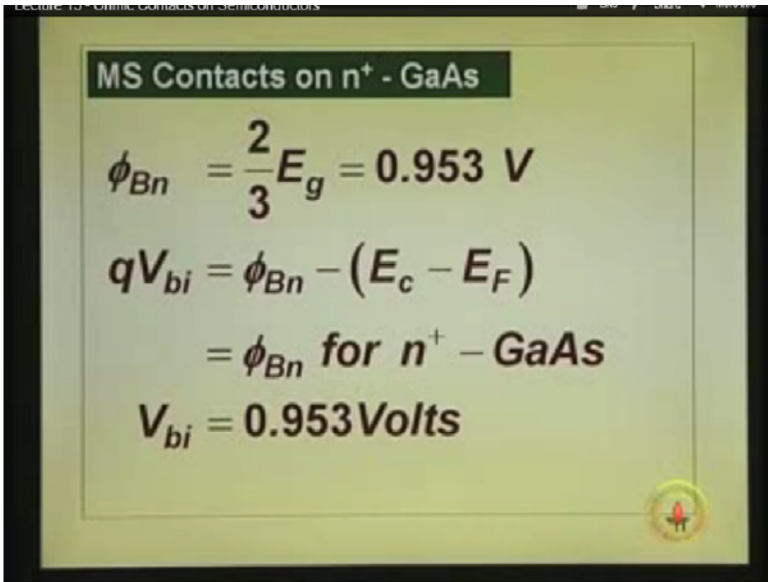
Do not put the metal directly on the n type substrate, form a n plus contact on the surface then put a metal contact on to that, then we will place a chance for that. We will see how that works out. In fact it is very simple, just look into the energy band diagram. Here what I am showing is the consequence of making the n layer n plus. When you make the n layer n plus what happens to the Fermi-level?

Fermi-level moves closer to conduction band but that is locked on to $E_{0.}$ If it is locked on to the neutral level E_0 that is two thirds of E_g, so the consequence of doping it heavily one is this quantity becomes closer and closer to phi_{Bn}, built-in potential is closer and closer to $phi_{Bn.}$ In the thing diagram that I have shown q phi V_{bi} is equal to $phi_{Bn.}$ In fact we will be getting closer and closer to that it doesn't move too much into conduction band, it will be just merging to this Ec. So what you will have will be, the phi V_{bi} which is about $phi_{Bn,}$ that is not what we are looking for, we are looking for more than that. Still the way it is drawn here it will be a rectifying contact.

Now what other things happen let us see, once we see that this phi_{Bn} is two thirds of E_g we know how much is that quantity that is .953 per electron volts for gallium arsenide that means V_{bi} is actually equal to .953 volts built-in potential.

qV_{bi} is .953 electron volts so V_{bi} is .953 volts. Now, this is to illustrate it further I have shown here phi_{Bn} and written the values corresponding to that.

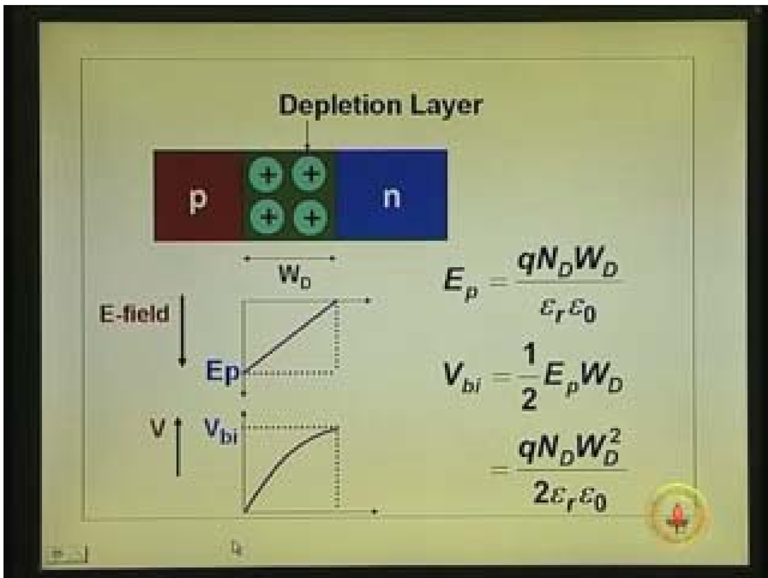
(Refer Slide Time 11:44)



For example here phi_{Bn} is two thirds of E_g, which is equal to .953. Now from the diagram we have seen that qV_{bi} is equal to phi_{Bn} minus E_c minus E_F where E_c is the conduction band edge; E_F is the Fermi level and in the heavily doped semiconductor E_c minus E_F is very small and extremely that can be said it is almost equal to 0. Therefore, qV_{bi} will be equal to phi_{Bn} minus 0 which is equal to phi_{Bn} for n plus gallium arsenide, so therefore V_{bi} is equal to .953volt.

If V_{bi} is .953 what will it tell, if you take a pn junction, instead of p you have to get a metal here, so all that you have done is, instead of p you have got metal p plus, you have got a metal here, it is exactly like the pn junction.

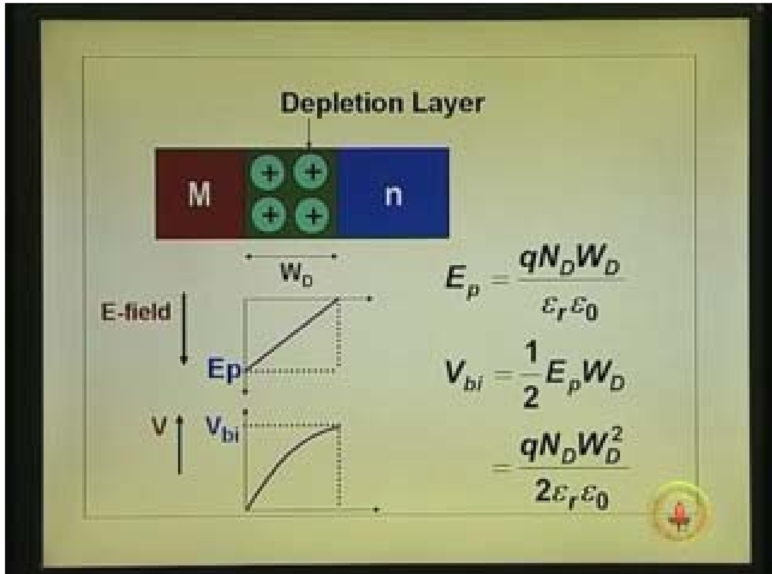
(Refer Slide Time 12:39)



So what you have on this side is the depletion layer and n type substrate; and this depletion layer the entire potential V_{bi} is appearing across the depletion layer. Now let me see whether I can take a just a look at this

This actually is metal, let me change it up, what I was intending to say was it is almost like a pn junction instead of p type material you have got metal there, then you have the depletion layer.

