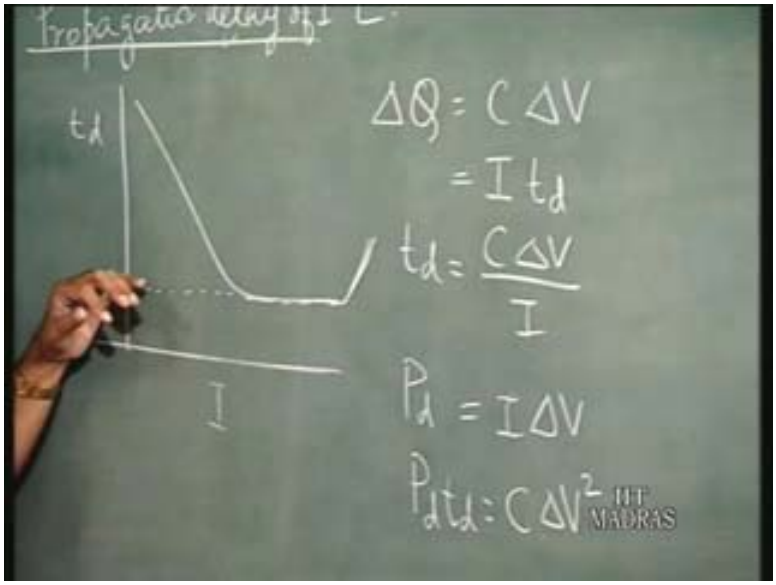
Digital Integrated Circuits
Dr. Amitava Das Gupta
Department of Electrical Engineering
Indian Institute of Technology, Madras
Lecture – 15
I2L – Propagation delay Self aligned Technology

Today we shall continue the discussion on I squared L logic circuits. Last class we saw the noise margin problems of I squared L. See I squared L was in the 1970's when it was first talked about, it was very popular in the sense that people thought it would be very popular logic circuit but it does not happen that way and one of the reasons of course is poor noise margin but the other major reason is that the propagation delays have not really been very low. In those days, may be 10 nanoseconds was a very acceptable propagation delay but no longer these days and it has been very difficult with the original structure to reduce the propagation delays. One can understand that because of the very large amounts of charge stored in the device.

So in today's discussion we shall take up the propagation delay problem and see how attempts have been made to actually reduce the propagation delays mostly with the help of improved technology. The propagation delay of I squared L; now the I squared L delay just like as a delay in a bipolar junction transistor depends on the capacitances and we know that there are two types of capacitances, one is the junction capacitances and the other is the stored chart capacitance or the diffusion capacitance. So initially when the current levels are low, it is the junction capacitance which dominates. Whereas at high current levels the stored chart capacitance dominates because the stored chart capacitance is proportional to current, you know that.

So as we go for higher and higher currents, the stored chart capacitance increases and that's starts to dominate. At low current levels it is the junction capacitance which dominates and we have already seen in a bipolar transistor that if you take the current and the delay, so if you increase the current initially the delay falls and then it saturate. That is because if you just take capacitance and find out how long it takes to charge or discharge, you know that the basic relation is if you have a capacitance and Q is equal to CV. So if you have to charge it through delta V voltage, basically you have to supply a charge of delta Q and if you are charging with the constants current say, what is Q? What is current? Current is the rate of flow of charge. So if you are charging with the constant current I, so it will take a time say t_d. So I into t_d is that delta Q, if you are passing current I through t_d, you require this. So the t_d is the time required for the voltage across the capacitance to change by delta V and then from this you get t_d is equal to C delta V by I that is very well known. That is if you have a capacitance C charging through delta V with a constant current I, you required time t_d.

(Refer Slide Time 10:19 min)



So which means that if you are increasing the current, it will take less time to charge. Obviously because basically that this two products is equal to the charge required and so if I is more t_d will be less. So this gives this part of the characteristic. That is if you are increasing I, it is going to fall but when the other capacitance, the diffusion capacitance starts to dominate, you will find that the diffusion capacitance itself is proportional to current. The junction capacitance is independent of current. So you get this type of relation but the diffusion capacitance is itself proportional to current that CD, so in this relation t_d will become independent of current, if C is proportional to current. So when the diffusion capacitance starts to dominate, it becomes the constant. So this becomes independent of current.

So what do you have is with current initially the capacitance the delays starts to fall with increasing current and then it becomes the constant. That is if you go on increasing the current further, there is no further reduction in the delay. So this is the minimum delay which you have and it is determined by the diffusion capacitance or the stored charge effects in the device.

