

High Speed Devices and Circuits
Prof. K. N. Bhat
Department of Electrical Engineering
Indian Institute of Technology, Madras

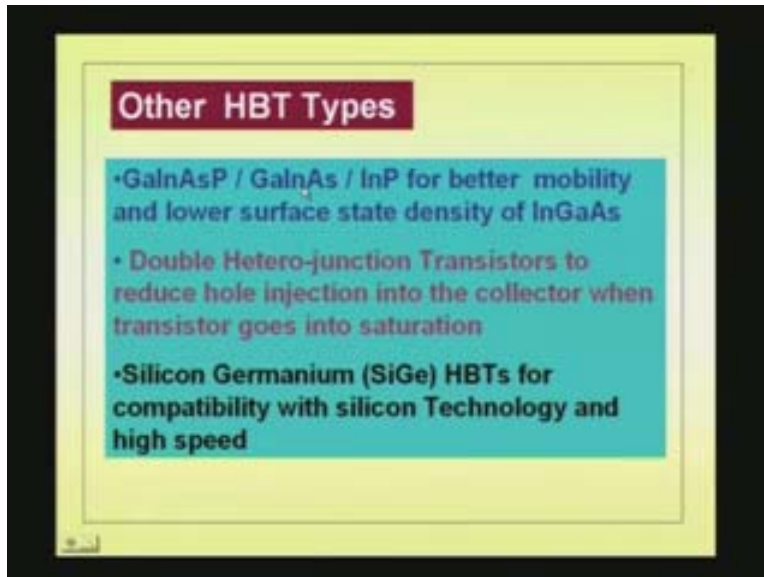
Lecture 41

Heterojunction Bipolar Transistors (HBT) – 4 (contd.)

We will have yet another session on HBT. We have had three sessions focusing mainly on gallium arsenide-based heterojunction bipolar transistors. The reason for switching over to this gallium arsenide-based HBT is to reduce the base width and still accommodate high doping in the base region. This is possible because in the emitter region, you can reduce the doping when the band gap is large, without hurting the emitter efficiency – you can still have very high emitter efficiency.

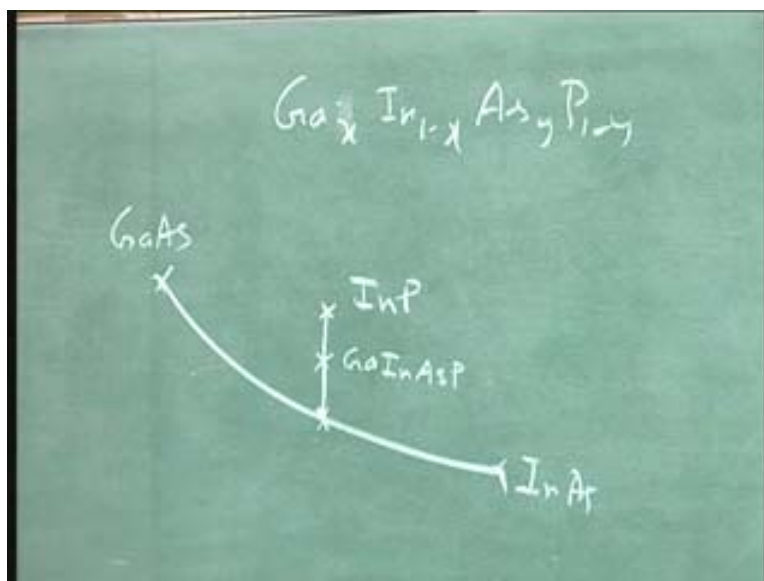
We also saw how the surface effects control the performance of this device. We saw that it is important to passivate the surface of gallium arsenide if we want to get reasonable beta, reasonable performance. Otherwise, the entire performance will be masked by the surface recombination effects. In fact, the first result that was done by Bell Labs, New Jersey was disastrous in the beginning – they expected beta of about 2000 but they got a beta of about 30; the whole reason was the surface effect. Once they passivated the surface, they found that the beta goes up by a factor of 60 right up to 1800, but there were problems with stability and that was sorted out over several years by switching over from sodium sulphide passivation to ammonium sulphide – ammonium sulphide with a coating of PECVD nitride or polymide. Things like that have been going on over the past decade now. I have brought that scenario almost up to date with gallium arsenide-based HBT.

(Refer Slide Time: 02:52)



The other types of HBTs are gallium indium arsenic phosphide – that is quaternary, it is gallium indium arsenide on a ternary and on a binary compound.... Here, all these are lattice match, that is, gallium indium arsenide matches with gallium indium phosphide – you add a little bit of phosphorus to that and the band gap goes closer to indium phosphide. In fact, I just wanted to put that diagram.

(Refer Slide Time: 03:26)



For example, we have... I am just drawing in the diagram indium arsenide, gallium arsenide. I am not putting the axis, etc. In the axis, you have got the band gap and here you have got the lattice constants. This is indium arsenide. Somewhere here, you have got gallium indium arsenide where one of them is 0.47 and the other one is 0.53. Now you add and here, you have got indium phosphide. The band gap of that is higher than gallium indium arsenide and the gallium indium arsenide band gap is lower than that of gallium arsenide.

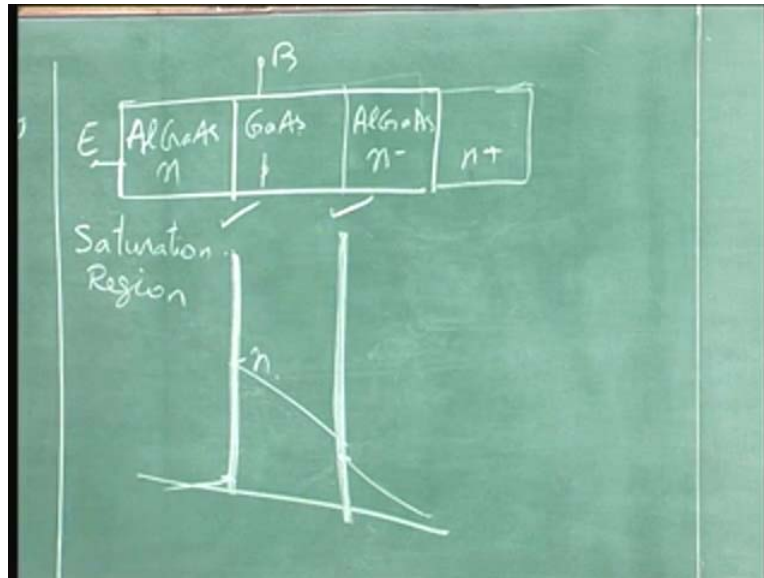
Now to this ternary compound, you add phosphorus – you can call instead of gallium_x, indium_x, etc., you can add phosphorus and arsenic_y. You can call it like this: gallium_x indium_{1-x} arsenic_y phosphorous_{1-y}. We are matching all those things, so if I add phosphorus to that, you move along this vertical line – keeping the same lattice constant x, you move up and so this will be actually somewhere here (Refer Slide Time: 04:42) gallium indium arsenic phosphide. Why I do this diagram is, you know that this lattice matches with gallium indium arsenide and indium phosphide, because after all, adding phosphorus does not change the lattice constant with respect to that, because in the same phosphorus atom, [05:01] lattice. Now why should you go for this?

This type of structure has a benefit. One is gallium indium arsenide as you know has much better mobility than gallium arsenide, because indium arsenide has got higher mobility. When you mix with gallium, gallium indium arsenide has got better mobility – but in between indium arsenide and gallium arsenide. The band gap is also different – lower than that of gallium arsenide, but we get a fairly decent band gap there. The main reason is that it gives better mobility and of course, you put this material that has got a wider band gap region on top of that.

Here, the advantage of gallium indium arsenide over gallium arsenide is not merely higher mobility but it also has lower surface state densities – that is what I have seen. Gallium indium arsenide is easier to passivate compared to gallium arsenide. This is one of the key reasons why people are trying to look into that. Of course, the added attraction for those who are working on indium phosphide substrates, it is the same material – that is of course on a lighter sense. Now there is another HBT. I am not going into details of that – it is only the change in the material but performance-wise, it gives slightly better performance compared to gallium arsenide. Of course, it has the disadvantage of lower band gap – you pay for that in terms of lower band gap and associated effect on the breakdown voltages, etc. Now you can have a double heterojunction. In

fact, when starting with the gallium arsenide HBT, I pointed out that the structure.... Let me just put it on the board here.