(Refer Slide Time 13:26)



The voltage across the depletion layer is equal to built-in potential that is V_{bi} and that is .953 now what we are trying to see is, what is the depletion layer width there, when you doping it? The doping is very high here.

Now just like in the case of pn junction, instead of p you have got the metal there, all the depletion layer is on this side, if it is p plus there is no depletion layer, if it is metal there is no depletion layer here, the electric field actually is plus here minus there from right to left. Now by solving poison's equation you know that, the electric field will be linear peak at the junction and linearly decaying it is 0 at the other end of depletion layer.

What is the peak value of depletion layer? From the [] law you can see, the peak value of electric field actually is equal to total charge on the right hand side of this peak, circuit value, total charge there that is q N_D into W_D, that is the total charge, divided by epsilon_r epsilon_0 this is from [] law per unit area of course.

Let us say, peak per unit area, how many field lengths are crossing there? The field lengths which are crossing that are q N_D into W_D, so charge is this quantity. Once you know what is the peak electric field is, we can find out what is this voltage is.

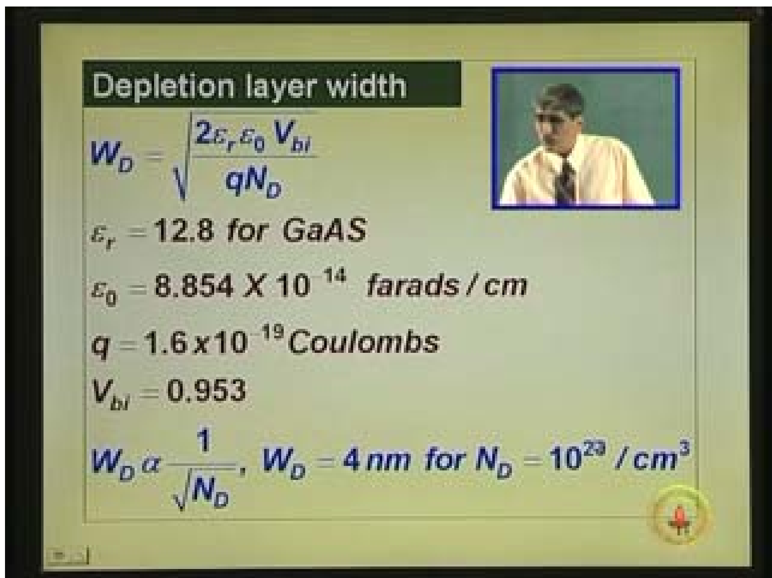
See what we are trying to find out is, we know the value of V_{bi} we know it to be equal to .953, what we are trying to find out is, what is the depletion layer width for a particular doping?

So for that we are making use of the expression. Peak electric field is like this and once you know the peak electric field, the voltage across that is peak electric field into the width divide by 2. It is actually the area under this curve, area under the E_p equal top E dx is the voltage integral, and the integral of electric field is the voltage so E_p into W_D divided by 2 that is the area under this curve.

That turns out to be and the plot of that of course is square law, it is like this integral of that linear function that is what I have plotted here. So this voltage V_{bi} that is actually equal to this quantity. This is a well known formula for a fully depleted region q N_D into W_D squared divided by twice of $epsilon_r$ $epsilon_0$ which actually is half of this into $W_{D.}$

That means once you know V_{bi} we know now, that is why we evoked that, we said phi_{Bn} is equal to V_{bi} almost that is .953 for gallium arsenide. Now what we want to see is, if I make this ohmic so that is fixed V_{bi} because once phi_{Bn} is fixed phi E_g minus phi_0 that is fixed and once that is fixed, I can control the width of the depletion layer by changing the doping.

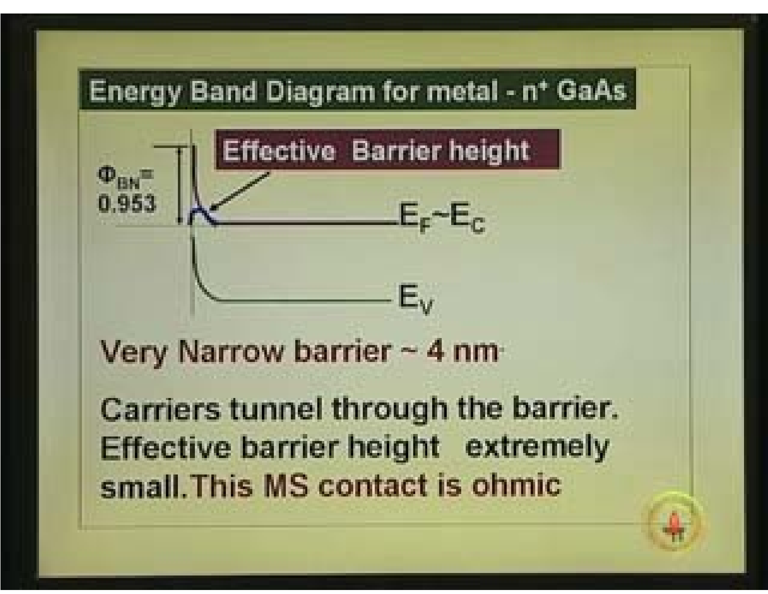
(Refer Slide Time 16:56)



Now I increase the doping W_D will fall. So just using that expression here, W_D is square root of that quantity. Now you can see V_{bi} is fixed at .953 a marginal change will be which will be the doping because logarithmically varying as doping, almost constant. So if I increase the doping W_D will fall. Let us take the case where the gallium arsenide material where $epsilon_r$ is 12.8; $epsilon_0$ is free space 8.854 into 10 to power minus 14 farads per centimeter; q is 1.6 into 10 power minus 19 coulombs substitute all those things there and V_{bi} which we have estimated keen because it is equal to E_g minus phi_0 .953 I plug in all these things and if I put N_D equal to 10 to power of 20, you know very heavily doped.

Then W_D turns up to 4 nanometers, very narrow, the width of the depletion layer is very narrow, now 4 nanometers 40 Armstrong that is the diagram now. So the depletion layer you see is decided by that variation potentially still .953 that is phi_{Bn} which you see.

