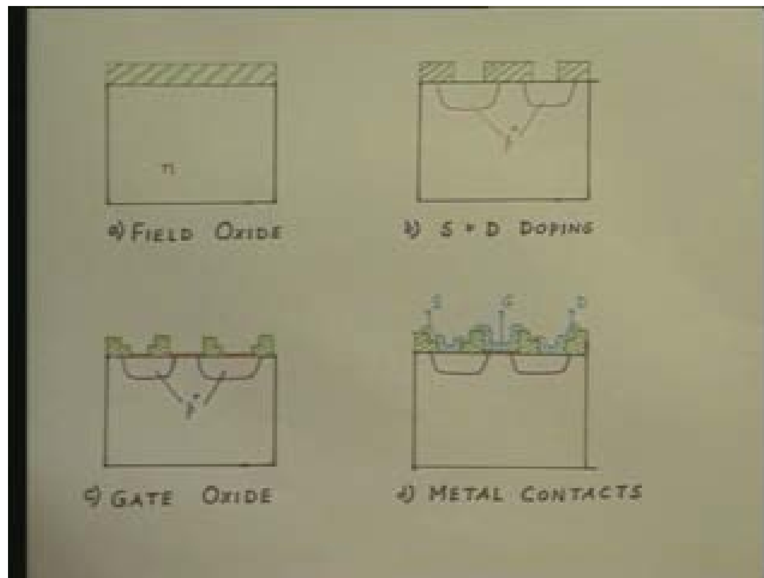
VLSI Technology
Dr. Nandita Dasgupta
Department of Electrical Engineering
Indian Institute of Technology, Madras

Lecture - 2
Bipolar Junction Transistor Fabrication

In the previous class, we have discussed the fabrication steps necessary to realize a bipolar junction transistor in an integrated circuit. In today's class we are going to see the other branch of the VLSI technology, namely the fabrication of a MOS field effect transistor. You know that the MOS field effect transistor is widely used particularly in digital logic circuits and as memory devices. That is because, we need very high packing density for the memory. If you remember, last class I showed you one chip from IBM; that was a 4 mega bit dynamic RAM, right. That means 4 million transistors in that chip. So, to realize that kind of high packing density, we have to go for a MOSFET.

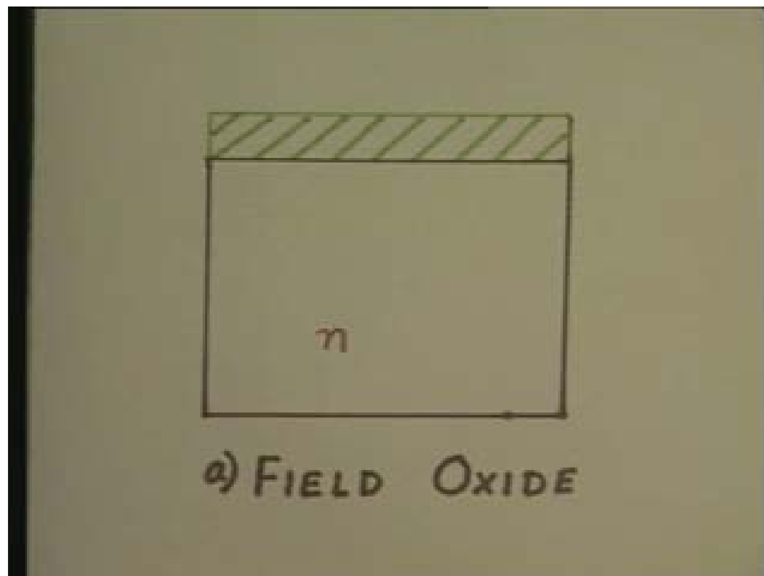
Now, we will discuss the fabrication steps of the MOSFET in a minute and also, let me tell you that you need fewer and simpler processing steps in order to realize a MOSFET compared to a bipolar junction transistor; lesser number of steps and also complicated steps. At least that is what it will seem when you look at the MOSFET fabrication technology. But, it is very difficult to precisely control the characteristics of the MOSFET with such simple steps. In fact, that is one reason why even though the theory of MOSFET was predicted by shortly even before than the theory of bipolar junction transistors, it took considerably long time for the MOSFETs to come in the market. That is because, even though on the face of it, when you look at the individual processing steps you find there are less number of steps, but it is very difficult to gain precise control over the MOSFET characteristics. We will discuss this in detail in today's class.