Similarly also in an I squared L, if you plot the delay versus current that is the total injector current or the current you are feeding through the power supply, if you are increasing that current you will get initially a reduction in delay and then it saturates. So normally you would operate the I squared L circuit in somewhere along this region. That is once you go to the minimum delay region, of course there is no point in increasing the current any further because what you are doing is you are not gaining in terms of speed but you are increasing the power

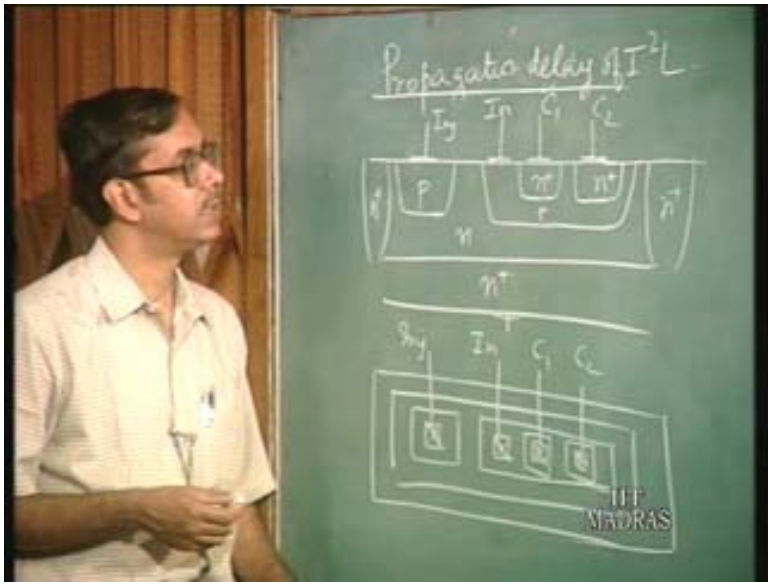
dissipation. Of course you have also seen that increasing the current, increases noise margin so there are other trades also.

In fact if you go on increasing the current too much, the delay goes up. That is because of some other effects called high injection effects in bipolar transistors which degrades the device operation. So basically you have this type of curve for td versus I and if you want the power delay product, power dissipation is given by I, if you want to say into nearly equal to delta V. That is basically if you assume that in saturation it goes very close to zero volts, the transistor then the power delay product is equal to and you are operating in this region where the junction capacitance is dominate, you get something like C delta V square.

Now this delta V is very small quantity in I squared L and power delay products are as low as 0.1 Pico joules and it can be very easily achieved. So that is one plus point of I squared L, the power delay product in I squared L is quite low but the problem is the delays are quite high. This minimum delay which you have from the normal structure is around 10 nanoseconds, it's very difficult to go below that because of the excessive charge stored in the device.

So we shall look at it, go back to the I squared L structure and now look at some modifications of the structure as such which has taken place from the technological point of view in order to improve this aspect that is propagation delay of the I squared L cell. So I will just draw the I squared L cell once more and so this is n region, n plus and then you have the collectors, this is the base, this is the injector and then you have the n plus guard ring (Refer Slide Time: 11:25). So you have this in the injector contact, this is the base contact, this is the collector contact and then you have the emitter contact from below. So injector, input, collector$_{1}$, collector$_{2}$ this is the well-known structure which we have been discussing. Again just to draw the top view, you have the injector then you have the base and you have the collectors and these are the contacts. So this is the injector contact, this is the input contact, these are the collector contacts.

(Refer Slide Time 12:42 min)



Now as I was saying that the problem with I squared L is see, there are large number of areas where you can have charge storage. For example the emitter itself is lightly doped with the result that you can have large charged storage in the emitter itself. Once it is lightly doped, you can have large minority carrier storage, all through this region. So when this base emitter junction of the NPN transistor becomes forward biased, you have stored charge in the emitter. Normally in the normal bipolar transistor, we don't talk of stored charge in the emitter because it's very heavily doped, N_I squared by N, its proportional to that. So that is going to be low but they are going to have large charge stored here. Also in the base regions, you have a lot of extrinsic base area which means that is the area which is not really a part of the transistor.

