VLSI Technology

Dr. Nandita Dasgupta

Department of Electrical Engineering

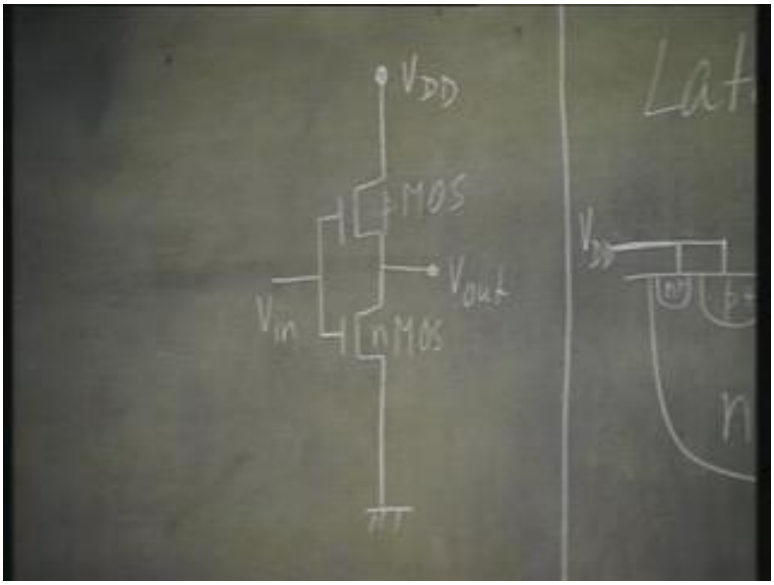
Indian Institute of Technology, Madras

Lecture - 39

Latch up in CMOS

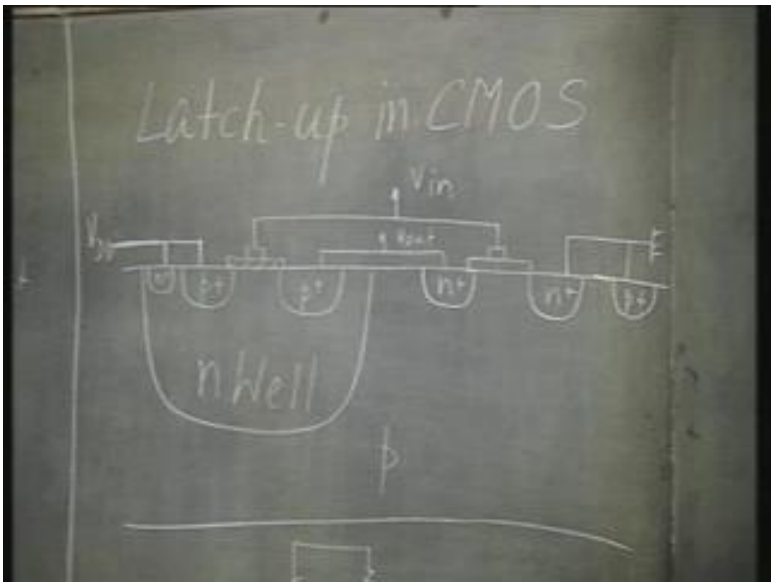
We have been discussing about the problems in CMOS, basic CMOS technology. Now, one of the serious problems in a CMOS circuit is the problem of latch up. This problem is inherent in a CMOS, because you know, we have both a p channel device and an n channel device, ok and a basic CMOS inverter is connected like this.

(Refer Slide Time: 1:52)



You have the pMOS device and you have the nMOS device. The gates are shorted together and you give the input at the gate and this is the point where you take the output; this is the source of the nMOS transistor and this is the pMOS transistor. Ok?

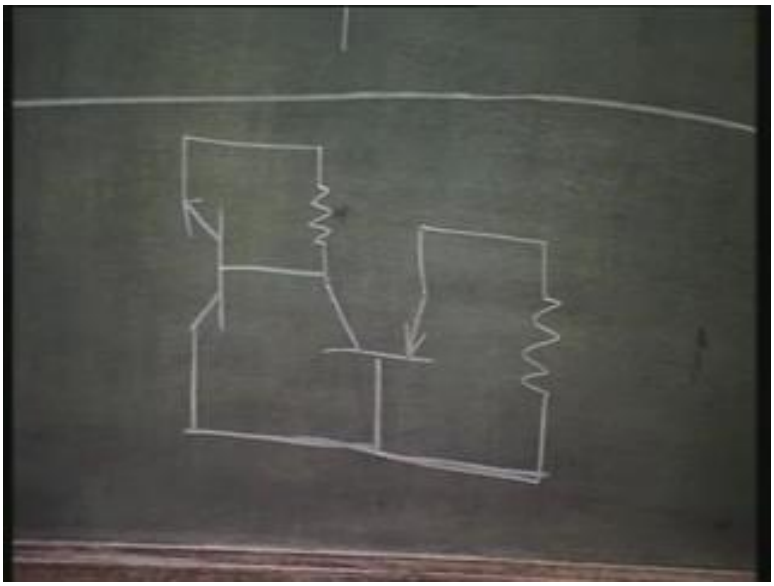
(Refer Slide Time: 2:23)



Now, if I draw a cross sectional diagram of such a CMOS inverter, I have taken an n well technology which is compatible with nMOS logic. Ok? I could have taken a p well technology or a twin tub technology, does not matter, ok? This is the basic CMOS inverter circuit. Let us look at this circuit. I have the n well here and the p substrate, right? So, the substrate doping is very low. In that I have an n well; inside the n well, my pMOS is housed, right? This is the p source and drain, p channel source and drain and I have a substrate contact. You know, all the actual devices will have a substrate contact where you can, if you want you can give a source to substrate bias also. In this particular case, I have shown that the source and substrate are shorted together, ok and similarly, I have an n channel device, where again the source and substrate are shorted together. Input is given between the, connecting the two gates I have given the input and this is the output point. So, this is the basic CMOS inverter.

Now, look at it from another angle. I have a pnp structure here. I have a pnp bipolar junction transistor. Vertically if you look down I have a pnp bipolar junction transistor, right?

