VLSI Technology
Dr. Nandita Dasgupta
Department of Electrical Engineering
Indian Institute of Technology, Madras

Lecture - 40
BICMOS technology

So, today we are going to have the last class on this VLSI technology and in this class, I just want to give you a brief outline about how the merger of the two technologies namely the bipolar junction transistor technology and the CMOS technology has been brought about in order to realize what is called the biCMOS technology. Now, biCMOS technology, essentially what does it mean? It is a CMOS based technology that is the core process flow that is the CMOS process flow. We are going to deal with the basic process steps needed to realize a CMOS device and in this core process flow we want to incorporate a high performance bipolar transistor.

Notice the words high performance bipolar transistor, right. So, how do we go about it is something like this. We first take a basic CMOS core process flow; we try to incorporate a bipolar transistor and when we talk about a bipolar transistor, it is almost always implied that is an npn bipolar transistor. So, we try to incorporate an npn transistor in this core process flow using as few extra steps as possible, because you want to keep your process flow economic; you do not want to incorporate too many extra steps, because extra steps means added complexity and therefore, additional cost. So, we will try to keep these extra steps minimal and try to realize a bipolar junction transistor. But obviously, this bipolar junction transistor is not going to be a high performance bipolar junction transistor.

So, what do we do? We keep on modifying the process steps, in order to enhance the performance of the bipolar junction transistor. Sometimes we will see that in doing so, we even enhance the performance of the CMOS transistor as well, in which case it is fine, because your entire process is becoming better and finally, we see how many steps we can actually use in common between the CMOS transistor as well as the bipolar junction transistor. So, this is essentially the principle of the biCMOS technology. In biCMOS

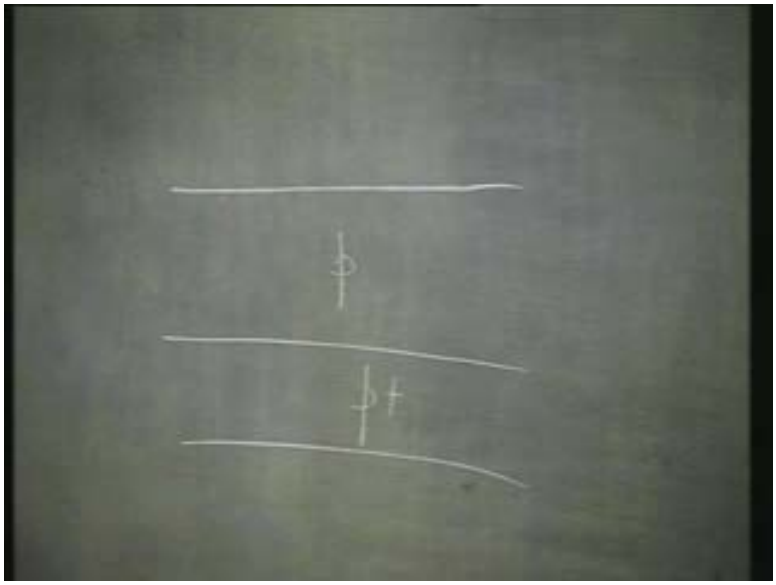
technology, we are just trying to get a bipolar junction transistor incorporated in the basic CMOS process flow and as I have already discussed in the last two classes that as the device dimensions become smaller, we have a lot of problems in CMOS; for example latch up problem, for example the subsurface punch through problem, for example the hot electron effects in the n channel devices, right. So, we have to actually make suitable adjustments.

Take for example, the question of latch up. You know if I want to prevent latch up, then we could use a retrograde well structure. If you use a retrograde well structure, it is actually coming closer to the bipolar junction transistor technology, because what we are using is the same buried layer, buried n plus layers, which will be used to cut down the resistance of the n well, as well as it will be used to cut down the collector resistance problem of the bipolar junction transistor. For example, let us say, the gate poly, we are using poly gates, silicided poly gates for CMOS devices. We could use the same poly deposition step for the poly emitter transistors. You know, high speed transistors, they always use poly emitters, right. We could use the same poly deposition step for the poly emitter transistor.

Even finer aspects, for example you see, this lightly doped drain structure, which we discussed, in order to make a hot electron resistant structure; remember, for these lightly doped drains, we had to use an oxide spacer layer, oxide side walls, right. Well, you could use the same oxide side walls in order to align the extrinsic base of the bipolar junction transistor to the emitter. So, a lot of steps judiciously can be combined together to be of use for both a CMOS transistor and a bipolar junction transistor, right and in doing that, we get a performance enhancement at little extra cost. So, you can get a high performance biCMOS technology, you can realize a high performance biCMOS technology, with only a few additional steps, so that you can get a performance enhancement say by a factor of 2, while the cost increase is only 1.3 to 1.5 times. That is a major saving as far as VLSI technology is concerned.