(Refer Slide Time: 07:00)

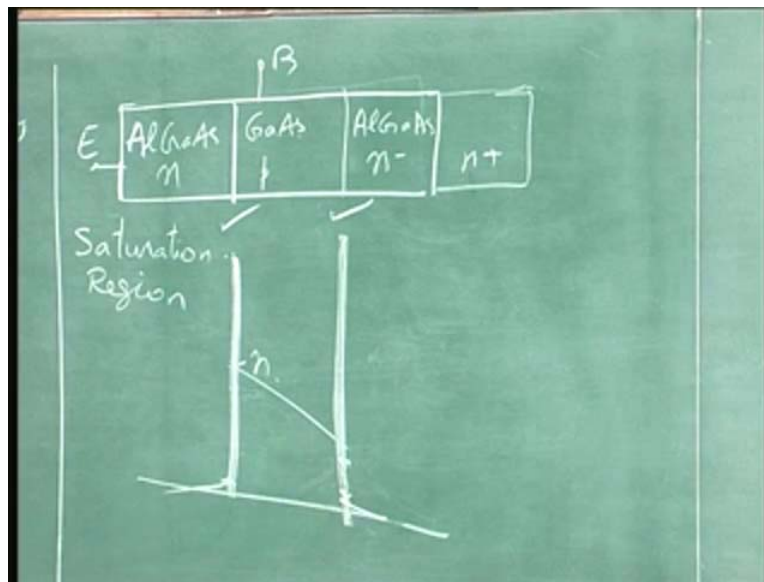


The structure of aluminum gallium arsenide. That is n and in fact lightly doped, then p-type gallium arsenide. This is the emitter, this is the base and the collector... in a heterojunction transistor, we have one heterojunction. In a double heterojunction, we have this also – AlGaAs. On that, you can have.... This may be n minus and then you have got n-plus layer. The n-plus layer can be gallium arsenide, but the main thing is that both this junction and this junction are heterojunctions. What is the merit of this? The merit of that is whatever you gain here (Refer Slide Time: 07:55), you gain here also.

When the transistor is in the active region, not a big role for this, because this layer will be depleted. But if the transistor goes into saturation as in switching circuits, if you are not using ECL, then in TTL and structures like that, the transistor goes into saturation, collector–base junction gets forward-biased and if the collector–base junction is forward-biased, what happens? You have the carrier concentration like that, stored charge there, this is n and of course here, you have got hole injection. Since this is a wide band gap material, we saw that the hole injection into this region is negligibly small: n_i square by doping; n_i square is very small, so the carrier concentration here is very low and that is why you get higher beta – emitter efficiency is very

high in spite of lower doping; here also, it is lower doping, but we have got wider band gap. So if it is forward-biased, the device is in the saturation region, if this were the same gallium arsenide, you would have got the hole concentration reached like that. Of course, I just draw a small region there to signify the transition region. You would have got high concentration hole injection like that if it is a narrow band gap material.

(Refer Slide Time: 09:44)



If it is a wide band gap material, what happens here? Same thing. The hole injection from the np region to the n region when it is getting forward-biased is very very small, because the boundary value here is $pn_0 - p$ in the n region; pn_0 is related to n_i square; n_i square is small, because it is a wide band gap material, so you will have very small injection there. That is what we mean by saying that the transistor will go to saturation. Still, you will not have the stored charge effect; in double heterojunction or switching transistors, this will be ideal. You do not have to use lifetime control techniques, etc. All that you have to worry about will be charge in that region. No stored charge here, no stored charge here, because this is a wide band gap material. We can minimize the stored charge here by shrinking the base width, so you are in business for high-speed circuits. That is one of the merits of double heterojunction transistors (10:37).

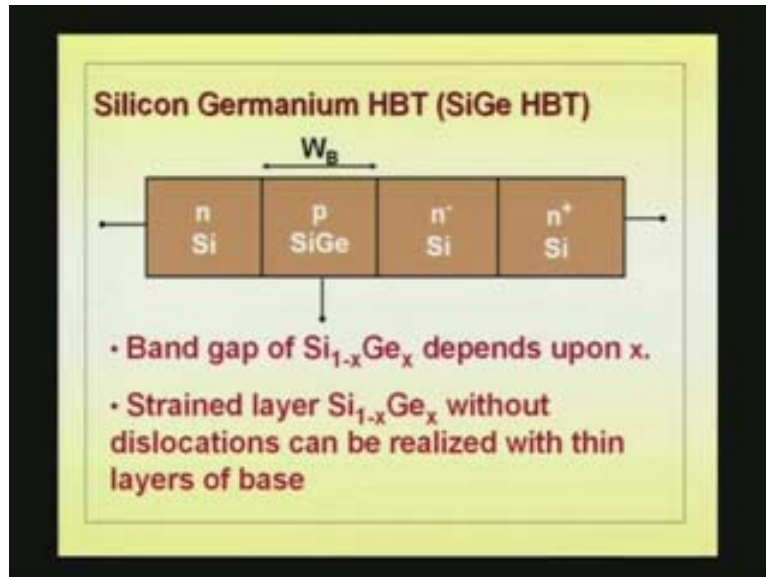
That is why I just thought I will point this out to you because this is something that people have been exploring, but if you are working on any emitter-coupled logic circuit, you do not have to

worry about that, because there is no forward bias on the collector–base junction and there is no saturation – it is only active. Still, you can make use of that (10:58). Without using ECL, you can use it for high-speed devices. So you can see that bipolar is a potential candidate for high-speed devices – HBTs. The other one that has gained popularity over the past decade is the silicon germanium HBT.

What is the merit of silicon germanium? The merit is apparently or evidently, you have got silicon germanium in the base, silicon in the emitter, because the band gap of germanium is 0.7, the band gap of silicon is 1.1 and silicon germanium will be somewhere in between, depending upon how much germanium is there. You have a narrow band gap material and a wide band gap material and if you can grow one over the other and still get a defect-free layer, then you can make heterojunctions with silicon. This is where the joy of the silicon technology person is, because silicon technology is so much advanced today – if we can bring in HBT in that, nothing like that. People are working on that and they came up with the silicon germanium transistor.

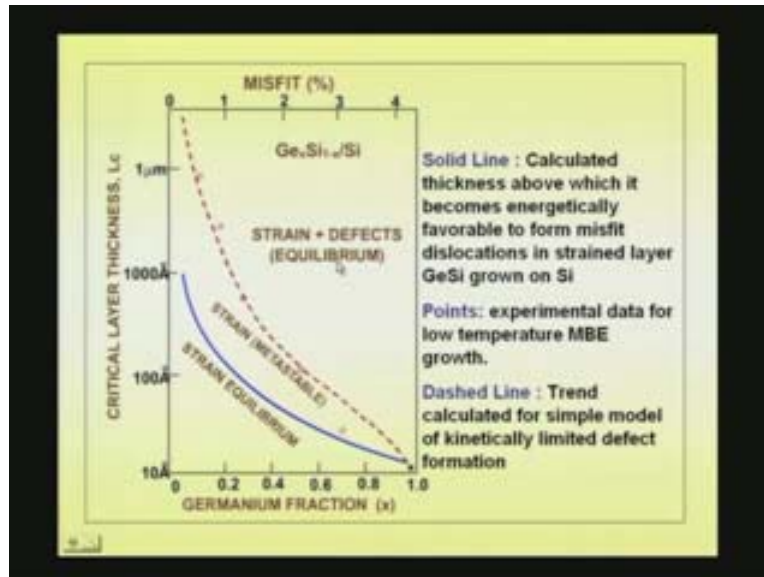
Compatibility with silicon technology: I can have a silicon germanium transistor HBT and I can have CMOS of silicon, I can have BiCMOS – bipolar and CMOS combination, which will give real high speed; in fact, IBM has been working on it and it has succeeded in doing this – combination of silicon germanium transistors and BiCMOS. Let us go through that and see why and how this gives better performance or comparable to that of compound semiconductors like gallium arsenide-based devices.

