High Speed Devices and Circuits

Prof. K. N. Bhat

Department of Electrical Engineering

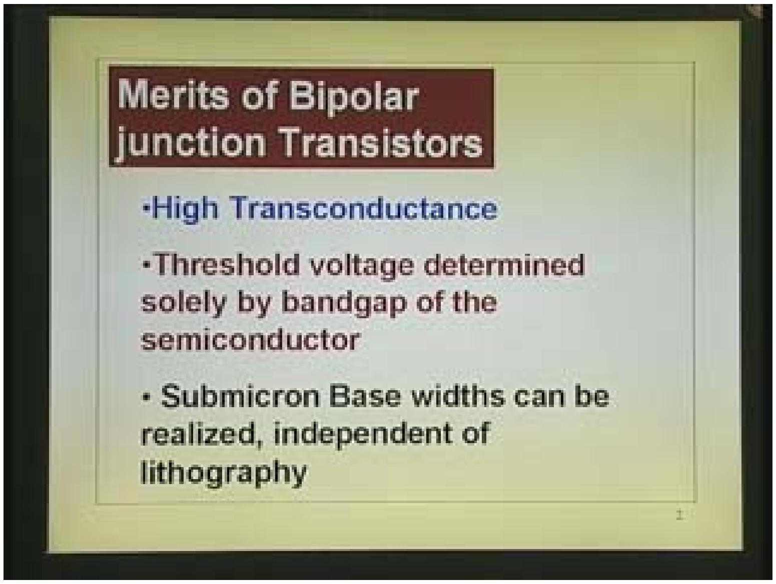
Indian Institute of Technology, Madras

Lecture - 38

Heterojunction Bipolar Transistors (HBT)

We have been discussing the merits of bipolar junction transistor and we were in the process of taking a look at the heterojunction bipolar transistor.

(Refer Slide Time: 01:24)



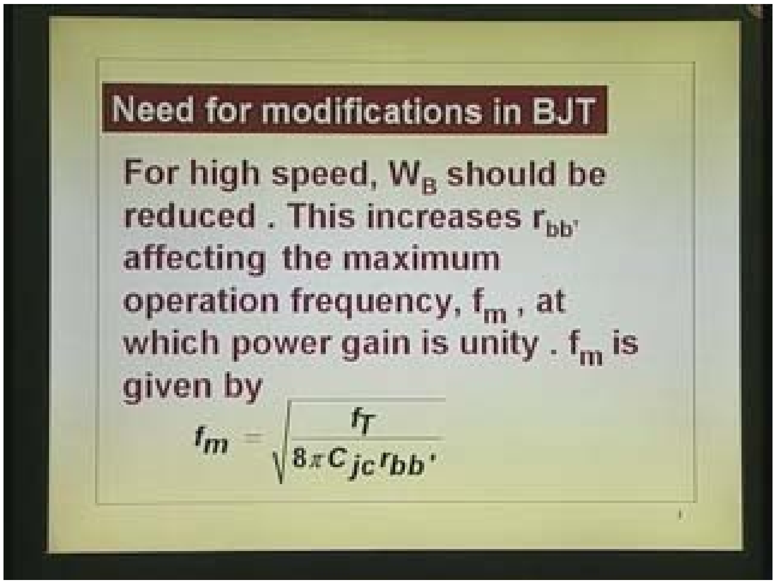
Merits of Bipolar junction Transistors

- High Transconductance
- Threshold voltage determined solely by bandgap of the semiconductor
- Submicron Base widths can be realized, independent of lithography

Just to see the summary of what we discussed last time; the bipolar transistor has several merits: one is it has got high transconductance which is actually exponentially dependent on $V_{BE.}$ So delta I_C by delta V_{BE} is I_C by V_T. For a given area, we get better transconductance compared to MOSFET and other things. Now, the threshold voltage is determined completely or solely by band gap of the semiconductor. For example, if we take silicon, the cut in voltage is about 0.65 to 0.75 Volts; people take it about 0.65.

If we take gallium arsenide, close to 1 Volt cut in voltage. We can call it as threshold voltage, because definition is actually the voltage at which current will begin to flow, at least 10 percent of the operating current that we want to do. Then, we also discussed that submicron base widths can be realized independent of lithography. We have that advantage, because we do not need to use lithography to control the base width. It is decided by the base diffusion and the emitter diffusion, if we are talking of diffusion or it is decided by the epitaxial layer thickness or by implantation doses. So it is that way controllable. I do not need to have electron beam lithography to get 0.1 micron technology base width. That is about the BJT.

(Refer Slide Time: 03:23)



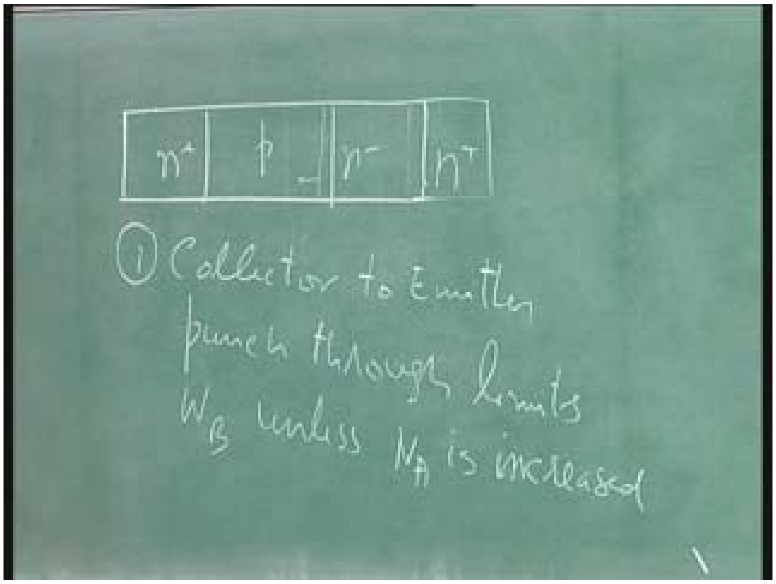
Need for modifications in BJT

For high speed, W_B should be reduced . This increases $r_{bb'}$ affecting the maximum operation frequency, f_m , at which power gain is unity . f_m is given by

$$f_m = \sqrt{\frac{f_T}{8\pi C_{jc} r_{bb'}}}$$

With all these advantages, what modification do we need to get to that performance? One of the things that we need to do with BJT is to increase the cut off frequency f_T we have to keep on reducing the base width; W_B the base width should be reduced. After all, the transit time decides the cut off frequency - base transit time. Transit time is W_B square by 2dn for n p n transistor, centimeter squared divided by centimeter squared per second. So smaller the W_B, smaller is the transit time, better is the cut off frequency. That is what is we are saying here. When we reduce the base width, there are a couple of things happening. One of them the doping is not too high, moderate doping, one of the things that happens I can just draw the diagram here.

(Refer Slide Time: 04:50)

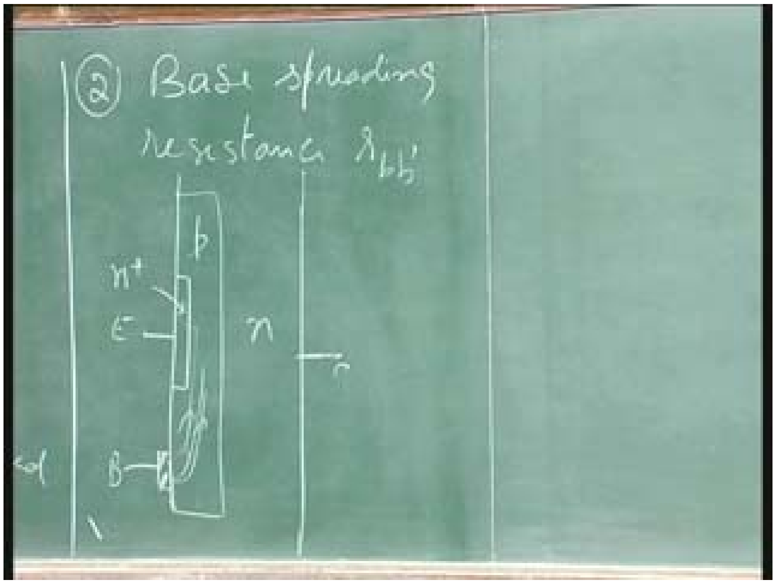


① Collector to Emitter punch through limits W_B unless N_A is increased

Keep on reducing base width. We become more and more ambitious say 0.1, less than 0.1, like that if we go, what happens is I have not written in the slide there, but we can see n plus, p, n minus and n plus. You will have problem. This is the one-dimensional picture, but even in the one-dimensional picture, one of the problems that we will have will be the depletion layer moves more into this region, but it also moves into this region. How much it moves into this region depends up on doping. If the doping is even 10 to the power of 16, then also this width can be about 0.3 microns, very comfortably, very easily.

