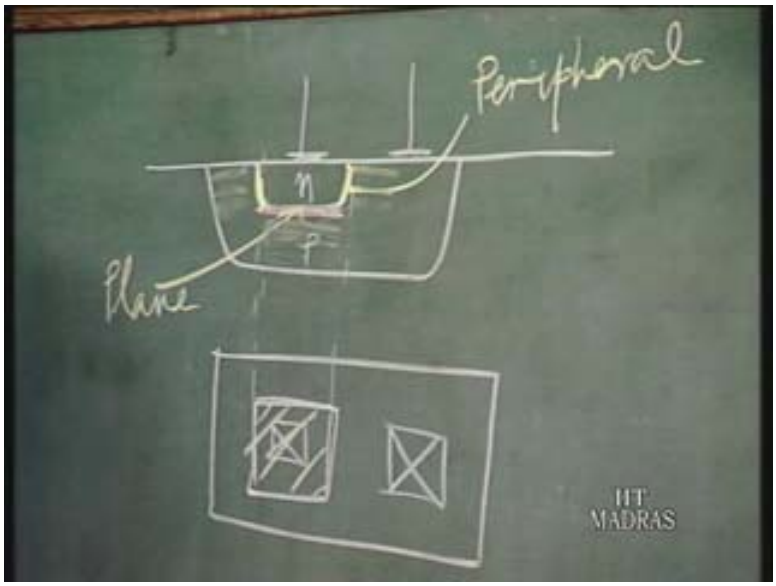
Digital Integrated Circuits
Dr. Amitava Dasgupta
Department of Electrical Engineering
Indian Institute of Technology, Madras
Lecture - 22
Polyemitter Bipolar
Transistor in ECL;
Propagation delay

Last class we were discussing the different circuit variations in ECL. Primarily aimed at reducing the power dissipation in ECL gates which is a primary drawback of ECL gates. Today's class we shall take up some technology related aspects basically to improve the speed performance of ECL gates. We know that ECL is the fastest logic circuit and there is a lot of effort to improve the speed and make it as fast as possible to reduce the delays and one way to do that is to have improved technology. We shall take up some of these issues in today's class.

Now if you look at a bipolar junction transistor, we have drawn the cross section quite a number of times. This is the emitter base junction, this is the base contact. So this is np, this is a emitter base junction and the delay in a circuit as we have seen is dependent to a large extent on the emitter base junction capacitors because the emitter base capacitance which means of course the junction capacitance as well as the diffusion capacitance. Of course in emitter base junction it is the diffusion capacitance which dominates. That is because there is large charge stored storage and the diffusion capacitance is many times larger than the junction capacitance, so that is the problem.

(Refer Slide Time: 00:05:17)

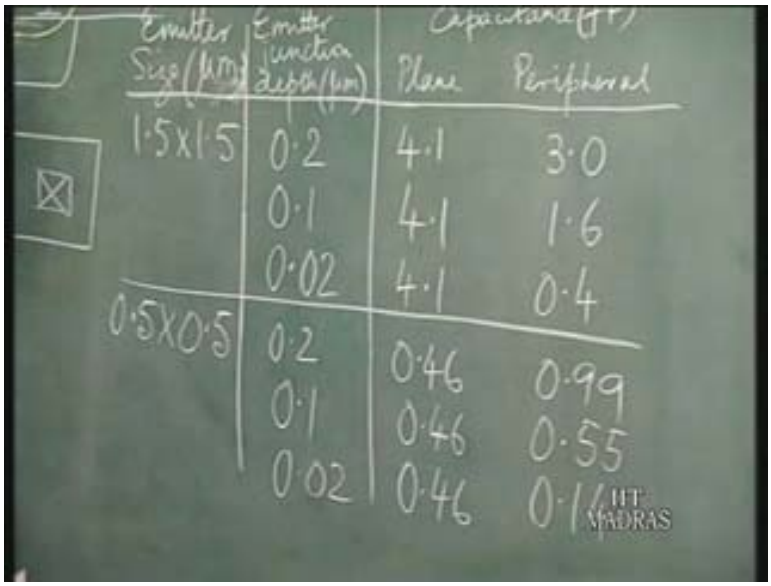


So we have to see how we can reduce this emitter base capacitance as such. One way to reduce the emitter base capacitance is to reduce the area. So if you look from the top that is you have the emitter region like this maybe and then you have the base and you have the emitter contact here and the base contact here. If you reduce this area of the emitter region, you will reduce the capacitance. So we shall see how the emitter base area affects the base emitter capacitances.

Now this capacitance, you can actually break it up into two parts. One is due to the charge stored in this region, this part of the junction which is called the plane region of the junction and there is another part which is also important which is the sides of this junction. This sides we can call it the peripheral region and this region is the plane region. So you have a component of capacitance due to the plane region and you have a component of the capacitance due to the peripheral region. Now the area of the plane region is the area of the mask, this is the emitter area.

Now what is the area of the peripheral region that is going to be the perimeter of this region into the junction depth. So this is the area of these sides. So you have charge storage not only here but also on the sides of this. So the capacitance is partly due to the plane region and partly due to the peripheral region.

(Refer Slide Time: 00:08:32)



I just give you some data and I think from this data, certain points will become very clear, emitter size in microns emitter, junction depth in micron, emitter base capacitance in femto farad. This is the plane component and the peripheral component, some data just going to write here. So emitter size in micron. So if you have a 1.5 into

1.5 emitter size and the emitter base junction depth is 0.2 micron. The plane component is 4.1femto farad and the peripheral component is 3 femto farad.

If you reduce it to 0.1 micron that is the junction depth, this one remains 4.1 because we are not changing the emitter area but this peripheral component reduces to 1.6. If you reduce the junction depth even further to 0.02 micron, this one remains unchanged whereas the peripheral part becomes 0.4 femto farad because the junction depth reduces, the area of the peripheral part is basically reducing.

Now suppose we go for a 0.5 micron into 0.5 micron emitter size and again the junction depth is 0.2 micron. This one has gone down to 0.46 femto farad. Of course the area itself is reduced from here. The peripheral part is also less at 0.99 femto farad. If you go to 0.1 junction depth, this remains the same and the peripheral part is 0.55. For 0.02, again 0.46 is the plane component and 0.14 is the peripheral component. Now if you look at this table what is the message it gives? If you compare this set and this set, you see that what we have done is the emitter size has been reduced as we had said, in order to reduce the capacitance, we have to reduce the emitter size. So emitter size has been reduced.