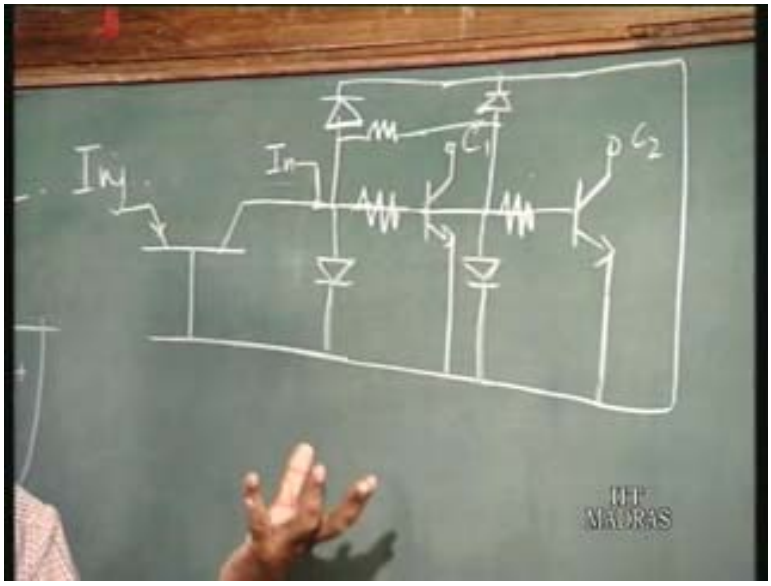
See we call this as the intrinsic carrier. See this is the NPN transistor where the transistor action is taking place, this is the intrinsic area this is the useful area of the device. The other regions here, these are the extrinsic regions, this is the useful area again for this NPN transistor but for the other regions these are the extrinsic areas, these are parasitic areas which we cannot avoid but you have a lot of stored charge there. Now if you look at the structure not only in between the transistors, you have all around the transistors. That is the top and bottom side of this NPN transistor, this is here. You have a lot of extrinsic area and you have charge stored here automatically when the device goes to forward bias, the junction goes to forward bias and so when you have to switch off, you have to remove that excess chart and it gives rise to a lot of delay. So that is the basic problem in the I squared L structure which gives rise to the poor performance in terms of propagation delay.

In fact if you draw the equivalent circuit of this particular structure, the equivalent circuit would look like this. You have PNP transistor and then this goes to the input contact, so this is the injector contact. When this is the input side then this goes to the base of this NPN transistors. Now here you have series resistance, this is one transistor say then you have another transistor but again you have series resistance. So this is collector$_{1.}$, this is collector$_{2.}$, this is the input not only that, you have parasitic diode sitting here. These diodes see in these

regions, this is the PN junction diode. So this is the PN junction diode here, it not only have series resistance say this is the series resistance from here to here.

So this resistance from here to here but you also have PN junction here. So whatever current flows in here, part of it is getting wasted because of the PN junction here and also you have series resistances and also you have what is called the rail region. So this region just under the device but you also have the rail region that is this sides, I just mark with the color chalk that these regions, this side and this side the top and bottom which is actually not contributing anything. So we can also draw the equivalent circuit like this, another set of diodes and this actually this is sorted here. So this is the contribution of the rail, the rails are the sides. So this is the equivalent circuit of this cell.

(Refer Slide Time 18:39 min)

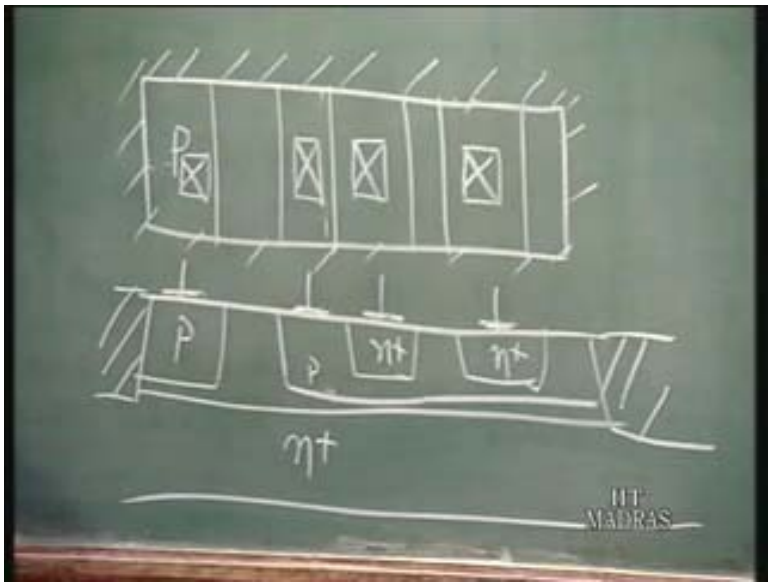


Now you see that because of this lot of parasitic elements there is going to be lot of this similarity between this device and this device. Normally even when we have obtained the

condition for the proper operation of the I squared L cell, what we did was we assume that when we have a multi collector transistor, we assume that there are so many transistors in parallel and all the transistors are similar but you see that this transistor is not going to behave in a similar fashion as this transistor. It is going to have less base current compared to this transistor. It should work through our analysis but it may not work because this transistor may go to saturation, this C one may go to saturation but the other transistor may not go to saturation. So this could be the problem. So what do you do? So we have to look at how to eliminate this parasitic elements.