(Refer Slide Time: 12:51)



That is the structure: silicon emitter, silicon germanium p type, silicon n minus, silicon n plus. You will be wondering why we are putting a lower band gap material, but that is the whole key here. Let us go through this quickly and see what benefits we get – other benefits. The band gap of silicon germanium, which is this base, depends upon x – the mole fraction of germanium; $x = 0$ is silicon and $x = 1$ is germanium. Now the lattice constant of germanium is different from that of silicon; germanium would match with gallium arsenide, 5.63 Angstroms – that is (13:38); silicon is lower than that – 5.43 or so; you may say it is just 4 percent, but 4 percent is quite large when you grow one layer over the other, because for each 25 atoms, there will be one mismatch. If you have to grow a layer over that, how do you do that?

(Refer Slide Time: 13:59)



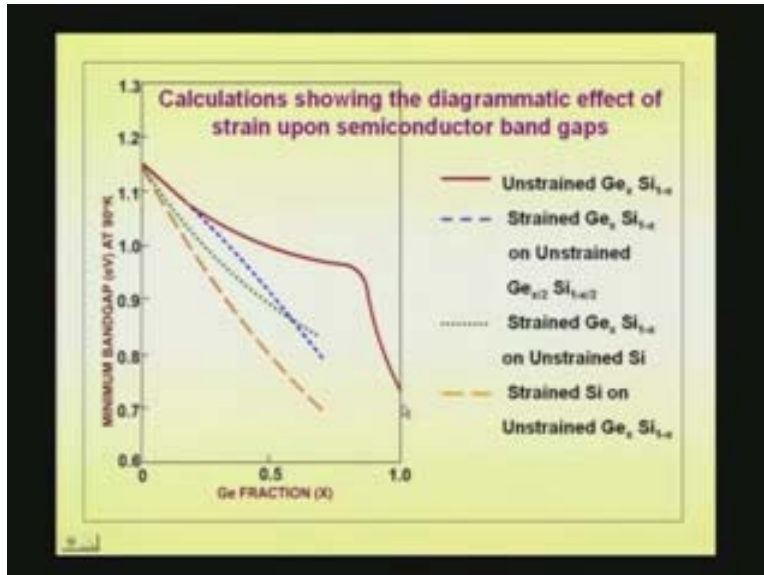
This diagram is a very popular diagram – silicon germanium people always flash this. What we have on the y-axis is the critical thickness of germanium on silicon or silicon germanium on silicon substrate. What is the maximum? I have discussed this earlier when I talked of chain layers. You can grow a lattice match layer, epitaxial layer on a substrate up to a certain thickness. I am not flashing the diagram again, but if you recall, if the lattice is wider, the top layer gets compressed like that and elongated that way so that there is no defect, there is no misfit, no dislocation, you get a single crystal but a strained layer – that is better than a defective layer.

Now if you keep on growing, the stress keeps on building up and at a particular point, the bonds will break and you get a defect. We discussed all this earlier – just recall that. There is a maximum thickness layer that you can tolerate, which does not have a defect but it is strained. You can see here the strained layer maximum thickness. This curve (Refer Slide 15:21) shows you one of the theoretical calculations that has made strain equilibrium – the maximum layer when the strain is in equilibrium. For example, I can grow germanium on silicon but the thickness of layer that you can grow is about 10 Angstroms. It may be just two atomic layers, but you keep on reducing x , go towards silicon. You can see that I can go even to 1000 Angstroms; x is 0.2 – 20 percent germanium, I can grow a layer about 100 Angstroms that is defect-free. This was the initial theoretical value, but if you do this growth by ultra-high vacuum techniques like MBE, etc., you can get thicker layers that are defect-free in the sense that as you grow, you do

not provide the high temperature required for them to break up the bond and so it grows without breaking the bond for thicker layers – that is what is shown here in the dotted line.

The dotted line is actually the one that is called metastable – you grow it at lower temperature; though it has a tendency to create defective layers at this point, if you grow at lower temperatures, you can get thicker layers; not too thick – you are not pitching up on very thick layers, you are talking of base width (16:47) microns and you are happy with that; if we are able to get 1000 Angstroms, that is quite sufficient. I do not know whether you are able to see but there are points one (Refer Slide Time: 16:55), two, three, four, five – those are the experimental points. People have grown silicon germanium by MBE and found that strained layers that are defect-free can be grown with that thickness – that matches with this curve, dotted line. What we are telling is I can have 0.2 x and still get a strained defect-free layer that is about 2000 Angstroms. Now that gives us confidence in this technology and people have gone ahead with making these devices.

(Refer Slide Time: 18:08)



Let us take a look at the band gap. There are a number of curves we have. All that you have to understand is if you just have a silicon germanium alloy, you will get a band gap some particular band – no strain, just make an alloy of silicon germanium, single crystal and grow a thick layer, the intermediate layer will be defective, but the top layer will be silicon germanium with no

defect – thick layer; that is not suitable for making HBT, but if you do that, that will be the band gap, the top curve.

Germanium is 0.72, silicon is slightly more than 1.1 and in fact, we will talk of 1.111 or 1.115 – that range we will talk of. This curve shows it is varying from about 1.15 to... it all depends also on what temperature you are measuring that at – 2.72, but what we are talking of is silicon germanium that is strained – the top layer is unstrained silicon germanium. When the layer is strained, this lattice is under stress, the energy band diagrams also undergo changes and because of that, you get different band gap – in fact, you get lower band gap and you can see that here and in fact, that suits you, you can mix very small quantities of germanium and still get a band gap that is lower than that of silicon.

The top curve here is unstrained silicon germanium. This particular curve (Refer Slide Time: 19:07) is actually strained silicon germanium on unstrained silicon. If you grow thick silicon substrate n type, grow silicon germanium, that is a strained layer, then this is the band gap that they have estimated. You get a band gap that follows this (19:29), which is much smaller than that of silicon germanium unstrained. You can see that for about 0.2, you will have about 1 electron volt; instead of 1.15 here, you have got 1 electron volt (19:44); you have got about 0.15, which is (19:46) about 0.2 – in that ballpark; you have got a material that is strained but with a band gap definitely smaller than that of silicon by about 0.1 electron volts; that is good enough – to get that e to the power of ΔE_{gB} by kT , e to the power of 4 – 50 times more. What is the benefit of using that?

Now let me go back to this diagram. You can put a layer of 1000 Angstroms very comfortably – with a base width 1000 to 2000 Angstroms, with a band gap that is smaller than that of the emitter. Compared to a silicon NPN transistor – silicon germanium base with changeover, what benefit do you get? We have seen already what benefit you get – you can dope that region heavily.

(Refer Slide Time: 20:41)

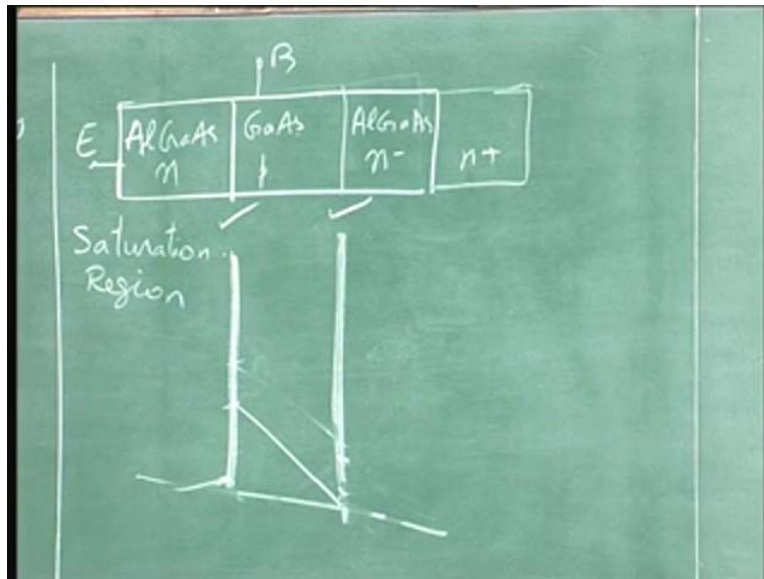
Benefits of SiGe HBT over Si BJT

- Collector Current I_C is larger for a given V_{BE}

$$I_C = \frac{qD_{nB}n_{iB}^2}{W_B \int_0 N_A(x) dx} e^{V_{BE}/V_T}$$
$$n_{iB}^2 \propto e^{-E_{gB}/kT}$$