(Refer Slide Time 18:03)



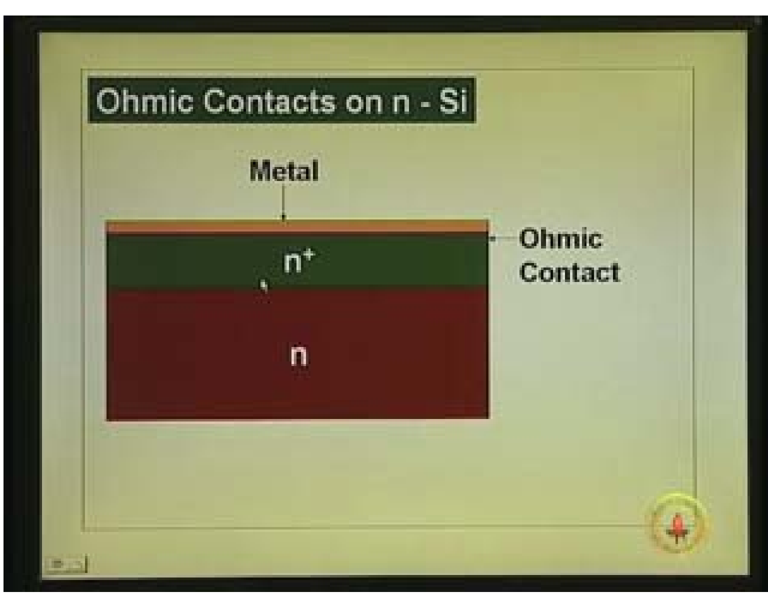
So apparently, it looks as if the barrier height is .53 but now what is happening is the total width here, maximum width is actually 4 nanometers that width, that is what we have calculated depletion layer width is up to this point that is 40 Armstrongs. And as you go towards this the barrier height becomes narrower and narrower less than 40 Armstrongs

which means above certain height here, which as if there is no barrier because the electrons can tunnel through from either side.

So the barrier height will not be phi_{Bn} the effective barrier height what I put here is by this line is very narrow, very very small, it will be much smaller than .953 that is the nice thing about heavily doping. When you dope very heavily, whether it is silicon gallium arsenide or indium phosphide whatever it is, when you dope it very heavily its barrier height becomes very small because whatever barrier is there that is very thin so carriers tunnel through that. For the electrons hardly there is any barrier.

So the carriers can tunnel through all through these portions and the effective barrier height is extremely small. That is the key. When the effective barrier height is very small immediately you can say now that this is ohmic contact. If you want to make an ohmic contact take a n type substrate make the surface n plus and put the metal on top of that, so n plus n is of course extension of n layer and the barrier which you are concerned here between these two if you have reduced it that becomes ohmic contact to that.

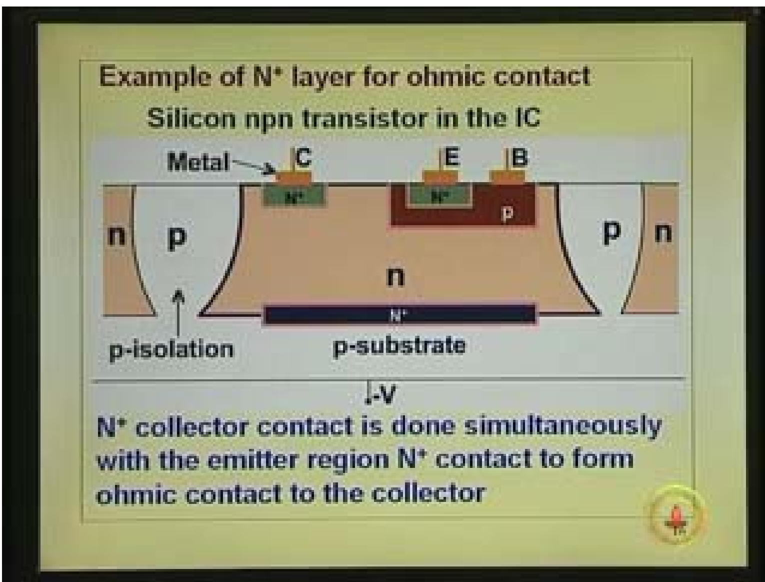
(Refer Slide Time 19:50)



So you have ohmic contact with n plus region and n plus sets more transitions to n layer. Whenever you want to make ohmic contact what you do is, take n type substrate put n plus layer by some means implantation or diffusion and realize a metal semiconductor

contact on the top of that. So that is the key to make it ohmic contact to n type substrate invariably everywhere we will see a classic example of this ohmic contact in silicon itself is this. What is this? This is actually the transistor in the integrated circuit in silicon.

(Refer Slide Time 20:34)



Now we can see, just for those of you who are not very familiar, we can see that is n plus layer on the top p layer n layer. There is an npn transistor. Just quickly go through the way this is made so that we familiarize with that.

Start with a p- type substrate and on that p type substrate you form this n plus layer in this portion, which actually is called a buried layer. The purpose of that is to make the n plus region, to dope the next n layer very heavily there.

Put a buried layer there, n plus layer that means you are providing a low resistance path for electrons so put that selectively, so that you can do by oxidizing and diffusing only in this portion then grow an epitaxial layer n type layer on the top of that- p type layer, n plus layer in selected portions, then n n layer may be 10 microns, 15 microns depending upon the integrated circuit technology that you have got, today they talk of that one or two microns this whole layer.

20 years back it must have been 20 microns now it has made on 1 to 2 microns. That is the n layer which actually is the collector region, and after that you do is you want to have the transistors isolated from each other. So these regions p region here through this window you diffuse p, I am not showing the masking layers etcetera just showing the final structure, just going through the process very quickly just to give an idea what is happening here.

You had the n layer all through we can see this is n this is n but in between you have put a p layer by diffusion. Now what you have got is this n layer the cross section of the entire layer that goes like this, the cross section actually goes like that, top view will be like this, what I am showing is a cross section. You have got the island of the n layer like that if you see top you have got through you have got a p region.

Another island is coming nearby separated by p layer, if you reverse bias the pn junction each island is isolated by the reverse bias junction, that is called pn junction isolation which is very popular in bipolar IC's, not very high performance but moderate performance IC's will have I mean very high performance IC's you must put a dielectric layer there because pn junction isolation never gets submission. Now notice here this p layer is connected to this p substrate and that is connected to minus V so that the pn junction is always reverse biased.

So in the IC, the most negative point, the substrate is connected to the most negative point. Now coming back to this device, in this region, this island is the region where you form your transistor. You can see what is done is diffuse a p layer diffuse n plus layer for the emitter and when you diffuse the n plus layer simultaneously, you diffuse this n plus layer for the collector.