If this width is 0.2 microns, there is a chance that the depletion layer will punch through; the collector base depletion layer will punch through and reach the emitter base junction, when we talk of base widths of 0.1 microns of that order. The only way that we can beat that is increase the doping here (Refer Slide Time: 05:53). Make this doping higher. So collector to emitter punch through voltage calls for higher doping in the base region and the base width is reduced. That is one thing which I did not note in the slide, but that is very important thing which one can see. So punch through emitter; employing the depletion layer punch through. Depletion layer reaching the emitter base junction. Limits W_B; the lower limit on W_B unless N_A is increased. In other words, what we are telling is, we must increase N_A, if we are going lower W_B to avoid this problem of depletion layer punch-through. We may say, I will have non-uniformly doped substrate, but how much is the punch through voltage depends up on the total doping concentration here. See after all, if we reduce the base width, unless we increase the doping concentration, then the integrated doping concentration will not be sufficient, because after all, when we have plus charges on depletion layer, it calls for minus charges. The minus charges are limited by how many dopants are present here. So, reduce the base width, number of dopants present are less unless you increase the doping concentration. Even 10 to the power 17, 10 to the power 18, that level, when we go to that level of 0.1 micron, then we will have lot of problems associated with that. This is following the usual scaling loss; whenever we scale down the dimensional doping must be increased, MOSFET, MESFET everywhere. What other parameters are there?

(Refer Slide Time: 08:25)

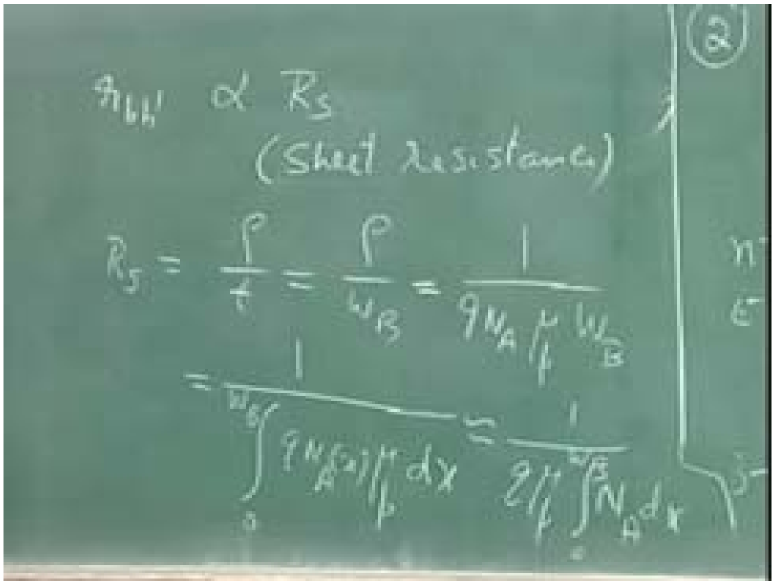


The other parameter that gets affected is the second one - the base spreading resistance. Usually, it is given by the symbol $r_{bb\ dash}$, just to note that b dash is some intrinsic point. So it is r_b base resistance, which is coming there which is external to the device which is the result of… I have a n plus layer here; then I have this p region, p layer, coming up here. I have contact here for the base and I am showing general structure that is bipolar transistor. Instead of showing vertically, I am showing it… put it like that.

Usually, we will have top to the bottom and this is the nth collector. It goes up substrate. Now here the current flows laterally from the base, there is current load which actually supplies the recombinational current and also which supplies whole current to the emitter. The base current is in lateral. That is the one which is seeing this lateral resistance, resistance looking down is seen there. So the resistance of that region depends up on, what is the sheet resistance of this layer. It is proportional to the sheet resistance of that layer and the sheet resistance of that layer would be dependant up on how much is the thickness there or to put it in simple terms, the resistance in this direction is more if the width is smaller r_{bb}. So, what I am trying to tell is r_{bbdash} dash increases. Let me just put that r_{bbdash} dash from this portion and also through, it is a spreading resistance.

As the current flows here, something spreads out to the base contact. That is why we call it as spreading resistance. It is not the bulk resistance of this portion. This entire region is contributing to the lateral flow, because after all the current flow to the base is not just one point. It is all going like that. So, we call it as spreading resistance. So that becomes more, if the area is smaller. This spreading resistance is proportional to…. Can I remove this now? So this is noted down.

(Refer Slide Time: 11:24)



$r_{bb'} \propto R_S$

(Sheet resistance)

$$R_S = \frac{\rho}{t} = \frac{\rho}{W_B} = \frac{1}{q N_A \mu_p W_B}$$

$$= \frac{1}{\int_0^{W_B} q N_A(x) \mu_p \, dx} = \frac{1}{q \mu_p \int_0^{W_B} N_A \, dx}$$

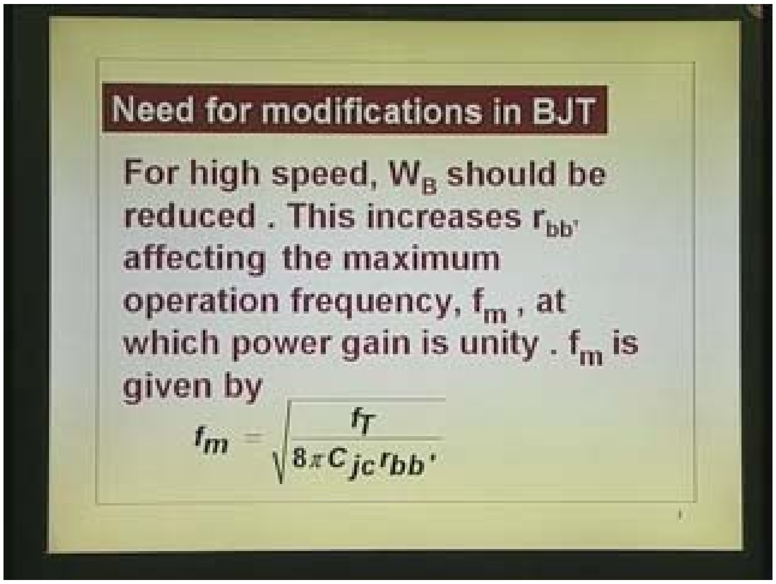
The resistance actually r_{bb}dash will be proportional to R_s where R_s is the sheet resistance. Just put it down here. Sheet resistance, R_s, is actually equal to rho divided by thickness. In this case, this is row divided by W_B - thickness. Sheet resistance looking from here is actually when we put a layer, sheet resistance is, what is the resistivity divided by thickness and thickness is W_{B} - base width in this case.

So what is rho? rho is 1 by q N_A into mu p that is sigma into W_B. Now we can see, if I reduce the base width R_s goes up linearly; if R_s goes up, r_{bbdash} goes up. So why I put this particular expression is, what matters is the product of N_A and $W_{B.}$ We can also put it in general as integral q N_A of x, it may be non-uniformly doped, into mu_p where a p type layer into d of x integrated 0 to W_B. This is W_B; that is the base width which I have put there and over the entire base width, if we take N_A of x into d of x - product - integrate it, that is actually equal to N_A into $W_{B.}$

So what I am telling is: the sheet resistance depends upon the total integrated doping. I can pull out the mobility and q term and what is left inside will be, is approximately, I can put it as 1 divided by 2 times mu_p. Actually mu_p will also be depended on mobility doping, when we go to high doping concentration. I am pulling out just to show you this will be equal to N_A times d of x, 0 to W_B. This is nothing but total integrated doping. Ultimately, what matters is how much is the doping concentration in the base, total number per centimeter squared.

Either in terms of resistance, we are talking of sheet resistance or in terms of voltages also, it is not enough, if we just increase the doping; the total doping we must increase. What happens if we… now both the things point out to the same thing. That is whether we talk of punch through voltage, limitation to avoid or reduce or do not allow this resistance to increase. Both of them point out to that we require to increase the doping in the base region. If we reduce the base width, we better increase the doping, if it is uniformly doped. If it is non-uniformly doped, of course integrate the total doping.

(Refer Slide Time: 15:28)



Need for modifications in BJT

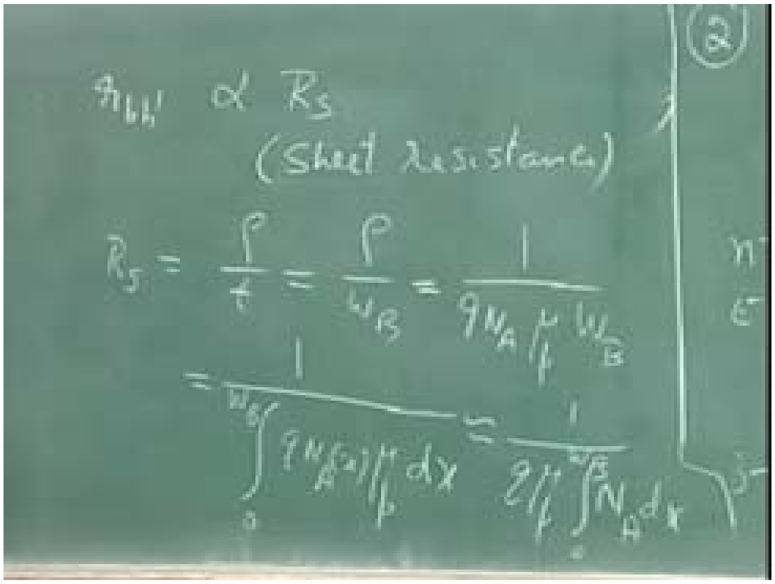
For high speed, W_B should be reduced . This increases $r_{bb'}$ affecting the maximum operation frequency, f_m , at which power gain is unity . f_m is given by

$$f_m = \sqrt{\frac{f_T}{8\pi C_{jc} r_{bb'}}}$$

There is one some equation, I have put it here just which actually is well known equation which comes up from the equivalent circuit. That is this r_{bbdash} dash, what parameter does it affect that I have put it here. The maximum frequency operation at which, power gains is unity. In fact, if we make an oscillator, maximum power gain that we can get is unity, when we go to that frequency.