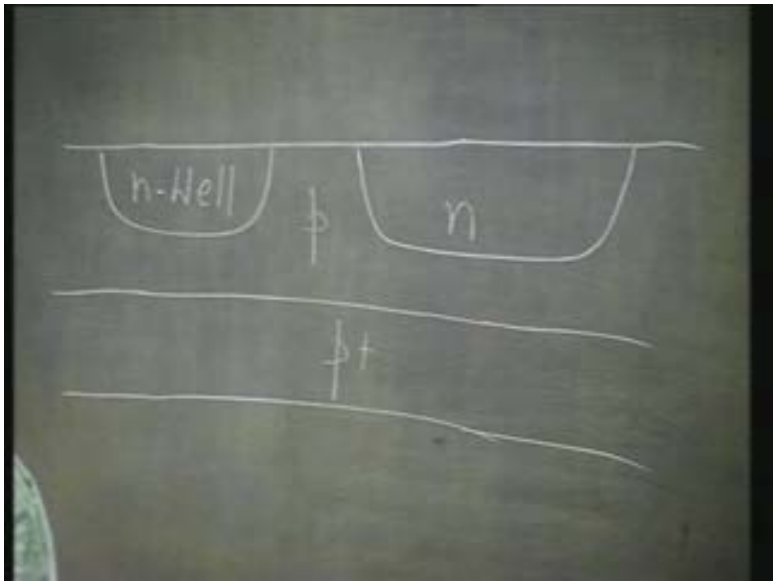
So, let us look at how biCMOS technology was evolved. First starting with a core CMOS process flow, how can we get a biCMOS technology? That is my objective is to realize both a CMOS as well as a bipolar junction transistor, an npn bipolar junction transistor. We will keep the basic process flow pretty simple, as simple as we possibly can. So, we are not talking about really anything very fancy. Let us simply start with a CMOS process flow. You know, we may have a latch up problem, so in order to reduce the latch up problem all we have done, all I am going to do in this is to use p on p plus substrate. The underlying p plus layer is going to cut down the shunt resistance of the substrate. So, we are simply starting with a p on p plus substrate.

(Refer Slide Time: 7:57)



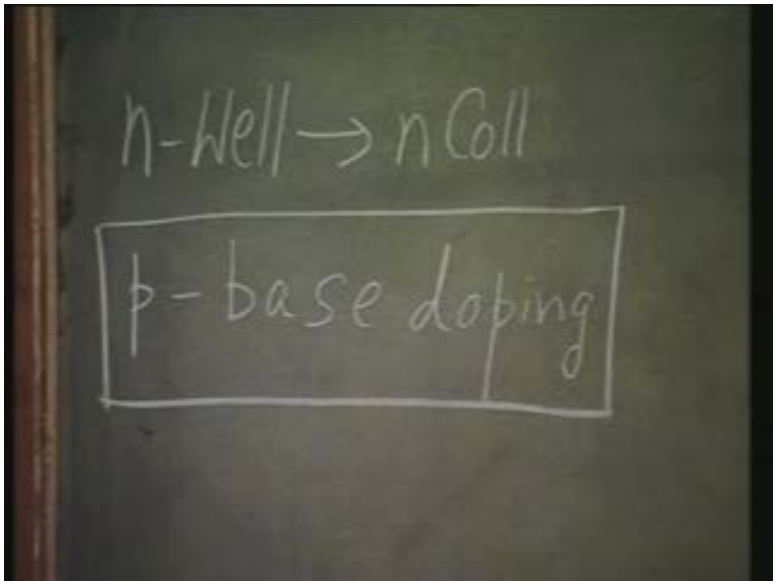
We have simply started with a p on p plus substrate. Now, you know, in order to realize a CMOS device in this, you must have an n well, I already have the p-type substrate. Actually I am talking about an n well technology, which is compatible with the n MOS process flow, so I need to have an n well in this. Let me have an n well.

(Refer Slide Time: 8:31)



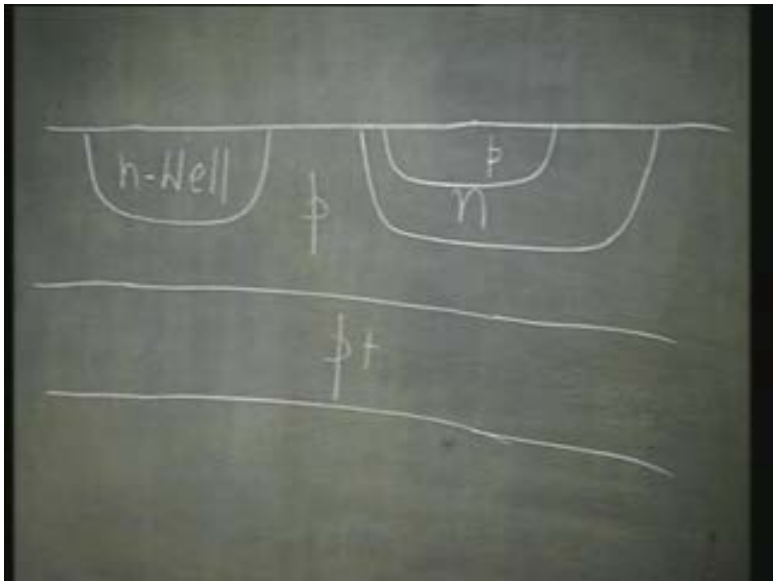
I am only outlining the basic process steps without talking about the field oxidation, without really talking about all the intermediate steps. I am just talking about the steps which are necessary to explain the biCMOS evolution. So you know, if I have a p on p plus substrate, if I want to realize a CMOS device, I must have an n well, right. This p-type substrate can be used for the n MOS device and inside this n well my p MOS device will be housed. Now, while doing this n well, I can have another n region implanted in the substrate, which can now house my npn transistor, right. In this n well, I could have my npn transistor, right. So, you see, first step is common.