What is the purpose? Finally, reverse put metal to make the contact ohmic contact p layer if you put metal it will form ohmic contact fairly well in silicon because one third of E_g that barrier height is low. So the phi_{Bp} will be small for this, no need to do any treatment there, you get automatically the ohmic contact whereas here naturally you have got this n plus layer which is the emitter layer. You have got the ohmic contact but you form this n

plus contact to the n layer simultaneously along with this so that you have ohmic contact here.

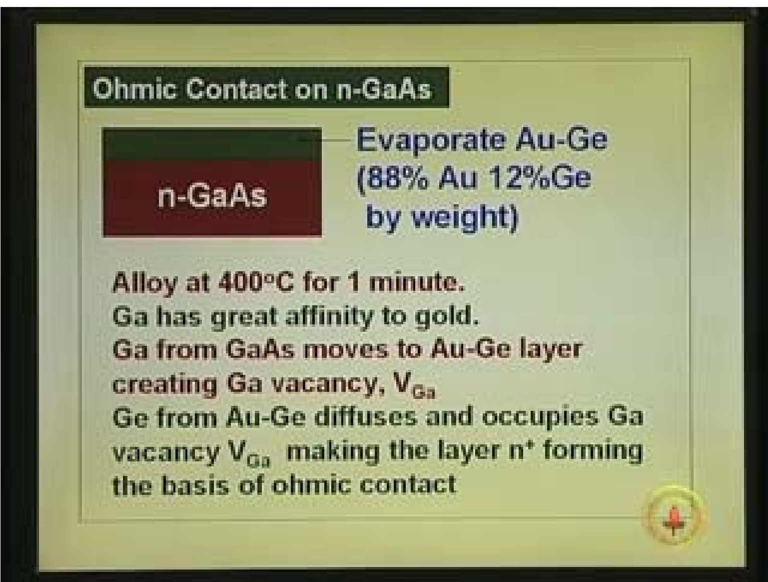
If I did not do this diffusion here what would happen? This metal and this n layer will form rectifying contact. You have a collector with a contact which is rectifying which is not ohmic, it will block the current flow in one particular direction. So that is why this ohmic contact is formed by a simultaneous diffusion of this n plus region along with the emitter. Now you can see if you want to make a pnp transistor you will have problem because we will have to do n plus contact to the n base region that has got to be separate step. Here, the step is automatic with that step but in a pnp transistor it should be p plus this is p region that p plus can be done, but what about this n region that is base? Instead of p there will be n so you must put an n plus region that has got to be the first step. It is one of the difficulties when you want to make a pnp transistors.

In fact if you carefully see the pnp transistor are costlier than npn because of that additional step of forming the n plus contact. You may say after all one more step, it is not just one more step, you have to oxidize open the window by lithography, edge the oxide then the diffusion. So many steps are there. When you do so many steps chance of defects coming in, contamination coming is more and the result will be E will go down that is why the cost is more.

This is just to illustrate the power of why n plus layer is put in this structure, the whole reason is that. That is the high D_{it} making the barrier height large in the n type substrate making that rectifying if you do not have the n plus contact.

Now coming to gallium arsenide, how do you form that? People have tried various tricks. One of the tricks which is working out universally is n type gallium arsenide, see after all what you need to do is realize a n plus layer on the top onto which we can put a contact metal.

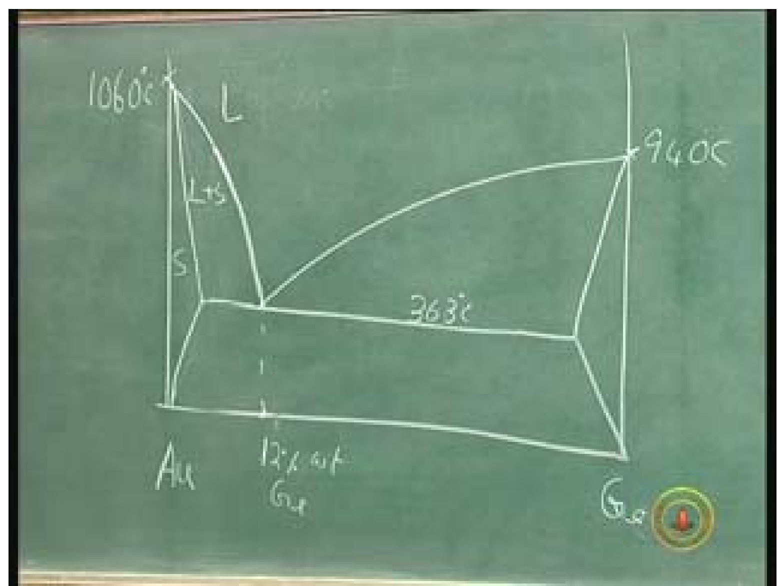
(Refer Slide Time 26:51)



So what is done is evaporate gold germanium onto gallium arsenide and gold 88% by weight germanium 12% by weight, in fact when a germanium 12% by weight would correspond to about 27 atomic percentage of germanium. You evaporate that and then heat up this in a nitrogen ambient or better with hydrogen and nitrogen to prevent any oxidation so alloy it at 400 centigrade all that happens in just one minute.

What happens here is the gold germanium, if you take a look at the phase diagram of gold germanium, gold and we are calling back gold into the gold germanium, the phase diagram that is germanium.

(Refer Slide Time 31:11)

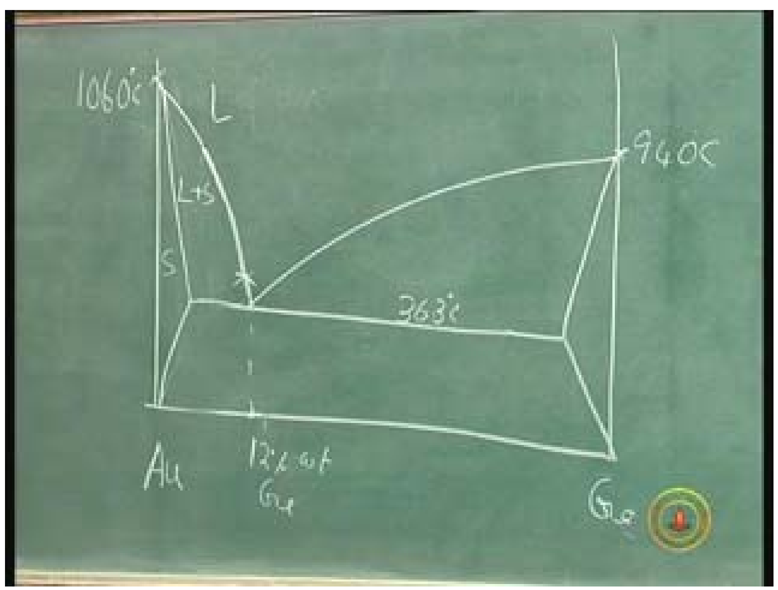


If you take germanium 100% gold that has a melting point of 1060 degree centigrade melting point and germanium has a melting point of about 940 degree centigrade around that 936 or so, that is the melting point, now when you mix them together the phase diagram is something like this, the fact we have aligned here, let me just see if I have those numbers here I thought I have noted down some of those numbers, 363 degree centigrade. If I remember just draw it once rather like this, almost comes like this or rather like this whatever way it comes, now that is the liquid; liquid plus solid is solid. when you come down from this side like this, there is a precipitation of solid and a composition of solid will be given by this curve, so the liquid plus solid and this is solid.