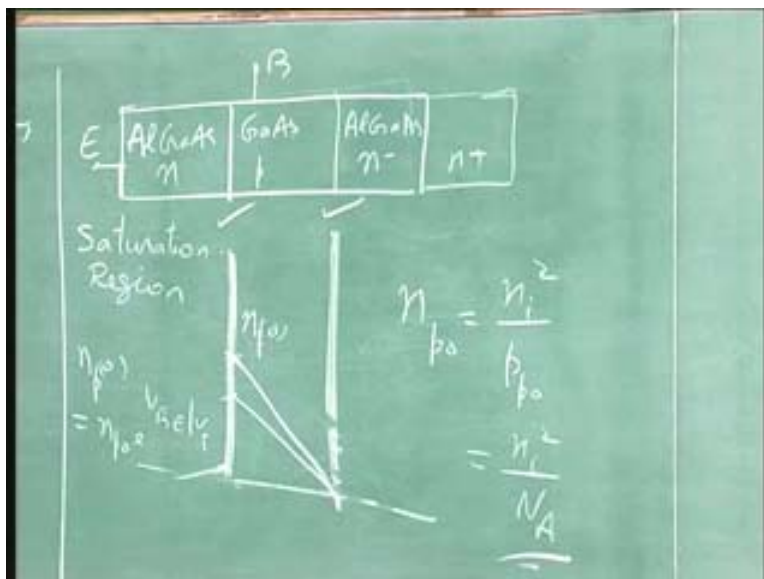
Apart from that even in gallium arsenide-based devices, that holds good. The collector current I_C is larger for a given V_{BE} . How do you say so? We have been writing this expression on and off for the collector current. That depends upon $q D_{nB} n_i$ square by total integrated doping into e to the power of V_{BE} by V_T – this is a very familiar formula for you. You know that you can increase the doping for a given current or you can actually.... When I reduce this band gap, n_i square goes up. If n_i square goes up for everything same, I_C goes up exponentially, n_i square is proportional to e to the power of minus E_{gB} by kT , so the collector current is larger for the same V_{BE} . What I am comparing is a silicon bipolar junction transistor and a silicon germanium base HBT. A silicon germanium base HBT will give higher collector current for the same V_{BE} . In other words, you can go to lower V_{BE} to get the same current.

(Refer Slide Time: 21:57)



We can go to the active region. How is that? That is the way the carrier concentration is. Now the band gap is smaller and you get a larger current. How do you do that in this diagram?

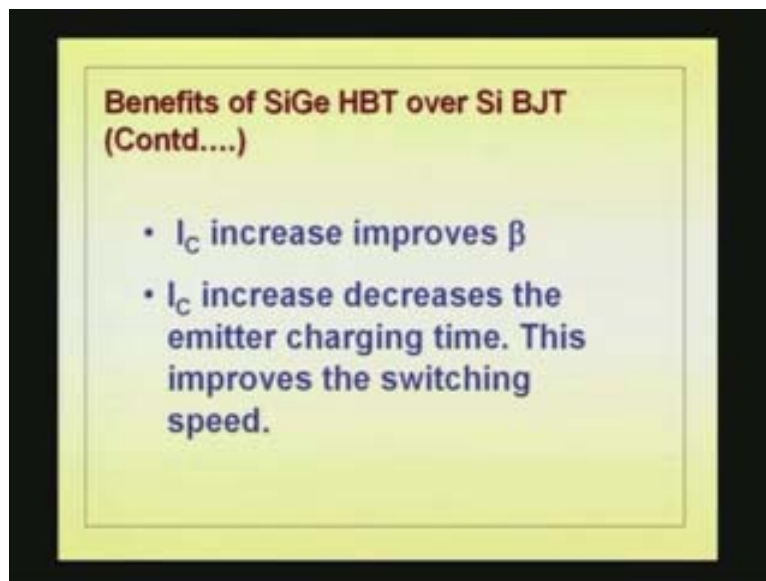
(Refer Slide Time: 22:12)



This is fixed, that goes up and it is like that for the same voltage because after all, $n_{p(0)}$ is equal to $n_{p0} e^{\frac{V_{BE}}{V_T}}$ and n_{p0} is actually the minority carrier concentration, thermal equilibrium. For a given doping, when n_i square is larger, this will be larger (Refer Slide Time:

22:47) and that is why that is larger. $n_{p(0)}$, the minority carrier concentration, is larger if the band gap is here, n_i square divided by p_{p0} . This is n_i square divided by doping. If that increases, that increases and if that increases, $n_{p(0)}$ increases, current decreases. It is a simple way of understanding from the formula we have put in the slide. So for a given voltage, the current is more; this current does not change – what you have injected to the metal, total current increases.

(Refer Slide Time: 23:32)

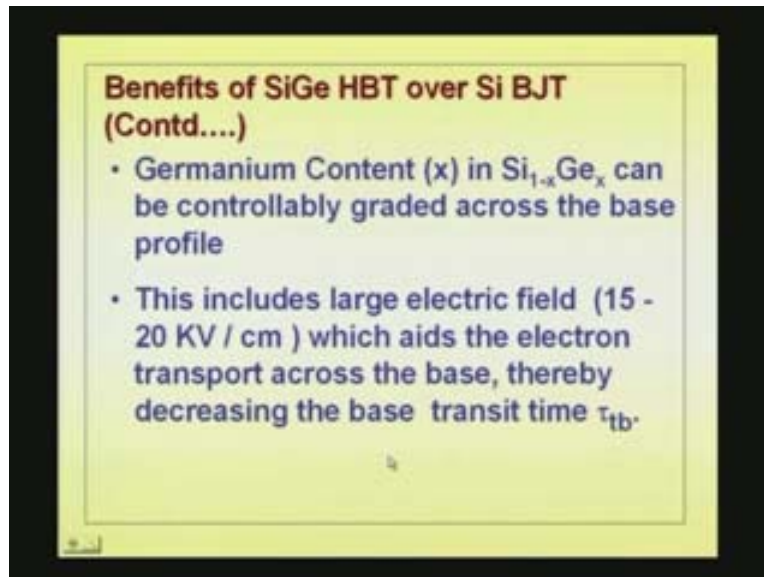


What is the impact of that? The impact of that is that beta improves – collector current increases means beta improves, the base current does not change much because the base width we are talking of is very small, volume recombination we saw is negligible in HBT – it all depends upon the emitter efficiency. Then, the increase in I_c decreases the emitter charging time or any charging time for a given V_{BE} for a given change in collector current; the collector current change is more for a given change in V_{BE} in this case compared to BJT and so, all the charging times are smaller; particularly, the emitter charging time is smaller and this improves the switching speed.

It is a well-known fact that if the currents are larger for a given change in voltage, the switching [24:26] are large; even an integrated circuit has a device (24:28), because after all, the cut-off frequency will depend upon the transit time in the base and the charging time in the emitter – emitter base capacitor charging time. The emitter base charging time depends upon how much

the current is – that is larger here for V_{BE} . You have to charge the smaller voltage to get a particular current or for reaching a particular voltage, you have a larger current available, so it charges faster. That is why the charging time of the emitter base capacitor is smaller, so (25:04).