Correspondingly the plane part of the capacitance has reduced from 4.1 to 0.46. So this has been reduced to one third so the capacitance would be reduced by one ninth area, one third on each side so the capacitance is just proportional to area but you see the peripheral component has not reduced that much. So in fact the peripheral capacitance is much higher than the plane capacitance but if you go on reducing the junction depth then only say if you go for 0.02 micron junction depth that is the emitter junction depth then only you find that the peripheral component becomes less than the plane component because again just to reiterate, the plane component is proportional to a square. If you say one of the sides is a and the peripheral component is proportional to a because it is proportional to the perimeter into the junction depth.

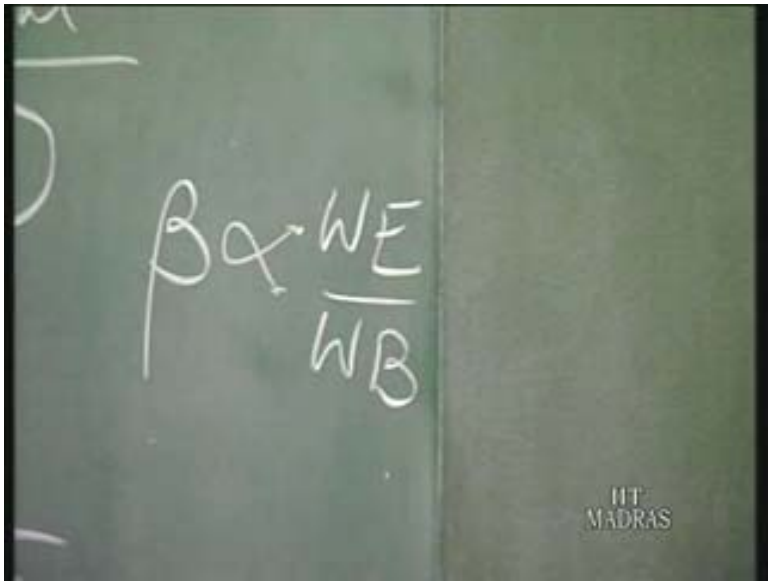
So if you reduce a that is one of the sides, if you call this a, this is a say. one of the sides of the emitter is a. If you reduce a, the perimeter to the area ratio goes up and the peripheral component starts to dominate. So what you have to do is when you go for smaller and smaller dimensions of the emitter, in order to derive the benefit of a lower capacitance you also have to reduce the junction depth otherwise it makes no sense because it is the peripheral part of the capacitance which is starting to dominate. You are just by reducing the area you are not getting enough benefit.

So now you have to reduce the junction depth that is this junction depth has to be reduced. Now what happens if you reduce the junction depth, the emitter junction depth or the emitter width? Now we have seen a relation for the beta of a transistor and we have also seen that the beta of a transistor, among other things is proportional to the ratio of the base width to the emitter width, I mean it's the other way round. for higher beta you require reduced base width. So it's proportional to emitter width by base

width, beta is proportional to WE by WB (Refer Slide Time: 00:11:56) where WE is the emitter width and WB is the base width.

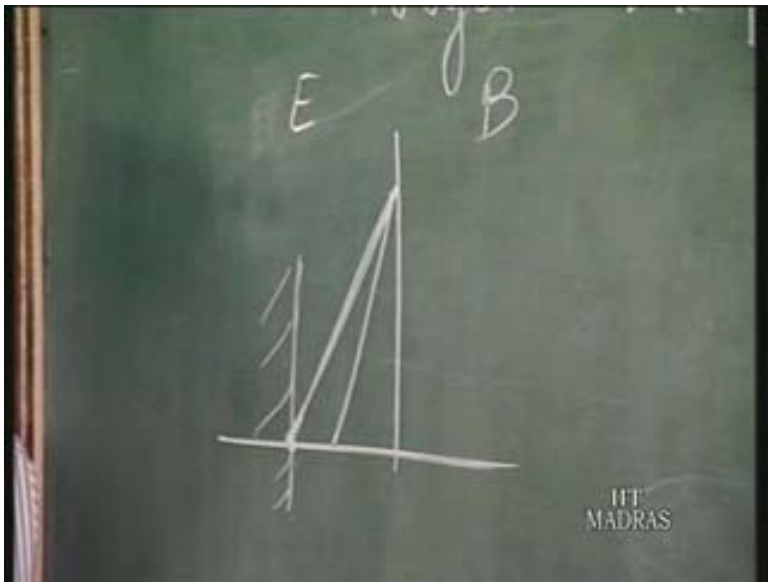
So if you reduce the emitter width in order to reduce the capacitance, the beta is going to get affected. So from the other consideration you cannot reduce the emitter width. So what is the solution? The solution came in the form of a particular type of bipolar transistor which is called the polyemitter bipolar transistor.

(Refer Slide Time: 00:11:56)



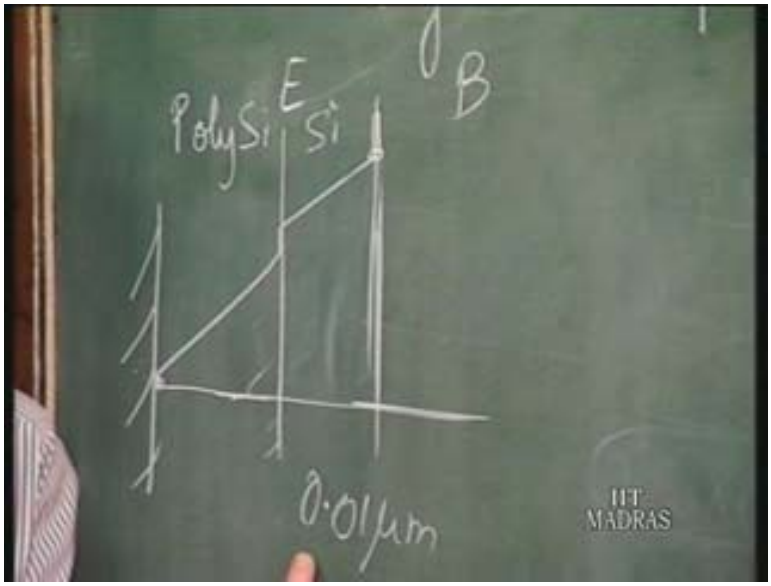
I am just going to discuss what is the polyemitter bipolar transistor. In a polyemitter bipolar transistor the advantage is that you can reduce the emitter junction depth without affecting the beta of the transistor. That is the major advantage and all modern high frequency bipolar transistors are in fact polyemitter bipolar transistors. So what do you do is sometimes it is called PET. Now what do you do is you know that in BJT, if this is the emitter region and this is the base region and this is the minority carrier profile and this is the metal, so this is the metal interface. The minority carrier profile in the emitter region, the minority carrier concentration goes to zero at the metal contact because here the recombination rate is very high and so the excess minority carrier concentration goes to zero.