But you do not have to worry about that, what I am focusing is this point. From there this is liquid in the sense whatever composition is there vary gold from 100% melting point is 1060 and if germanium alone is there 100% melting point is 940 degree centigrade. If the composition is 12% by weight germanium at that point, that is the melting point, what we are telling is there is a melting point for gold germanium which is lower than either gold or germanium when the composition is 12% by weight or 20 per 27% atomic percentage. That is 27 atomic percentage of germanium or 12% by weight the melting point is around 363 degree centigrade, nice thing that happens is you go to 400 degree centigrade somewhere up there slightly above.

These are the two things that happen. The mechanism by which you get or ultimately what happens is when you do this alloying at 400, in just one minute you get the ohmic contact nicely with gallium arsenide. The mechanism is like this: the gallium from the substrate gallium arsenide moves towards the surface in fact when you have when you go to temperature slightly more than you take t.

(Refer Slide Time 33:24)



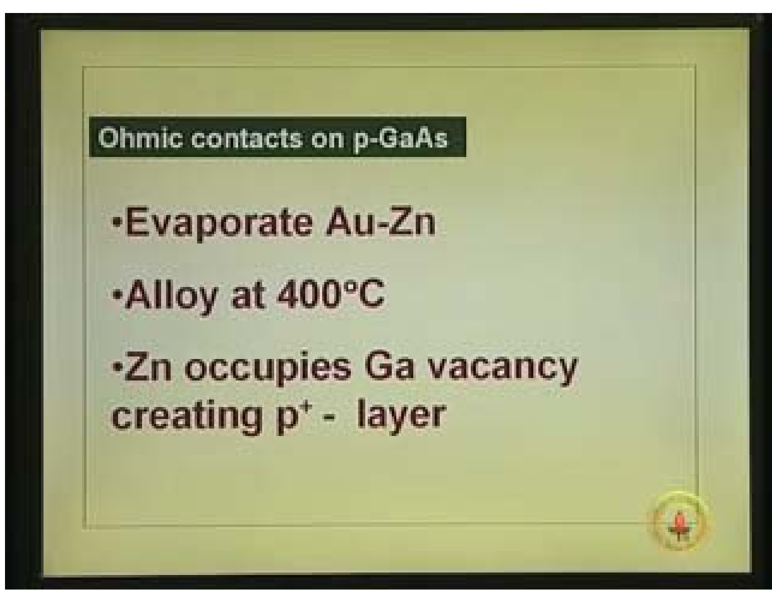
This is the melting point, if I go slightly more than that you are going on to this point I am filling this point because this becomes less gallium there less than germanium, germanium concentration comes down here, what happens is the gallium from the substrate moves up to the molten layer, this is the molten layer on the top, it moves on to this top layer and from the top layer germanium moves down so gallium moves up because gallium has got lot of great affinity towards gold let us put it that way. So it moves towards the surface where the this is actually in a molten stage so we go slightly up above that temperature 400 degree centigrade if still it is liquid, the germanium concentration can come down from there, still it be liquid state and that germanium concentration it has come down, why it has come down, till gone down to the gallium arsenide. All that had happened is, gallium has moved up to the gold germanium and little quantity of germanium has moved down to gallium arsenide.

And where is it sitting? Germanium is sitting where the gallium has moved out that is what we said here germanium from gold germanium diffuses and occupies gallium vacancy $V_{gallium}$ making the layer n plus forming the basis of ohmic contact.

Germanium is fourth group element. Gallium has moved out to the surface that means gallium vacancies are created, so germanium goes and occupies the gallium vacancy so it makes it n type n plus because lot of germanium is there, 10% is large number compared to total number of gold germanium is present there of the order to 10 to the power 20 so you will have this ohmic contact formed.

The key thing is this is the most successful ohmic contact that has been made on gallium arsenide. What about p type? In fact even in p type if you want to make a good contact you better make it p plus.

(Refer Slide Time 34:29)



The moment you make it p plus the barrier thickness becomes very small. As a result the effective barrier height becomes very very small. So what you do is instead of gold germanium you put gold zinc.

Zinc some percentage is there it is not exactly 12 slightly different, you get actually alloys of gold and zinc you can buy it, you can buy gold germanium alloys if you do not have how to make gold germanium contact it would not sit idle.

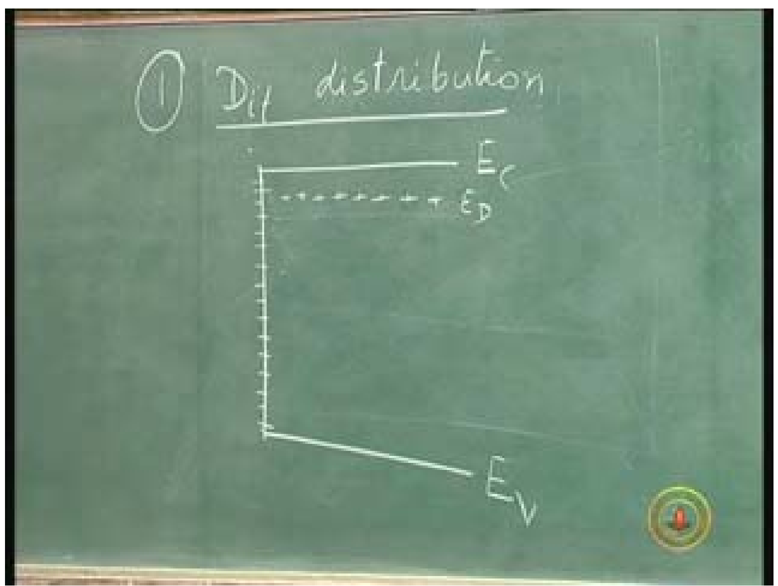
You weigh gold, you weigh germanium 12% of that, put them together in a bowl, melt them together, evaporate you get gold germanium. Similarly, we can do with zinc. So gold zinc, alloy it for 400. Zinc occupies gallium vacancy. The key thing is when you have the molten gold compound, gold germanium or gold zinc, gallium moves up to the layer creating gallium vacancy whether it is becoming n plus or p plus would depend upon what is moving down; if it is gold germanium, germanium moves down making it n plus; if it goes gold zinc, zinc is p type, so that makes it p plus. So you make p once you make it p plus, it is very good ohmic contact. What about indium phosphide? Indium phosphide also takes same trick. After all gold germanium if gallium moving up indium will move up and germanium will go and occupy that position to make it n plus.