(Refer Slide Time: 25:04)

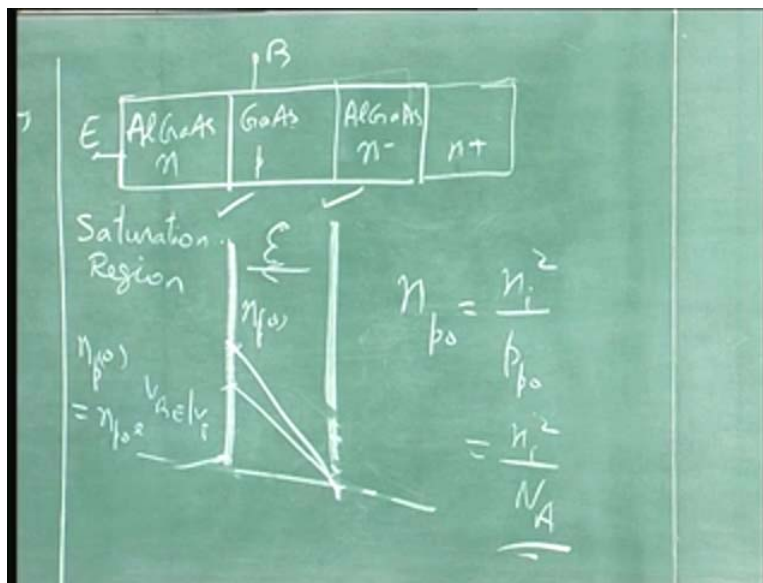


Another very important advantage that is there for all HBTs but is mainly exploited in the case of HBT silicon germanium transistors is this factor – germanium content (x) in silicon germanium can be controllably graded across the base. I do not know whether it rings a bell. As I move from here to here (Refer Slide Time: 25:35), I can vary the germanium content, I can make the germanium content more here and I can keep on changing that; that means I can vary the band gap across the base – narrower band gap, wider band gap here. You can do that by just growing, varying the mole fraction of germanium by adjusting the flow rates or by adjusting the germanium content in MBE – you can vary the band gap. What do you get out of that? You get of course the band gap change; if you change the band gap, it introduces an electric field.

Do you remember? When I draw a diagram like this, the carrier concentration is linear varying – injected carrier concentration. What you are assuming in this carrier transport is through diffusion – there is no electric field here and the transit time is w^2 by $2D_n$. You can increase the speed or reduce the transit time by not only reducing w but if I can bring in an electric field, the transport will be faster. The velocity by diffusion is smaller compared to

velocity by drift. If I introduce drift by some means, I am in this (27:01). I can increase the speed and the transit time can be reduced by a factor of 5 to 10. From 10 GHz, you may go to 100 GHz – people go there; 100 GHz is not a joke – it is a reality, it is not a myth; (27:21) reported 100 GHz silicon germanium transistors, HBT. One of the reasons is this – you can introduce that. I will just quickly go through the analysis of this. Now I have put this as a statement here: you have an electric field that aids the electron transport across the base, thereby decreasing the base transit time.

(Refer Slide Time: 28:00)



When I say an electric field is present here aiding the transport of electrons, the field direction is plus here, minus there – electrons are moving in that direction and it **attracts towards positive (27:58)**, so there is a field in this direction, a built-in field comes in and in the next diagram, you will see it clearly.

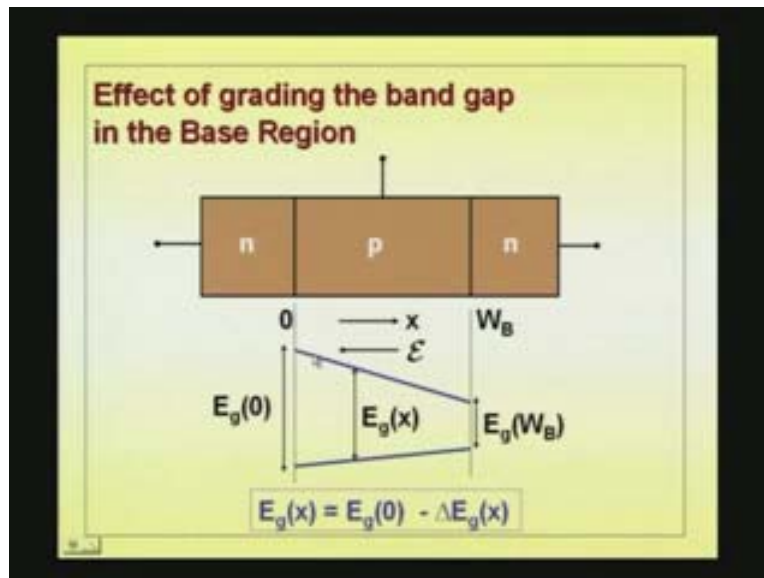
(Refer Slide Time: 28:08)

Benefits of SiGe HBT over Si BJT (Contd....)

- Germanium Content (x) in $\text{Si}_{1-x}\text{Ge}_x$ can be controllably graded across the base profile
- This includes large electric field (15 - 20 KV / cm) which aids the electron transport across the base, thereby decreasing the base transit time τ_{tb} .
- f_T is better due to smaller τ_{tb}

The τ_{tb} is of course reduced, f_T is better due to smaller transit time.

(Refer Slide Time: 28:13)



That diagram tells us the whole story of grading the band gap across the base. I just put of course both the conduction band and valance band. It could be that the valance band is not changing, only the conduction band (28:27), but I have shown both. Now what we are interested in seeing is the impact of this change in the conduction band and the valance band – you can see smaller

band gap E_g at W_B , larger band gap E_g at 0. I hope you would understand that the x that I am using here is for the distance – I tried to change it but I am stuck with the thing, so I am just keeping that; this x is the distance – it is not the mole fraction that we are talking of here now when you write the equations.

x silicon germanium, there we can put x is the mole fraction, but this is actually the distance. So you can see that if I have an energy band diagram like this, conduction band like this, what does it mean? Whenever there is a tilt in the band gap, there is an electric field. A tilt in the band gap like this would mean that this is positive here, minus here (Refer Slide Time: 29:30). The electrons that are injected here find themselves in an electric field that pushes them in that direction towards the collector – drift velocity.

You do not worry about what is happening to the hole here, because we do not care about the holes. In fact, in this type of thing, the electric field, the nice thing about this added thing is if you have a band gap like this, the field for electrons will be in this direction, the field for the holes will be in the other direction. The holes are being prevented from moving into the emitter and the emitter efficiency further improves. This is a classic thing – the nice thing about this type of structure band gap varying continuously across the base region. E_g is a function of x . In fact, what you can also understand is.... We will come back to this if we have time to see the physical meaning of this, but right now from this diagram itself, you can see that there is an electric field like this. Whenever there is a band gap changing like that, the field is plus on the lower energy side and minus on that side.

(Refer Slide Time: 30:48)

Electric Field on Electrons

Thermal Equilibrium Condition

$$J_n = qD_n \frac{d}{dx} [n_{p0}(x)] + qn_{p0}(x)\mu_n \mathcal{E} = 0 \quad \dots\dots\dots (1)$$

n_{p0} is the thermal equilibrium value of n_p and is a function of distance, x , from the emitter base junction

$$n_{p0}(x) = \frac{n_i^2(x)}{p_{p0}} = \frac{n_i^2(x)}{N_A(x)} \quad \dots\dots\dots 2$$

Now let us go back and see what is the magnitude of this field – we will rush through that very quickly, very simple equation. Instead of writing the current as due to diffusion alone, since you know that if there is electric field, you have to write diffusion current and drift current, it is $q D_n \frac{dn}{dx}$. I am writing this equation at thermal equilibrium; no bias – we are not applying any bias, so it is $q D_n \frac{dn}{dx}$. I am putting n_{p0} instead of n – thermal equilibrium and it is a function of x and $q n_{p0}$ (31:27) (31:28), E is the electric field and that should be equal to 0. No electron current under thermal equilibrium conditions. n_{p0} is actually the thermal equilibrium value of n_p – electron concentration in the p region.

If I remove the 0, it will be actually injected carrier concentration and it is a function of distance x from the emitter to the base, from the emitter–base junction. Why? Because n_{p0} is n_i square by p_p and p_p is the doping; doping may also be function of x ; even if it is constant, you will have the n_p varying; even if it is constant, if the band gap is varied, n_i is varying, n_i is actually increasing as you move from the emitter to the other end and so n_p actually increases. The thermal equilibrium concentration value is more in that region as x goes on. Now to find out the field, you have to equate this current to 0, because it is thermal equilibrium. There will be a concentration gradient existing, but if the current flow has to be 0 when the concentration gradient is present, the only way possible is presence of electric field and that comes because of

this charge redistribution. So that electric field is the one we are going to find out. Equate this diffusion current to drift current and plug in this n_{p0} equal to n_i square by $N_A(x)$.