(Refer Slide Time: 00:13:58)



Now if you reduce the emitter junction depth, this is the emitter junction depth, the slope is going to go up. So basically the injection of holes from the base to the emitter is going to increase and that increases the base current. What we are saying is that in fact for high beta, the emitter thickness must be larger to reduce the base current but that gives rise to problems of peripheral capacitors. What is done is, if this is the silicon part, you have another part which is polysilicon. So you have a silicon polysilicon interface and then you have the metal. If you have such an emitter structure, the entire emitter now consists of two parts one is polysilicon, the other is silicon. The profile is something like this, here you have a small difference because there is some recombination at this interface but the recombination is very small. So in fact the carrier concentration is going to go to zero at the polysilicon metal interface.

(Refer Slide Time: 00:15:56)

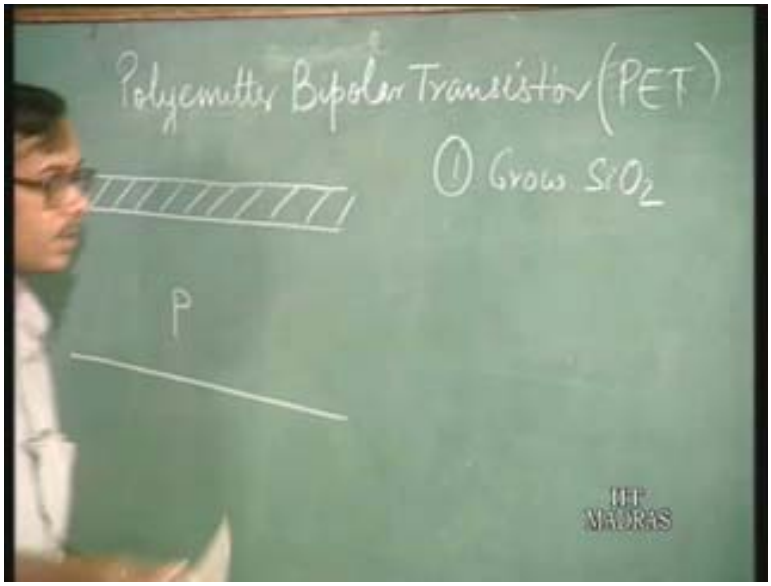


The profile is almost like this, so the total emitter width for calculation of beta is actually the sum of the polysilicon thickness as well as the silicon thickness which can be made quite large whereas when we look at the transistor structure as such the peripheral capacitance is going to be determined by the silicon thickness which can be made very small. In fact it can be made as small as 0.01 micron which for all practical purposes, it's going to eliminate the peripheral component. So you make the silicon part very small thickness, make a very wide polysilicon thickness, so the peripheral capacitance which depends on the silicon thickness is reduced whereas the beta which depends on the emitter thickness, the larger the emitter thickness the larger is going to be beta. So you have a high beta.

In fact in this polyemitter transistors, you can get a very large gains means current gains that is beta because of this property of using polyemitter and in fact again because you get a large, you can have a large WE by WB ratio. Again you can trade it off with the doping concentration ratio and in fact you can also increase the base doping concentration which has other effects like reducing the base resistance and all which is advantages from the circuit point of view because the time constant to charge the base would be reduced. So these are the advantages of a polyemitter transistor.

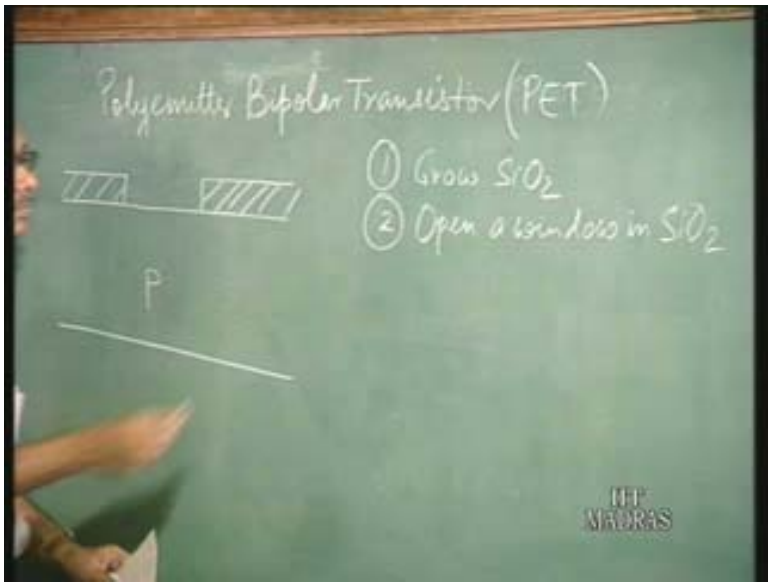
Now we shall see how actually a polyemitter transistor can be fabricated or what does it look like finally. So that is the next thing we shall take up because it is very important especially from the ECL point of view because ECL you know is a high speed circuit and almost all modern ECL transistors are in fact polyemitter transistors to derive the benefit of high speed. I shall take up now the structure of the polyemitter transistor and how one can actually fabricate a polyemitter transistor.