(Refer Slide Time: 9:58)



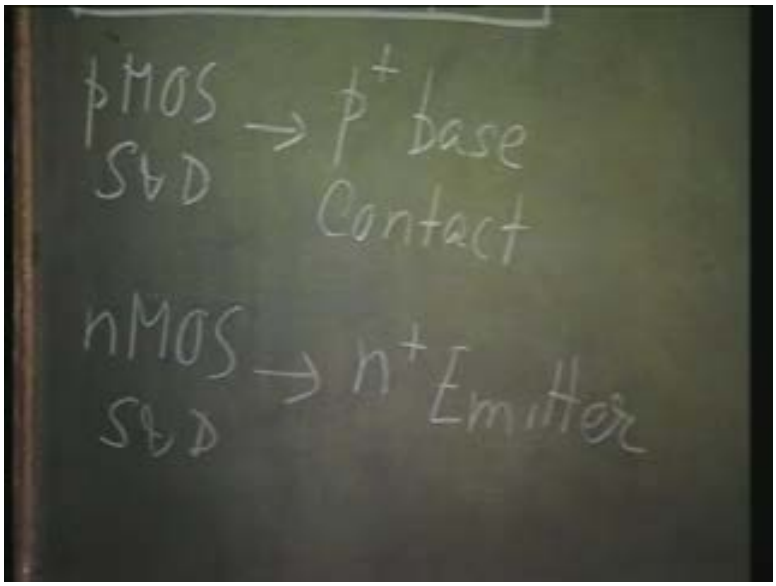
The n well can serve as the n collector. In order to realize this npn bipolar junction transistor, what do we have to do next? We must have the base. Now, you see, the base doping must be low, low p-type doping, right. Therefore I cannot use the p MOS source and drain implantation for this base doping, right, because the p MOS source and drain need to be necessarily heavily doped. So, I must have one extra step that is the p base doping. I am going to put it in a box just to signify that this is an extra step. This p base doping is needed exclusively for the bipolar junction transistor. It does not serve any purpose as far as CMOS device is concerned. I cannot merge any two steps here, right. So, I need one exclusive step for bipolar junction transistor, which is the p base doping.

(Refer Slide Time: 11:54)



Now, you see, once I have the p base doping, I have the p base here, if now I want to use a base contact, a base p plus contact, now I could use the p MOS source and drain for the base contact realization as well. All I have to do is together with opening the p channel, together with opening the p MOS source and drain, I have to open a window here, so that a small p plus regions gets doped here, which will provide my base contact.

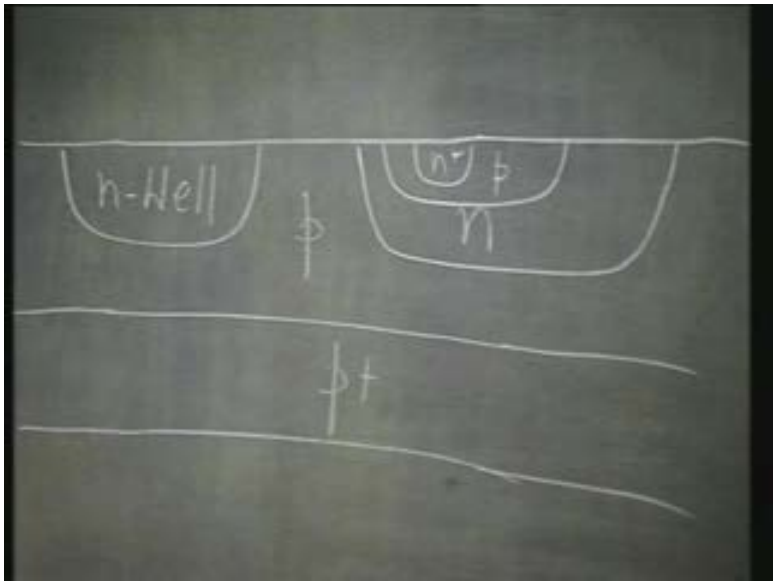
(Refer Slide Time: 12:42)



So, these two again can be clubbed together. You can use it for your p plus base contact, which is actually acting as the extrinsic base. You are trying to cut down the parasitic base resistance, not for the intrinsic base; intrinsic base, you need one exclusive step, p base doping. This is for the active transistor, the p base doping. But, next to that, if you want to have a p plus region to cut down the base resistance that is your extrinsic base doping, that step can be clubbed with the p MOS source and drain and then of course, the emitter of the npn transistor, which is n plus and this can be clubbed very easily with the n MOS source and drain.

So, these are the basic steps needed to realize the bipolar junction transistor in a CMOS process flow. Of course, I have not talked about the isolation which you already know. We can use the same local oxidation, LOCOS technique both for isolating the bipolar junction transistors as well as the CMOS; so, that is any way going to be a common step. Now, if we look at this, the basic requirements for an npn transistor, we see that we have needed only one extra step, the step in the box that is the controlled intrinsic base doping, for which we need a low Gummel number, right. The total charge in the base region must be fairly low, so that the beta of the transistor is kept high and this doping has to be carried out separately. We cannot use any common step with CMOS. So, essentially you have needed only one extra step.

(Refer Slide Time: 15:33)