(Refer Slide Time 36:04)

Here zinc will move down to occupy the indium slightly, so you get good ohmic contact to n type and p type. Now few things more I want to discuss in fact I am deliberately kept it open here today because some of the things which I have promised that I will discuss I do not have this is in the slide but I will go to board and explain those things to you.

We have some time left on to which we can discuss very clearly, number one what we said is the D_{it} distribution, I promised in the last lecture that we will discuss this sometime so I have got some time to discuss this.

(Refer Slide Time 39:05)



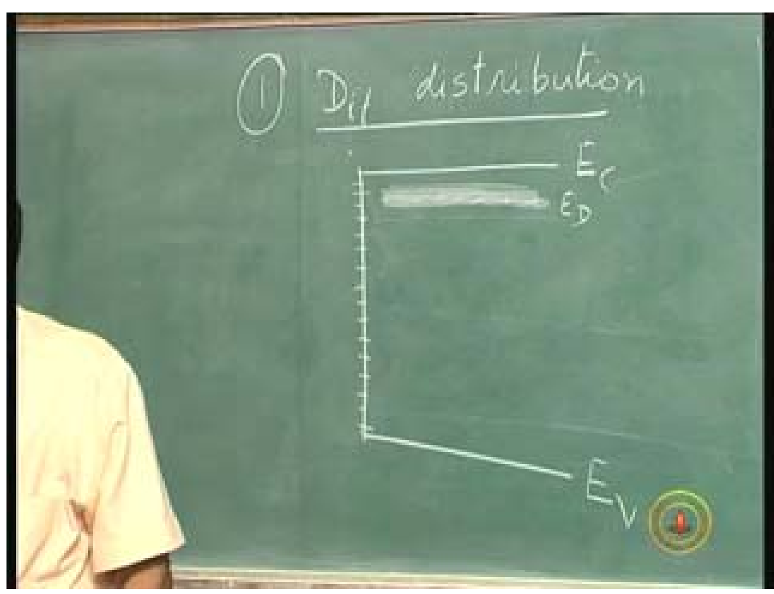
We said on the surface this is E_c and this is E_V. What we said is this interface state density will be distributed all over this, all over the band gap and this is due to the tangling bond. So it is like introducing an impurity in the bulk of semiconductor if I introduce doping what will you get? You get a level any deviation from periodicity gives a level so in the bulk what you got here in fact you draw it like this, these are the donor levels. you do not draw a continuous line there; you draw either like this or dotted lines and all these levels each one of those correspond to one donor and the entire thing corresponds to one level.

Now this actually is against Pauli's exclusion principle. What does Pauli's Exclusion Principle say? No two electrons can occupy same level but here if you see when it is occupying it occupies the same level; it is valid under certain conditions. It is valid so long as these atoms are far away. The atoms will be far away if the number of atoms are less, that is if the doping level is low, low meaning 10 to power of 16, 10 to the power 17, 10 to the power 18 till it is low.

Because silicon number of atoms per centimeter cube is 10 to power 22, so compared to that 10 to power 16, 10 to power 17 is low but if you increase the concentration what happens? The atoms come closer and closer. Once they come closer and closer what happens? The wave functions interact with each other and instead of the one level, we get multiple levels. So when you dope heavily the semiconductor that single level which is donor level splits up into number of levels.

You can note when it is a heavily doped semiconductor, you can no longer talk of a single level we will talk of a band of energy, and it can even merge with conduction band that is the situation that you have when you have that.

(Refer Slide Time 40:02)



So multiple levels are there; it is just like the energy band formation. How is the energy band formed? How did you get the conduction band? The conduction band is formed because there are silicon atoms which are nearby the electrons in the outermost shell interact with each other as multiple levels you get.

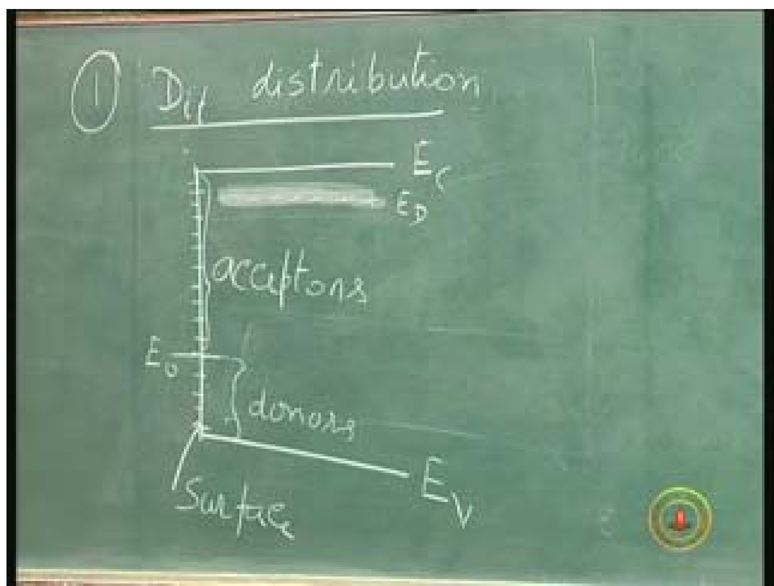
Same thing happens with impurity when the impurity concentration is high. Now let us take a look at it. So that is the same type of impurity can give a energy band split over as you keep upon increasing doping the impurity band become wider and wider.

Now take a look at this surface. This is the surface and we said there are surface states. What are these surface states? They are due to loss of periodicity, due to the tangling bond. Now this tangling bond or this particular surface atom can either donate that electron be a donor or accept an electron further because it needs some electrons, it can accept an electron just to complete the bond and be an acceptor. That means actually it can act as donor or acceptor. Some of them will form donors, some of them will form acceptors because depending upon the situation some of them will accept electrons and some of them donate electrons.