In fact, from the equivalent circuit, this is a very serious equation we put. So, once the equivalent circuit is shown, that it is related to cut off frequency. We can improve the cut off frequency by reducing the base width; so that we can improve, but maximum operation frequency where power gain is unity is limited by f_T and also C_{jc} - collector junction capacitance, we can reduce it and if I have n minus layer on the collector, automatically depletion layer width becomes longer and the C_{jc} becomes smaller, r_{bbdash}. So, RC combination between the collector base capacitance and the base spreading resistance together bring down the maximum operating frequency. So this is the general formula; I just borrowed it straight away; so there is no point in spending time in deriving from equivalent circuit. So, it hits the frequency of operation finally. So we need to keep this r_{bbdash} dash low; we cannot allow it to increase.

(Refer Slide Time: 16:58)



$r_{bb'} \propto R_S$

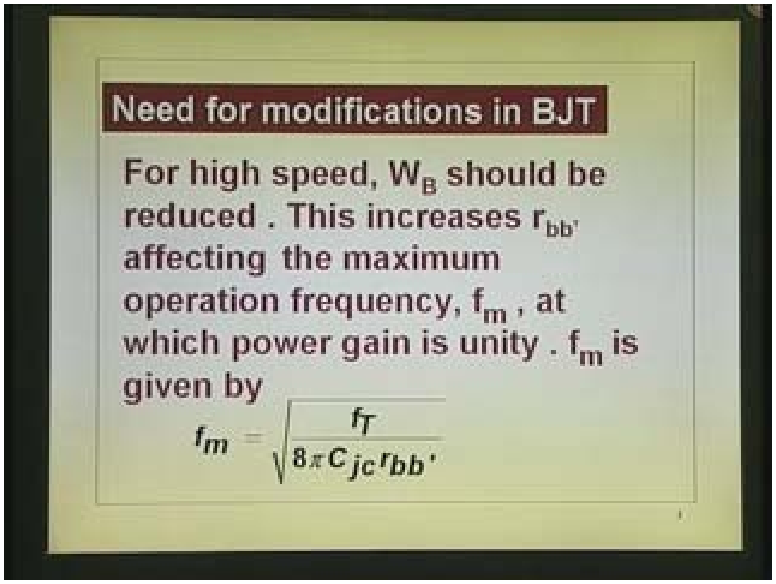
(Sheet resistance)

$$R_S = \frac{\rho}{t} = \frac{\rho}{W_B} = \frac{1}{q N_A \mu_p W_B}$$

$$= \frac{1}{\int_0^{W_B} q N_A(x) \mu_p \, dx} \approx \frac{1}{q \mu_p \int_0^{W_B} N_A \, dx}$$

What we have gained by reducing the base width in terms of transit time, we would have lost in terms of maximum operation frequency, because of increase in r_{bbdash} dash.

(Refer Slide Time: 17:10)



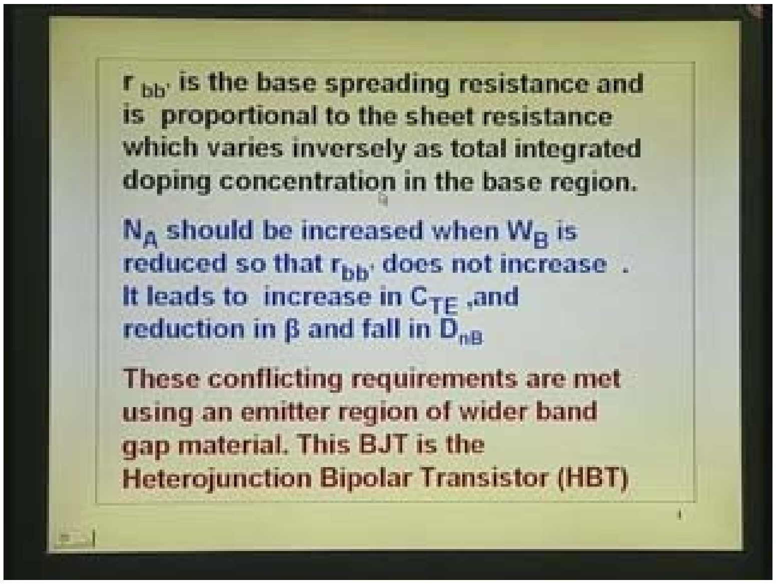
Need for modifications in BJT

For high speed, W_B should be reduced . This increases $r_{bb'}$ affecting the maximum operation frequency, f_m , at which power gain is unity . f_m is given by

$$f_m = \sqrt{\frac{f_T}{8\pi C_{jc} r_{bb'}}}$$

So, that is why, from the point of view of punch through voltage and also from the point of view of base spreading resistance, maximum frequency operation, we need to increase the integrated doping in the base.

(Refer Slide Time: 17:27)



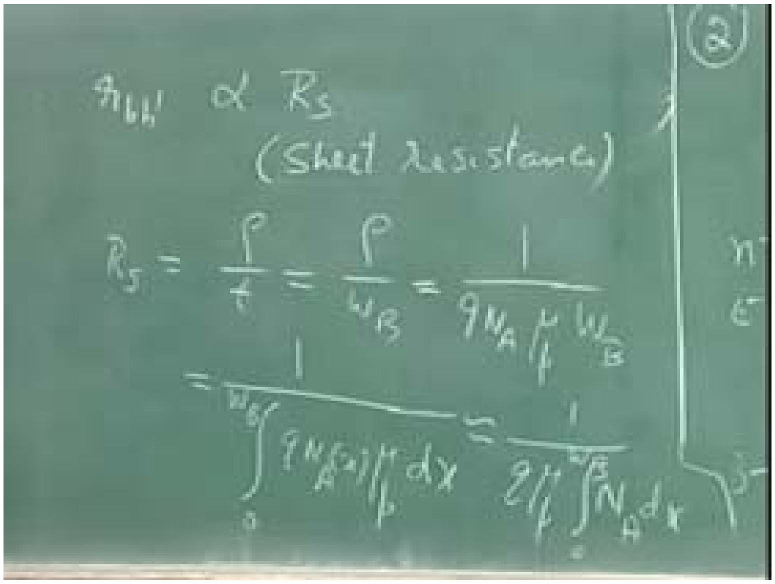
$r_{bb'}$ is the base spreading resistance and is proportional to the sheet resistance which varies inversely as total integrated doping concentration in the base region.

N_A should be increased when W_B is reduced so that $r_{bb'}$ does not increase . It leads to increase in C_{TE} ,and reduction in β and fall in D_{nB}

These conflicting requirements are met using an emitter region of wider band gap material. This BJT is the Heterojunction Bipolar Transistor (HBT)

So this is summing up what we have said now. r_{bbdash} dash is the base spreading resistance and is proportional to the sheet resistance which varies inversely as total integrated doping concentration in the base region.

(Refer Slide Time: 17:43)



$r_{bb'} \propto R_S$

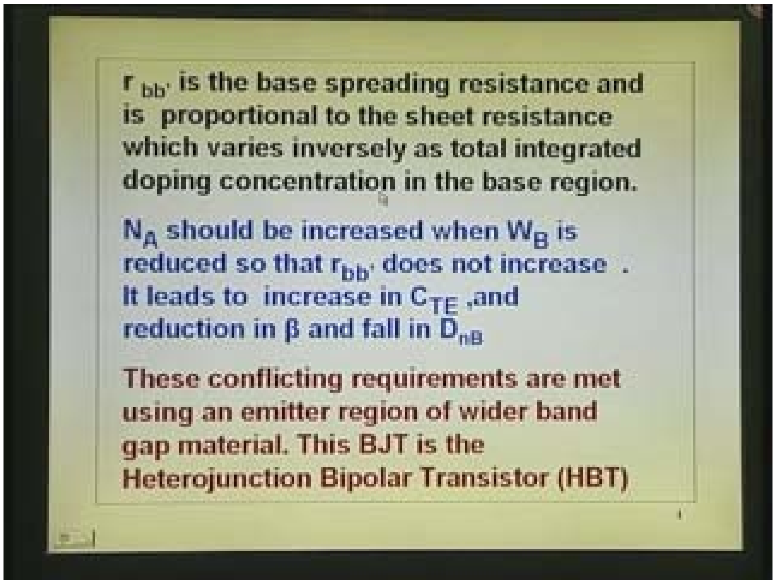
(Sheet resistance)

$$R_S = \frac{\rho}{t} = \frac{\rho}{W_B} = \frac{1}{q N_A \mu_p W_B}$$

$$= \frac{1}{\int_0^{W_B} q N_A(x) \mu_p \, dx} = \frac{1}{q \mu_p \int_0^{W_B} N_A \, dx}$$

R_s varies as inversely as the total integrated doping concentration in the base region. That is what we have said in the statement. We must actually increase this quantity to reduce it and finally reduce r_{bbdash}.