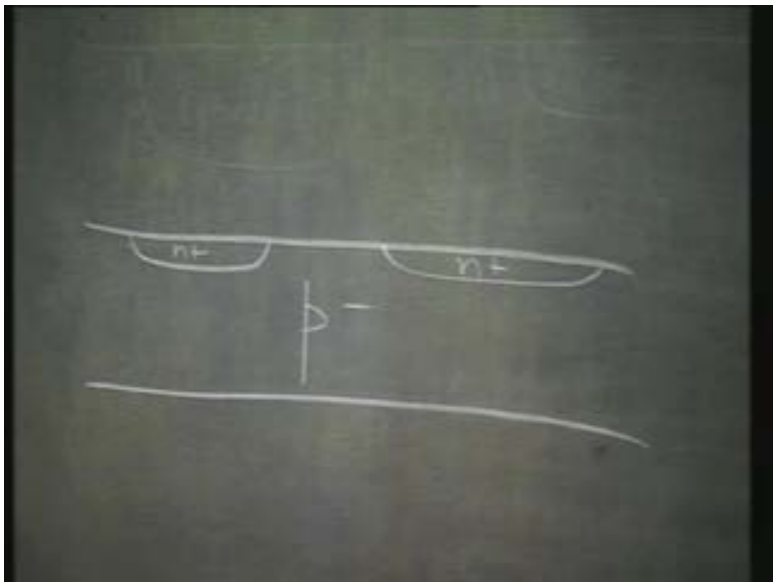


But, this transistor cannot be called a high performance transistor, because of the very simple reason that the collector is low doped. The collector doping is common with the n well doping and you know that I may need an n well doping fairly low, right. The n well doping must be fairly low, therefore the collector doping is low; therefore, the collector resistance is very high. So, even though I can realize a basic npn transistor in this process flow, the major problem of this transistor is that the collector resistance is very high, since it is controlled by the n well doping and the sheet resistance of this n well doping is about 2 kilo ohms per square; 2 kilo ohms per square is much too high for the collector resistance.

So, what do we do? We have to find some ways to cut down the collector resistance. Remember, what we used to do in bipolar junction transistor? We used to have a buried layer; we used to have a buried layer to cut down the collector resistance. Can we do the same thing in a CMOS process flow? Yes, we can; because, if you remember, if you have a buried layer aligned to the n wells, this will also cut down the latch up problem, right. So, what we should do is instead of using a p on p plus substrate, let us use a low p doped substrate, a p minus substrate, have the buried layer first and then have the wells aligned to this buried layers, right. That will surely enhance the performance of the bipolar

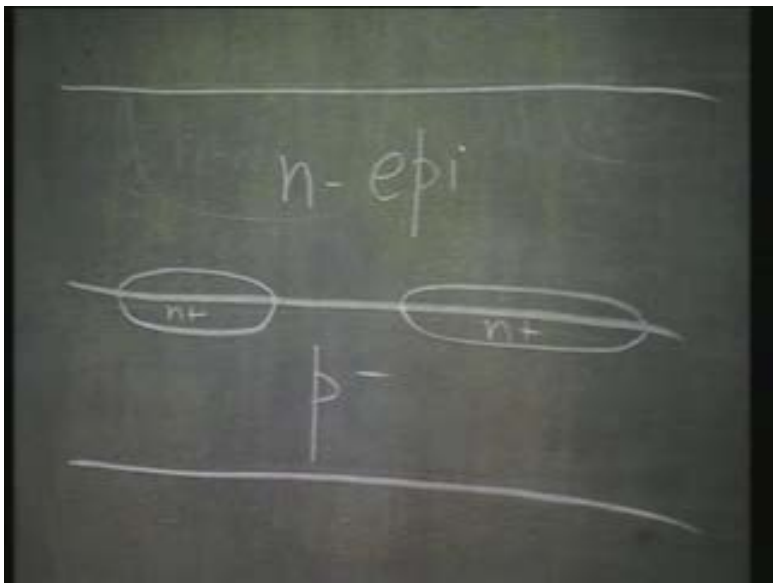
junction transistor, because of the n plus buried layer, to reduce the collector resistance, at the same time this will create less likelihood of the latch up problem for CMOS. So, we are trying to use this buried layer for, you know, mutual advantage; the advantage of bipolar junction transistor as well as advantage to CMOS.

(Refer Slide Time: 18:20)



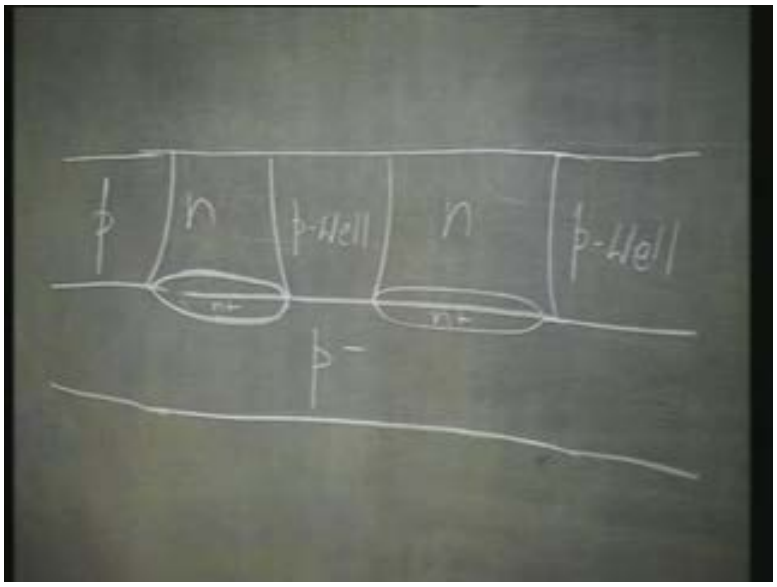
So, in that case, we start with a lightly doped p-type substrate and we first have the n plus layers, buried layers doped here and next step, we carry out an epitaxy, an n epitaxy.

(Refer Slide Time: 18:55)



Now, all we have to do is to have the p wells put in between.