(Refer Slide Time: 00:18:48)



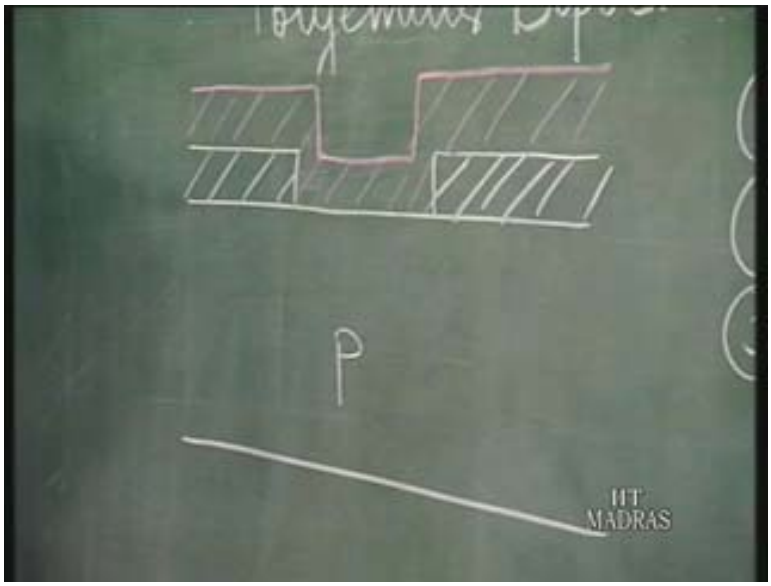
The steps. Suppose this is the p region which is the base of the transistor. Now we want to fabricate the polyemitter on top of this poly pn junction, the emitter junction on top of that. So the first step is grow silicon di oxide which is grown thermally. So you open a window in silicon di oxide using a mask.

(Refer Slide Time: 00:19:24)



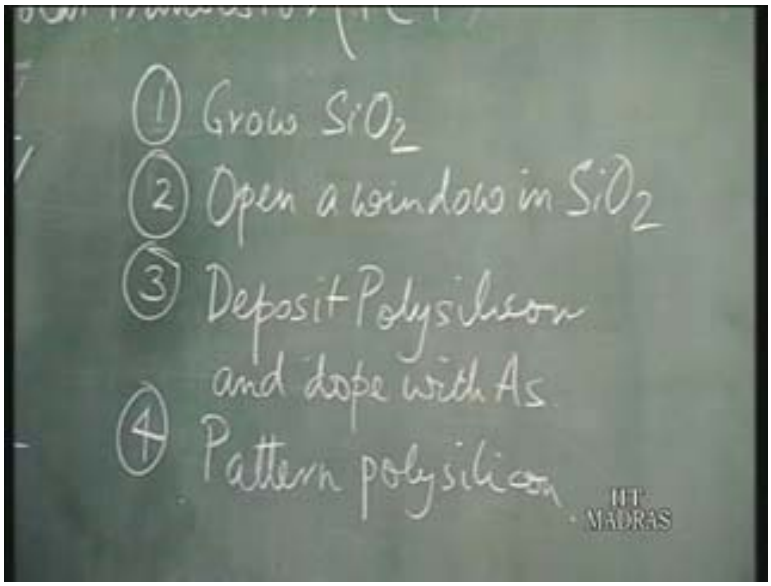
What you do is you open a window in silicon di oxide, we have the structure like this. It is here where you are going to form the emitter base junction.

(Refer Slide Time: 00:20:52)



Then what you do is deposit polysilicon and dope with arsenic. Arsenic is an n type dopant in silicon, it has the property that it has got a very small diffusion coefficient that is, it does not penetrate much into silicon. So if you want to form shallow junctions, you have to use arsenic; with phosphors you cannot form shallow junctions because it diffuses very fast into silicon and you cannot control, the control is not so good enough to form shallow junctions. So arsenic is usually used when you want to make shallow junctions.

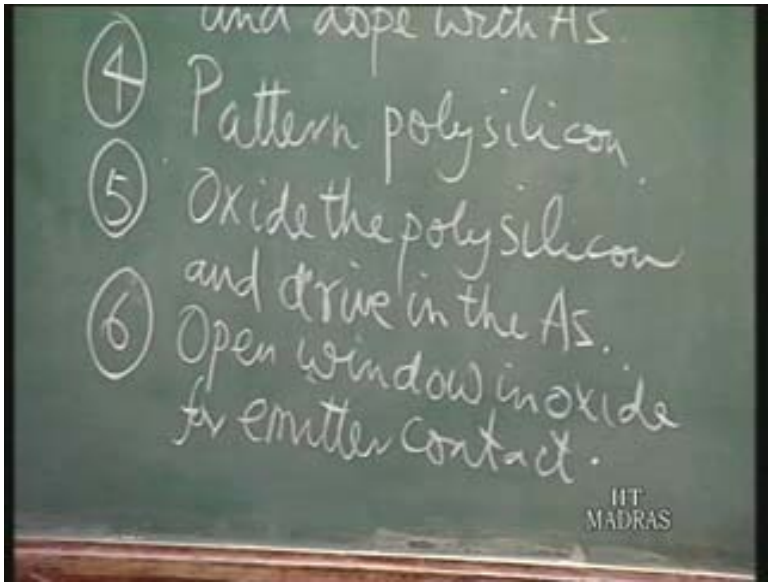
(Refer Slide Time: 00:21:06)



Basically what you do is you deposit polysilicon on top of this and this is doped with arsenic that is you introduce arsenic impurities here from the top. This is actually doped with arsenic so it is n type polysilicon which is heavily doped. The doping concentrations are made very high. Now the next step is you pattern the poly again by a lithographic process. So you are left with some of this, oxidize the polysilicon and drive in the arsenic. So this is done simultaneously.

