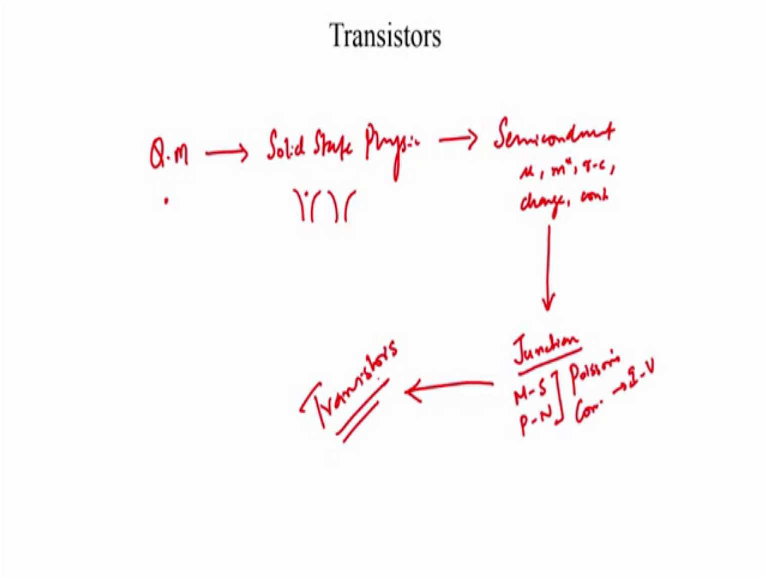
Semiconductor Devices and Circuits
Prof. Sanjiv Sambandan
Department of Instrumentation and Applied Physics
Indian Institute of Science, Bangalore

Lecture - 29
Bipolar Junction Transistors (BJT)

(Refer Slide Time: 00:15)



So, far we have, you know, just to summarize, you know, our journey so far in this course, we initially started off with basic ideas of quantum mechanics. And then we extended this to develop ideas on solid state physics ok which is basically what happens as you go from one electron sitting in a potential well, to an electron in a solid, where in you have a periodic potential value because of all the atoms.

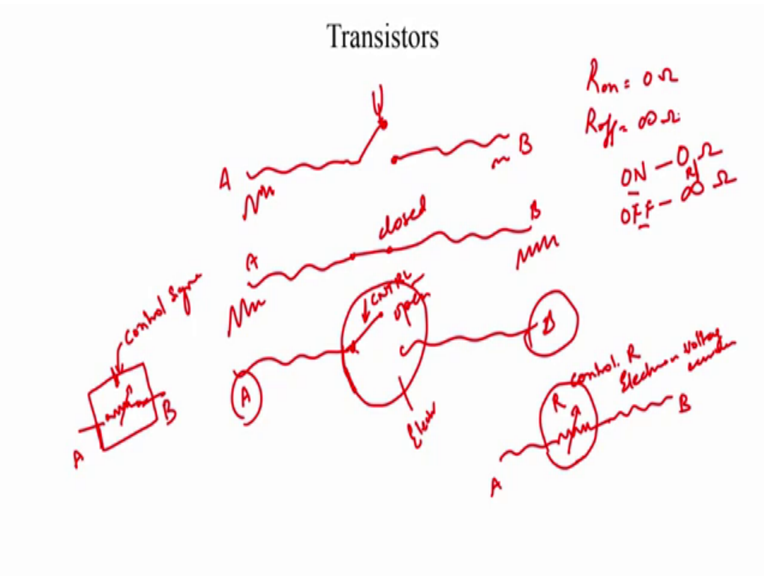
And then from here, we brought in basic ideas of semiconductor physics ok, and we spent a lot of time discussing concepts of mobility, effective mass, you know, recombination generation, the different charge transport mechanisms and the continuity equation. And then we continued with our journey towards looking at junctions. So, this was a very important step when and we looked at 2 kinds of junctions the metal semiconductor junction and the P-N junction.

And we saw how we saw the tools and techniques right off using solving for electrostatics using ideas like the Poisson's equation, and then trying to use the continuity equation to identify the current voltage characteristics etcetera. And now we are going to take a next big leap in our journey, and we are going to start talking about transistors ok.

So, these are so, we have now reached a point where the flavor of the course will slowly shift from more fundamental physics and device physics towards more of circuit design, ok. So, we are not here there, but the course will start taking that taking on that flavor. With regards to transistors, you know, we you know firstly, we will just look at what are transistors.

So, let us say, you know, just to just to sort of introduce this idea, these are essentially in a simple language we saw the building blocks of all your modern day integrated circuits right, along with all the diodes etcetera. So, the idea of a transistor is to construct a switch or a resistor that you could control whose resistance you can control electronically.

(Refer Slide Time: 02:52)



So, let us say you have you have a point A and, you know, you have a wire and you want to send a signal or you want to send a message to point B. Now one way to do it is to have your regular old switch which is probably used for controlling the lights at in your homes. And you have a signal and whenever there is a signal that needs to be transmitted to this side, you have somebody push this switch down and the switch is closed, and you have you have your signal whatever your signal is being transmitted to this point B, ok.