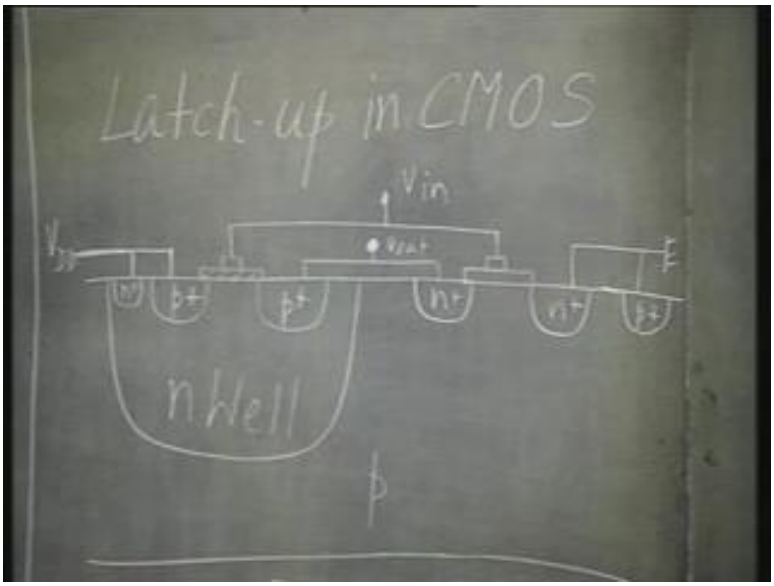
(Refer Slide Time: 4:03)



This is my pnp bipolar junction transistor, ok and I also have an npn bipolar junction transistor; a pnp transistor and an npn transistor. Again you see, these two transistors are, you could say, sort of merged together, right? The pnp transistor base is actually the, this base is actually the, this is the pnp transistor; this base is the collector of the npn transistor and again you see npn transistor, the base p is actually the collector of the pnp transistor, right? So, this base is shorted to this collector, this base is shorted to this collector, right? Ok? Ok. So, I have along with the CMOS inverter, two bipolar junction transistors as well - one pnp and the other npn; so far, so good.

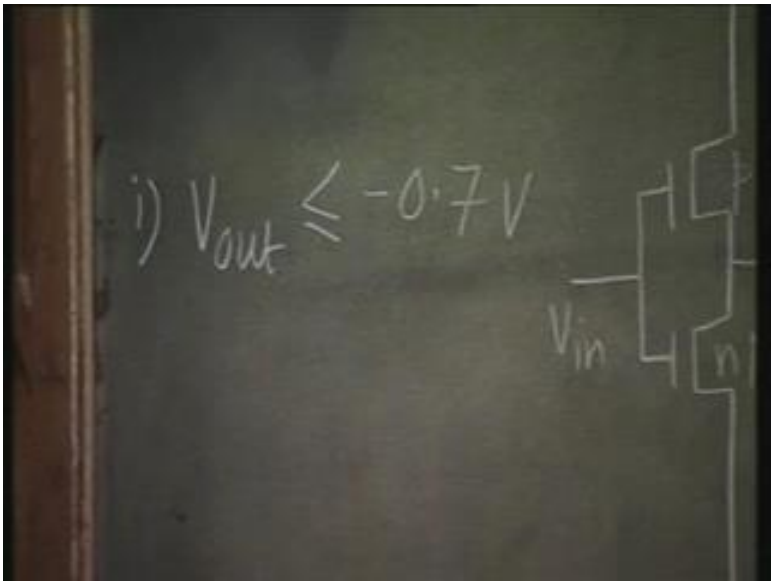
Now, let us say that at the output point, ok suppose there is a noise signal. Ok? Suppose your output voltage was at low level and then there was a noise, so that the output actually goes below the ground potential, goes below the ground potential by 0.7 volt. Ground potential or whatever you give at the source V_{SS}, ok it goes below that potential.

(Refer Slide Time: 6:11)



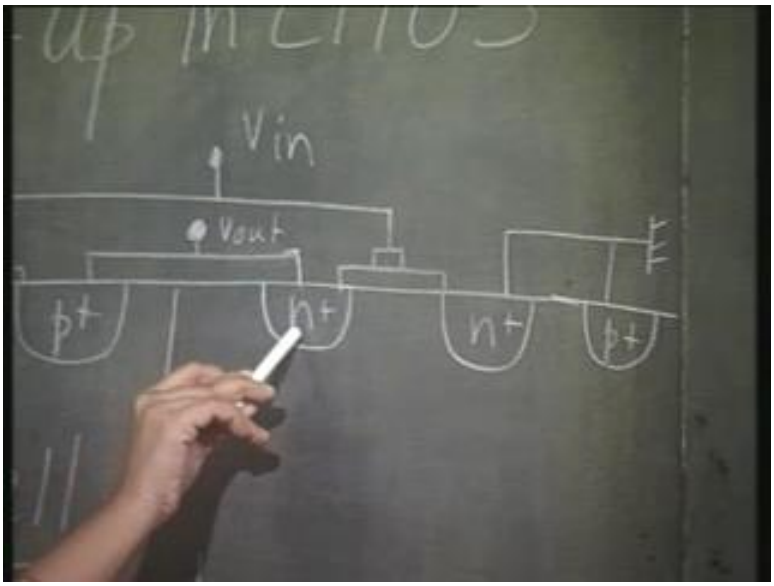
In this particular case, I have shown it to be ground, so it is below the ground potential by 0.7 volts, at this point. Ok?

(Refer Slide Time: 6:18)



So, let us say, to begin with V_{out} is …., right? Now, what is going to happen?

(Refer Slide Time: 6:37)