(Refer Slide Time: 33:00)

From 1 and 2

$$0 = qD_n \frac{d}{dx} \left(\frac{n_i^2}{N_A} \right) + q \frac{n_i^2}{N_A} \mu_n \mathcal{E}$$

$$= qD_n \left[\frac{1}{N_A} \frac{d}{dx} (n_i^2) - \frac{n_i^2}{N_A^2} \frac{dN_A}{dx} \right]$$

$$+ q \frac{n_i^2}{N_A} \mu_n \mathcal{E}$$

I have rewritten that equation. Instead of n_{p0} , now I am putting n_i square by N_A , so $q D_n d$ by d of $x n_{p0}$, this is n_i square by N_A , $q n_p$ is n_i square by N_A (33:21). That is 0 in thermal equilibrium condition. That is what we mean by equation 1 and 2. Now, all that you do is differentiate this. When I differentiate that, the first term, I am retaining the second term retaining as it is here – $q n_i$ square by $N_A \mu_n E$ is retained as it is; remember that N_A may be a function of x , so we will take that also into account and E is a function of x .

In the first term, if I take N_A also as a function of doping, then I get $q D_n$, I pull out 1 over N_A , qD_n formula: first function into differential of second function – it is 1 over N_A , differential of second function is d by dx of n_i square I am not differentiating it right now but keeping it intact plus second function is n_i square into differential of 1 by N_A , which is minus 1 by N_A square into dN_A by d of x , because N_A is a function of x ; if N_A is independent of x , that term goes off. This term is written here (Refer Slide Time: 34:34). Now, E is given by equating to 0. What I do is I find E equal to.... I take all these to the other side, divide by $q n_i$ square by N_A . Let us take this term and divide by $q n_i$ square by N_A .

(Refer Slide Time: 35:06)

This gives,

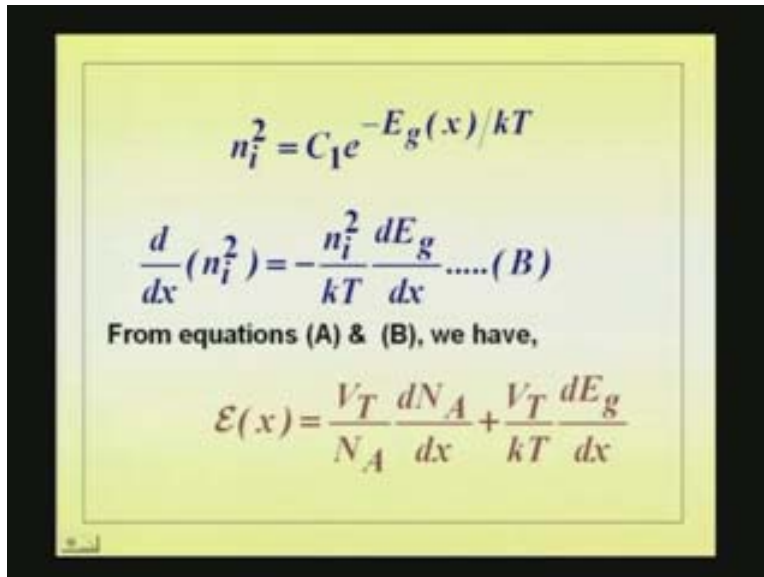
$$\mathcal{E}(x) = \frac{D_n}{\mu_n} \frac{1}{N_A} \frac{dN_A}{dx} - \frac{D_n}{\mu_n} \frac{1}{n_i^2} \frac{d}{dx} (n_i^2)$$

$$\mathcal{E}(x) = \frac{V_T}{N_A} \frac{dN_A}{dx} - \frac{V_T}{n_i^2} \frac{d}{dx} (n_i^2) \dots (A)$$

n_i square by N_A ... when you do that, you get that. Just rearranging the things, what you get is actually that the electric field is equal to D_n by μ_n . We can just see it here. You get D_n by μ_n when you divide the whole thing. D_n by μ_n is there for both the terms and in this term (Refer Slide Time: 35:22), n_i square cancels and you get 1 by N_A and since it goes to other side, it is plus, so you get D_n by μ_n into 1 by N_A dN_A by d of x . Similarly, the second term becomes D_n by μ_n into 1 by n_i square d by dx of....

You can just see for yourself by rearranging those – this is what you get. Now, the electric field. D_n by μ_n is actually equal to $V_T k_T$ by q , so V_T by N_A into dN_A by d of x – this is a very familiar formula for some of you who have taken more courses. If there is a doping gradation, there is an electric field due to the built-in field due to that. Wherever there is doping, gradation will be there, that is the first term. You get now additional terms due to the band gap variation. That is V_T by n_i square d by dx of n_i square. Now let us take a look at that. Now this n_i square is actually proportional to E to the power of minus E_g by k_T . I differentiate this (Refer Slide Time: 36:19) and divide by n_i square.

(Refer Slide Time: 36:11)



$$n_i^2 = C_1 e^{-E_g(x)/kT}$$

$$\frac{d}{dx}(n_i^2) = -\frac{n_i^2}{kT} \frac{dE_g}{dx} \dots (B)$$

From equations (A) & (B), we have,

$$\mathcal{E}(x) = \frac{V_T}{N_A} \frac{dN_A}{dx} + \frac{V_T}{kT} \frac{dE_g}{dx}$$

So d by dx n_i square is actually... When I differentiate that, you get minus 1 by k_T n_i square, (36:33) is the exponential term. When you differentiate this with respect to x, you get n_i square itself, because it is exponential e to the power of minus k – minus E_g by k_T , but 1 by k_T into differential of dE by dx, df by dx is a function you are differentiating once and then...

Simple differential equations, differential formula, so you get that because when you differentiate, you get whole term on the numerator, by k_T , differential of this, minus is there because of that. I plug in this quantity into the previous equation, which actually has V_T by n_i square into dE_g – we will just go back to that. See here, the first term is V_T by N_A into dN_A by d of x – that I am retaining as it is; second term, I am substituting for d by dx n_i square, which is nothing but minus 1 by k_T into dE_g by dx into n_i square; n_i square cancels, so I get that.

What you get now is V_T by N_A into dN_A by d of x plus V_T by k_T into dE_g by d of x. It tells you totally now that the electric field is decided by doping gradation plus gradation in the mole fraction – that is gradation in the band gap. Now let us take a look at that. That was missing there. This is the formula that we have got now, which tells us what is the electric field thermal equilibrium. Now when you forward bias the emitter–base junction and inject carrier electrons, that electron experiences this built-in field. What is the direction of this field? D_n by d of x. If the doping is changing from here to here, more doping and less doping there (Refer Slide Time:

38:46), if there is doping naturally and if you do a diffusion, it will have doping here, n_0 e to the power of minus x by L, something like that – exponentially it is falling. Then, the field actually will be.... D_n by dx is negative (39:01) in that direction. What about the second term – plus dE_g by d of x? If the band gap is going on falling here, E_g is going on falling and that also in that direction. Both are in this direction (Refer Slide Time: 39:18). Both the electric fields actually support the faster movement of electrons from the emitter to the collector – this is the added benefit that you get for silicon germanium HBT compared to BJT with silicon alone. You can also put this term particularly as... sometimes, you will see this put as V_T is k_T by q, so that becomes 1 by q.

(Refer Slide Time: 39:56)

$$\mathcal{E}(x) = \frac{V_T}{N_A} \frac{dN_A}{dx} + \frac{1}{q} \frac{dE_g}{dx}$$

- First term is due to doping gradation and gives electric field component aiding electron flow.
- Second Term is due to bandgap gradation and aids electron flow if dE_g/dx is negative.
- $E_g(W_B)$ should be less than $E_g(0)$

It is 1 by q dE_g by d of x. The first term is due to doping gradation and it gives the electric field component aiding the electron flow, provided doping gradation is in that direction from the emitter to the base collector, which is usually the case when you do diffusion. Take a substrate, diffuse from top, doping is higher on the top and so as you move down towards the collector, doping is reduced – that is aiding the electron flow. What about the second term? The second term is due to the band gap gradation and it aids electron flow if dE_g by dx is negative. It means actually that the band gap at the collector edge, at W_B should be less than the band gap at the emitter–base junction (40:44).