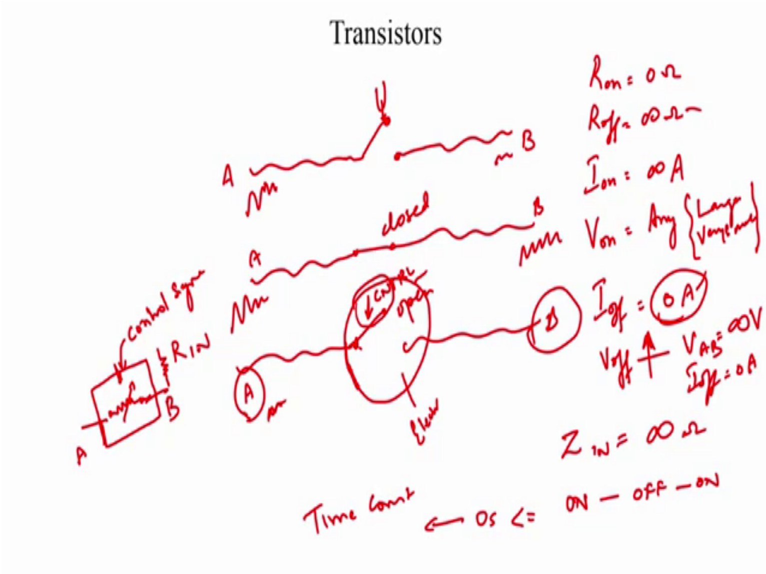
And then we do not want the signal to be transferred you keep the switch open ok. So, here the switch is closed and here the switch is open, and your message at point A will now not be transferred to point B. So, how do I make the operation of the switch very fast? If I have a mechanically operated switch, it is going to be a it is going to be a very slow process; because your now dependent on that mechanical time constants ok, which are which are quite slow ok.

So, the idea of a transistor is to develop a switch which can be controlled electronically ok, an electronically controlled switch; which means that this control of the switch going from on to off is done electronically, and your signals heading from A to B are all electronic. And it is a very nice way to interface the real world with the world of electronics, by having sensors and then having a bunch of switches performing different operations. To take this one take this idea one step further, what we need here is; so, this switch isn't either it is got 2 states he is either got a state of on or off. And ideally what do you mean by on is it is called a 0 resistance between A and B, and what do you mean by off is that it is got an infinite resistance between A and B. So, that is that is our idea of the switch.

But let us say that we would like this resistance to have any arbitrary value, and that we would will not control this value also electronically. So, what we are talking about is now a device that is a variable resistor that sits between A and B, and we would like to control, control the resistance of this resistor electronically ok. So, maybe by applying a small voltage or a small current etcetera and that is the idea of a transistor. So, this device that we are talking about is the transistor. So, essentially the transistor should have 3 points right. So, corresponding to the point A here, corresponding to the point B here and corresponding a terminal which corresponds to where the control signal is given, ok.

So, that is an between A and B you have this variable resistor. So, that is the idea of a transistor ok, and that is what the transistor is. Now there can be different kinds of transistors, but all of them would have certain properties. Since we have plenty use the transistor for, you know, we imagine the transistor to be used in such an application. What are the ideal properties of a transistor ok? So, the first one is that if I am using it in on and off state, let us say I am using it as a bi stable device ok. It is either conducting or non-conducting. What we would like is that, the on resistance of the transistor is 0 and the off resistance of the transistor is infinite ok. That is what we would like. And then if it is on, we would like the transistor to be able to carry an infinite amount of current.

(Refer Slide Time: 06:54)



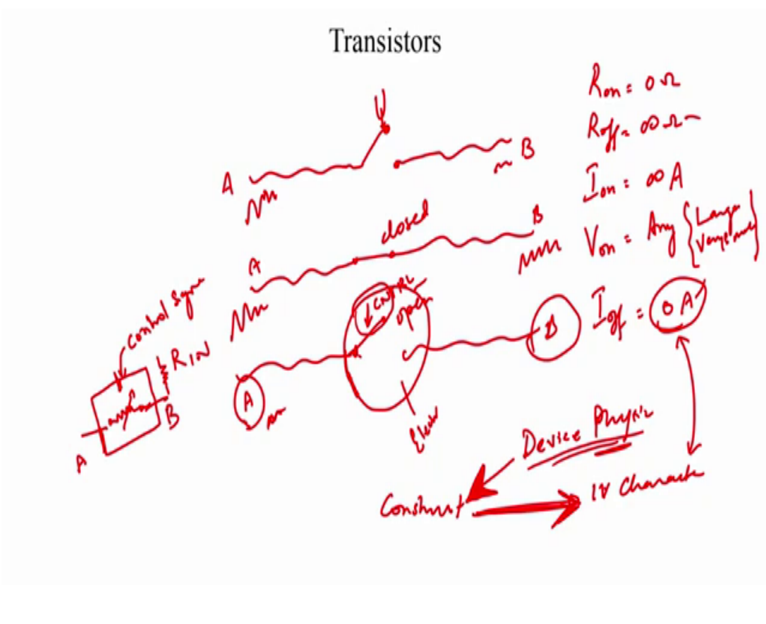
We do not want any limitations on the current density current density. And we would like it to operate at, you know, all kinds of voltages, you know, you should be able to take very large voltages or very small voltages as and how it desire, ok. So, this a bit arbitrary, but you will understand the meaning of this later. And if you look at the voltage across if you if you look at the off current ok, if when transistors off we would like it to have 0 current which is basically a consequence of the resistance being and off state resistance and it should take a value of 0 amps at any voltage. So, we want no matter how large the off voltage goes we would like this current to be 0 amps ok. So, by V off here I mean the voltage difference between A and B.

So, even if V AB is infinite volts I would like my device to carry 0 amps of current ok. That is probably asking for a lot, but that is what we want. And the other thing is that we would not like the control signal. So, here you want to send a signal a message between A and B, between from A to B right. You do not want any signal that you used to control the transistor to influence that message. So, essentially what we want is that if there is an if there is a resistor here ok. So, let me call that is R in, we would ideally want that resistance to be infinite; or let us say if it has an impedance we would like that impedance to be infinite ok. We do not want the control signal to impact the operation or the message it is not corrupt or of influence the message being transferred between A and B.

And we can continue listening a lot of these properties, but the other one is that we would like the switch or the transient between on to off state, or the time taken to go from on 2 off and back to on to be nearly 0. We would like all this to happen instantaneously. So, we want all this operation to be having a 0 time constant, 0 second time constant ok. So, these are some of the points that, you know, or some of the dream assets; so, the dream properties that a transistor must have ok. But realities quite different and depending on how you construct a transistor, you get some advantages and you lose and you lose some of the benefits ok. So, what we will talk about is we will focus our attention not only on these very general features of a transistor.