(Refer Slide Time: 19:28)



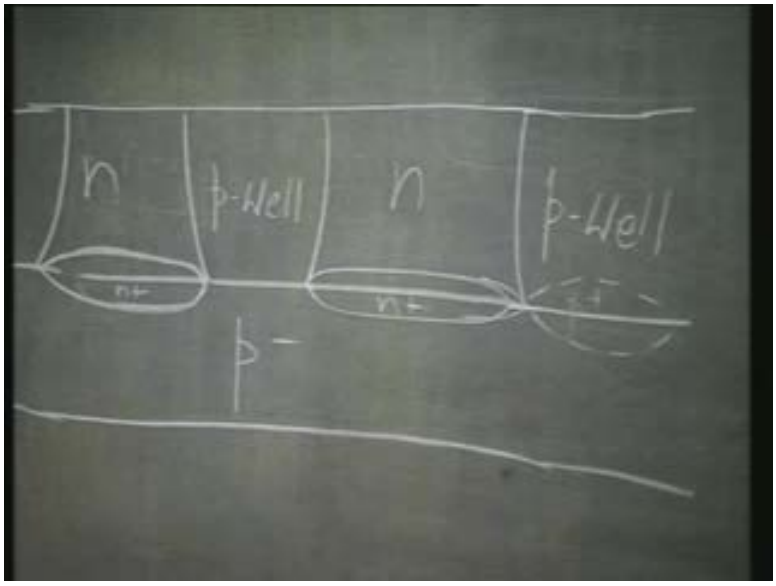
This n region can house the p MOS device, this n region can house the npn transistor, bipolar npn transistor and I have the buried layer, which is going to cut down the collector resistance for the bipolar junction transistor and this buried layer is going to cut down the substrate resistance for, to suppress or prevent the latch up problem. So, even

though I have actually added two steps - one of the buried layer and the other of the epitaxy – this is in fact, if you remember, this is actually your retrograde well with epitaxy, which can be used both to prevent the latch up in CMOS, as well as to cut down the collector resistance in bipolar junction transistor.

So, the features of this particular process flow is this buried n plus layer and you can also have a deep n plus collector here and here. That will also cut down the shunt resistance of your CMOS substrate, as well as will reduce the collector resistance, right. If you have a deep collector doping going all the way down to the buried layer, then your collector resistance will be much reduced and also if you have a deep n plus region here, this will cut down the shunt resistance to a large extent. That will be your substrate contact for the p MOS device. So, the substrate contact for the p MOS device and the collector contact of the npn transistor can be one and the same step and that will further reduce the resistances.

Now, in this particular case of course, we have a small problem. That is if I want to have another npn transistor on this side, it must be separated well from the adjacent npn transistor. Why? Because, you see, this p-type substrate is very low doped and necessarily, you know, your substrate is connected to the most negative point, right, so that this pn junction is actually acting as the isolation. But, since this p is very lightly doped, there is a possibility of punch through; unless it is separated well, there is a possibility of punch through. So, you have two choices. One is of course, you could add p plus buried layer.

(Refer Slide Time: 23:22)



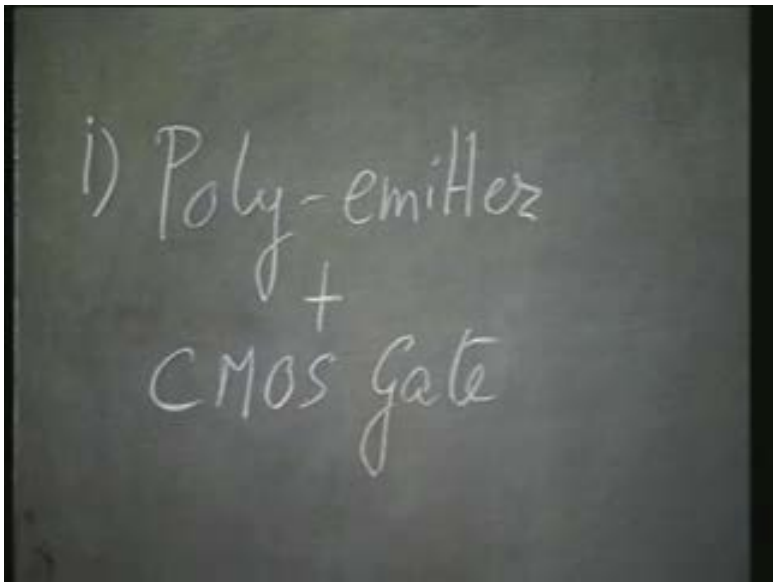
Just as n plus buried layers are given here, you could have p plus buried layer aligned to p well and the other possibility is instead of this n epitaxial region, you could have nearly intrinsic epitaxial region and then you see, this one is basically a p well technology; we have realized an n epitaxial layer and in that we have put in the p wells. Now, instead of doing that, what we could do is we could have a nearly intrinsic epi layer and then go for a twin tub technology. That is we realize both n wells and p wells, so that the packing density of the bipolar junction transistor is increased.

See, how our requirements are changing. To begin with, we were interested in just realizing a bipolar junction transistor. We saw that only by having one extra step that of the base doping, we could realize a bipolar junction transistor. So, the next criterion became how to enhance the performance of this bipolar junction transistor and we realized that the major problem in this bipolar junction transistor is because of the collector resistance. So, we added the buried layer and the epitaxy that is the retrograde well structure. So, basically we killed two birds in one stone. We prevented the latch up problem for the CMOS, as well as we reduced the collector resistance problem of the npn transistor. But now, what we want to do is we want to achieve greater packing density. We have enhanced the performance of the bipolar junction transistor, we now want to put

more bipolar junction transistors, we want to increase the packing density and in that we found the bottleneck is the low doping of the underlying substrate, because there is a possibility of punch through.