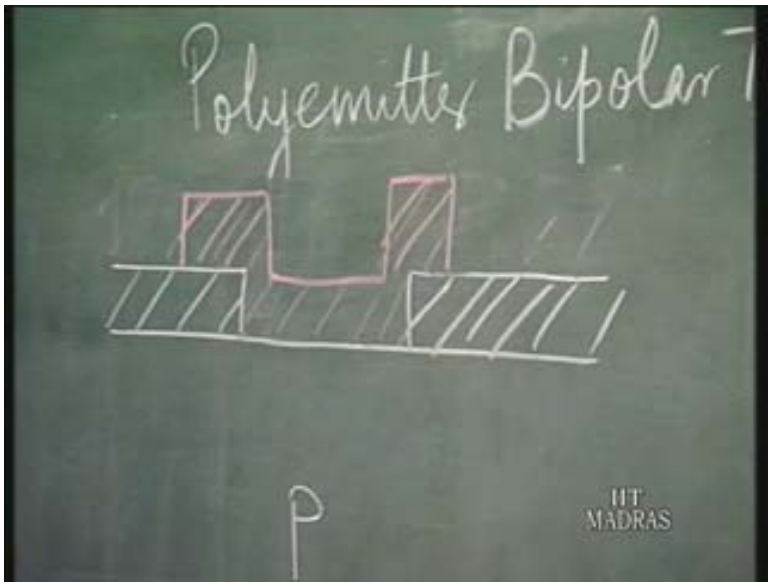
Oxidation of polysilicon is a high temperature process and as you do oxidize the polysilicon. What happens is the arsenic impurities in polysilicon also go into the silicon. They diffuse into the silicon.

(Refer Slide Time: 00:23:06)



What you have is you have another oxide layer on polysilicon. Since we have oxidized it and simultaneously you form an np junction here. This pn junction is formed, this is the n type. I am sure you cannot see this, so I write it this is n type. This is what you have after step 5.

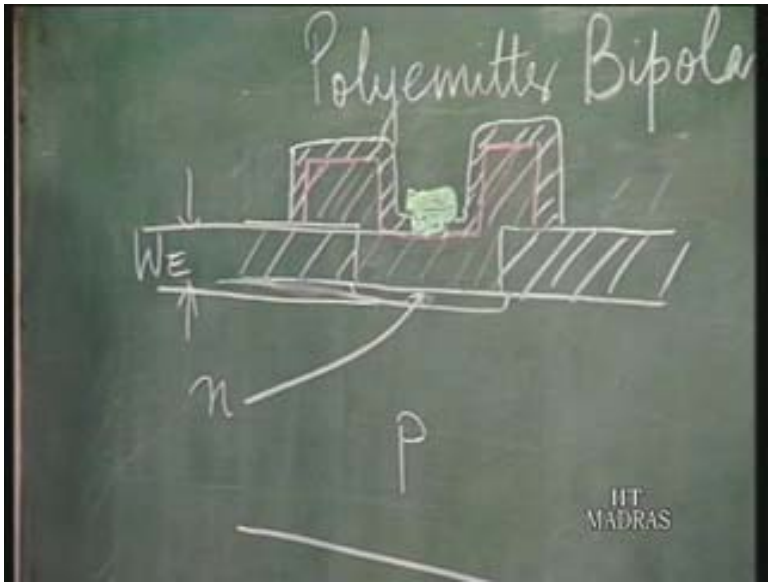
(Refer Slide Time: 00:21:22)



Again then the next step final step open widow in oxide for emitter contact that is you again open a window here in this oxide here and then you deposit metal. So this is the emitter contact.

So you see this is the emitter base junction structure here polyemitter of a polyemitter transistor. So as i said for calculation of beta, the total emitter thickness is going to be taken as this thickness inside the silicon of the n region. The n region starts from the junction here right up to the metal contact here. So the total emitter thickness WE is actually this. This is WE (Refer Slide Time: 24:16). It is the total thickness because the minority carrier concentration in the emitter which has its peak here is going to go to zero at the polysilicon emitter contact whereas the peripheral capacitance you see is due to this part of the junction here.

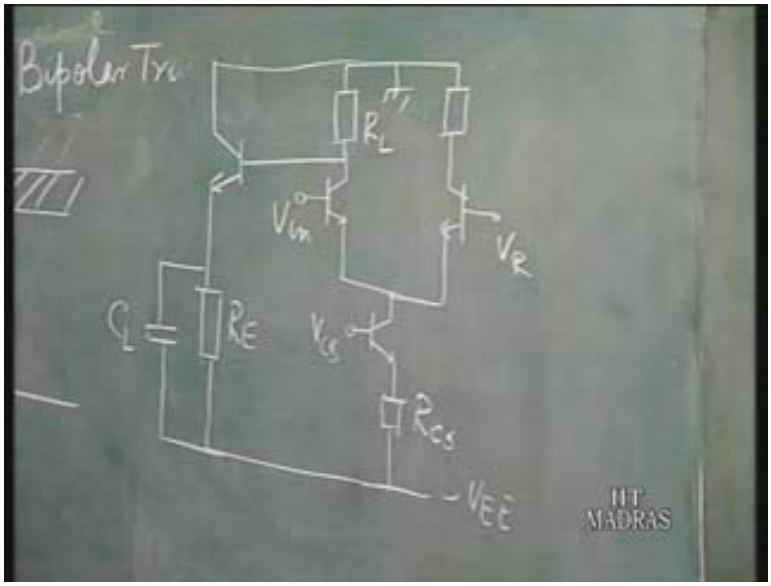
(Refer Slide Time: 00:24:18)



So this you see is very small, the part of the junction inside silicon is very small. The peripheral component of the capacitance is virtually eliminated. So you have high speed transistor as well as a high beta. This is the emitter base structure of a polyemitter transistor that is how you make a polyemitter transistor and as I said that virtually all high speed transistors nowadays, silicon transistors are polyemitter transistors. So we shall go ahead. Now with this background knowledge we shall go ahead and see how to evaluate the delay in ECL gate. I just leave it here.