But we will try to look at, you know, how do you design some kinds of transistors, and what are particularly the once that are very commonly used. And we will look at the device physics of these devices, we will use all the tools and techniques we had so far and look at the device physics of all these devices, and then extract the current voltage characteristics. And finally, try to build in equivalent circuit ok. So, we are going to follow the same protocol, we have already done this through metal semiconductor junctions and P-N junctions we are going to firstly construct a transistor

(Refer Slide Time: 10:25)



Then use all our tools of device physics on this construct and then from here extract from here extract your IV characteristics. So, the current voltage characteristics, and see whether it matches these properties that we desire ok. And if it does not we will try to improve upon it, or try to find the reasons as to why it does not ok; so, dones the exercise that we are going to pursue ok.

Now, once again all these topics are going to involve a lot of mathematics, there is going to be a lot of derivations around. Because as the device gets more and more complicated your device physics is also going to get more and more involved as it going to be more involving ok. But the point of the courses not nobody is going to test you on whether you can remember the derivation, that is not the point of the course. The point to the course is, you know, how to approach the device physics ok.

So, that is why we will go through this painful probably some people it is painful excursion of me talking about the derivations and me writing down the equations. But at the same time keeping in mind as to what the concept is, and what the physical intuition behind all that is ok. So, it is in some sense necessary it is a part of this course, but nobody's going to test you on whether you actually can derive this particular model for a particular device ok.

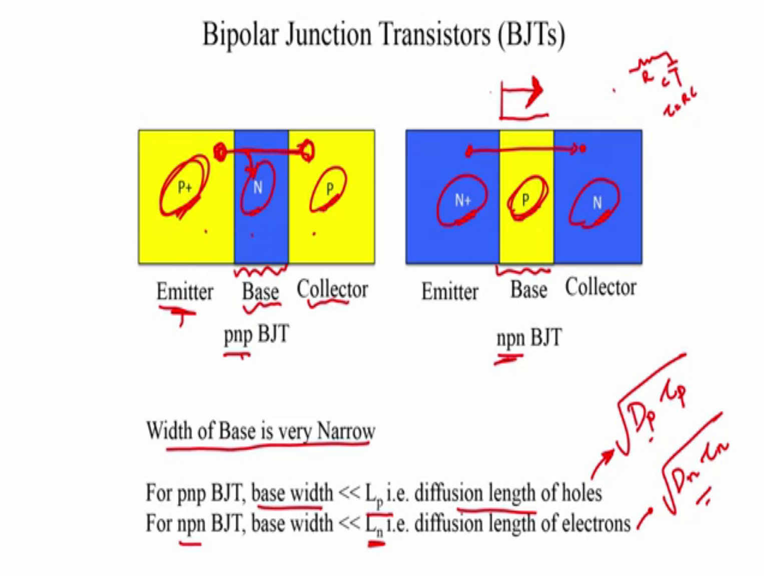
So, the point is to give you the toolkits and show you how it is done; so, that you will be able to build your own device in future and perform this kind of an analysis on that particular device ok. Now there are many kinds of transistors ok.

(Refer Slide Time: 12:17)

You will hear of many and I am probably I have listed a few of them here, but I am sure there are many others. Again this is something the design and development of transistors in the materials used for the transistors is very much the forefront of research in semiconducted devices. But these are some of the ideas behind transistor constructs that some of you might be familiar with, but in some sense are also very important.

So, you have bipolar junction transistors or BJTs. You have field effect transistors, and there are many kinds of field effect transistors, you have metal oxide semiconductor field effect transistors which are MOSFETs. And in MOSFETs you have something called as a depletion mode MOSFET or an enhancement mode MOSFET. You have something called as a MESFET which is a metal semiconductor field effect transistor. You have thin film transistors, you have tunnel FETS, you have junction field effect transistors etcetera. So, there are many kinds of transistors.

(Refer Slide Time: 13:21)



Let us start looking at the BJT ok. So, many of you might be familiar with this because the it is it is a it is a widely taught topic in your undergraduate studies, but nevertheless we will go through this entire exercise again ok. Particularly focusing on the device physics, and as I mentioned whether you like it or not a lot of the mathematics in the derivation. So, the BJT is got a construct like this. The geometry of it is simply it is it is there are 2 kinds ok, let us call it as a pnp BJT and an npn BJT. And the idea behind that is you have a construct where in you have a heavily doped P type semiconductor a moderately doped N type semiconductor and moderately doped P type semiconductor. So, you have a sequence of these 3 layers for a pnp BJT.

And for an npn BJT you have a heavily doped N plus region, you have a moderately doped P type and a moderately doped N type region. So, in some sense, it might look like these are 2 diodes right. It might look like these are 2 diodes. So, and look like this. But there is a significant difference ok. So, this region which is in the middle ok, where which is the N doped region here in the P doped region here has to be extremely small or extremely narrow. I will say not small, but extremely narrow for you would have a properly working BJT.

So, in some sense, just blindly using a circuit equivalent of 2 diodes is not probably the right representation of a BJT. So, these 3 regions are given very specific names. The heavily doped region is called as the emitter. The region in the middle is called as the base, and the region the remaining semiconductor bit is called as the collector. So, once again here for the npn you have the emitter, the base and the collector. And the key construct for the BJT is at the width of the base has to be very narrow. And what do you mean by very narrow? What we mean by very narrow is if it is a pnp BJT we expect. So, the idea of the BJT is to get the majority carriers from the emitter to migrate through the base and get into the collector ok.