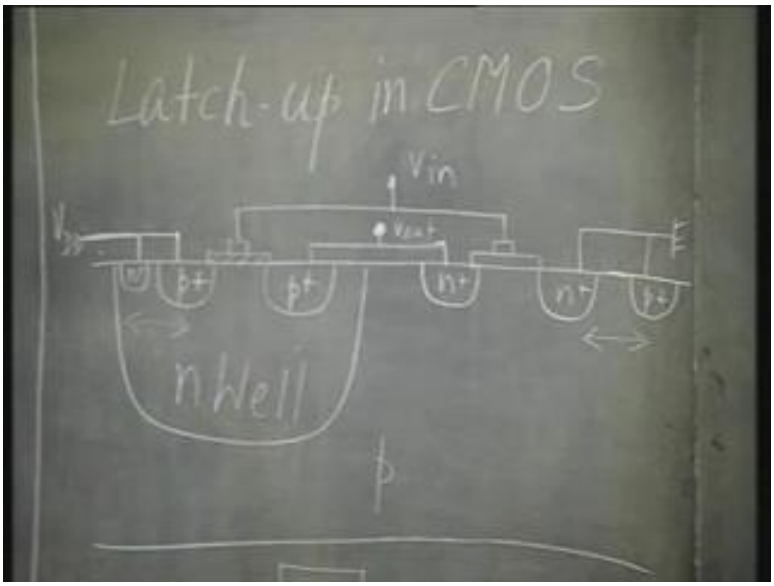


This n + therefore, is at a potential of - 0.7 or lower. So, look at this pn junction. What is going to happen to this pn junction? It is going to get forward biased. That means if I look at this npn transistor, the emitter starts to inject electron in the npn transistor, right? Where will these electrons go? It will flow to the collector, agreed? So, it is going out here, npn and then, it is getting collected here, going out. This is positive, $+V_{DD}$; electrons are getting collected and going out through this. Ok? Now, you see, I cannot afford to have the n well doping concentration high, right, because that will adversely affect my pMOS device performance. So, the n well doping is low. Ok?

Now, if the n well doping is low, when the current is flowing through this n well, there is considerable voltage drop in the well, right? In other words, I could say that this is my npn transistor, this is my pnp transistor. So, in this npn transistor, ok alright, so I have a considerable voltage drop here. Now, because of this voltage drop, ok this point is at V_{DD}, positive voltage, but this point here, it is at a considerably lower potential. Now, this point here is at a considerably lower potential, if I have a voltage drop between this, agreed? So, the n well potential here, at this point underneath the p + region, can again be more than 0.7 volts.

So, you see, your $p+$ region is connected to V_{DD}; this can be at a potential of $V_{DD} - 0.7$. Ok? Then, what happens? This pn junction also gets forward biased. Ok? So, the emitter of this pnp transistor now starts injecting holes, it starts injecting holes. Where will these holes go? It will go out here and get collected here. Now, in doing so, again you see, my substrate doping obviously, is very low.

(Refer Slide Time: 10:09)

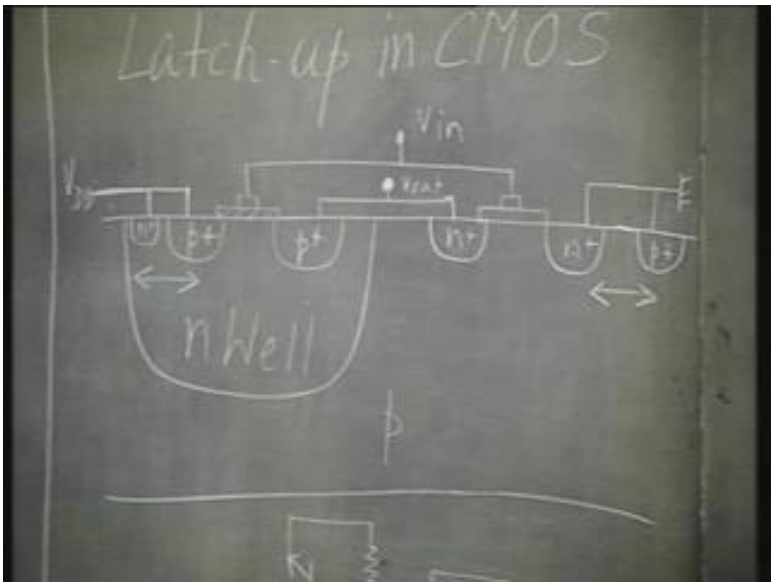


So, again there is a possibility that here there is a sufficient potential drop, right, so that this point here, underneath the $n+$ source, is at a potential of 0.7 volts higher than the ground potential. Holes are flowing in this direction, right? So, this point is at a greater potential, higher potential. So, there is a possibility, if this voltage drop is sufficiently high, that if the substrate doping concentration is sufficiently low, then there is a possibility that this pn junction can again get forward biased, right? Then, what happens? Again this one starts injecting holes, sorry, again this one starts injecting electrons.

So, what is happening? Once, just once, the voltage at the output point has fallen below the ground potential by 0.7 volts, right? Only once; that is enough to start a chain reaction. Because this pn junction is forward biased, this, the emitter of the npn transistor starts injecting electrons. These electrons go out, go into the n well, get collected through

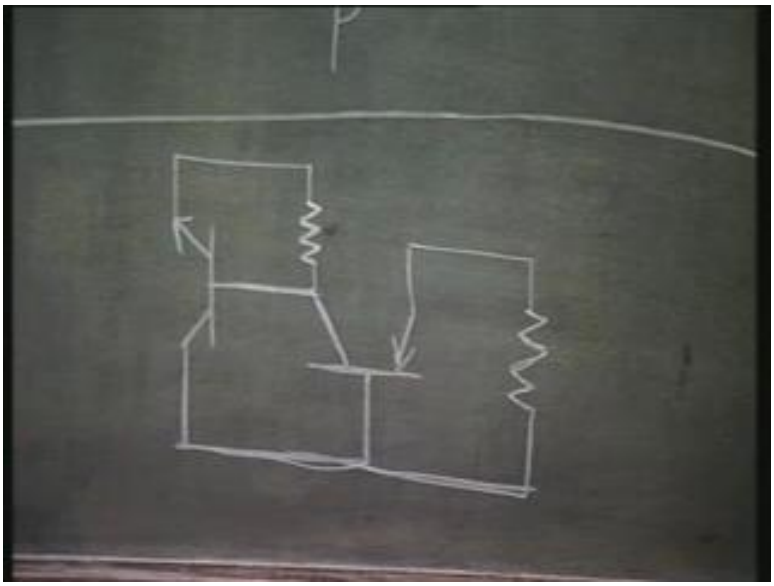
this. Now, because of the n well resistance, you see, this is the path of the npn transistor. Ok? In this collector region, this is the npn transistor, ok here I have a resistance. Ok?

(Refer Slide Time: 12:14)



In the npn transistor, here I have a resistance similarly and because of this resistance, here I can have a voltage drop which may be sufficiently high to turn ON this pn junction. Once this pn junction is forward biased, the pnp transistor gets activated and starts injecting holes. Ok? Once the holes are injected, it has to flow out of this path. When the holes are flowing in this direction, now, there can exist sufficient voltage drop here to turn ON this, this pn junction. Again, this one starts emitting electrons. Even after you have removed the initial spike, this process can continue forever provided your n well resistance and the p substrate resistance is sufficiently high, right?