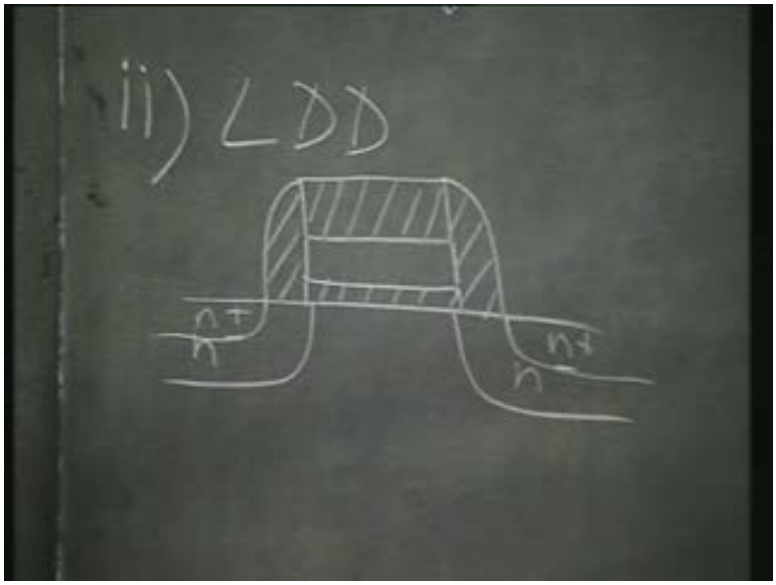
So, we say, fine, let us add one more step. Let us have, just like we had n plus buried layers, let us also have p plus buried layers, so that the punch through problem is reduced and we could have better packing density. In this way, as the technology evolved, we found that a lot of steps can be put together. For example, so far I am not talking about the, I have not talked about the finer aspects of the bipolar junction transistor technology. I am satisfied only with realizing a basic npn transistor with low collector resistance. Now comes, the finer features of the bipolar junction transistors. What are these finer features? First of all, let us have poly emitter transistors. You know, for ECL, we use high speed transistors and whenever we talk about high speed transistors, we want poly emitter transistors, right. Poly emitter transistors, they have much higher gain; we want to have poly emitter transistors. So, the gate in the CMOS transistor can be used to define the emitter region of the bipolar junction transistor, right. We could have the CMOS gate and the emitter merged together, the poly deposition can be the same step. So, let us see further refinements.

(Refer Slide Time: 27:19)



These two can be clubbed together, poly emitter and CMOS gate. Next you see, for high performance CMOS transistors, we must have silicided gate; polysilicon gate will not do, because of its larger sheet resistance, because the R C time delay product will become large, right. So, we must cut down the resistance of the gate as well as the interconnections. So, what do we use? We use, on top of the polysilicon we use silicided gate. The same thing we could also use to take the emitter contacts, right. So, the silicided gate can be used as the emitter contacts. Then comes, other finer aspects; for example, the LDD structure, lightly doped drain for the n MOS transistor.

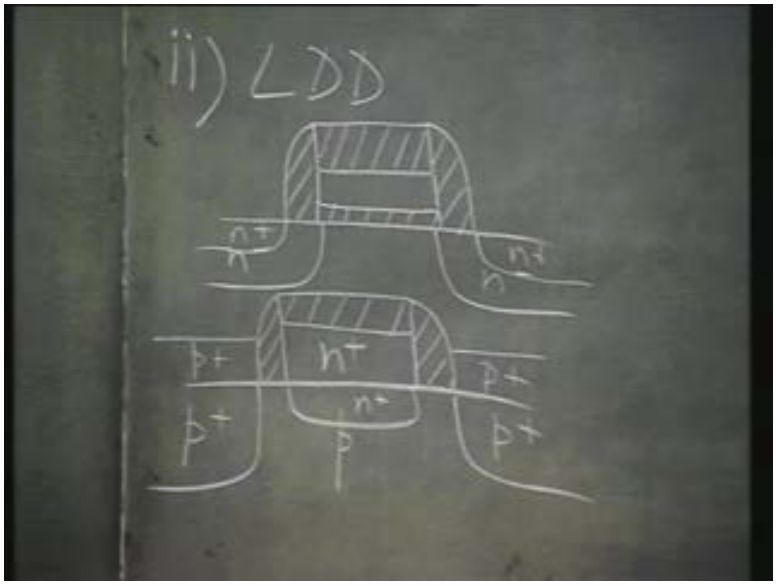
(Refer Slide Time: 28:35)



Do you remember how we got this lightly doped drain? We had defined the gate, right. Then, in the first step after that, we had done the source and drain diffusion. But, these are n minus diffusions and then, we deposited a layer of oxide and patterned that oxide and a second layer of oxide, so that finally what we had was a structure like this, right; side wall spacer for the LDD structure. So, the next implantation gave us ….. This was the LDD, lightly doped drain structure, so that keeping the channel length the same, we have actually brought down the electric field. There is a potential drop between this, in this n minus region. That was the lightly doped drain structure; we used this oxide spacer.

Think back for a moment to how the extrinsic base was self-aligned to poly emitter. What did we do?