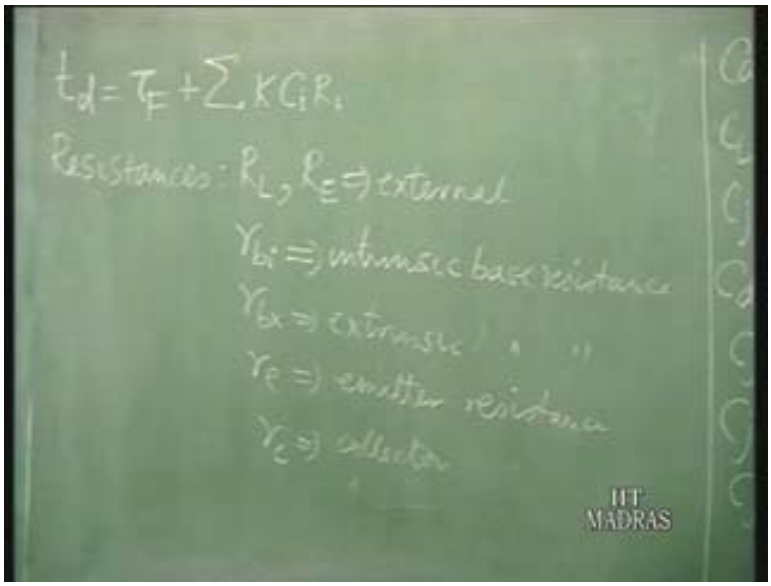
I will just draw the ECL gate once more. We have a current source here say this is the input. This is the reference voltage and this is the emitter follower say I am just taking the nor output. These are the resistances, ground and this is minus V_{EE}. So this is actually R_E say emitter follower resistance you call it C_L, this resistance you call it R_L. So you have this R_{CS} which is the current source resistance. The delay that actually calculated in ECL gate is like this.

(Refer Slide Time: 00:27:14)



The total delay t_d is taken to be the sum of the forward transient time of the transistor tau_F say plus a weighted sum of all the capacitances and resistances in the circuit. So that is the total delay, the sum of the forward transient time of the transistor plus a weighted sum of all the time constants in the circuit, all R_C time constants in the circuit. Now what are the different capacitances and resistances in the circuit?

(Refer Slide Time: 00:32:02)

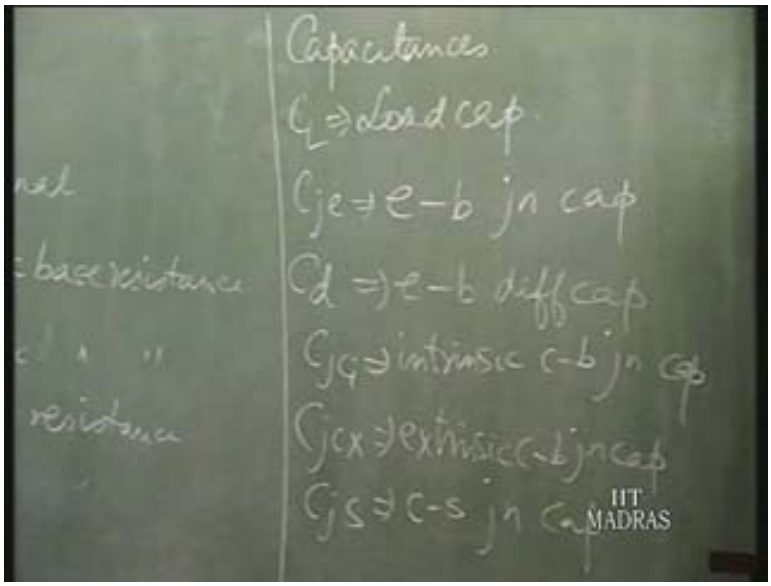


Now if you take up the resistances, you have the external resistances. If you look at the circuit R_L and R_E, these are the external resistances in the circuit and then you have the internal resistances in the transistor.

They are usually written as r_{bi} that is the intrinsic base resistance, I will explain what it is, r_{bx} is the extrinsic base resistance. Then you have the emitter resistance and the r_c is the collector resistance. So you have 6 different types of resistances in the circuit intrinsic base resistance, extrinsic base resistance, emitter resistance, collector resistance. These are all internal to the transistor and then you have the external resistances R_L and R_E in the circuit. Now capacitances, which are the capacitances you have in the circuit. You again have a number of capacitances in the circuit. I shall write here.

You have C_L the load capacitance which is the external capacitance, which is the load capacitance and then you have the internal capacitances. you have c_{je} which is the junction capacitances of the emitter base junction, I will write e-b junction capacitance, cd which is the e-b diffusion capacitances and then you have the c_{jc} is the other junction that is the collector base junction, c_{jci} which is the intrinsic collector base junction capacitance.

(Refer Slide Time: 00:34:53)



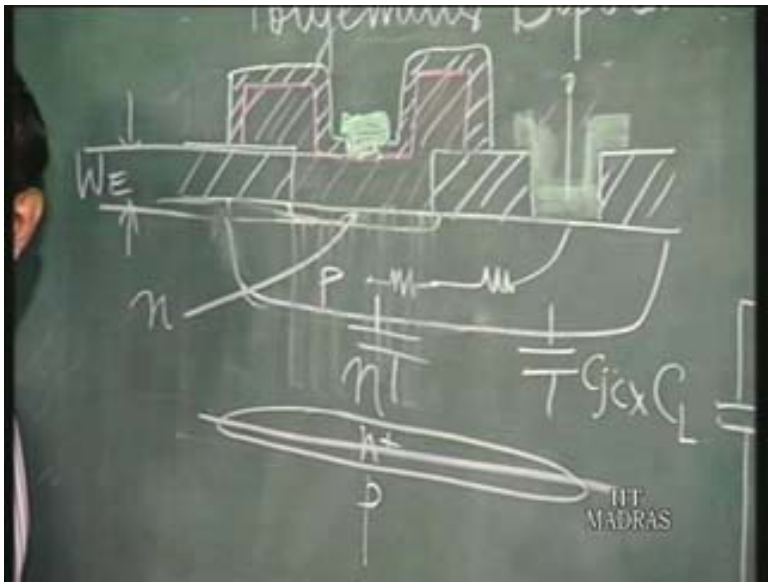
Then you have c_{jcx} which is the extrinsic collector base junction capacitance and then you have one more c_{js} which is the collector to substrate junction capacitance. So again you have 6 capacitances. In fact here you have 37 terms. So what people actually have done is evaluated the effect of all these capacitance and in fact developed waiting factors for all these capacitances in order to obtain the delays.

Now just before we go ahead what are these components we will just have a look by taking up the transistor structure as such. We will go back to this transistor structure here.

So this is the emitter base junction. Suppose you have the base region like this then you have the base contact somewhere here. So you will have the base metal here, so this is the base contact.