(Refer Slide Time: 13:12)



So, you see, the key factors in this latch up problem are these two resistances; the key factor associated with this are these two resistances. Ok? Now, we are at a disadvantage here, because as far as the, as far as the MOSFET performance is concerned, we would like to have these two doping concentrations low, right? But, if we keep these doping concentrations low, we have the parasitic npn and pnp transistors getting activated, more chance of them getting activated, right? So, we are at a, we are in a dilemma. This is the essential latch up problem in CMOS. So, what are the remedies? What do we want to do? We do not want to adversely affect the MOSFET performance, at the same time we want to cut down the gains of these two transistors. We must cut down the gains of these two transistors, right?

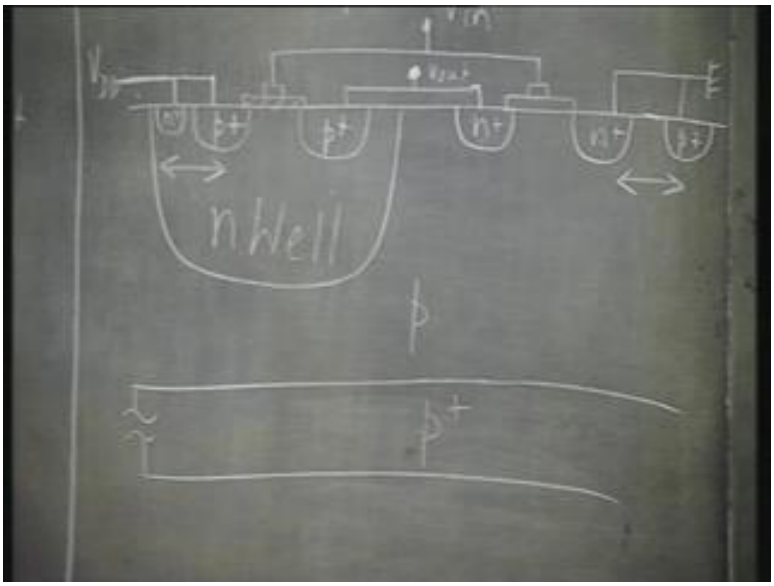
How do we cut down the gains of these two transistors? The older technique tried brute force technique. That is you deliberately irradiate the devices with neutron irradiation or you give gold doping. Gold or platinum, they are very good lifetime killers, because they introduce deep levels in the semiconductor, which kills lifetime. Ok? So, that is how the beta of the transistor was killed, but this technique is very difficult to control. So, instead of using these older techniques, later day technology tried to reduce these two resistances without affecting the performance of the MOSFET, right and that is possible because

now you have a lot of sophisticated techniques, for example ion implantation. You can make sure that you have a very heavily doped region underneath the n well or deep inside the p type substrate, right, so that the current finds little resistance. When the current is flowing in this direction; it finds little resistance. But, the MOSFET performance is not affected, because the doping is not coming anywhere near the depletion layer. As far as the MOSFET performance is concerned, this is too deep to affect the MOSFET performance, but it can kill the parasitic BJT action. Ok?

So, then we come to a more sophisticated CMOS technology and when I say more sophisticated, what do I mean, really? What I mean is, you see, so far we have seen there is a great difference between the bipolar junction transistor technology and the MOS technology, in the sense that for MOS devices, you have to use only bulk wafers; so far we have not used any epitaxy, right? So far, we have not done, have not used any epitaxial layer for MOSFET, right? So, we say that MOSFET is always a simpler technology, because you know, epitaxial wafers are horribly expensive. If you, if you have to grow it yourself, you must have the epitaxy facility, otherwise if you want to buy it from the vendors, an epi wafer is much more expensive than a bulk wafer. Ok? So, that was, so far we have seen that was an advantage with MOSFET, we did not need any epi wafers, right? You could do everything with a bulk wafer, alright?

Now, we see, because of this latch up problem, it might be better if we used epitaxial wafer. How is it going to be better? I will give you a very simple modification to this.

(Refer Slide Time: 17:42)

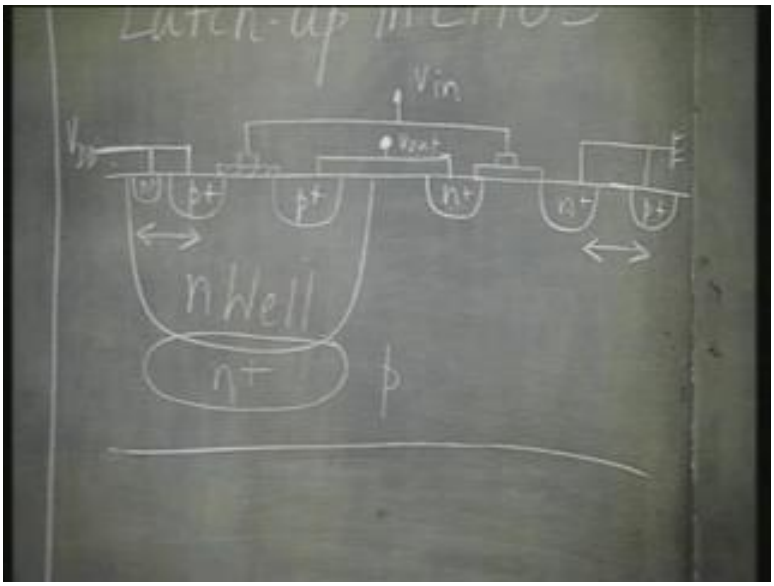


Instead of using a bulk p-type substrate, let me use a p on $p+$ epitaxial wafer. Do not get misled by my drawing. What I mean is this $p+$ region is much thicker than the epitaxial p region. I should actually put a break here, to signify that this $p+$ region is actually much thicker. This is the bulk wafer on which you have grown a p-type epitaxial layer; you have grown or you have bought. So, these are called p on $p+$ wafers. As soon as I have a p on $p+$ wafer, my latch up problem gets drastically reduced. Why? Because, the path of the current, look at the path of the current; it is now flowing through this $p+$ region, right? This is, this height is much smaller.