(Refer Slide Time: 30:12)

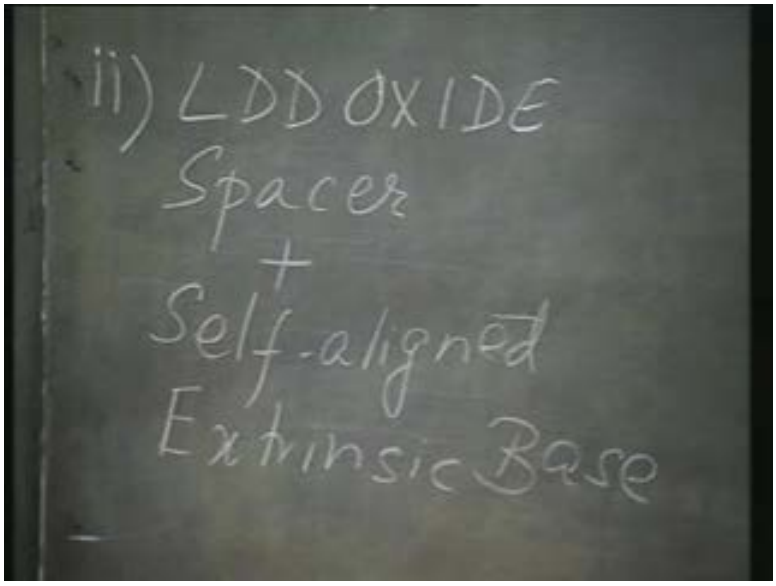


We had the base p-type region. We have deposited a poly and then there also we had used a spacer, right. We had this n plus poly to realize the emitter and then, we had used … and then we had used p plus poly from which we doped the … So, in doing the same annealing step, what do we have here? We had the p-type intrinsic base, on top of which we had deposited poly and implanted it with n plus. Then, we had protected that poly with oxide on top and also oxides on the side. Then, we had a poly deposited again and doped it p plus, so that after doing the annealing from this p plus poly, we had the p plus extrinsic base; from this n plus poly, we had the n plus emitter which was self-aligned to each other, right. The n plus emitter was self-aligned to the p plus extrinsic base.

Now, compare these two structures - the LDD needed for the n MOS in CMOS devices and the extrinsic base required for the npn bipolar junction transistor. In both cases, you find that we need an oxide spacer layer, right. The same kind of structure we are using, both to realize a lightly doped drain as well as to self-align the extrinsic base with the

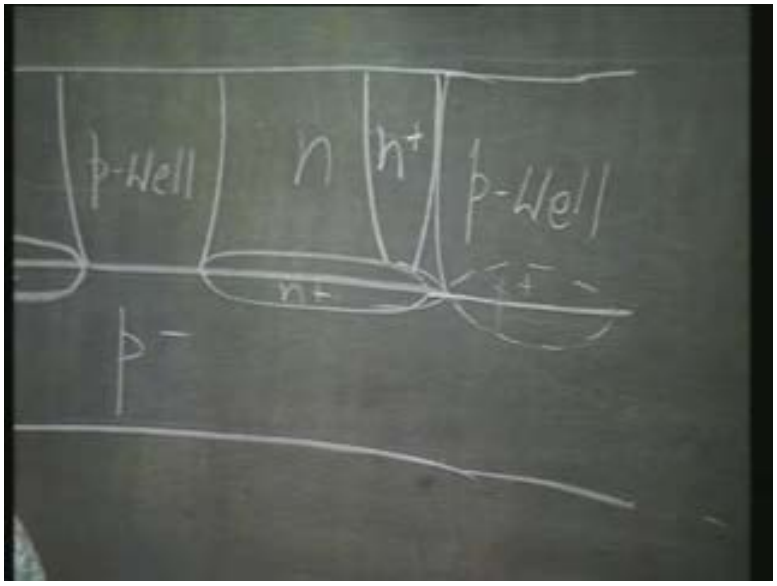
emitter. So, even this oxide deposition and the spacer formation can be done in a single step. We can merge these two steps together.

(Refer Slide Time: 32:56)



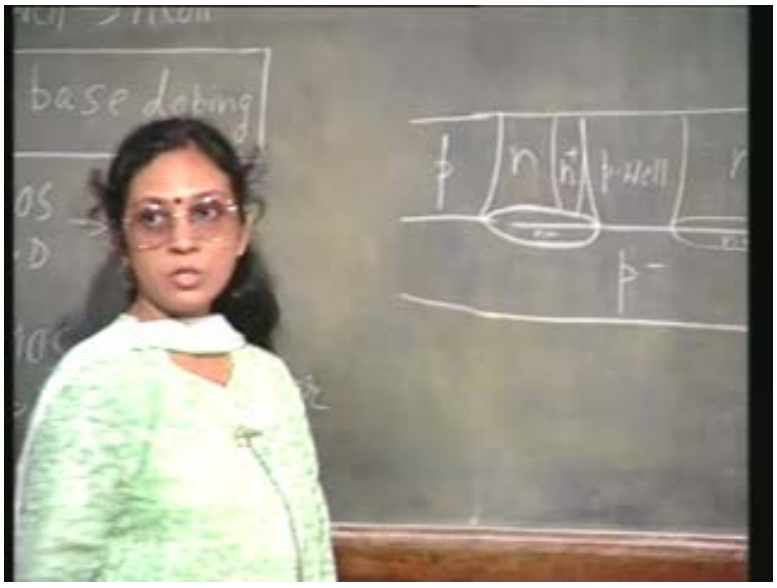
So, you could use the same oxide spacer both for the LDD structure as well as for the extrinsic base to be self-aligned with emitter. If you want to further refine your process, further improve the performance, then of course, you could go to trench isolation; instead of using LOCOS, you could go for trench isolation, right. Trench isolation will increase the packing density to a large extent. At the same time, you will not have any bird's beak or bird's crest problem, right.

(Refer Slide Time: 34:30)



So, you could use trench isolation to improve the packing density, both for the CMOS transistor as well as for the bipolar junction transistor and then, remember in this case, I said that you can have a deep n plus region extending all the way down to the buried n plus in order to reduce the collector resistance.