(Refer Slide Time: 3:48)



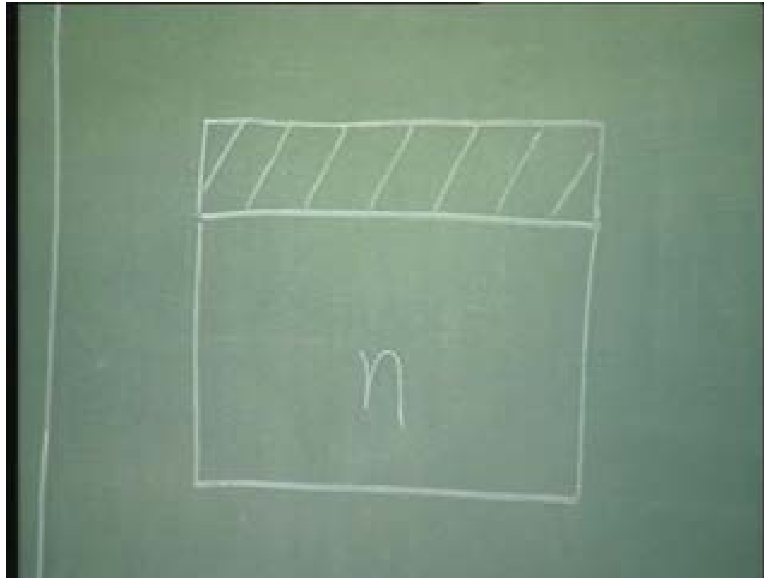
First of all let us look at card one and you can see the fabrication steps of the MOSFET.

(Refer Slide Time: 3:54)



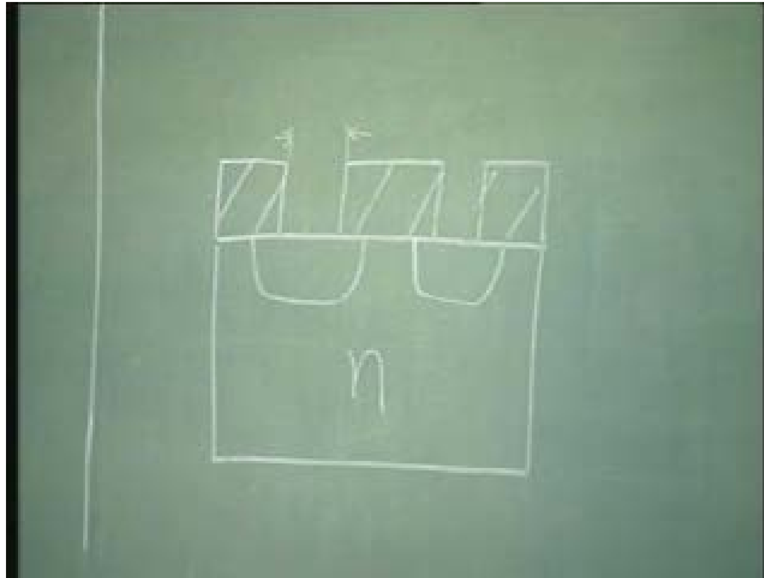
You see in the figure a, we have started with a p-type substrate.