In fact, if you want, you can have this $p+$ region all the way extended down to this $p+$, so that your current finds little resistance. There is no question; even if, to begin with, this one injected electrons because of V_{out} falling below 0.7 volts, even if this one injected electrons and this one injected holes, by the time the holes reach here, there is going to be no regenerative mechanism, right? So, you may have an initial fluctuation, but it is not going to get latched up, because you have drastically reduced the resistance in the path of the current without sacrificing your MOSFET performance at all. Of course, you have

increased the process complexity; you have increased the cost of the process, by using an epitaxial wafer instead of using a bulk wafer. I could suggest also another remedy.

(Refer Slide Time: 20:21)



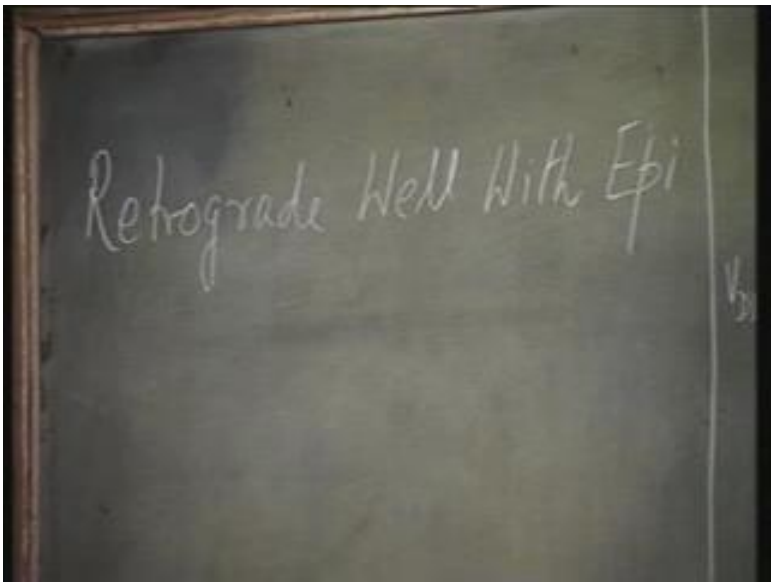
By judicious use of ion implantation, let me have what is called a retrograde well. A retrograde well is one, where the peak doping concentration lies deep inside the wafer; not at the surface, but deep inside. Ok? Again, you see, you are achieving the same thing, because when electrons are injected from here, it is actually flowing through this n+ region, where it encounters very little resistance. There is no question of having a sufficient voltage drop to turn ON this pn junction, right? You can have this n+ extend all the way to this n+, so as to cut down the shunt resistance by a large extent. So, there is going to be no question of any regenerative feedback and of course, the best thing will be to have both these - you have a retrograde well, you also have, have it combined with an epitaxial layer.

So, let me just give you in brief, the steps of such a technology. Then, we will show you how in this process, the CMOS technology has actually come pretty close to the bipolar junction transistor technology. So our objectives are therefore, to have a well doping, which should be pretty low. At the same time, we do not want to have any latch up and

remember, there is also another problem that is the subsurface punch through problem. Whenever you are having the well doping low, there is always a possibility that the source and drain depletion regions could merge, because the width of the depletion region will be considerable, if the substrate doping is low, right? So, you do not want that also to happen.

So, you see, you have two conflicting requirements. One is, for the optimum MOSFET performance, the well doping as well as the substrate doping must be low. To prevent latch up, I must not have a high shunt resistance, right? That means the path of the current must encounter little resistance and I must not have a subsurface punch through. So, you see, I can have all these requirements met, provided I have a retrograde well. That is the peak doping concentration is lying inside the surface. So, I have a heavily doped region lying inside the surface, which is going to prevent both subsurface punch through as well as preventing latch up and in order to have the optimal MOSFET performance, I should have an, a low doped well sitting on top of this heavily doped region which I can achieve by epitaxy, right? So, what I am going to have now is retrograde well with epitaxy.

(Refer Slide Time: 24:19)

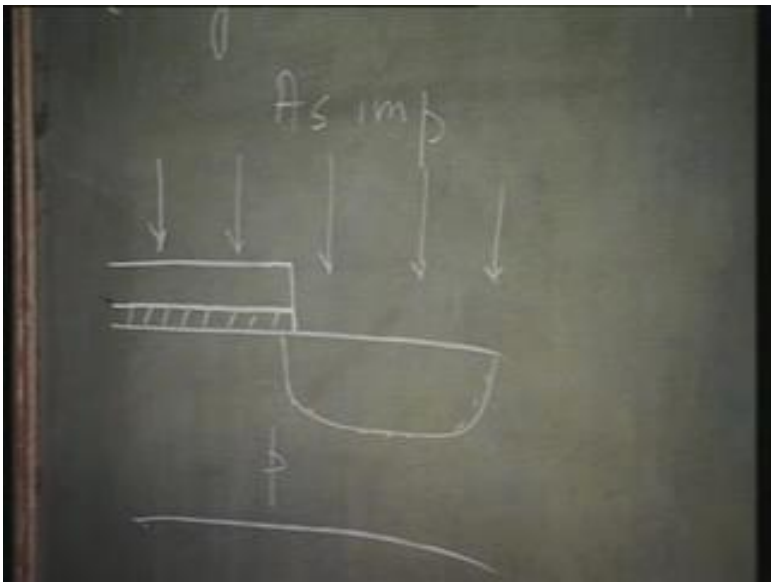


Let me to begin with start with a low doped p-type substrate, low doped p-type substrate. Ok?

(Refer Slide Time: 24:49)

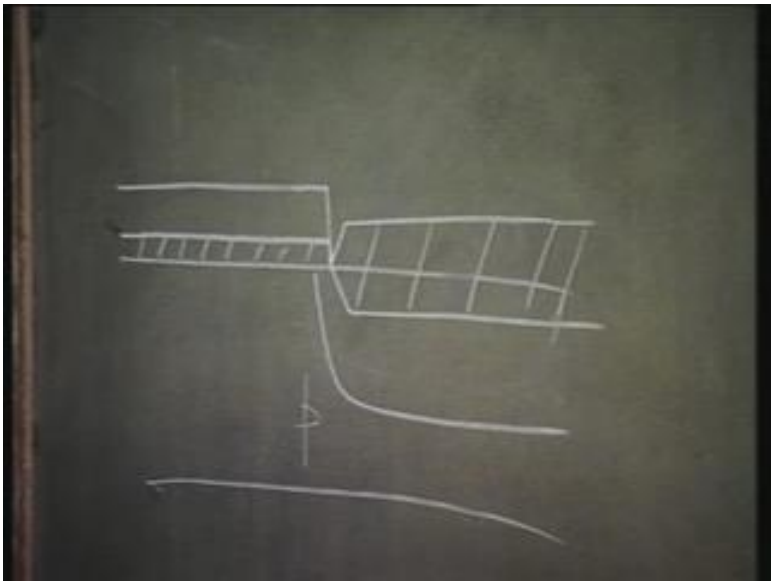
Now, in the first step I am going to have an n well technology. Ok? So, in the first step, I must demarcate the regions for the n well and remember, this is going to be a retrograde well. Ok? So, first of all, I must have an ion implantation to have this heavily doped region lying underneath the n well. Ok?