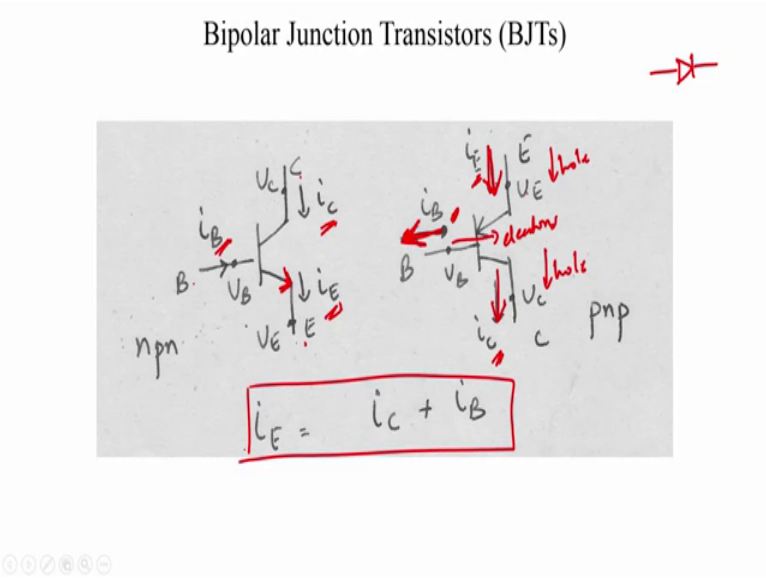
So, in this case we are talking about getting the holes from the emitter to go through the N doped base and get on to the collector. So, the key is for this to happen very successfully, there should not be too much of recombination in the base ok. And therefore, the base the width of the base has to be made narrow and it should be less than the diffusion length of the carriers we are talking about.

So, for a pnp BJT we want holes from the emitter to get to the collector and therefore, the width of the base should be less than the diffusion length of the holes. And if you remember if you have forgotten what diffusion length is it is the square root of the diffusion coefficient into the minority carrier lifetime, a square root of this product ok. We have we have looked at it when we studied P-N junction diodes.

Similarly, for an npn BJT we would like the electron from the emitter to successfully get through the base and approach the collector. And therefore, the width of the collector should be less than should be less than the diffusion length of the electrons ok and why less than the diffusion length? Because the diffusion length is a measure of the distance a minority carrier can migrate before it is recombined. So, if you remember we just like in an RC circuit, you have you have the time constant being represented by the product of RC which is which is a measure of how quickly the circuit responds. Just like that you have a diffusion length it is a measure of the distance traveled by the minority carrier before it recombines ok. So, that is the idea of a diffusion length.

So, with this understanding of the geometry and the construct let us just go ahead.

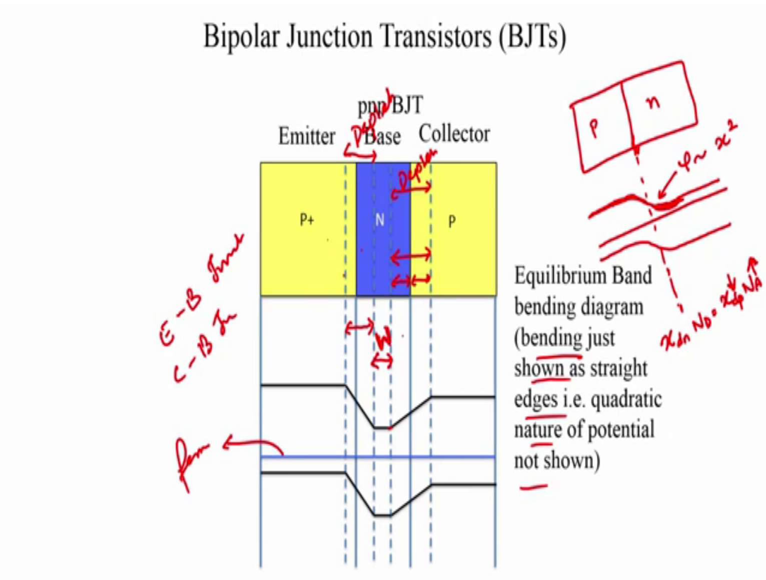
(Refer Slide Time: 18:06)



So, there are some circuit symbols ok. So, we had a circuit symbol for the diode as this. So, this is the circuit symbol for the BJT, you have for a npn BJT the key is you watch out for this arrow. The arrow is pointing towards the migration or the current direction and it is always pointed I mean it is always between the base and the emitter.

So, for the npn BJT you have the collector the base and the emitter and you have the collector current, the emitter current and a base current. And similarly for the pnp BJT you have the emitter current the collector current and the base current, and in either case you will have this relation holding true ok. We are probably left into this a little bit early, but it is just for me to introduce the symbol the circuit symbol for the BJT, all right.

(Refer Slide Time: 19:03)



But now let us let us head off an talk about something a bit more familiar; which is to draw the band bending diagram. So, here you have a P-N junction and you have another N-P junction. So, what is the equilibrium band bending diagram for this? Ok, so, for a P-N junction diode if you remember; so, when we looked at a P-N junction diode, at equilibrium the first thing that happens is the fermi level a lines ok, and then since the N region has given electrons to the P, they will find that the conduction band edge bend in the P bends towards the fermi level near the junction. And it bends away from the fermi level in the on the N side. And similarly the valence band follows ok. So, there was the band bending diagram for a P-N junction.

So, it is the same diagram that we need to draw for these 2 P-N junctions ok. So, here you have a P plus region and here you have a moderately doped N. So, these dashed lines indicate the depletion boundaries. So, this is the total depletion width so, that is the depletion region. So, similarly this is the depletion region for this junction. And if you watch carefully since this is heavily doped, the depletion width here is expected to be much lesser than the depletion width on the N side.