Now the first thing which one can do is in fact came about with proper isolation techniques that is you use oxide isolation. That is silicon dioxide is used nowadays in semiconductor technology for isolation that is isolating between individual devices or individual different areas on this semi-conductor chip. That is one can remove this rail portion that is basically you have a structure like this. I am just drawing the top view, I think it will be clear. You have the P, so this is the PNP region and these are the collectors. This is one collector, this is another collector, this is the base contact. So the entire region is P and this is the oxide region, this is oxide all round. So basically if I draw the cross section across this region, you have silicon dioxide, so this is n, this is n plus wafer, you have P which goes all the way, this is the base region, this is the collector and this is the other collector. So this is the device, you have one more contact here that is the injector contact.

(Refer Slide Time 21:44 min)



This is the I squared L structure, so this is the oxide regions isolating. It goes all around so this is the base, these are the two collectors. It is almost identical to this except what we have done here is if I just modify this, we have removed this portion and just made this oxide and it goes all the way here (Refer Slide Time: 22:10), that is the same as this. Not only on this side but also on the other side, this side also what we have done is we have all this region, we have reduced the area up to here and all around you have oxide. So we have removed a lot of this extrinsic carrier, these are the contacts. So that is going to help in one way that is most large portion of the extrinsic carrier has been removed, this side which you are having on the basically the rails. Rails have been totally removed.

If in the equivalent circuit what you will see is this portion has been totally removed. So we have to sort of try to improve the performance but this in fact gives rise to some extra problems that is because if you look at the structure original structure; now the resistance so for example the series resistance for this particular device say C_2, the current has to flow all the way here. There is a large series resistance and if you have a transistor like this... Which region is the series resistance going to be more? It is going to be under the collector diffusions, resistance is proportional to cross sectional area.
So here you see in this region basically the base width, this is the base width of the transistor although it is not drawn to scale.

So this is going to be very small, this going to be a very narrow pinched region and the current has to flow under this to reach this transistor. So this resistance, this portion is going to be very high because of the very narrow region here. Here of course again since the cross sectional area is large, the resistance is smaller. If you look at the rail region which we have eliminated, in the rail regions the current did not have to flow under any of these collector diffusions, in the rail region which we have eliminated.

So the rail resistance which we have in fact removed here. That resistance is much less compared to these resistances because here these resistances, the current has to flow under the collector diffusion areas where the cross sectional area is much less. In the rail region you do not have any collector diffusion so the cross sectional areas are much larger. So these resistances were much lower resistances so in fact if they were there the effective resistance for this is going to be less.

So basically we have removed the low resistance path so that is the problem. Again this did not really work out, so finally what came about was a self-aligned I squared L process which we shall go through now which in fact removed most of these problems. So I shall go through the process technology for the self-align process I squared L process and we shall then understand how it improved the device performance. Again in that self-aligned process one uses oxide isolation. So the first step is you define the active area which means where you want to make the I squared L cell and in the surrounding regions you have oxide isolator, silicon dioxide to isolate. In fact you may have cells, other cells also side by side, other active regions which are separated by these silicon dioxide. So this is one active area in which you are going to make

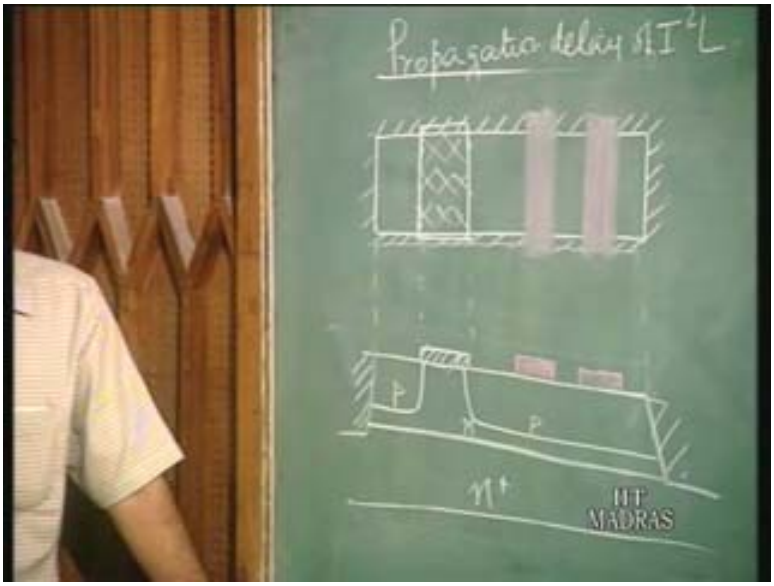
one cells. So this is the top view, so if you look at the cross section you have again the starting wafer is n on n plus and this is silicon dioxide surrounding the active region.