Actually that means, you cannot represent it by one level, we cannot represent that either donor level or acceptor level by one level. Even if you say it is one donor level since there are 10 to power of 15 atoms per centimeter square on surface, that many levels are there and they all cannot stay on same level because uncertainty principle. You cannot have the same level, so that splits up because number of atoms is so large on surface. See when you say 10 to power 20 per centimeter cube it is in the volume; same number is there on the surface when you talk it is 10 to power 15.

So now, actually you have got very large number of defect states on the surface that means actually very large number of levels are there, all them were split up and that is why you get the entire thing splitting upon these band gap instead of one level you have got these states spread out. Just like this splits up with a band, this has split up with a band on the surface. So you have got those surface states large density of states distributed all over the surface now as I said it can be donors or acceptors, where will the donor levels will be?

(Refer Slide Time 43:40)



We said that there is a neutral level E_0 somewhere, which is characterized that particular material and we said below that there are donors above that there are acceptors. How have we justified this? This we said as acceptors, this as donors. We said let us put it here E_0 is that level all these are donors and all these are acceptors. How do we say that? Usually when you talk of semiconductors, shallow donors we say it is near conduction band, shallow acceptors we say it is near valence band. What is the meaning of that? The meaning of that is when you put the donor level near conduction band the meaning is the energy required to remove that electron and donate to the conduction band is very small

That is possible when you introduce that fifth group in silicon or fourth group in the gallium side so one extra electron is present which is free to move and that can be very easily removed. That is the idea. That is why you locate the donor level there. Similarly, whereas acceptor level you will locate it here shallow, that is called shallow level because it can easily accept electrons from valence band to that.

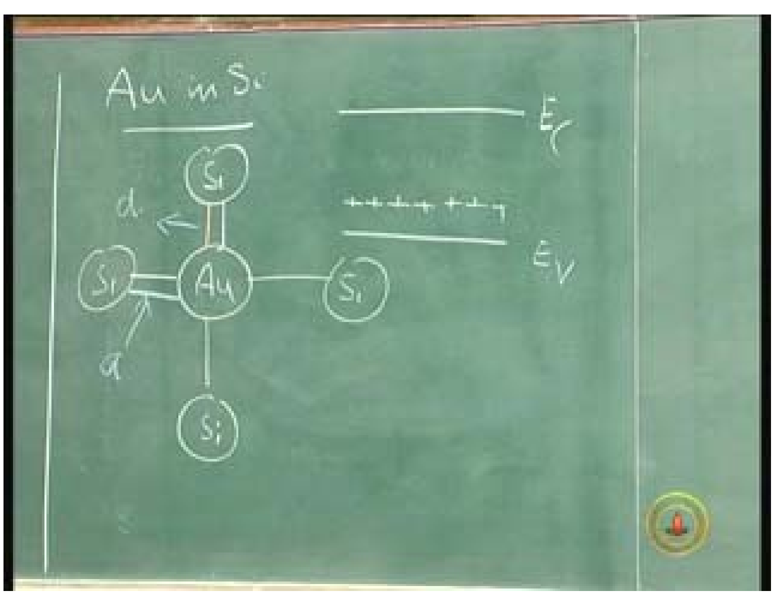
That means when an electron is accepted from valence band to that level it has donated a hole to the valence band that gets ionized. When you say donor is ionized what you say is it had donated electrons to conduction band. When you say an acceptor is ionized it has accepted an electron but donated a hole to the valence band both are donors. You will

look into it one donates electrons one donates holes, one donates negative charge other one donates positive charge, that is the key.

So when the donor level donates electrons it will become positively charged, when acceptor level donates a hole it becomes negatively charged, when it takes an electron. Now here we are talking of the reverse situation where donors are here.

I will illustrate that with an example of gold in silicon. Gold in silicon gives multiple levels because that is very simple to understand, it is all once you understand that you can easily understand this.

(Refer Slide Time 49:24)



Now let us take a look at gold in silicon. Gold has got one electron, valence electron and it can go and occupy the lattice side in silicon very comfortably. So if it can occupy what happens? Gold-silicon-silicon. I am just taking this example because it is very easy to understand. This is the thing which is encountered often, silicon. Now if you see that is the electron available for gold and then that is the silicon-silicon-silicon-silicon, electron from silicon are four are there.

If this were silicon atom, what would happen? One two three bonds would have been formed totally. All the bonds are formed if this were silicon. Unfortunately, this is

occupied by gold which has only one electron. Silicon has put four electrons which satisfies all the bonds so only one bond is there. Now this gold can behave the way it like in the sense depending upon situation it is again an amphoteric material gold in silicon. It can behave as an acceptor or a donor.

Now let us see this can donate this electron. Now to donate this electron you must remove the electron take it to conduction band. How much energy will be required? If it is a silicon atom if I break this silicon atom, actually to break the bond and donate it to conduction band, how much energy was it required? 1.1electron volts

Now for this gold atom, its size is different and etcetera so this energy required is not exactly 1.1 if it is a donor it would require less than 1.1 electron volts not 50 milli volts and all that. You cannot break this with small energy because it is not the fifth electron of the donor, of the phosphorous also. So what happens now is E_c and E_V the donor level due to donation of this electron will be located far away from the conduction band, somewhere here, it is not actually exactly equal to 1.1, if it is 1.1 electron volt it is something like .3 electron volt that means this is about .8 electron volts.

So almost large energy is required to break this but it can still donate. Now, let us again take a look at this, this explains now that your donor level is closer to the valence band. It is a deep donor. Why it is said deep donor? Not because it is like deep in semiconductor it is a deep donor because it is deep in the energy band diagram far away from the conduction band. If it is very close to conduction band, it is called a shallow donor; if it is very far away from the conduction band it is called deep donor, so deep donor. It takes a lot of energy from there to there in the conduction band.