(Refer Slide Time: 25:28)



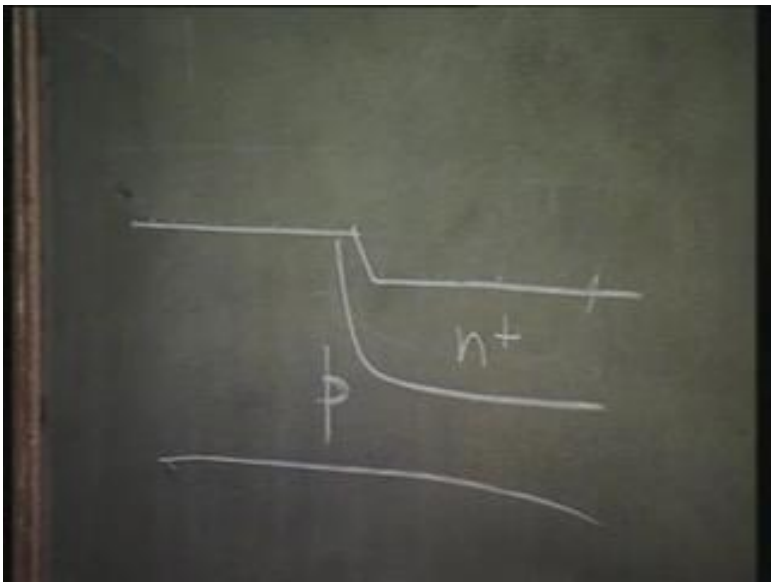
So, to this end, I first protect certain region with an oxide-nitride cap, ok and I do a blanket implant. Ok? One uses arsenic for this blanket implantation, ok because you remember that for all buried layers, you prefer arsenic or antimony rather than phosphorus, because phosphorus will go deep. It actually should not be so deep, so I should have a thinner region, never mind. Ok? Now, after this implantation is over, your next step is to carry out an oxidation, right?

(Refer Slide Time: 26:42)



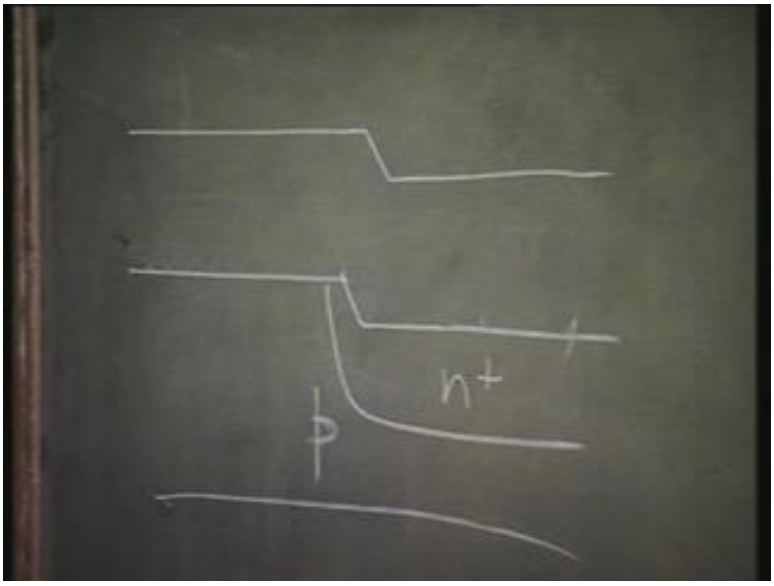
So, I do an oxidation and in the process this arsenic has gone in a little bit, fine. Now, I am going to strip off both the oxide and the nitride.

(Refer Slide Time: 27:18)



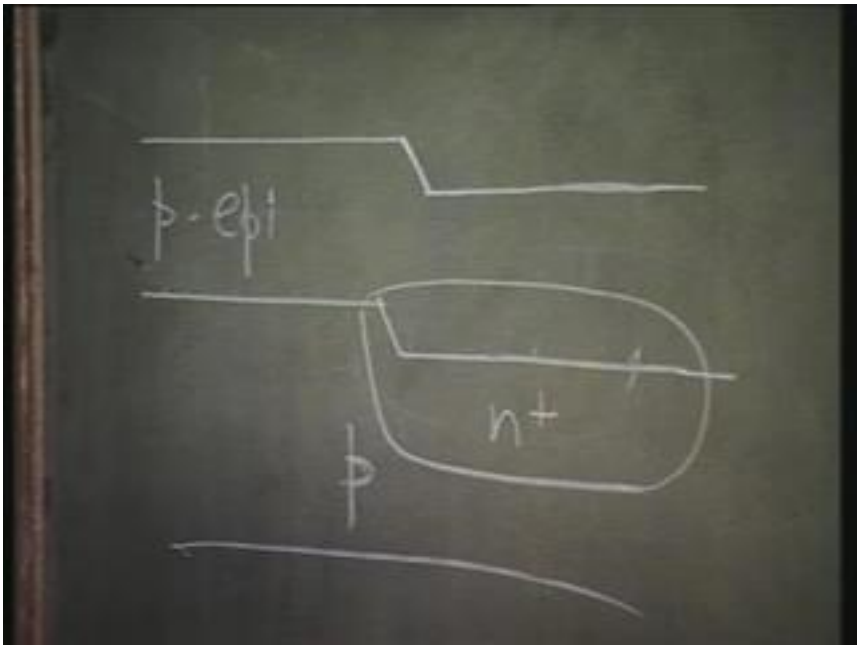
So, this is your n + region; this is the p-type substrate. Ok? In the next step, now you carry out an epitaxy.