(Refer Slide Time: 17:59)



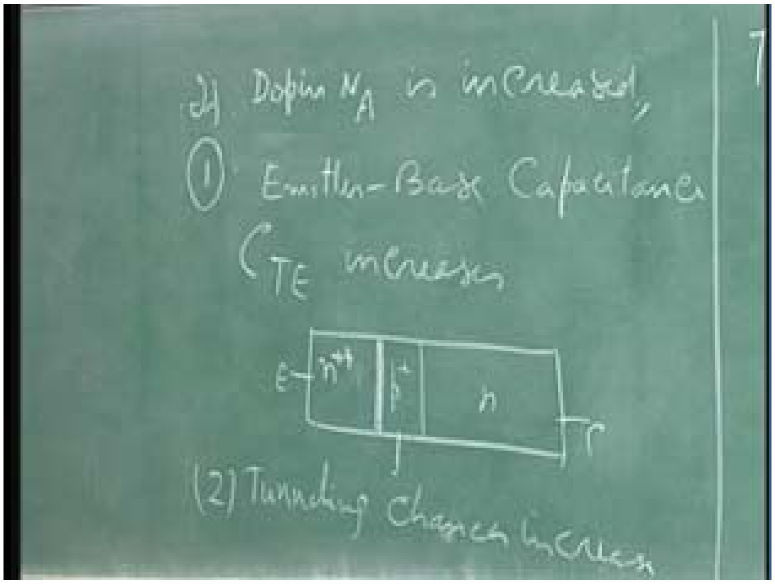
$r_{bb'}$ is the base spreading resistance and is proportional to the sheet resistance which varies inversely as total integrated doping concentration in the base region.

N_A should be increased when W_B is reduced so that $r_{bb'}$ does not increase . It leads to increase in C_{TE} ,and reduction in β and fall in D_{nB}

These conflicting requirements are met using an emitter region of wider band gap material. This BJT is the Heterojunction Bipolar Transistor (HBT)

N_A or rather total integrated doping N_A d of x should be increased when W_B is reduced, so that r_{bbdash} does not increase. Keep it within the limit. What does it lead to? Do this doping increase in the base region, what are the parameters which are affected?

(Refer Slide Time: 18:29)



(i) Doping N_A is increased,

① Emitter-Base Capacitance C_{TE} increases

(2) Tunneling Chances increase

Let us just take a look at that. So when we increase the doping in the base region, several things can happen. One of them is doping N_A is increased. One is emitter base junction capacitance, transistor capacitance. C_{TE}; I have not expanded that there. So C_{TE} increases, why? Just put it here. N plus already, we have 10 to the power 20 and I make this base width very small and increase this p. Let me put it this way, so as to differentiate n plus plus 10 to the power 20.

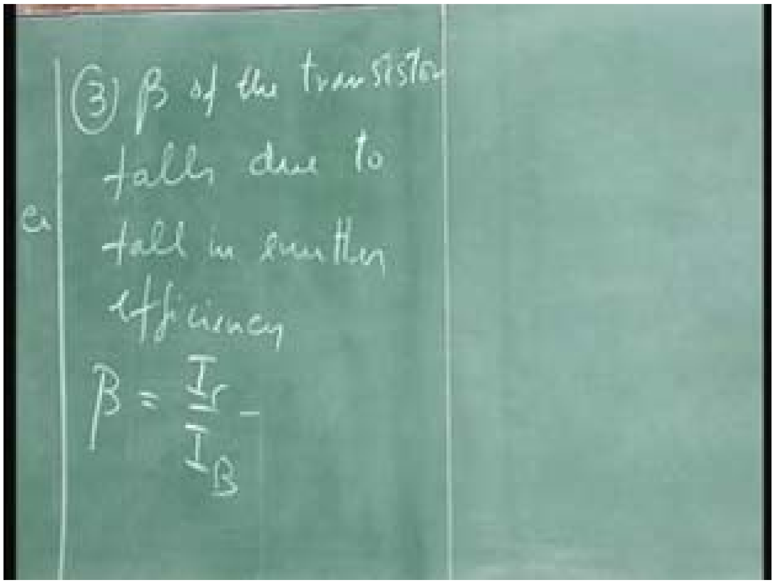
This I will put as p plus 10 to the power 18. I take it up. In the anxiety to reduce the base width and also to reduce the r_{bbdash} and punch through, I will make it 10 to the power 17, 10 to the power 18, all through. Then what happens? The depletion layer width here is actually very small. So result of increasing doping in the base reduces the depletion layer width which actually increases the capacitance. That is this. (Refer Slide Time: 20:14) The moment we increase the capacitance, it is going to affect you through a charging time required for charging this capacitor. So we would have reduced the base transit time, but we have increased the time required to charge this capacitor. So we can talk of that as a transit time through this charging time required. So the cut off frequency will come down, because this will go up.

Base width, we have reduced to reduce the transit time in the base, but we have to follow it up with the increase in the base doping which actually increases capacitance - C_{TE} - which will introduce delay in charging this voltage. Unless we charge this voltage, charge this capacitor, there is no transport taking place. So we have delays introduced in charging the emitter based capacitor. This is one effect; what other effects will be there? This is one of major effects.

Other effect is we go too much into this doping increase here, both of them, the width becomes small. That is equal to 10 to the power 19. We know the upper limit comes up because of that. We go to 10 to the power 19, the C_{TE} becomes worst - one. Even if we do not care about C_{TE}, there is one more parameter which comes up. The depletion layer becomes so narrow that there will be tunneling current. Once we have tunneling current, junction behaves entirely differently the way we want. It can behave like, if we take a tunnel junction, it behaves like a ohmic contact initially. In fact, it will almost behave like ohmic contact there. We will have negative resistance characteristics etc., but that is ultimate. So these are two effects plus what of the transistor? beta of the transistor what happens?

So second one is tunneling chances increase. When I say tunneling chances, I am saying tunneling the electrons here that increases. That is when we go to doping to 10 to the power 18 and above, because we have to go to depletion layer widths which are something like 30, 40 angstroms; otherwise it does not.

(Refer Slide Time: 22:58)



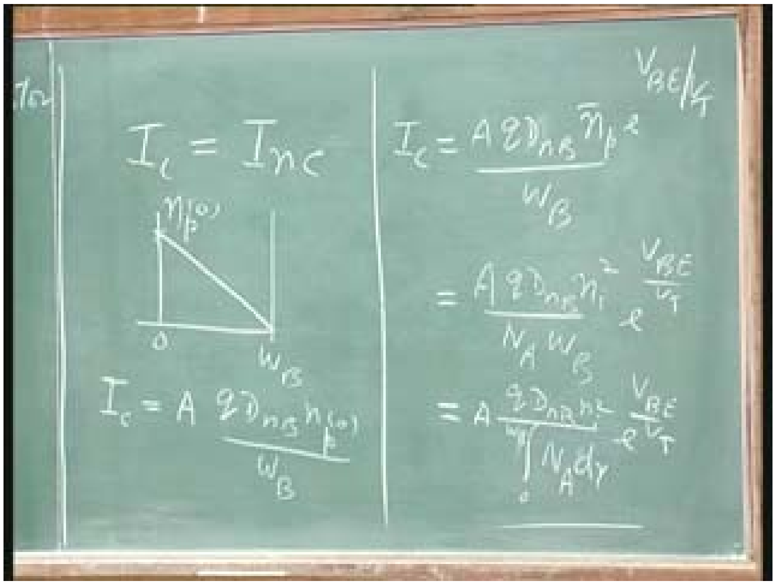
③ β of the transistor falls due to fall in emitter efficiency

$$\beta = \frac{I_C}{I_B}$$

The transistor beta, the third one, is the beta of the transistor. It falls due to fall in emitter efficiency. If we do not recall, I will just write down some of the equations related to that and the base width is so small - 0.1, 0.2 microns - you may not worry about the recombination in the base; you may not worry about transport from the emitter to the collector. Transit times are very small compared to the lifetime. So we will not have… we will hardly have any recombination. So the current gain of all the transistors today, modern transistors are limited by emitter efficiency rather than the base transport factor.

In the olden days junction transistors, base widths were 25 microns. The beta was 40, 50, controlled totally by the base transport factor. We do not have to worry about that thing. So beta will be actually equal to current collected here divided by whole current injected there. So what is beta now? So the beta is I_C divided by I_B. I am neglecting, of course, the reverse saturation current there. I think I will just have to go one step further down here.

(Refer Slide Time: 24:35)



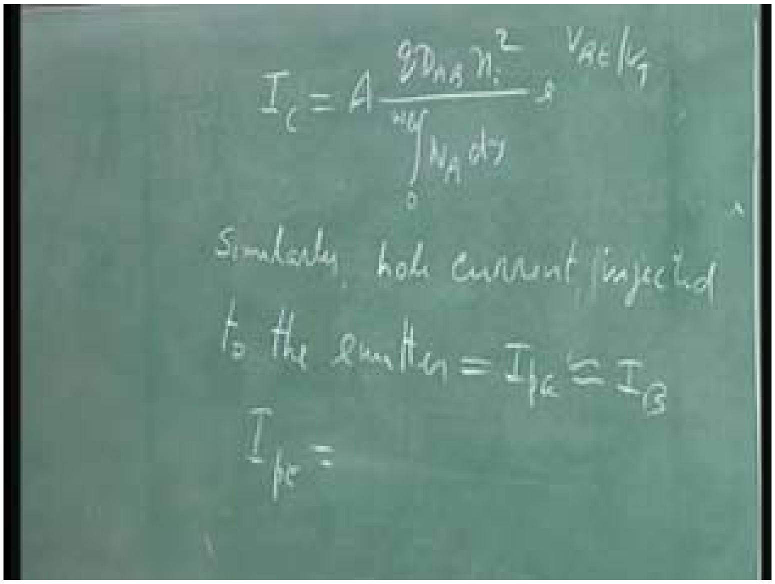
I_C is equal to I_{nc}, electron component of current reached in the collector, standard symbol. We take the base region, 0 to W_B, uniformly dope this region. Let us consider that for simplicity is actually n_p 0, electron concentration at x is equal 0 and this is 0. So I_C, therefore is equal to area q D_{nB} into n_p 0 divided by W_B. n_p 0 is actually n_p thermal equilibrium e to the power of V_{BE} by V_T. So we have got I_C, therefore is equal to area into q D_{nB} into n p bar e to the power of V_{BE} by

V_T divided by W_B. That is the value. This actually is equal to… this is the standard equation that I am rewriting it, emit area q D_{nB} into n_i squared divided by N_A in the base region. n p bar is n_i squared by N_A, thermal equilibrium value into e to the power of V_{BE} by V_T and this I can write also as area A into q D_{nB} n_i squared divided by integral N_A into d of x. This quantity is the total integrated doping.