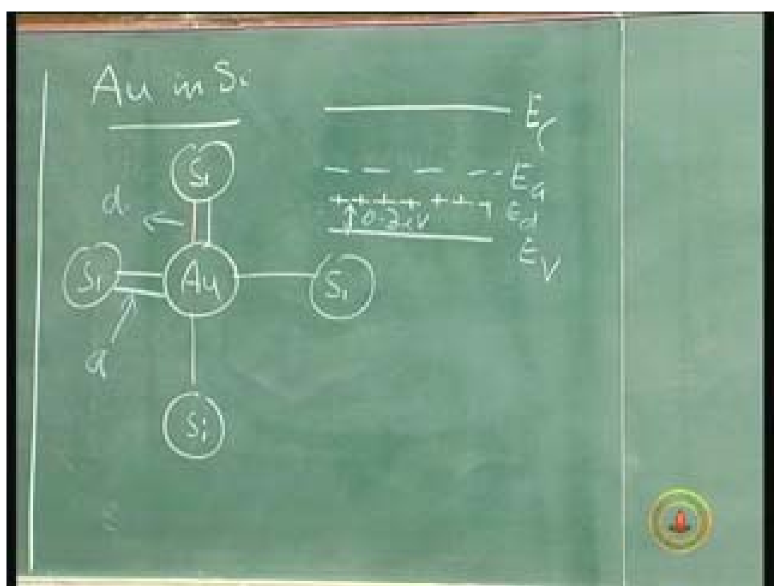
Now there is a chance that instead of donating, it may find it easier to accept an electron. Supposing it accepts an electron what happens? This is the donor and this is a. Instead of donating it accepts an electron. That is another way and when it accepts an electron where will I put the energy level diagram now here? Corresponding to that there is a level when you say there is an acceptor level. You must be able to locate in energy band diagram.

Now I have identified this particular location closer to valence band because it requires lot of energy to remove this electrons, because like the bond silicon. Now once you have identified this level with those valence bands here .3 electron volts above the valence band here. Where will you identify that acceptor level? Will you identify it above or below? That is the only question I am asking. It will actually you have to look at it from the point of energy.

If it has donated then it has got charged positively, it requires certain amount of energy to take electron from infinity to that point. It is easier to take to a plus charge. Suppose it is accepting an electron, will it require more energy to take electron to that point compared to this or less energy? Because when it is accepted, the charge state of this is negative; if it is donated the charge state of that is positive.

Please understand, when a donor donates it becomes positively charged and acceptor accepts it becomes negatively charged. So that is negative charge if it is accepted instead of donating if it has accepted an electron it has become negatively charged to situations.

(Refer Slide Time 52:17)



So this level corresponding to this acceptor level will be above this, I will put that E_a E_d. So I will go through that now, once again because if it has donated I have identified that

energy level is closer to valence band, because it requires more energy to break that bond. So energy level is .3 electron volts around those electron volts in amount to balance that.

Having said that, suppose instead of donating if it were accepting it would have negatively charged. So between the two states, the donor state and acceptor state if I am trying to bring an electron that is why there are two atoms, one atom donated; one atom accepted, it is not the same atom that is doing the job; one atom donated one atom accepted, there are 10 power 15 gold atoms. If you take the electron from infinity to that point to the donated gold, you have to spend a certain amount of energy that is marked by the E_d there.

Now there is other atom which it has accepted, between the two of them that is one is donated, the other one is negative accepted. It becomes easier to take electron to this point compared to taking electron to this point. To take an electron to negative charge side it will have to spend more energy, so negatively charged side will have more energy compared to that.

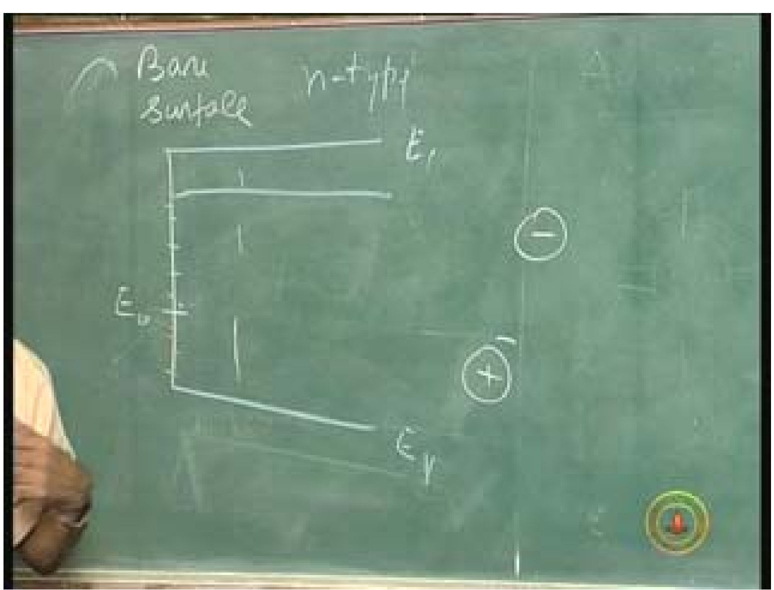
So what I am trying to point out is, you have got a situation now where the deep donors are below or the same types of impurity if it is donating and accepting the donor levels should be below and the acceptor levels will be up. Now, that has come up here. The donor levels are below and the acceptor levels are up. You have got some of them donating, some of them accepting but all those which are donating if it is only one which is donating it will be only one level, if there are hundreds of thousands of them that are donating then they spread up into donor band that is the donor band. And all the acceptor levels are above that. It is exactly the same situation. So you get a donor band and an acceptor band separated by a neutral level; neutral level in the sense if the Fermi-level is coming to this point, the levels above that will be not occupied really and the levels below that will be occupied and that the whole thing will be neutral.

Let us go back to that diagram and see now what happens if I take a semiconductor which is a surface with the no metal.

Let me just draw that I still have few minutes left so I can just discuss that quickly if I have not finished it I will come back to it in the next lecture.

I have a conduction band, valence band, n type semiconductor, this is the surface that is the neutral level, all these levels are occupied and all these levels are not occupied now, if it is not occupied neutral.

(Refer Slide Time 56:27)



If the Fermi level is there this is a free surface no metal, absolutely no metal. I have taken a semiconductor surface like that if I take a look at that what will happen will be, because of high density of states no oxide nothing, just a bare semiconductor. We are talking of a bare semiconductor surface. What will happen now? This is n type. What is the charge here now? What is the charge on the surface? Fermi level is above the neutral level. All the levels below the Fermi level are occupied. What is the charge state of this donor? Donors are occupied neutral, these are acceptors occupied by electrons negative, and so this is negative charge.

All these states are negatively charged. Where has the negative charge come from? It has accepted electron from the semiconductor. If it has accepted electron from semiconductor what will happen? n type semiconductor electrons have been removed from some region

here. So this region has contributed to the electrons here. That means this region will be depleted.

I think I would discuss this next lecture because this is a very important concept even if you take what emerges from here is, even if you take a free semiconductor surface do not put any metal that will have depleted surface, because these electrons will have to supply it from this one, so entire energy band diagram will bend. That aspect is a very interesting aspect which hurts the MESFET when you make the MESFET in enhancement mode that is the difficulty people have. So you will have depleted surface this concept of some of these things I will bring in next lecture before I go on to the IV characteristic of that.

So we will see you next time.