(Refer Slide Time: 27:46)



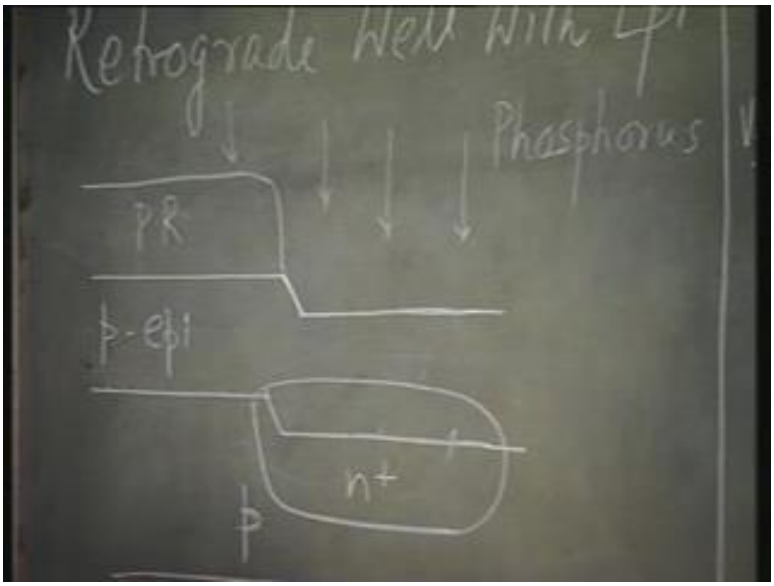
You do a very lightly doped p epitaxy and you know, provided there is no pattern shift I am going to have an epitaxial layer also like this, right? This step will be transferred here provided, of course, I have taken care not to have any pattern shift. So, I have a p-type epitaxial layer.

(Refer Slide Time: 28:07)



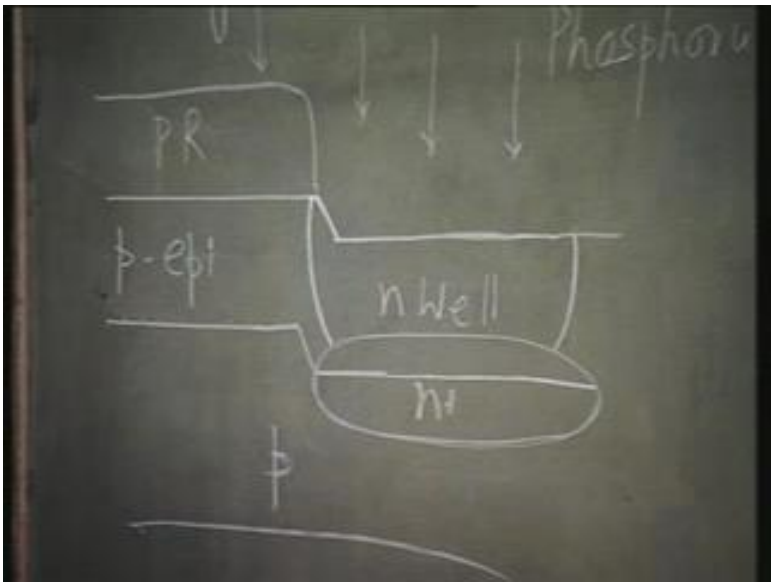
In that, during that process, this will also fill out a bit and I will have this n + buried layer also spread a little bit into the epitaxial region, is it not? Ok. Now, what do I want to do? Now, of course, I have to do my n well implantation, because you see, this is also p-epi layer and this indentation is going to serve, to signify, which is the region where the n well should be formed. This n well must be aligned to the underneath buried n + region. Ok? So, that is all you have to do.

(Refer Slide Time: 29:08)



Next step, you just protect this with photoresist and carry out, this time you do a phosphorus implantation, because you want to have a deep n well.

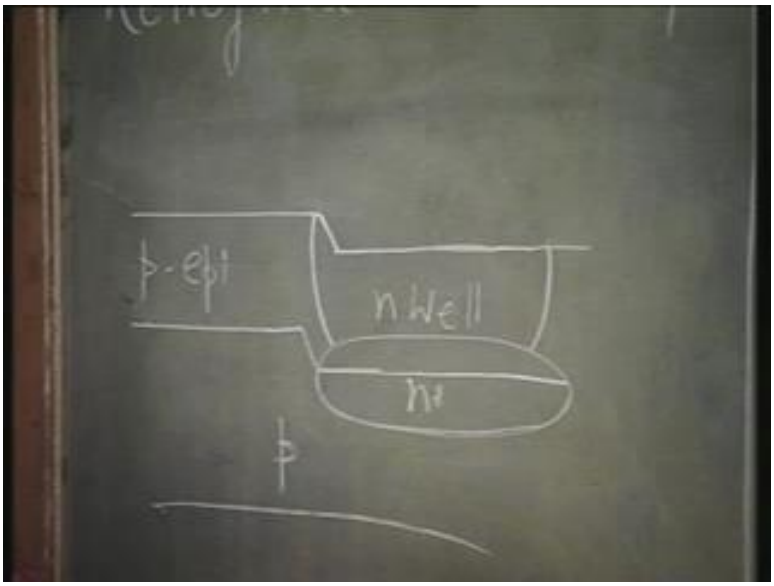
(Refer Slide Time: 29:28)



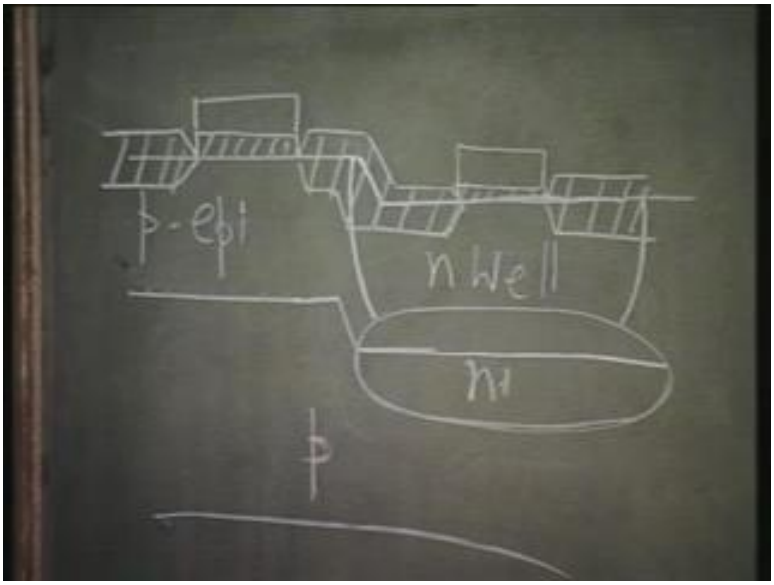
So, you have, let us draw it in a more realistic manner, so that what we have is something like this. So, now you have both your p region, you have your n region, ok and all you

have to do is to form the field oxide before proceeding with the source and drain diffusion. Ok? So, again you have, after removing this photoresist, this is what you have.