The next step what we do is you grow silicon dioxide on top and etch out then leaving as certain region here which is going to separate the P diffusion. So basically if you look at the mask on this, you have a region like this. This is looking from the top, mask for P plus diffusion and then what do you do is you do boron diffusion or basically implantation is done, boron implantation through this mask. So here silicon dioxide protects this region when you do an implantation. So you have P region here and another P region here, this is the exposed regions. So this is the P and this is the P, so you can identify that this is for the injector P diffusion and this is the base region for the NPN transistors, so this is done.

I will also write down the steps. So first is active area definition then injector and base diffusion or implantation. This is done simultaneously, both are P region. Then what do you have to do is we have to create the collectors. For that the technique which is used is poly silicon which is poly crystalline silicon, this material is very interesting in the sense that it can be deposited on silicon at relatively low temperatures. Relatively low means around 600 degree centigrade whereas most other processors usually take place of 1000 degree, so even 600 degrees is considered low.

So you can deposit poly silicon and the advantage of poly silicon is that you can dope it also, just like silicon and if it is very heavily doped with impurities it behaves almost like a metal. That is you know that in silicon if the doping concentration is high, the resistivity is low. So if you make the doping concentration, the impurity concentration very high, the resistance becomes very low. It starts to behave almost like a metal. So what is done is poly silicon is deposited, actually it is doped with arsenic which is the n type impurity. Poly silicon is deposited and again patterned and you have a pattern like this, I shall draw it here. So may be these are the poly silicon regions and from the mask I just draw it like this, top view it looks something like this. This is the top view and this is how it looks.

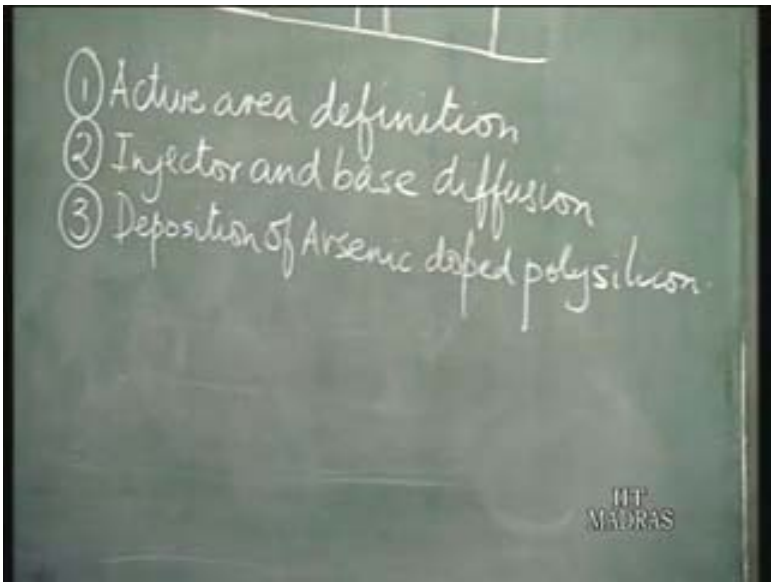
(Refer Slide Time: 32:48)



So this is arsenic doped poly silicon which is n type, deposition of arsenic doped poly silicon and this is followed by what is called a driving. That is this poly silicon consists of a lot of arsenic which is an n type impurity. Now if you subject it to high temperature, this arsenic is going to diffuse into the silicon. Diffusion is the process by which impurities move from region a higher concentration to region of lower concentration in this case. So this is full of arsenic so it is going to move into the silicon and create an n region here, just it is going to be self-aligned in the sense it is going to be created just below the arsenic regions.