In fact, if we write for general case also, we will derive that equation, total integrated doping. Area into q D_{nB} n_i squared divided by N_A into d of x 0 to W_B e to the power of V_{BE} by V_T. So this is just to remind you; you all know it. I have just reproduced whatever we have been discussing earlier, in many of the lectures, we must have heard about these things. This is the total integrated doping. I can call it by N_A T B, N_A total in the base, the denominator.

Now, please note, at this point, the entire thing depends on n_i squared. If I have a device which has got the same integrated doping concentration and if n_i is smaller, that current is smaller. If I take germanium device to use the same equations, this n_i will have much more current. We will not jump into germanium device, because the band gap is too low, it cannot go to higher temperatures. That is I_C. What is I_B? So let me just write down now. Let me take up this whole portion, come back to this.

(Refer Slide Time: 28:50)



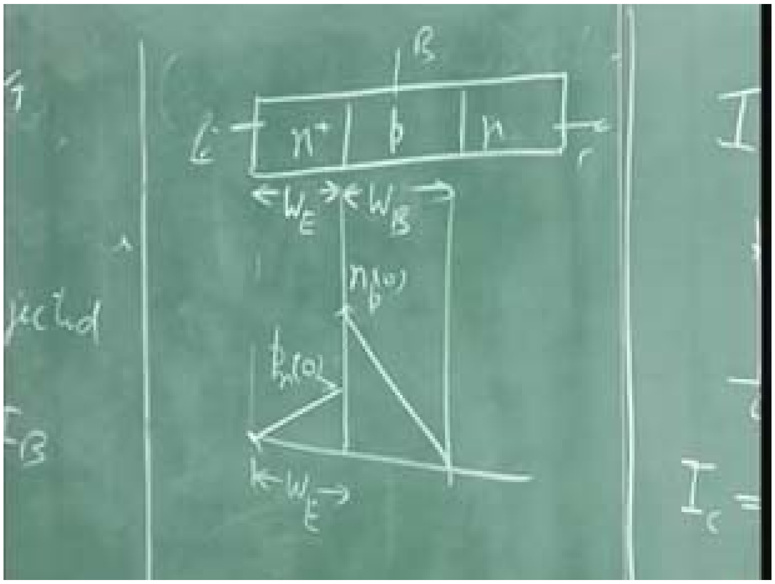
$$I_C = A\frac{qD_{nB}n_i^2}{\int_0^{W_B} N_A\,dx}e^{V_{BE}/V_T}$$

Similarly, hole current injected to the emitter $= I_{pE} \cong I_B$

$I_{pE} =$

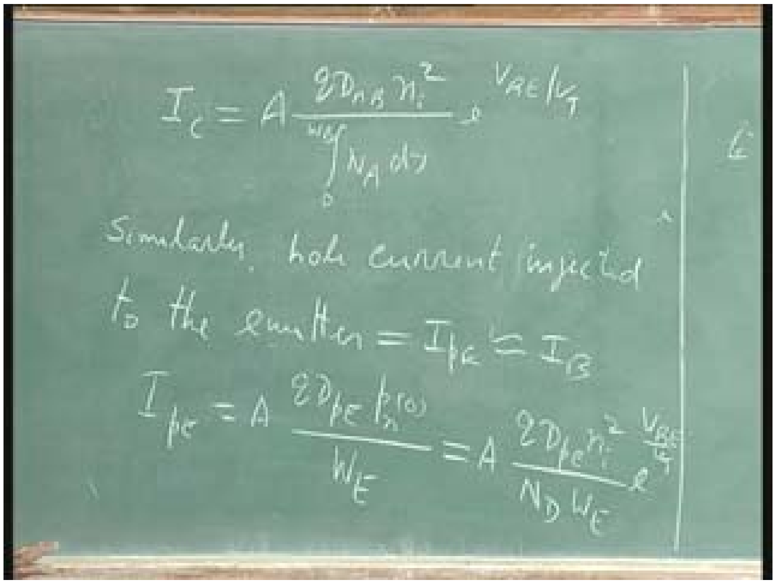
I_C is area q D_{nB} n_i squared divided by integral N_A d of x 0 to W_B. Similarly, this is I_C, whole current injected from the emitter to the collector. Similarly, I_{pE} whole current injected to the emitter, is I_{pE} approximately equal to I_B, because recombination current is the base is negligible. Otherwise, we have to add that also. Base current will be this whole current injected to the emitter plus recombination current in the base. I am neglecting that because base width is very very small, nothing much recombined. I_p will be how much?

(Refer Slide Time: 30:25)



Let me put this across to you by drawing a diagram here. N plus, p, n minus, collector, base, emitter. That is the current distribution here, n_p 0 and everything was related to n_p 0 divided by W_B and here, I will take this width W_E small and this is W_B. So this current component we have written now. All that we did was n_p 0 divided by W_B into q D n into area and n_p 0 is related to e to the power of V_{BE} by V_T. Here if this width W_E is very small compared to depletion length, how will that be? This is linear because base width is small compared to the depletion length and this width is small, today's transistor that width is very small. So that will be like this. So this will be actually p_n 0, this quantity (Refer Slide Time: 31:56).

(Refer Slide Time: 31:59)



$$I_C = A\frac{qD_{nB}n_i^2}{\int_0^{W_B} N_A\,dx}e^{V_{BE}/V_T}$$

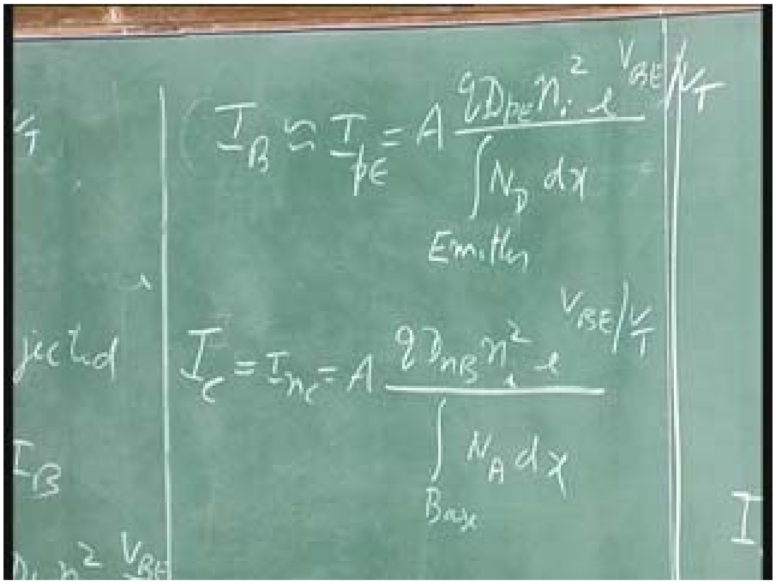
Similarly, hole current injected to the emitter = $I_{pE} \simeq I_B$

$$I_{pE} = A\frac{qD_{pE}p_n(0)}{W_E} = A\frac{qD_{pE}n_i^2}{N_D W_E}e^{\frac{V_{BE}}{V_T}}$$

So this will be actually equal to area q D_{pE}, we are getting this. We are getting closer to HBT; now, p_n, 0. I am just putting 0 to indicate that it is the edge of the depletion layer divided by W_E, because that is the diffusion current, uniformly doped p_n 0 divided by W_E that gives me D_p by d of x; D_p is this one in the emitter; D_p by d of x is the slope, p_n 0 by W_E. So this I can write it as p n 0 e to the power of V_{BE} by V_T; p n bar is actually n_i squared divided by N_D. So this will be area, I will skip steps and put it as D_{pe} n_i squared divided by N_D into W_E.

We have got the I_{pE} which is actually the base current. Area q D_{pE} n_i squared by N_D, n_i squared by N_D is actually p_n 0, the thermal equilibrium. This is p n bar e to the power of V_{BE} by V_T. V_{BE} by V_T, I put here. p n bar is nothing but n_i squared by N_D [this is doping] ; n_i squared by N doping, p n product is n_i squared. Thermal equilibrium p n bar is n_i squared by NN bar. NN bar is nothing but N_D. So what is this quantity? This is similar to that. This can be written as…so since this is already there, I will remove from here go back into this.