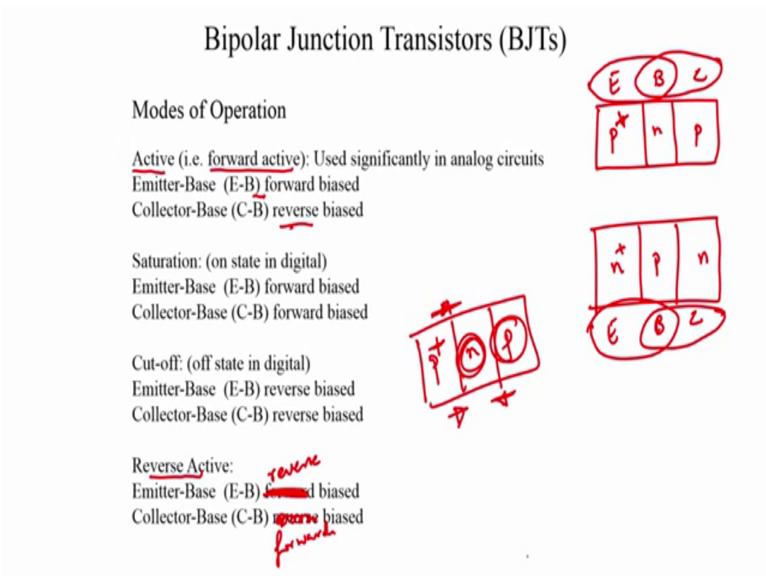
And why is that? Because for a P-N junction we had this condition, because of the continuity of the electric field at the junction if you recollect we had the condition that x d n into N D was equal to your x d p into N A; which meant that if your as the doping concentration increases your x d p has to come down. And therefore, your depletion width here is smaller as compared to the depletion width here because the doping concentration here is larger as compared to the doping concentration here.

So the, I have tried to represent that as accurately as possible, all right. So, here you have the depletion region on the emitter base junction. So, this is called as the emitter base junction. And then similarly you have the depletion region on the collector base junction. So, since both of them are moderately doped I have sort of represented both these depletion widths as almost equal, and that is the total depletion width of the collector base junction.

So, in this region you have your P-N junction your P plus N junction giving you a band bending, and then you have your N-P junction which gives you a band bending of this kind and therefore, this is your equilibrium state band bending diagram, with the fermi level being aligned and shown by this blue line here.

So, that is what happens in equilibrium. And since all of you are now familiar with solving Poisson's equation, we will not go through that exercise again, since it is just solving Poisson's equation for 2 P-N 2 different P-N junctions. With the with the only exception that though that is not really required for the electrostatics with the exception that this N region is very narrow, and N particular this region which is called as the width of the base is very, very narrow it has to be less than the diffusion length of the minority carriers, all right.

(Refer Slide Time: 22:59)



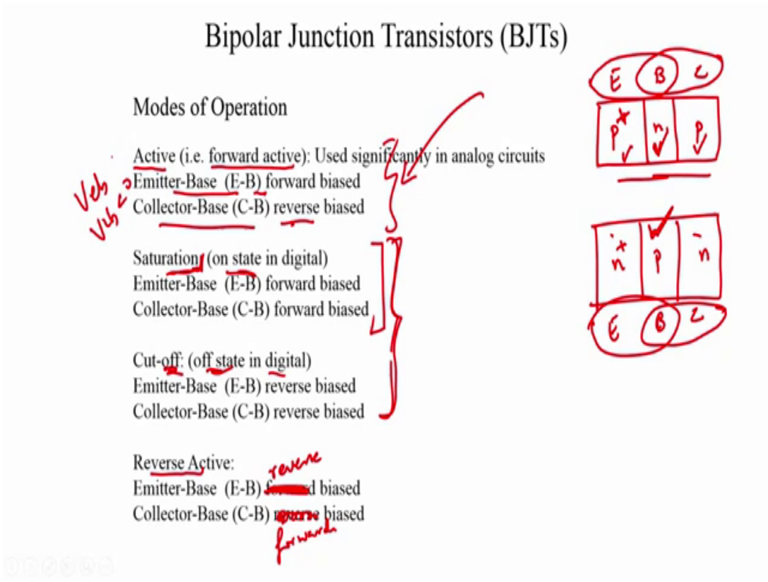
So, the BJT operates in several modes right. So, here let me just redraw the circuit, let me redraw this device structure. So, we considered a pnp, but all this is equally true for an npn ok. And what? We have several different modes of active of operation. So, the first is something called as the active mode ok, which is a more appropriately called the forward active mode of operation.

And this mode of operation is used in amplifier circuits, which we where use the BJT as an amplifier in a and in any analog circuit you will probably be using this mode of operation quite a bit ok. And in this mode of operation the emitter base junction is forward biased ok. So, you want to forward bias this diode, you want to forward bias this diode, and you want to reverse bias the collector base junction.

I mean we will see intuitively as to why this is needed ok, what happens in this mode ok. So, you want to forward bias the emitter base junction and reverse biased the collector base junction. And since we have talked about the forward active we can also talk about the reverse active ok, which is for the inverse active if you like. In which case you just flip these 2, which is the emitter base junction is sorry, it is a typo here the emitter base junction is a reverse biased, and the collector base junction is forward biased. Apologies for this typo, but please do correct it ok. So, the idea is since the m if not for this x is doping the emitter and collect this device structures completely symmetric. So, for example, let us say we did not excessively dope them it let us say it is pnp.

The device structure completely symmetric; so, whether you forward biases and reverse bias that or reverse biases and forward bias that it is the same ok. But the difference comes about because of this excess doping ok. And that is why in the forward active region, we are very careful to forward bias this diode which is the excess doped P and the moderately doped N. And reverse bias the moderately doped P-N the N. But in reverse active we do the opposite; which is P forward bias this diode and reverse bias this diode ok. So, that is the only difference between active and inverse active, all right.