Now the next step what is done is oxidation. They oxidize, silicon plus in presence of oxygen it is going to be converted to silicon dioxide. So you create another region of whatever the exposed silicon is there, it is going to be converted to silicon dioxide. Poly silicon is also going to be converted to silicon dioxide, it is the process of oxidation but interesting property is that poly silicon oxidation rate is much higher than that of monocrystalline silicon. So the oxide thickness on poly silicon is going to be much greater than on monocrystalline silicon. So you have say an oxide here but here it is going to be much thicker and then you have something like this, maybe here it is going to be much thicker. So you have a silicon dioxide all over, here also you have.

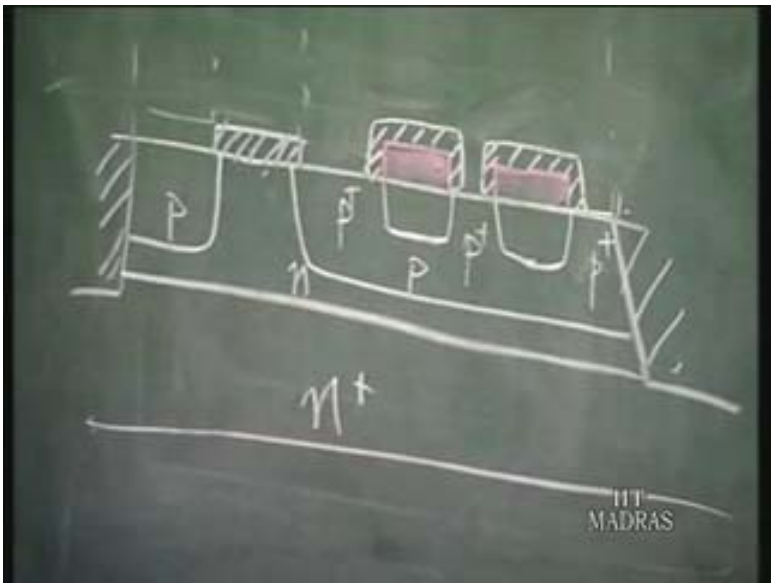
(Refer Slide Time: 33:17)



So what you do is you oxidize and then etch back the oxide. You oxidize plus etch back then what is the advantage of this is; you are doing oxidation and then you are etching back part of that oxidize and you control that amount which you are etching back. Since the oxide thickness on poly silicon is more than that on other regions. Once etch back, you will be left with some oxide covering the poly silicon whereas in the other regions you have removed the extra oxide. You will be finally left with oxide covering for these poly silicon regions and that is very important for this process. So what you have done is grown oxide and etch back but you still retain some oxide on the poly silicon, so you have oxide here and here and here. Silicon dioxide is an insulator that is the important property of silicon dioxide, you should utilize it. Basically you did not require any mask here for this process.

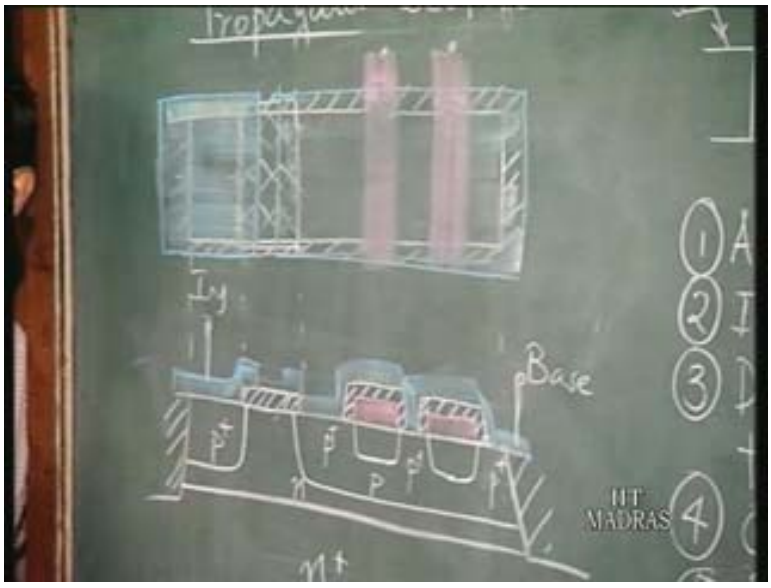
Now the next step is what you do is again you do a boron implantation. Boron is a P type impurity. So once you do that, on the exposed regions what do you do is you make it P plus, here the exposed region is made P plus, heavily doped regions. The exposed region which you have now are the extrinsic regions. You see here the exposed regions, here it is going to become P plus also. Again you don't require a mask. Whatever you expose, you just doing a P plus implantation, here it is covered with silicon dioxide nothing is happening. So only the exposed region you are covering, making a P plus. So what is the idea? The idea is number one, you are going to reduce the series resistance. These extrinsic regions are not taking part in any device action or device operation. So we must have to reduce the series resistance. Also if you have a P plus region, the stored chart in those regions is going to be less. In the extrinsic region the stored charges must be reduced and in P plus the stored charge is going to be less because again the stored charges proportional to N_L squared by N. So automatically this region becomes P plus whereas under the collector it is still P. Only in these regions and then what do you do is you do the metallization.