(Refer Slide Time: 34:13)



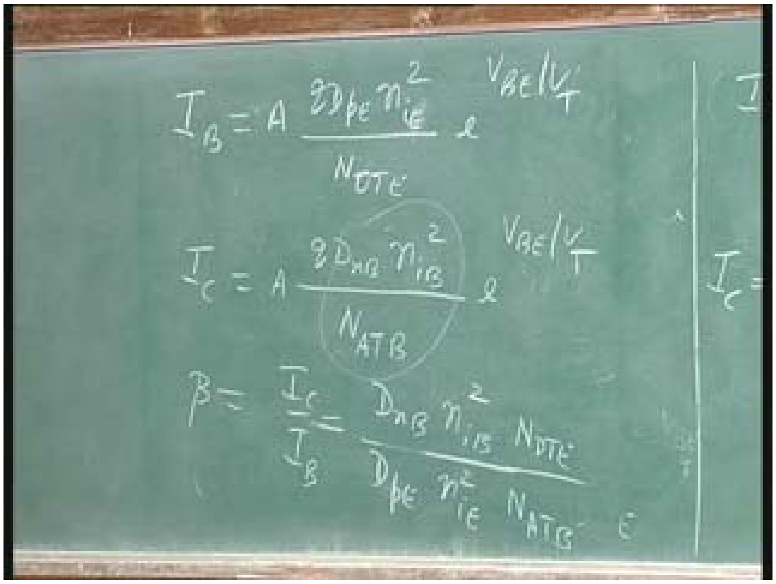
$$I_B \simeq I_{pE} = A\,\frac{q D_{pE} n_i^2 e^{V_{BE}/V_T}}{\int_{\text{Emitter}} N_D\,dx}$$

$$I_C = I_{nC} = A\,\frac{q D_{nB} n_i^2 e^{V_{BE}/V_T}}{\int_{\text{Base}} N_A\,dx}$$

So I_{pE}, let me write once again these two equations: I_B is approximately equal to I_{pE} which is actually equal to area q D_{pE}. I am using the symbol suffix E there again, because we have now different regions. We have the n type emitter, n type collector. I am just talking of whole diffusion coefficient in the emitter there. So I have put E there into n_i squared divided by integral N_D into d of x integrated over the emitter, 0 to W_E, I am just putting integrated over the entire emitter length. That is actually this quantity, total integrated doping concentration.

So I_C is approximately equal to I_{nc} which is equal to I_{nE}, area into 0.2 e to the power V_{BE} by V_T. Similarly, I can write down, I could have skipped this, but I thought it is better: I_{nc} area q D_{nB} ni squared e to the power V_{BE} by V_T divided by integrated over the base N_A into d of x. Beta now is of these two. So, I can remove all these things which we have been discussing here, because finally we have got down the general equations for the collector current and the base current. Beta is ratio of these two. Notice here we have taken n_i here and n_i here is the same thing. What I will do now is, I will just assume that they are not equal. Suppose they are not equal, what happens? That point onwards I think I will some of the slides which are put there.

(Refer Slide Time: 37:06)



We will go to that equation provide a continuity from here. I_B is actually area into q D_{pE}. I put it as n_i. This n_i refers to emitter, whole current injected to the emitter. We are talking of doping the emitter. I will put it as n_{iE} squared divided by N_D, total in emitter, where N_D, total emitter is this one; doping concentration total in the emitter - that is the required quantity - into e to the power of V_{BE} by V_T. Notice, that I have rewritten this equation, whatever I have written earlier, replacing n_i by n_{iE} square.

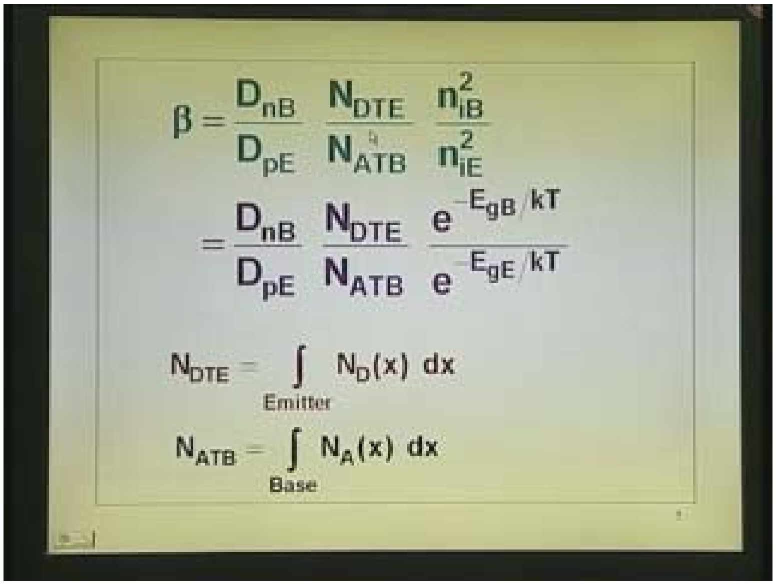
I just want to distinguish between the base width, base doping, base region intrinsic concentration and emitter concentration, if it is different what happens? I_c is area q D_{nB}, electron coefficient in the base into if we go back to this quantity here, this current is the current flowing in the base region. All the properties are related to the base, the whole current injected, all the properties are related to the emitter. That is why we call n_{iE} there. This one will be ni_B. So that is ni_B squared divided by acceptor in the base, total in the base region, N_A accepter, total acceptor concentration. Just the way of putting the thing, instead of writing integral N_A d of x every time, we put that total doping concentration integral into e to the power of V_{BE} by V_T.

In fact, when we find beta, we divide this by that. Area goes off. So beta which is I_C divided by I_B, this cancels out and we have got now D_{nB} ni_B squared. They are remaining. This cancels V_{BE} and in the denominator, we have got N_{ATB}. So I just put this quantity divided by, q gets cancelled

$D_{pE,}$ I will write one below the other one. D_{nB} by D_{pE} and we have got ni_B squared here divided by n_{iE} squared and we have got N_{ATB} and here N_{DTE}. It is the same equation that we have been writing in the first course or so, but the difference is I have added these two things, I have put these two.

If the band gap is the sam, n_i will be the same in the emitter and the base that will go off and beta will be decided by D_{nB} by D_{pE}, total integrated doping in the emitter divided by total integrated doping in the base. In fact, that is what I have put it here.

(Refer Slide Time: 41:54)



$$\beta = \frac{D_{nB}}{D_{pE}} \frac{N_{DTE}}{N_{ATB}} \frac{n_{iB}^2}{n_{iE}^2}$$

$$= \frac{D_{nB}}{D_{pE}} \frac{N_{DTE}}{N_{ATB}} \frac{e^{-E_{gB}/kT}}{e^{-E_{gE}/kT}}$$

$$N_{DTE} = \int_{\text{Emitter}} N_D(x)\, dx$$

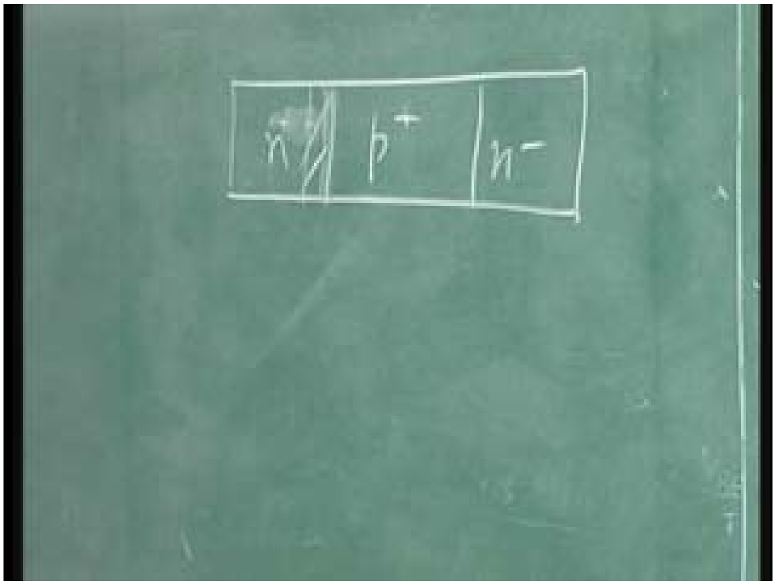
$$N_{ATB} = \int_{\text{Base}} N_A(x)\, dx$$

Let us go back to these equations. So all this effort has been to explain this particular equation and I have written D_{nB} by $D_{pE,}$ N_{DTE} by N_{ATB} multiplied by this quantity. This ratio will be 1, if the band gap in the emitter and the band gap in the base are same and we will get a beta decided by this. Now we can see the beta of this transistor, if these two are the same will be dependent upon total integrated doping in emitter divided by total doping in the base. We have been arguing out just now is I need to increase the total integrated doping in the base. Why?

If I reduce the base width, I must increase the doping concentration and the total integrated doping will also be going up, because of that. To take care to ensure that after all, we saw the punch through voltage and also the r_{bb} depend up on total integrated doping. So this ratio, if I increase this concentration N_{ATB}, the ratio will go down. The beta will go down. Just we can take an example of 10 to the power 18 and 0.1 micron and here 10 to the power 20 into 0.1 micron. Then we have got100, but if I take it even beyond that point, beta falls down.