(Refer Slide Time: 25:43)



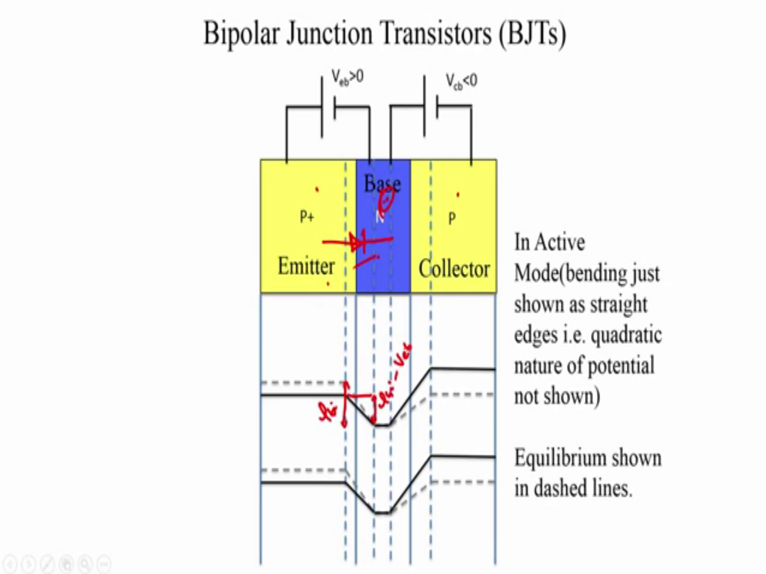
In saturation mode operation the forward bias both sides, both junctions. So, this diode is forward biased and that diode is forward biased; which means that this voltage is going to be higher than that, that voltage is going to be higher than this and in this case this is going to have a larger voltage is compared to the other 2 sides. And in the cutoff mode so, before I move on the saturation mode is the mode that is used as the on state in any digital circuit. So, if you want to completely close a switch, we are talking about the saturation mode of operation. And the cutoff mode is the off state in any digital circuit. So, if you are building digital circuits, you will probably be using the saturation and cutoff mode a lot, while maybe also using the active mode for certain special purposes.

So, the cutoff mode is where both junctions are reverse biased and do not permit any current ok. If you just reverse bias both junctions, you do not permit any current at all through the circuit, because these are the different modes of operation. And in this particular lecture, I will we will derive the general current voltage characteristics for the BJT from which you can actually then go in and specifically apply these conditions of making V eb large or V cb negative etcetera, etcetera, and derive your current voltage characteristics for the active saturation or cutoff or reverse active as an when you desire.

But I will pay special attention to the active mode of operation. Because this is something that we might use in the future towards the end of this course, all right. So, these are the different modes of operation of the BJT.

So, particularly let us look at the active mode ok. So, I am going to pay some attention to the active mode of operation.

(Refer Slide Time: 27:41)

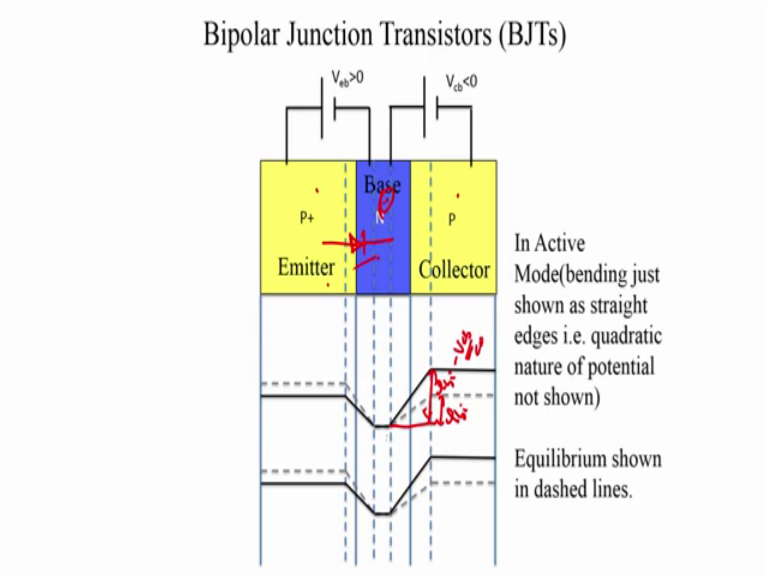


So, what is the active mode mean? The active mode means that the emitter base junction is forward biased and the collector base junction is reverse biased; which means that this P is at a higher potential is compared to N, and this P is at a lower potential compared to the N region. So, since this diode is now forward biased, the depletion width will reduce a little and the band bending will become a little bit more shallower ok.

So, which is sort of represented here. So, the dark line shows the active mode of operation, while the dashed line is the equilibrium condition. So, you can see that, you know, this built in potential is reduced as compared to that. So, that was phi b i and this become phi b i minus V eb.

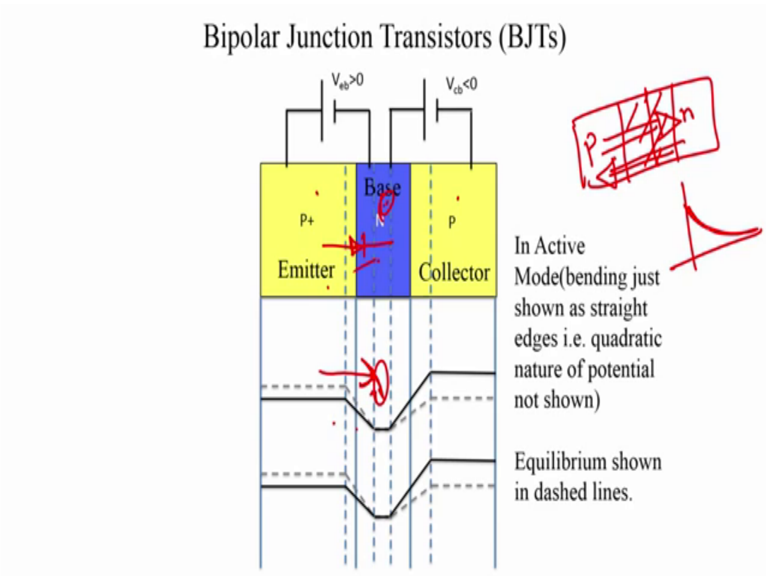
And on the other hand since this diode is now reverse biased this built in potential is increased.