Now if you look at the values of the resistances R_L and R_E are external resistances. Now you have r_{bi} and r_{bx}, r_{bi} is the intrinsic base resistance which is usually this part of the resistance that is intrinsic. So this is the intrinsic transistor region that is this is the region where you have a transistor action. This part of the device is called the intrinsic region.

(Refer Slide Time: 00:35:40)



Here you have the transistor action taking place. So this is the intrinsic base resistance and then you have an extrinsic base resistance that is from the contact right up to the intrinsic device region. Then you have emitter resistance, so here this is the p region for the base, we have the n region and then also you will have the substrate which is again a p. You will have a collector resistance of course you will have a buried collector n plus here as in a bipolar resistor but there is going to be again a resistance for the collector resistance from the intrinsic part of the device to the contact and also you have an emitter resistance also. So these are the internal to the device. If you look at the capacitances, you are aware that one is the load capacitance the external capacitance, you have the emitter base junction, you have the two types of capacitance one is the junction capacitance and the other is the diffusion capacitance. One is due to the depletion region and other is due to the stored charge.

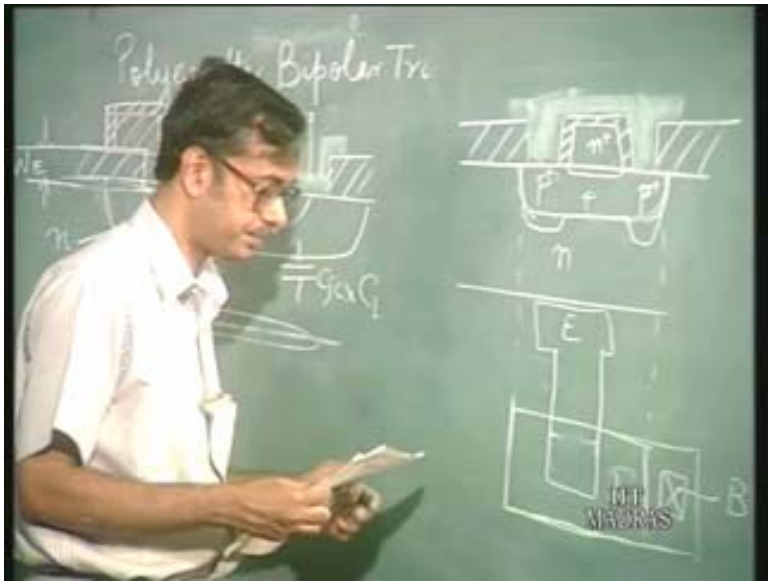
Now c_{jci} and c_{jcx}; c_{jci} is the capacitance for the intrinsic part of the junction. Again this is c_{jci} will be the capacitance for these parts of the device, this is this capacitance and

c_{jcx} is basically the extrinsic capacitance, this is the extra region you have a junction here (Refer Slide Time: 35:30). The junction goes all the way, base collector junction. So this is extrinsic, this is actually not taking part in the transistor operation, transistor action takes place here but you cannot avoid it in this region so you have all the capacitances there, you will have the junction. This is reversed bias junction and so you have charge here.

So that is the extrinsic capacitance and again you have another capacitance from the collector to the substrate. So these are all the capacitances. So one has to see how to reduce these capacitances or as well as the resistances in order to improve the speed performance or to reduce the delay. Again what it is done is you must have a self-aligned structure to reduce the values of the parasitic components. All the parasitic components that is which are not actually taking part in transistor operation. For example the extrinsic components but which has an effect on the performance.

So just to give you an idea, there are many complicated self-aligned processes. I am just going to give you show you a simple process something which should not be much difficult to follow. So basically you have the n region here then this is the oxide. What you can do is you diffuse the p region which is the base region, you diffuse the base region of course this junction depths shows very big p diffusion but these are not to scale anyway.

(Refer Slide Time: 00:42:25)



So then what you do is once you have the p region, you deposit polysilicon all over and oxidize it and etch back. This is something similar to what you have done in the i squared l process if you recall. Finally you have n plus polysilicon here say and you

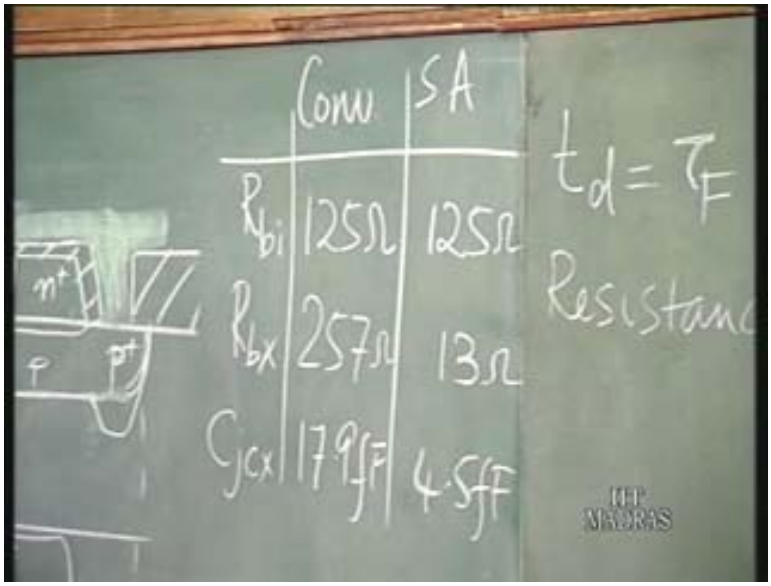
have oxide around it, this is polysilicon, n plus polysilicon is deposited. You oxidize the polysilicon and you etch back, it will leave us the oxide around it because of the fact that polysilicon oxidizes at a faster rate than silicon. So all the oxide on the silicon is going to be etched back leaving behind oxide on the polysilicon and then what you do is you again do a p diffusion and this diffusion is going into the extrinsic regions. So you make these regions p plus. Why are you doing this? To reduce the extrinsic base resistances. So this region becomes p plus so basically the structure becomes something like this. We have p plus regions and then you can deposit metal right away or if you can also deposit p plus polysilicon, it will take contact from the base.