Not only that mobility also falls down in the base, but now what we want to do is ability to increase this unlimited by this punch through, unlimited by this tunneling etc. Only way, I can do is reduce the doping in the emitter. If I reduce the doping in the emitter, what happens? What I want to do is I want to reduce the doping in the emitter.

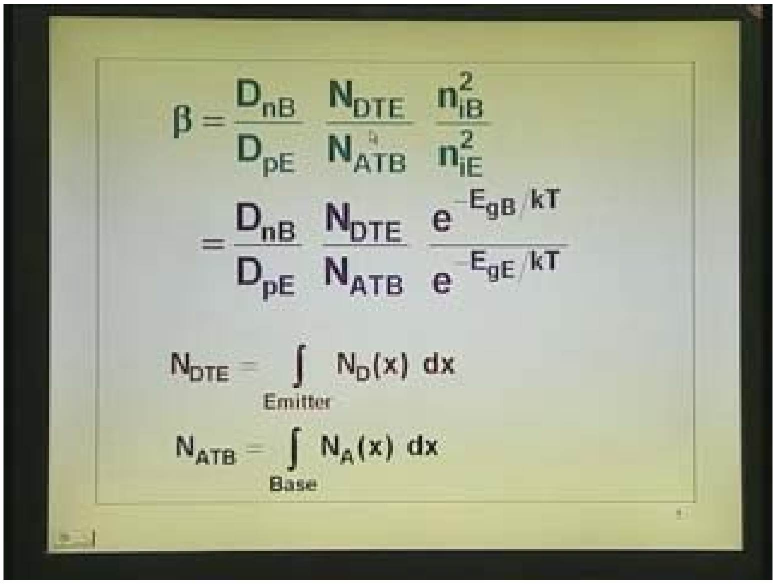
(Refer Slide Time: 43:58)



If I have the transistor fraction like this n, p, n minus. I am keeping this n plus plus, n plus, p plus layer. Now this depletion layer width becomes narrow. I can increase the depletion layer width by reducing the doping here. I put this n. Now we see the whole thing is reversed. Base is much more doped than the emitter. Total integrated doping in the emitter is lower than the total integrated doping in the base. I can have this depletion layer going up like this, wider depletion layer; C_{TE} will go down, if I am able to decrease the doping, but the worry will be if I reduce the doping here, if the band gaps are same in the two regions, emitter efficiency will go down. We will get a very poor transistor.

So with this reduction in the doping here, we achieve everything that we want: we can increase this doping as much as we like; we can reduce this base spreading resistance; we can prevent the punch through and we can reduce the doping here considerably, so that the transistor capacitance is reduced.

(Refer Slide Time: 45:19)



$$\beta = \frac{D_{nB}}{D_{pE}} \frac{N_{DTE}}{N_{ATB}} \frac{n_{iB}^2}{n_{iE}^2}$$

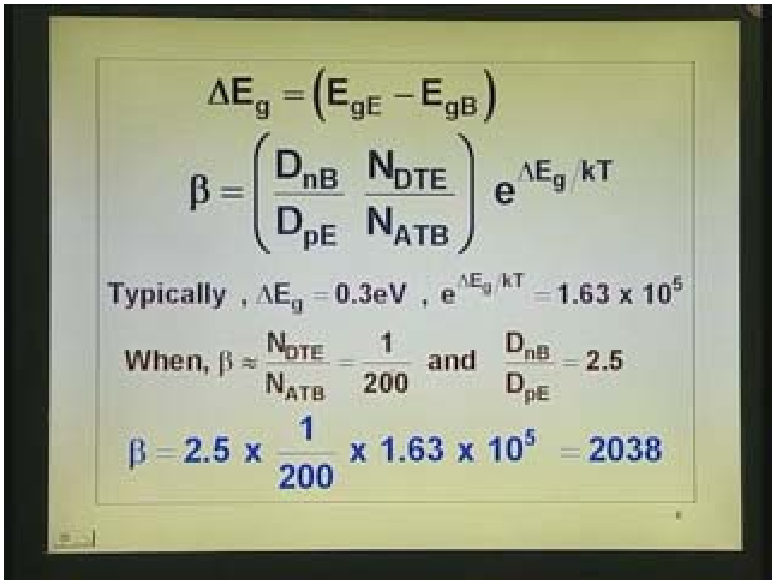
$$= \frac{D_{nB}}{D_{pE}} \frac{N_{DTE}}{N_{ATB}} \frac{e^{-E_{gB}/kT}}{e^{-E_{gE}/kT}}$$

$$N_{DTE} = \int_{\text{Emitter}} N_D(x)\, dx$$

$$N_{ATB} = \int_{\text{Base}} N_A(x)\, dx$$

Now this is the situation described by this equation. The intension is to increase the total integrated doping in the base and reduce the doping here, at the same time keep beta high. How can we keep the beta high? Take a look at this. Only way that we can do that is, use the materials which have different band gap n_{iB} and $n_{iE,}$ should be different. What way do you want it? ni_B squared is some constant e to the power of minus E_{gB} by $k_{T,}$ proportion exponentially to E_{gB}, band gap in the base and this is e to the power of minus E_{gE} by k_T. So what will we get here? These are some of the explanations that I given here, integrated doping given the symbol there. Let us say, this is wider band gap. Emitter is wider band gap compared to the base. So E_{gE} is greater than $E_{gB.}$

(Refer Slide Time: 46:06)



$$\Delta E_g = \left(E_{gE} - E_{gB}\right)$$

$$\beta = \left(\frac{D_{nB}}{D_{pE}}\,\frac{N_{DTE}}{N_{ATB}}\right) e^{\Delta E_g/kT}$$

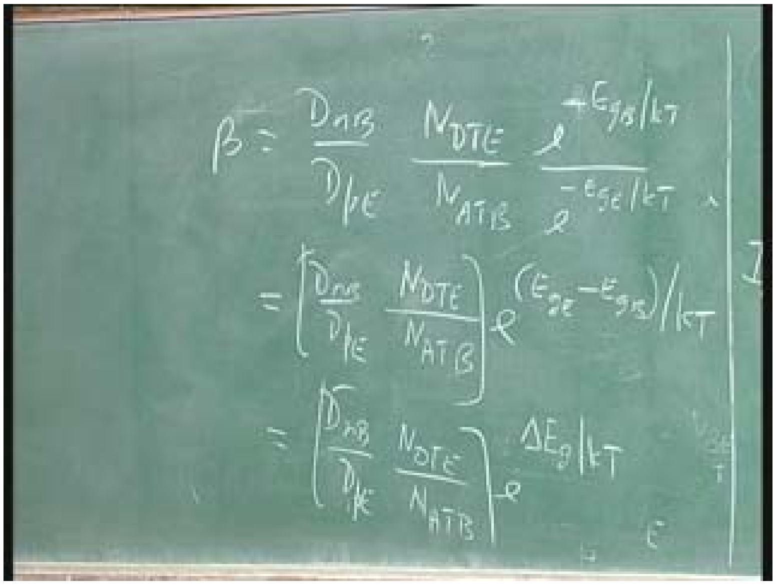
Typically , $\Delta E_g = 0.3eV$, $e^{\Delta E_g/kT} = 1.63 \times 10^5$

When, $\beta \approx \frac{N_{DTE}}{N_{ATB}} = \frac{1}{200}$ and $\frac{D_{nB}}{D_{pE}} = 2.5$

$$\beta = 2.5 \times \frac{1}{200} \times 1.63 \times 10^5 = 2038$$

So I will call it as delta E_g equal to E_{gE} minus E_{gB}. So now what will happen?

(Refer Slide Time: 46:26)



$$\beta = \frac{D_{nB}}{D_{pE}}\,\frac{N_{DTE}}{N_{ATB}}\,\frac{e^{-E_{gB}/kT}}{e^{-E_{gE}/kT}}$$

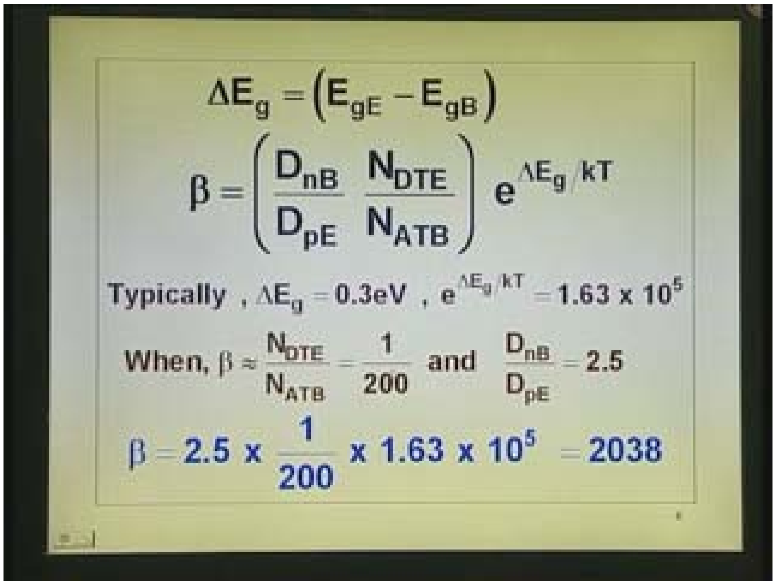
$$= \left[\frac{D_{nB}}{D_{pE}}\,\frac{N_{DTE}}{N_{ATB}}\right] e^{(E_{gE}-E_{gB})/kT}$$

$$= \left[\frac{D_{nB}}{D_{pE}}\,\frac{N_{DTE}}{N_{ATB}}\right] e^{\Delta E_g/kT}$$