(Refer Slide Time: 30:45)

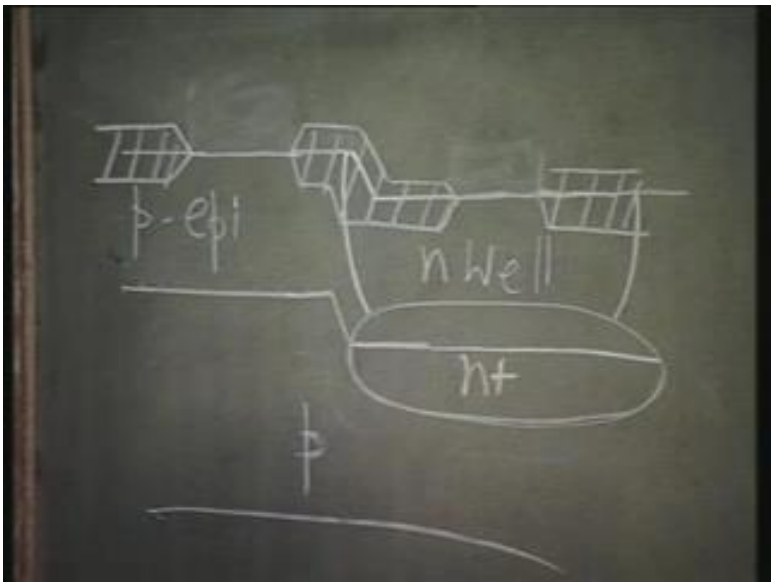


(Refer Slide Time: 30:48)



Now, you define the active regions by the oxide-nitride mask as before, subject it to oxidation.

(Refer Slide Time: 31:38)



Remove the oxide nitride mask. So, this is the region where you have your nMOS; this is the region where you have your pMOS, right and you have the retrograde well. This is the retrograde well, the n+ region lying below the n well, so that the latch up problem will be minimized, as well as the subsurface punch through problem will be avoided. Ok? If you want to be even more particular, you could also add a p+ region here; you could also add a p+ region here. Ok? So, you see, this is actually a more sophisticated CMOS technology, where I have used retrograde well with epitaxy.

Now, let us pause for a moment and see, in this process has the whole thing come closer to the bipolar junction transistor technology? Well, I have n well clubbed with n+ buried layer, which I am going to use for my pMOS, right? I could also use this identical region to form my npn transistor. The n+ buried layer would cut down the collector resistance. The n well with its low doping will serve as the collector. I need to do just another p doping for the base and an n+ doping for the emitter. Now, again you see, this n+ emitter doping, I could club it with the n+ source and drain doping. When I am carrying out the source and drain doping for the nMOS device, I could do the emitter of the npn transistor in the same shot. So, essentially I need exactly one extra step, that of the p base. With this one extra step, it is possible for me to realize an npn transistor in the same

process flow. Well, this is looking at it from really a very optimistic angle. As a matter of fact, it is not just one step. You need to do other things to improve the performance of the bipolar junction transistor, but the basic point remains unchanged; that is it is possible. As the CMOS technology is getting more and more complex, so that we are going to use epitaxial layers and buried layers, it is going to be possible to incorporate a bipolar junction transistor in the CMOS process flow, by only putting in a few extra steps.

But, the point is why would you like to do it? That is because, the technology or the performance of both these devices are nearing the saturation level, right? You have already achieved enormous packing density with CMOS. So, now you see, what happens is to improve the performance of these devices by a factor of 2, the process complexity has to be increased enormously, because we are nearing the saturation level of their performance. But, if we can think of circuits where, you know in a circuit, essentially what do we have? We do not need high speed throughout the circuit, right? We only need high speeds at certain bottlenecks. Ok?

So, if we can conceive of a circuit, where in these bottleneck regions we are going to put in high speed devices, for example, high speed bipolar junction transistors, you know, ECL logic is still the fastest logic and you know, for ECL transistors, we use the poly emitter, poly extrinsic base transistors. Ok? So, suppose I could put in ECL logic in those bottlenecks and the rest of the portions, where I do not need that high speed, I do with CMOS, where I can realize very high packing density and very low power dissipation, because as the circuit complexity increases, you see, finally what puts the limit is how much power it needs to be dissipated in a given area, right? So, there CMOS scores heavily. So, you have therefore, a combination of bipolar junction transistor and CMOS in a circuitry and in doing this, without sticking to one particular technology, but in judiciously using a combination of both these, you can achieve a performance enhancement with little extra cost.

That is to say suppose you can double the performance with may be a cost enhancement of 1.3 to 1.5 times, that is what makes biCMOS so attractive. Ok? So, biCMOS is

essentially going to occupy a niche market, for very high performance circuitry at a competitive cost, right? You are going to use ECL or bipolar junction transistors at the high, at the point where you require high speed and you are going to use CMOS in the rest of the circuit. So, essentially, all biCMOS circuitry, they have this one common principle. That is you have the basic CMOS process flow; you have the basic CMOS process flow and in that you want to incorporate a bipolar junction transistor, it is not the other way round, right?

You want to have a CMOS process; you have a CMOS process flow and you want to incorporate a bipolar junction transistor, npn bipolar junction transistor in that same process flow by using few extra steps.

So, this is essentially a biCMOS technology, which is possible because of the complex requirements in today's CMOS circuitry. That has made the concept of biCMOS feasible. So, in the next class, I am going to give you a biCMOS process flow, ok where you will see how in the basic CMOS technology we can incorporate a bipolar junction transistor, what are the drawbacks of this basic process flow, basic bipolar junction transistor, because you see, you are just incorporating a few extra steps, so the bipolar junction transistor you are getting may not be very good. Ok? So, how to fine tune the performance of the bipolar junction transistor and then we will see finally, what emerges as a biCMOS process flow.