(Refer Slide Time: 40:46)

Electric Field due to bandgap gradation is

$$\frac{1}{q} \frac{dE_g}{dx}$$

For a linear gradation,

$$\frac{dE_g}{dx} = -\frac{E_g(0) - E_g(W_B)}{W_B} = -\frac{\Delta E_g}{W_B}$$

For $\Delta E_g = 0.15$ eV and $W_B = 0.1 \mu\text{m}$
Electric Field = $0.15/10^{-5} = 15$ KV / cm

Now let us get an idea of what sort of fields are there in the transistor. You see here this $1/q$ by dE_g by d of x – that is the second term; we are seeing what is the magnitude of the field due to the band gap gradation – that is that. For a linear gradation, you have got the linearly varying band gap – ΔE_g varies as a function of x linearly, then you have got ΔE_g by d of x , the negative sign is there, that minus that by W_B . Your dE_g by d of x is negative, that is why it is put it there. Magnitude is that quantity – it is ΔE_g by W_B .

Now if ΔE_g is 0.15 electron volts – (41:54) vary the band gap from the collector to the emitter by 0.15, then for a 0.1 micron base width, we can still get a strained layer that has band gap and has no defects. The electric field is 0.15 by 10 power of -5 , 15 KV per cm. This is the benefit that you get. Current (42:21) by drift and diffusion. Now, there are a couple of things I want to point out here. If you take the electric field in the case of doping gradation, what sort of electric field do you get?

(Refer Slide Time: 42:43)

$$N_A(x) = N_{A0} e^{-x/L}$$

$$E = \frac{V_T}{N_A} \frac{dN_A}{dx}$$

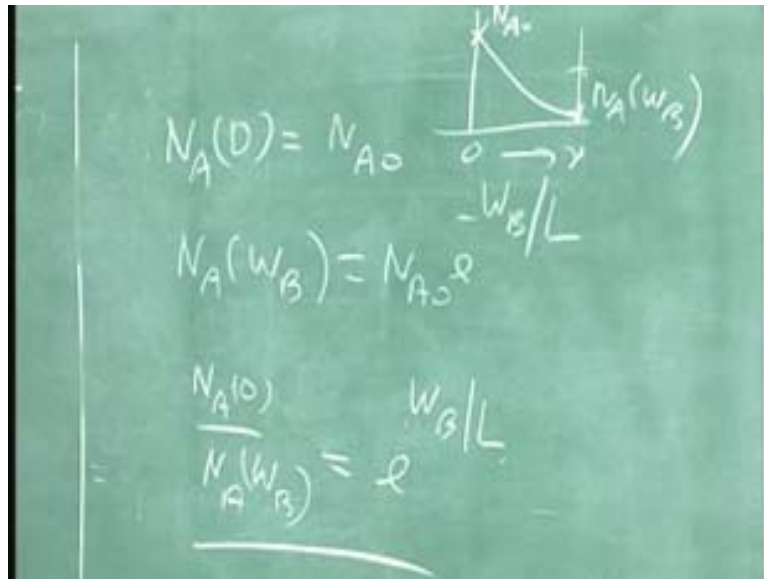
$$= -\frac{V_T}{L}$$

E [A/cm]
 Saturation Region
 $n_i(0) = n_{i0} e^{-x/L}$

Let me put that. If I take the $N_A(x)$ equal to N_{A0} into e to the power of minus x by L , what is the electric field? No band gap gradation. You get this all the time in the case of BJT – all modern BJTs have an electric field built in the base because of this gradation that you saw just now. What is that actually? V_T by N_A of x into dN_A by dx . We just now saw this; go back and see that term – that is due to doping gradation. In fact, you can derive that equation right from the fundamentals when n_i is constant – we will get only that one term. The way we started, we have taken n_i varying and n_A also as varying. When you take n_i constant, you will get the second term dropped out, if ΔE_g by dx is 0.

If that is the thing, what is the electric field? L is a constant, the decay constant for doping. What is the electric field? V_T by L . dN_A by d of x is this (Refer Slide Time: 44:12) into divided by L and that is the quantity. You will get (44:20) minus – that is minus. From here, the minus sign comes and this tells you that the electric field is in that direction – the minus sign is right and it must be there; otherwise, whatever we have been telling is not correct. The minus sign comes from there and in fact, it tells that dN_A by dx is negative and it is a constant electric field. What will be the magnitude of this quantity V_T by L ? Some idea we must have. That depends upon L .

(Refer Slide Time: 45:15)



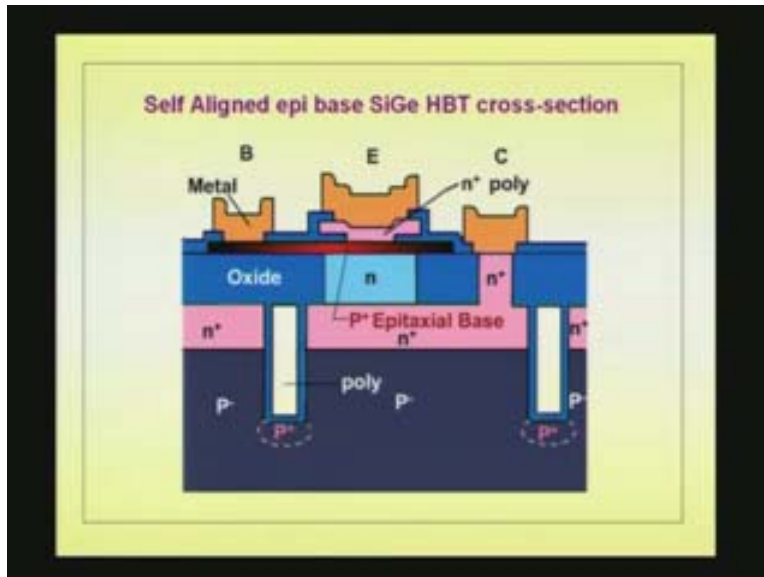
How much will L be? I go back down here. $N_A(0)$ is N_{A0} . N_A at W_B is N_{A0} into e to the power of minus W_B by L . What I am getting at is this: that field depends upon the base width and L . Therefore, $N_A(0)$, the doping at the emitter base (45:50), divided by N_A at W_B is actually equal to e to the power of W_B by L . This W_B by L therefore depends on logarithm of that (Refer Slide Time: 46:09). Supposing W_B You will get the benefit of this only if the doping here is smaller compared to doping there – gradation. When we have doping concentration like this – 0 and x , some variation like that, N_{A0} and that is $n_A(W_B)$, that is what you are talking of – exponential doping concentration gradation, you will get the doping variation like this. Now W_B

Supposing this ratio is 100 or let us say this ratio is 54 – doping at the emitter and doping at the collector edge, that ratio is 54, which is reasonable to assume, what will be this factor? 4, e to the power of 4 is 54; if this ratio is 54, you get this as e to the power of 4, so W_B by L is equal to 4. If W_B by L is 4, if W_B is let us say 0.4 micron, I am just giving some example, L will be equal to 0.1 micron – I am just computing some number to see what is the order of the field here. If this ratio is 54, this is e to the power of 4; that means W_B by L is 4 and if W_B is 0.4, L is 0.1.

(Refer Slide Time: 47:50)

If L is equal to 0.1, I am just computing the (47:52). What is the order of this L ? It may be 0.1, 0.2, whatever. If it is 0.1, what is the V_T by L field now? 25 millivolts divided by 0.1. So for this case, this is actually minus V_T by L , which is actually equal to 25 millivolts into 10 to the power of -3 divided by 10 to the power of -5 . How much is that? This is 2.5 KV per cm. I just took this example to illustrate that you will have in the conventional transistor itself a built-in field of this order. In fact, the factor W_B by L is called the field factor, which actually decides what is the potential change there due to that electric field. You have this much field and you also saw that in this case, I have taken an example where the base width is 0.4. If I take the same example with base width is 0.1, one-fourth of that will be 4 times that -10 KV per cm. I took a wrong example there of L equal to 0.1, taking W_B as 0.4, but if you have to compare with W_B equal to 0.1 microns and if this is 4, L will be 0.1 by 4 and this will be 10 KV per cm. It just depends upon the base width and the doping gradation there. In other words, it depends upon the W_B by L ratio and that L depends upon the doping concentrations.