(Refer Slide Time: 28:49)



The equilibrium let us see that was phi b i it is now increased to phi b i minus V cb where V cb is less than 0. So, that is that is the change in the band bending diagram. So, based on your understanding of P-N junction theory, what do you expect would happen here? Since this diode is now forward biased, you will have a huge diffusion of carriers from this P to this N.

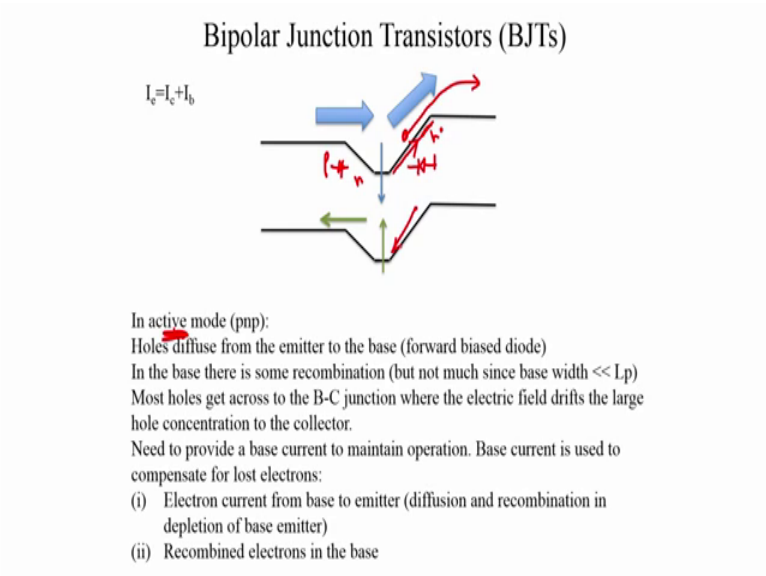
(Refer Slide Time: 29:16)



And these carriers would now start in your, in your P-N junction diode theory, we saw that there was a depletion region ok. And you had you had your P you had your N and when the junction was forward biased, you had a huge migration of the holes from the P to the N a migration of the electrons from the N to the P. And the holes became the minority carriers on the N side and you had a concentration gradient which went like this, which we saw from solving the, which we found to be exponentially dependent from on distance from the solution of the continuity equation. So, that was the P-N junction theory.

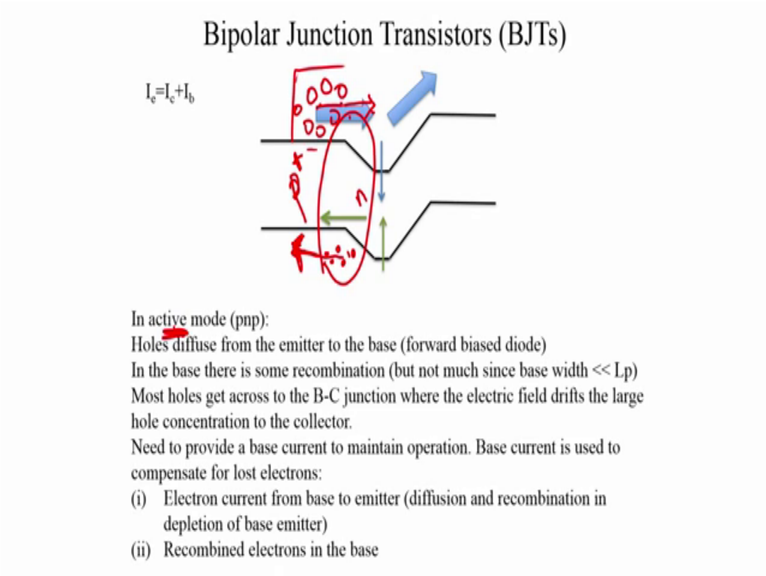
It is the same thing here, but the only difference is, on this N side all these carriers are not recombined, ok. Because the diffusion length is now the base width is now less than the diffusion length. So, therefore, you have a lot of these my carriers coming in, but now all of them are not recombined your base width is very, very small, you recombine just a few of them. And most of these carriers now get into this depletion region. And you will see a massive flux of these carriers so, the other side ok. So, let us to look at that in more detail.

(Refer Slide Time: 30:39)



So, in active mode we are talking about the active mode of operation, this P-N junction is forward biased. And this P-N junction is reverse biased ok which means that there is a huge electric field here ok. This electric field is going encourage all the holes if the hole gets in here is going to very rapidly move up and if there is an electron here it is going to very rapidly move down help.

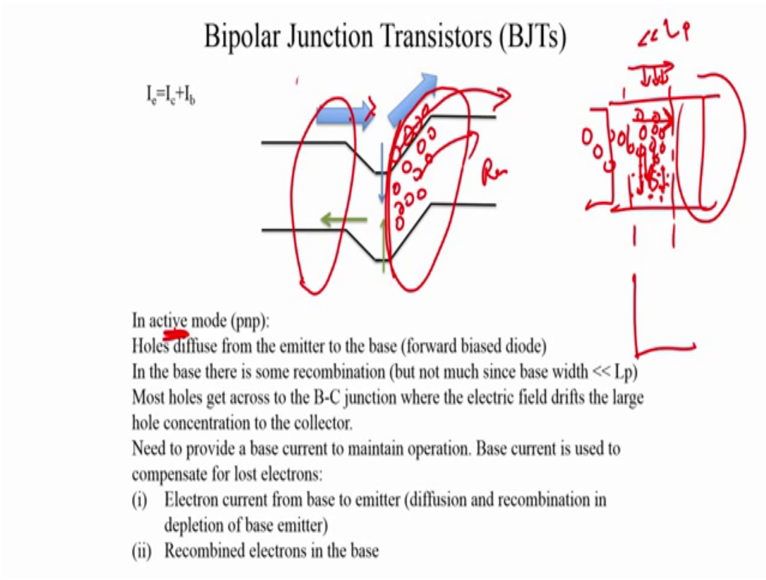
(Refer Slide Time: 31:22)



And because this junction is forward biased you have a large hole diffusion from the P plus to the N side. And you also have a significant electron diffusion from the N to the P, but the hole diffusion is much larger than this electron diffusion simply because this P junction this P side is more heavily doped than the N. So, you do have a large electron migration or the diffusion of electrons from the N to the P, and from the emitter to the base because of this forward bias, and just at the whole movement from the emitter to the base is going to be much larger it is going to dominate the current total current. Now all these electrons are going to enter the base and in the base. So, what is the base made of the base is N doped?