(Refer Slide Time: 34:49)

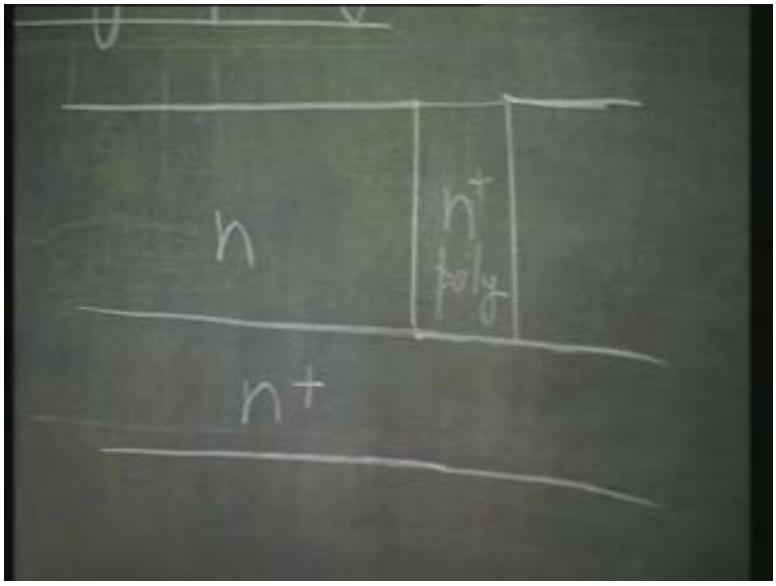


At the same time, here also you could have a deep n plus region, which is going to act as your substrate contact for the p MOS device, the n well substrate contact. Now, always whenever you are trying to have these deep junctions, it is a better option to use a poly-plug, what is known as a poly-plug.

(Refer Slide Time: 35:32)

A poly-plug is somewhat similar in concept to the trench isolation. Do you remember what we had in a trench isolation case? What did we have? We first etched a trench in the semiconductor material, right and then after growing the thin side wall oxide, we filled up the trench with undoped poly and then, we sealed the trench with a top thin oxide layer. So, what happened essentially? The trench is isolating two regions of the semiconductor by means of this undoped poly, sealed on all sides by the oxide. Now, suppose if you have a trench etched and fill it up not with undoped poly, but with heavily doped poly, then what you have is called a poly-plug, right.

(Refer Slide Time: 36:43)



So, you see, essentially what you have here is I am just expanding the bottom of the n well; I am just expanding the bottom of the n well. So, what you have is the n and the underlying n plus buried layer. This n well is going to house the npn transistor as well as the p MOS device. Now, we do an etching; etch all the way down to the n plus region by means of dry anisotropic etching, to have perfectly vertical side walls and now suppose we fill this by means of an n plus polysilicon, we fill this trench by means of n plus polysilicon. What do we have?

We have realized an n plus contact extending all the way down to the buried layer, which could serve both as your collector contact as well as for the substrate contact for the CMOS device. Why am I preferring this, because I have less lateral spreading. You know, whether you use ion implantation or diffusion, the deeper your junction is, the more is the lateral spread, right. So, if I am talking about a deep junction, if the junction has to reach all the way down to the n plus buried layer, it has to go through the entire epitaxial region. That means on both sides it must have spread considerably; about, for diffusion, you know, it is about 80% percent. For ion implantation also it is approximately the same, 75%, 70 to 80%, right. So, if you want to increase the packing density a better idea might be to open a window of the required dimensions and then fill

it up with n plus poly, so that the lateral spread is reduced and the packing density is further improved. This is called the poly-plug, n plus poly collector plug as well as the n plus poly substrate contact plug.

So, we have seen both the bipolar junction transistor technology as well as the MOS technology, particularly the CMOS technology which is of great dominance today and then we have seen, though very briefly, how a biCMOS technology can evolve from the CMOS technology and the interesting point is as the CMOS technology becomes more and more complex, as a matter of fact, it helps in realizing a bipolar junction transistor in the same process flow. It helps to some extent, because a lot of the process steps now become common to each other. So, with very few additional steps, it is possible to realize a bipolar junction transistor in the same CMOS process flow.

But, remember one thing. There are a few steps on which no merger is possible. That is major problem is the base region. There is no compatible step in CMOS process flow. The base of the npn transistor must be a separate step. Of course, the collector region can share the n well. You can reduce the collector resistance by having an n plus buried layer underneath the n well, which will also help in cutting down the latch up problem in CMOS. In that sense, these two steps can be merged together. The buried layer can serve both purposes. The same way you can use either an n plus deep junction or an n plus poly-plug to realize both a collector contact as well as the substrate contact for the p MOS device.

The extrinsic base can be realized by using the p plus source and drain of the p MOS device and in order to have the extrinsic base self-aligned to the emitter, we could use the same oxide spacer layer as used for the LDD of the n MOS transistor, lightly doped drain of the n MOS transistor, to make it a hot electron resistant structure and of course, the emitter could be either if you are simply using n plus emitter, it can straight away be the source and drain of the n MOS transistor; the same implantation step can be used. If on the other hand, you are planning to use a poly emitter transistor, merge it with the gate, gate of the CMOS transistors, right. So, you could use the same poly deposition and poly

doping step for the emitter. To further cut down the emitter resistance, contact resistance you could have the same silicide process flow that you have adopted for your CMOS, in order to cut down the interconnect resistances. So, all these steps could be common. The one essential extra step however remains the base doping. So, this is essentially the biCMOS process flow and how it has evolved from the basic CMOS process.