(Refer Slide Time: 37:12)



Now you have another mask. I think this will be more visible. So this is one region and another mask right from here up to here. This actually goes out from here. So this entire region you deposit the metal. So what happens? In the cross sectional view, you have metal like this and then you have metal going from here all the way. So this becomes the injector contact.

(Refer Slide Time 40:51 min)



The base contact you can take from somewhere here maybe and the individual collector contact is actually taken from somewhere here, it goes there, along this you cannot show because it is covered. So this is the collectors, it goes this way from the top view you can take the contact from somewhere here. Now what you have achieved? See this is the collector, this poly silicon itself is acting as the contact, much in the same way as in the MOSFET technology you have poly gates. You have the poly silicon acting as the gate material as well as for metal lines. So this is the collector contact, they can through the poly silicon it goes this way and this is the base contact. Now what is happening is see this is the base, the base contact is here. It comes and it makes contact here also but because of the intermediate oxide region the collector and the base are not shorted.

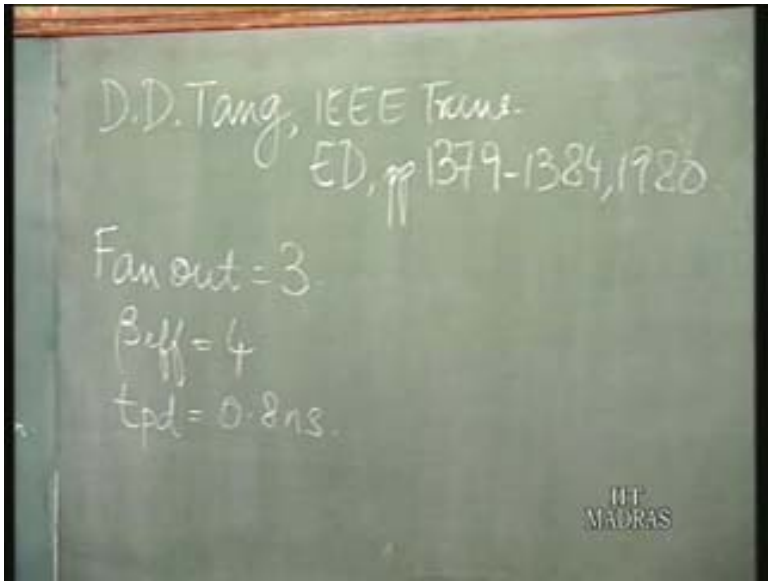
The collector has a covering of silicon dioxide which acts as an insulator. Now what you have done is you have basically shorted the base here as well as here as well as here and you have intermediate P plus region. So all this regions, these resistances are basically shorted out because you have a contact here and in the equivalent circuit here this and this has been shorted. This has been shorted, all the base regions even if you have three or four collectors now all the regions will be sorted and this is P plus. So the series resistance to this region is going to be very small because of the P plus in the extrinsic region. So this is the final structure you have. So you have the injector contact, the base contact and the collector contact taking out.

This is the self-aligned process where self-aligning means you have done a lot of processors without actually happening to align because all these regions the extrinsic regions and the collector they all aligned by themselves with respect to each other, almost automatically when you have to do an alignment you always have to maintain a tolerance so that increases the area but here because of the fact that you have a self-align, the area requirement is much less. So boron implantation this is where we have stopped and then metallization so that you complete the process.