The beta is D_{nB} by D_{pE} N_{DTE} emitter divided by N_{ATB} base. I want to cut that down. e to the power of minus E_{gB} by k_T divided by e to the power of minus E_{gE} by k_T. That is what we have written in the previous page which actually is D_{nB} divided by D_{pE} into N_{DTE} divided by this suffix. This looks clumsy, but it tells you doping total concentration in the emitter, acceptor total concentration in the base, donor concentration total in the emitter, acceptor concentration total in the base; e to the power of E_{gE} minus E_{gB} by $k_{T.}$ E_{gE} is greater than $E_{gB.}$ I am taking it up. This is actually equal to, what is this quantity? This is the quantity that we get beta. Band gaps are same; E_{gB} equal to E_{gE} that is the beta of the homojunction transistors. Now let me just put it like this, D_{nB} divided by D_{pE} into N_{DTE} by e to the power of delta E_g by $k_{T.}$

Now if the band gap here is smaller than that. we are gone; this becomes negative. If emitter band gap is smaller than the base band gap, this will be negative quantity and beta will be even poorer. If the band gap in the emitter is larger than this, we can see even, if it is just 100 milli Volts larger, 0.1 electron Volts larger, then the 100/25, e to the power of 4, this is a factor of e to the power of 4; that is about 54. So we get a big factor multiplying to this which gives you the ability. If we choose the band gap of the emitter larger than that… we see the numbers that are here.

(Refer Slide Time: 49:10)



$$\Delta E_g = \left(E_{gE} - E_{gB}\right)$$

$$\beta = \left(\frac{D_{nB}}{D_{pE}}\ \frac{N_{DTE}}{N_{ATB}}\right) e^{\Delta E_g/kT}$$

Typically , $\Delta E_g = 0.3eV$, $e^{\Delta E_g/kT} = 1.63 \times 10^5$

When, $\beta \approx \frac{N_{DTE}}{N_{ATB}} = \frac{1}{200}$ and $\frac{D_{nB}}{D_{pE}} = 2.5$

$$\beta = 2.5 \times \frac{1}{200} \times 1.63 \times 10^5 = 2038$$

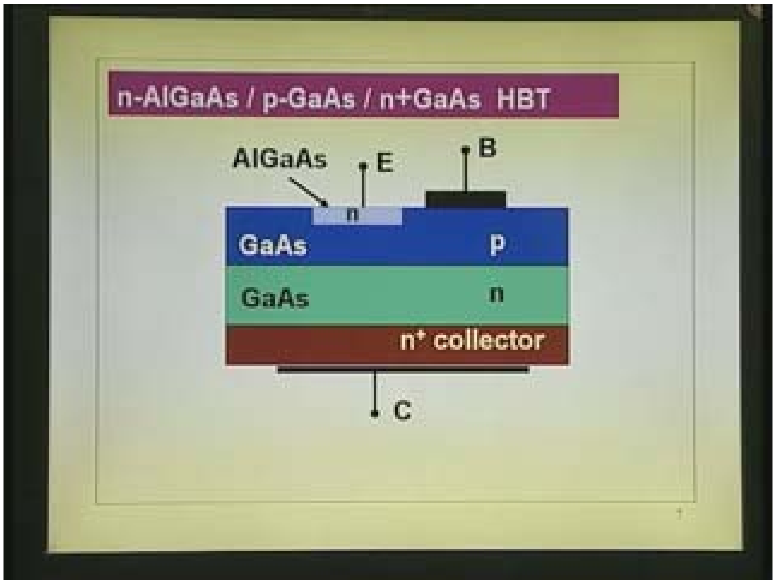
If we take a look at whatever I have written on the board there. What I am trying to tell is if E_{gE} is larger than $E_{gB.}$ This is large number. If this is large number, I can keep the beta high, even if we reduce N_{DE}. Now we can see, I have chosen number to illustrate what sort of beta we can get. By taking delta E_g equal to 0.3 electron Volts, make that emitter of the aluminum gallium arsenide. Make the base that of gallium arsenide. Delta E_g is actually, we can get 0.3 very easily. We can get 0.3, even if we get aluminum content is much lower.

We do not have to go to 0.4 aluminum content, x equal to 0.4. Even much lower than that, if we take delta E_g 0.3 we get very easily, a wider band gap compared to that. Then delta E_g by k_T is about 10 to the power of 5, 0.3 delta E_g that is 300 milli electron Volts divided by 25 milli electron Volts, e to the power of 12, that is the voltage. So this number is 10 to the power of 5. Now we can see what we have got by using a emitter which is not of gallium arsenide, but which is of aluminum gallium arsenide and base of gallium arsenide. So that is 10 to the power of 5.

Now, I will choose N_D total, 200 times smaller than that of N_A in the base. I can choose 10 to the power 19 doping in the base. I can choose emitter doping 10 to the power 17. It looks funny when we think of the homojunction transistor, but here it is the reality, 10 to the power 17 doping. Then, we get a factor of 200 or I just took a number is 200. It is that one. These are some of devices which people have fabricated with that ratio. That is I am just putting these numbers; that will be D_{nB} by $D_{pE,}$ because the doping is very high, we will not get all high mobility there; we will not get all that large value of D_{nB} 2.5, we have taken, but we do not have to worry about that ratio now, because everything is covered by this quantity.

So 2.5 here for this ratio and for this ratio, it is 1/200, 2.5 there. The ratio of N_{DTE} by N_{ATB} is 1/200; base doping is higher than the emitter into 10 to the power 5, we get 2038. So we get a beta which is about 2000 with a doping in the emitter 200 times smaller than doping in the base. All those are achieved, because we have made a transistor whose emitter is wide band gap material compared to that of base.

(Refer Slide Time: 52:04)



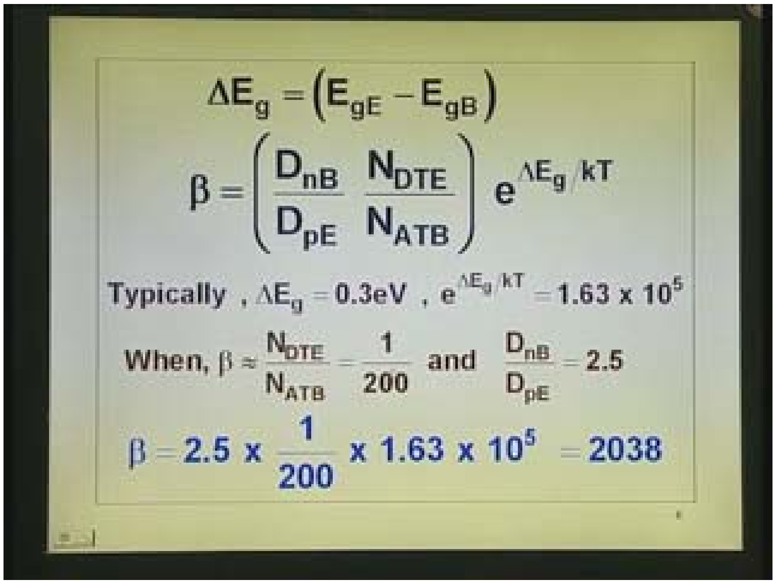
n-AlGaAs / p-GaAs / n+GaAs HBT

So what you see here is the n type region, doping of that of 10 to the power of 17, aluminum gallium arsenide, x you choose. You do not want to go too much of x, because then we will have problems with the aluminum content there and we will also have problem making Homi contact here. So we choose 0.1, 0.2 x. We will get that high band gap difference between the emitter and the base. So it is aluminum gallium arsenide emitter.

This is actually we can call it as generic HBT that is if we use first time HBT ever used or first time heterostructures used are AlGaAs GAas because there is good lattice match. So gallium arsenide 1.43 and this can be 1.73, then 0.3 we get. So we get beta which is very high. This is the schematic diagram which I have shown.

Now we go and make a device. We just make a device like this with all these doping concentrations. Will we get the 2000?

(Refer Slide Time: 53:17)



$$\Delta E_g = \left(E_{gE} - E_{gB}\right)$$

$$\beta = \left(\frac{D_{nB}}{D_{pE}} \frac{N_{DTE}}{N_{ATB}}\right) e^{\Delta E_g / kT}$$

Typically , $\Delta E_g = 0.3eV$, $e^{\Delta E_g / kT} = 1.63 \times 10^5$

When, $\beta \approx \frac{N_{DTE}}{N_{ATB}} = \frac{1}{200}$ and $\frac{D_{nB}}{D_{pE}} = 2.5$

$$\beta = 2.5 \times \frac{1}{200} \times 1.63 \times 10^5 = 2038$$

Before we leave today, I will just mention one or two sentences about that. Actual devices when they made in BELL lab, way back about 15 years back, they made this device with those doping concentration etc. They found couple of things. One the beta is not 2000. Beta is about 30, very bad beta and make the area larger and larger, beta becomes larger and larger.

Apparently it is a two-dimensional effect that comes into the picture. If I see these equations, if I jump with joy, I can get beta with emitter doping of no value, but unless we take care of something else which is happening, we will end up with low beta, much poorer homojunction. So we will see what the effect that is playing role is in the next lecture.