(Refer Slide Time: 49:48)

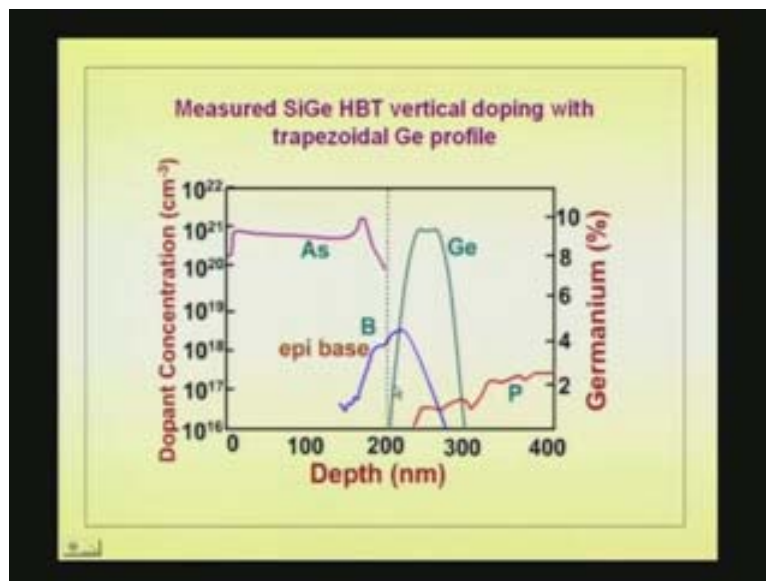


Now, let us go further down. This is the structure that.... Now, we can run through the results that people have reported. They have made devices that can go right up to 100 GHz. You go to the Web and say 'silicon germanium HBT', you will see IBM results put across – they have made BiCMOS, they have made ECS and all sorts of things using the silicon germanium technology. It is a big contender for gallium arsenide technology – silicon germanium. This is the answer, challenge thrown by silicon technology towards gallium arsenide-based HBT, but still the high electron mobility transistor stays there and there is no contender here – that is one thing that you have to remember. But there are other issues: when you want to go to harsh environments like higher temperatures, you cannot use base materials that have smaller band gap. Silicon germanium has a smaller band gap compared to gallium arsenide, etc., and so for those harsh environments, you still have to look at gallium arsenide-based devices.

This is the type of structure that you see. In a bipolar technology, you have the integrated circuit, you have the p minus layer substrate and n plus collector contact (51:08) from the top, then n-type layer, collector and you can see the narrow p plus epitaxial base made of silicon germanium and contact is taken laterally there. Of course, I am not going through the technology of this – this is a self-aligned technology where you can grow layer by layer without too much difficulty in aligning single alignment (51:32) defined already, you can realize this. This is that. This is the emitter, base, collector; the emitter is polysilicon, n-plus polysilicon – this is another thing that is

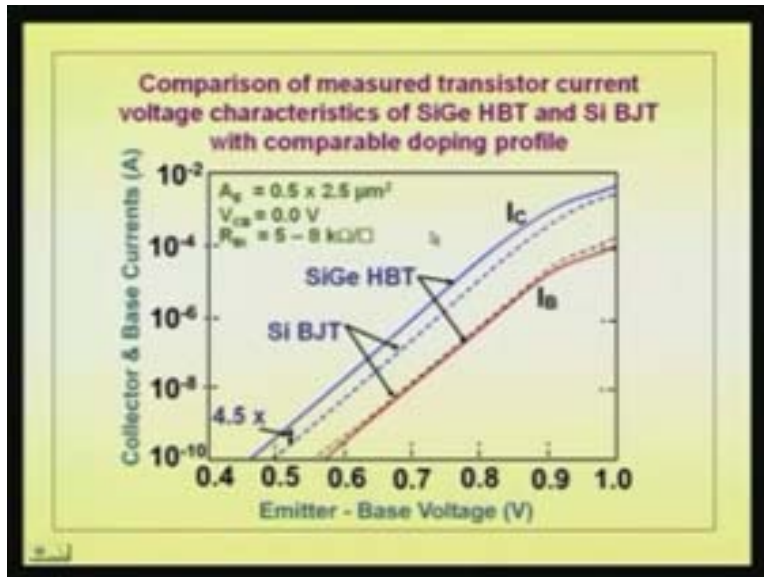
actually very attractive for self-aligned techniques; you also put n plus contact – it is slightly diffused into that region; then you make a single crystal – a very shallow n plus junction or n-type junction and then a contact is taken on the n plus layer. This is the structure here. Of course, people use silicon technology for trench isolation, etc., where you put a trench, oxidize it, put polysilicon... all those techniques are there for isolating devices integrated circuits. (52:25) IBM (52:26) and the layers, the device structures are microns.

(Refer Slide Time: 52:33)



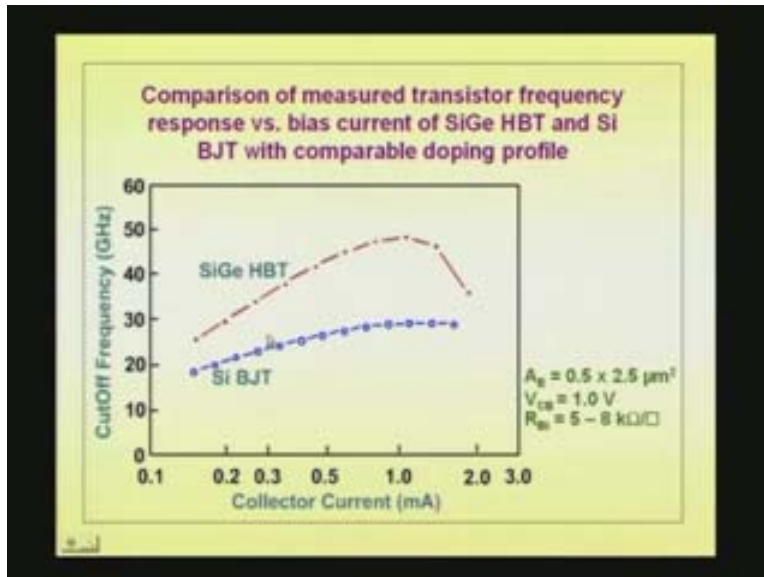
Doping profile. I will just quickly run through this because there is no more explanation needed. This particular structure that is used here does not show doping concentration changes; arsenic emitter doping (Refer Slide Time: 52:45), base doping gradation and germanium; we have not shown that gradation that much in this case, but you can see from here to here that the gradation is there; germanium gradation is there, that is, the band gap is varying from larger to smaller value (52:59) other way and there is a built-in field coming up; the percentage of germanium goes right up to about 10 percent, 0.1, so band gap is narrower compared to that.

(Refer Slide Time: 53:12)



As I said right at the beginning, the collector current is larger in the case of silicon germanium BJT compared to the.... There are two curves: collector current in the silicon germanium HBT (Refer Slide Time: 53:21), collector current in the case of silicon BJT – that is about a factor of 4.5 larger; one is due to the field and the other one is heterojunction effect – both. The base current is the same and the device size as we can see is 0.5 by 2.5 micron – very small strip for the emitter, which reduces all the capacitance (53:42).

(Refer Slide Time: 53:44)



Then, cut-off frequencies. Once you get actually larger current for a smaller voltage, the charging times are all smaller and you get much higher cut-off frequencies. For this particular device that has been reported quite some years back, the cut-off frequency depends upon the collector current, because that gives better charging capability for the emitter-base junction and that is why as you go to higher currents, you get higher cut-off frequency, but what you have to do is compare with the single BJT – much higher. For 20 or 25, we get about 50. This is an older result that I am putting here, but I have seen 100 GHz; these all are GHz, 50 GHz.

In the emitter-coupled logic circuit they have made, there BJT, if we get 30 picoseconds, the gate delay – unloaded of course, (54:45) of 1, fan-in by fan-out 1, that is 30 in the case of BJT if you have minimum and here you get 20 for HBT (54:53). In fact, they have gone down further below that picosecond range (54:58) – much lower than that when you go to GHz. These are the merits that we get in the case of silicon germanium. In fact, I think I do not need to discuss more about these things. In summary, silicon germanium is playing a key role today and a lot of people are putting in efforts, because we can realize silicon germanium transistors bipolar along with CMOS integrated circuits. With that note, I will close today. We will take on the circuit aspects in the next lecture.