So you can have the metal going around like this, you have the base contact metal going around like this. Now if you look from the top, what does the structure look like? This is the base region then you have the emitter region somewhere here. Now the emitter poly actually goes like this from the top and you take the emitter contact somewhere here say. The emitter poly which is shown here, it goes like this in this way. This is the top view and you take the emitter contact here and all this emitter poly is covered with oxide.

On top of that you have the base metal or polysilicon and that you can take a contact somewhere here or even you can take it out this way. This is the base metal, you can even take it out and have an contact somewhere here. So you have the emitter metal running this way, the base metal running this way and because of the oxide covering the emitter poly, there is no connection. Once you do a drive in, one step I missed out.

So this is the n plus poly, so once you do the drive in you have a pn junction in the silicon. This is basically a polyemitter transistor. This is a self-align transistor and in a self-align transistor, the parasitic components will be very much reduced. So just to give you an idea of the values.

(Refer Slide Time: 00:43:25)



See r_{bi} for example. This is conventional, this is self-aligned say for a particular process this is 125 ohms. Then r_{bx}, this is say 257 ohms this can be as low as 13 ohms. This has been published for a particular process, I am just giving you the results. The other interesting thing is c_{jcx}, so it is 17.9 femto farad, this goes down to 4.5 femto farad. Because the extrinsic part of the junction, this junction here if you look back at this circuit, there is a large extrinsic part for taking the base contact. Here because of the self-aligned nature, the emitter and this base are self-aligned and the base contact is just separated from the emitter region by this thickness of the oxide here which is just about maybe 0.4 or 0.3 microns.

So you see that the extrinsic region has been cut down drastically and so high speed transistor must have a self-aligned structure so that the extrinsic regions are reduced. The extrinsic capacitances are reduced which gives rise to lower delays. In fact a self-aligned transistor is going to be much faster compared to conventional transistor. Now I shall just take up one more issue here, I shall also discuss the hetero junction bipolar transistor but maybe before I go into that I will just like to discuss something about the delays.

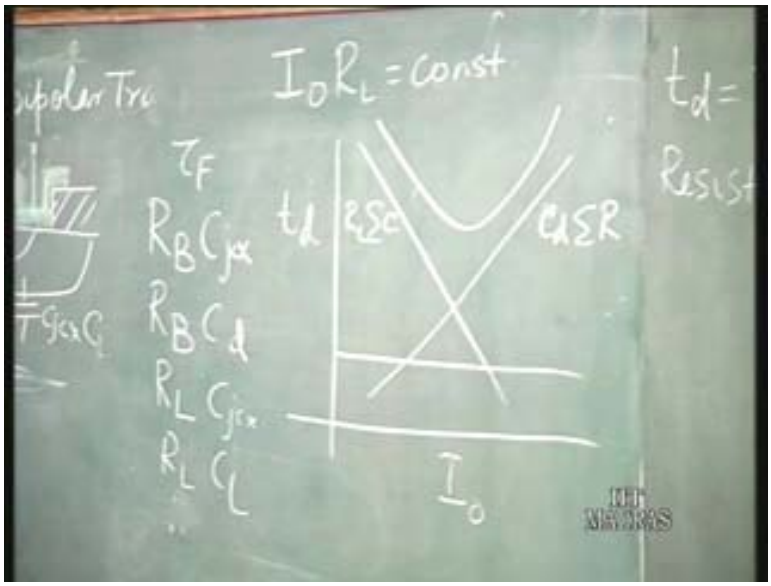
We have always said in our discussion that we talked of I_0 into R is equal to, we have said v_{on} sometimes we said I_0 into R is 500 millivolts because I_0 into R in a ECL gate that is a current source value into the load resistance that is the logic swing. That is the amount by which the output voltage changes and we have never said how to choose I_0 a and r. What is important is the drop across the collector resistance. I mean is there anyway or you can just choose any value of I_0 and any value of r.

Now if you look into this expression here there are many components of delay, many time constants. There is the tau_f and there are 36 other components in that summation sign. Now out of that the major components has been seen consists of R_B C_{jx} R_B C_d R_L

C_{jcx}. This is c_{jcx} also, R_L C_L then of course R_B C_D is already there then of course tau_f. So these are some of the major components which give rise to delay. There are other time constants but there contribution is less.

Now if you look at these values, you see that suppose you are designing for a particular logic swing. That is in the ECL gate, the I_0 into R is say a particular value. Suppose you say that I_0 into R should be 0.4 volts that is the drop across the resistance when the full current is flowing in a branch is 0.4 volts. Say that is I_0 into R say R_L is a constant. Now if I plot the delay versus I_0 that is basically what I want to say is does the delay depend on the value of the current source that is what is the current which you are using in the ECL gate.

(Refer Slide Time: 00:50:15)



So if you look at these components you will find the tau_f is independent of I_0. It is going to be something like this a constant and then you have another component say C_D. Does C_D depend on I_0? It is dependent on I_0, so C_D is proportional to I_0. So all the components which have cd that is we should say cd to all r, all the components, C_D here it shows C_D R_B that means basically R_B x R_B intrinsic and also the other components all these components will increase with I_0 and then you have the other components R_L which has R_L in them. If you increase I_0 what happens to R_L? I_0 into R_L is a constant, so R_L is going to reduce basically. All the components which have R_L in them they are going to reduce.

These components which have R_L into all the c's, all the capacitances so these components are going to reduce. What you get? You see that the total delay is going to be the sum of all these delays. The delay is going to be something like this which basically means what we want to say is that if you want to keep the logic swing constant

that is I_0 into R_L constant then there is a particular value of I_0 which is going to give you a minimum delay.

You have to bias the ECL gate for that particular I_0 to get the minimum delay and then you choose R_L accordingly so that I_0 into R_L is equal to the logic swing. So if you are chosen 0.4 as your logic swing, you can have all sorts of combinations of I_0 and R_L to get 0.4 volts. Now what you are going to choose as the value of I_0 and what the value of R_L? So you have to see at which I_0 the delay is minimum. So depending on that you bias, you choose the value of I_0 and accordingly the value of R_L is going to be fixed. The delay is not independent of I_0 but there is a minima and the ECL gate must be biased at that I_0 to get the minimum delay. We shall stop here today.