(Refer Slide Time 44:23 min)

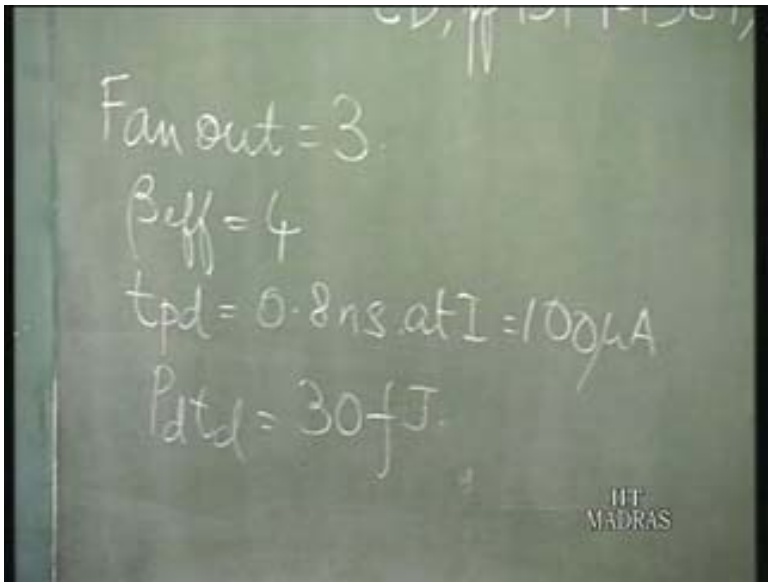
So this is the self-aligned I squared L process and you see that this is of course much superior compare to the previous thing, equivalent circuit wise also, lot of the parasitic are basically getting removed and the charts stored in these diode should be much less, its extrinsic region because of the P plus doping in the extrinsic regions. So this particular structure was actually proposed in 1980 I think, it came out in a paper and of course it was far superior to other technologies. I shall give you the reference also, just if you want to go through it, it is D. D. Tang, IEEE transactions on electron devices, page 1379-1384, 1980.

(Refer Slide Time 46:02 min)



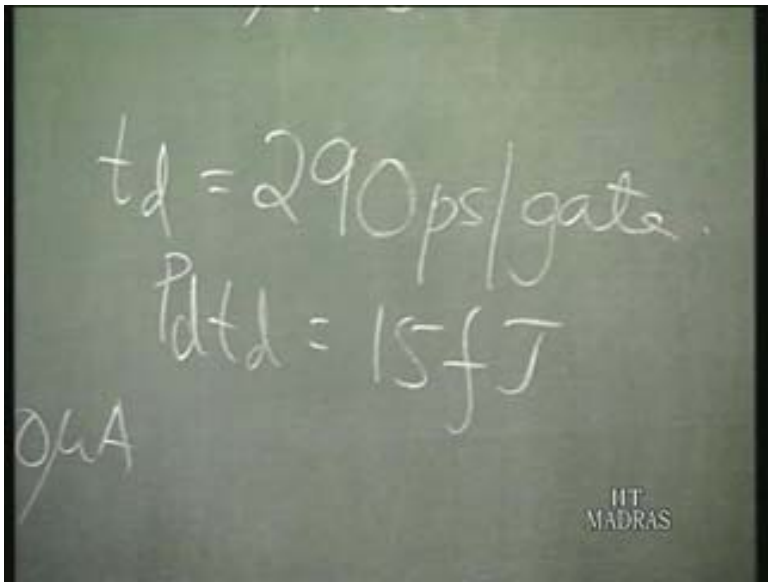
The results which they gave, came out was for a fan out of 3, for this particular structure for fan out equal to 3 that is for 3 collectors. They had a beta effective equal to 4 and interestingly the propagation delay of 0.8 nanoseconds. At collector current of I_0 equal to 100 microampere.

(Refer Slide Time 46:29 min)



So power delay product was 30 femto joules. So you see that of course it is very remarkable result. So much better than the previous results. So this is really sub nanoseconds device. In fact later on still further improvements have been made in technology which makes it very complicated in fact and most of the simplicity inherent in I squared L structure you know that has been lost but it has been 1985 there was another paper by another group which achieved two ninety picoseconds per gate delay. So that is I think a power delay product and power delay product of 15 femto joule.

(Refer Slide Time 47:36 min)



So this is for the best results which have been reported for I squared L. It's quite good you know for silicon I squared L. Presently lot of work is going on in I squared L made out of hydro junctions. I shall take the topic sometime later, few classes later because I have to introduce the advantages of hydro junctions and hydro junction I squared L, the delays have been even less than 100 picoseconds and in fact that is the area where lot of work is going on a I squared L instead rather than silicon I squared L presently.

In fact even Texas instruments have fabricated a microprocessor using hetero structure I squared L. So we shall go to that later on but this is the present condition, the delays which is being achieved using I squared L. We shall take up the next class some special I squared L structures or some special modifications of I squared L and then we shall actually go into hetero structure I squared L and see the latest performance before we wind up the topic of I squared L and go on to other logic families.