So, you now let us let us focus on the base.

(Refer Slide Time: 32:18)



So, here you have the base region, and let us say that those are the depletion boundaries. So, you have all these holes that are now diffused into the base, you have a large number of these holes now this base is made of a lot of it is got a excess number of electrons are the majority carriers in the base and the holes are the minority carriers. So, this excess whole population is going to start recombining with the electrons in the base. But and they are also going to diffuse from the higher concentration to the lower concentration end. So, they are diffusing, but they are also recombining.

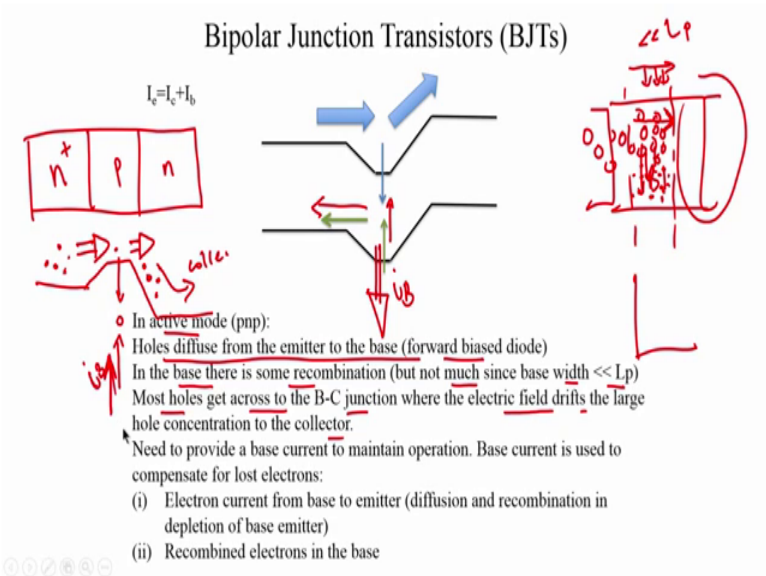
Now, if the base width was very large, then they would all recombine and very little will get on to the other side. But the base has been specially designed to have a length less than L p and therefore, all these holes will now make it to a across the base and get into this depletion region at the base collector junction. So, all the holes will successfully make it through perhaps not all, but most of them would successfully make it through, and get into this depletion region.

And once they are in this depletion region they see a very large field because this junction is now reverse biased. So, they see a very large field and all these holes will run into the collector. So, we have successfully transferred a large number of holes from the emitter through the base and to the collector. And that is the idea of active mode operation, and that is why you forward bias one junction and reverse bias the other.

So, just to summarize, you know, I have just written down these points hole, in active mode of operation holes diffuse from the emitter to the base because the forward base bias diode, in the base there is some recombination, but not much since the base width is going to be much less than L p. Most holes get across to the base collector junction where the electric field drifts the large hole concentration to the collector.

So, all the holes here are now drift to the collector and essentially the collector collects a large a whole population from the emitter. But to maintain all this operation in steady state, you know, if you want to keep this phenomena going continuously, you need to compensate for all the electrons lost in the base.

(Refer Slide Time: 34:46)



Whereas, the base lost electrons, a lot of electrons have migrated from the base to the P side because of the diffusion. And a lot of the electrons have recombined.

Now, if you do not replace this electron, this mechanism would stop, you know, it will reach a new equilibrium and everything would stop. So, in order to keep this going you need to provide a steady base current, you need to keep injecting electrons into the base. If you think of the current in the conventional sense, you have a base current that flows outward.

So, that is why in your so, we are discussing a pnp junction, and that is why if you look at the circuit diagram, you will say see that for a pnp the base current is drawn this way.

So, here you have the holes moving from the you have hole migration, you have holes moving here, and you have electrons moving in to compensate for the electrons that are recombined that have diffused to the emitter and therefore, the base current moves here.

So, you have a emitter current, you have a collector current, and you have a base current, with this relation being followed. So, that is the essence or the qualitative discussion of the working of the BJT. So, we have looked at the pnp BJT, the npn BJT is the same argument. So, he want to draw this properly so, let us take the npn junction. So, you have the N plus P N n the band bending is now going to look like that, ok. So, that is the heavily reverse biased P-N junction and this is the forward bias P-N junction. And if you are going to have a large number of electrons that are going to head to the base, some of these electrons will recombine, and a large number of these electrons will now start drifting away into the collector and will be collected by the collector.

So, in order to recompensate for these holes in the base to keep things moving for an npn junction, you need to supply a you need to add more holes; which is being which is in which implies that you need to add a base current that is directed this way ok; so, that is the npn operation. So, we discussed both the pnp as well as the npn operation, in active mode.

So, we are only looking at active mode at equilibrium we know what the band bending diagram is I am being special attention to active mode operation although we will derive the general current voltage characteristics. So, it is this mode of operation that is used in analog circuits. And the idea is that a small fluctuation in your base voltage will result in a small fluctuation in the base current; which would in turn result in a large amplification of this current because it the collector current is going to it is going to be quite large. So, that is the idea that is the idea of the amplification. And we will get to this when we talk about circuits, we will come to this later on.