(Refer Slide Time: 4:05)



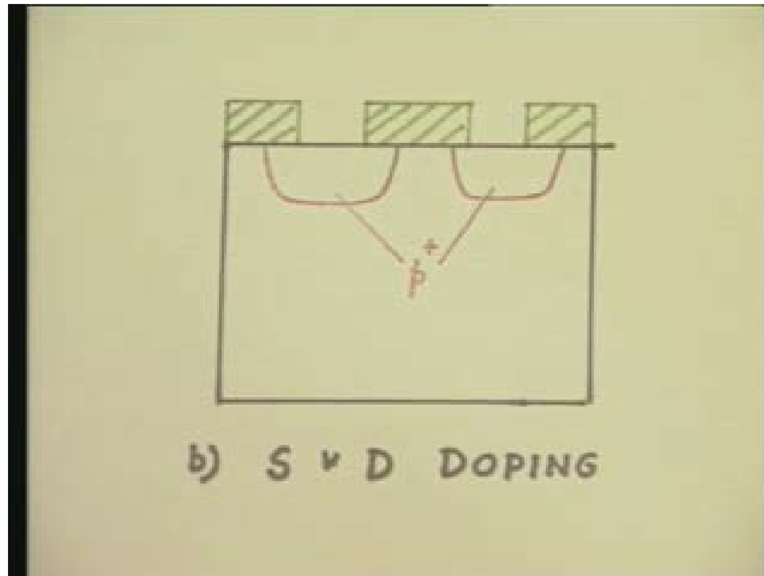
I am merely replicating on the board what is already there in the card. So, we have started with a, let us start with an n-type substrate; I think that is easier. We have started with an n-type substrate. First step is identical to what we have done even for a bipolar junction transistor that is you grow an oxide. This step is called the field oxidation. You can see in the card, it is written as the field oxide - step a. That is the first step in the fabrication of the MOSFET is the growth of a thick field oxide. Once this thick field oxide is grown next step is photolithography and opening windows in that oxide. So, what do I do?

(Refer Slide Time: 5:26)



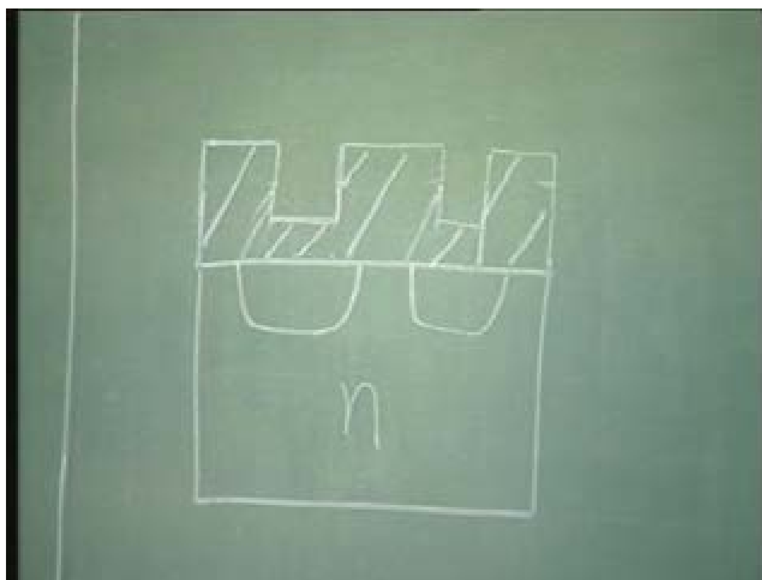
We have opened windows in this oxide and then through these windows you do the doping of source and drain. I want to draw your attention to one point here. If you look at the figure carefully, you will find that the window is here. This is the dimension of the window. But, when I dope through this window notice that the doped region extends laterally to some distance. So, you cannot really dope. When you are doping through a window, you cannot really dope precisely. According to the dimensions of the window there will always be some lateral encroachment and this lateral encroachment will be more and more if your junction is deeper. For a shallower junction, the lateral encroachment will be comparatively less. For a deeper junction the lateral encroachment will be more.

(Refer Slide Time: 7:13)



This is actually the second step or step b in the card. So, you have now your source and drain doping. You have opened windows in the field oxide and you have done your source and drain doping. After this step is over, usually there is another oxidation step in order to build up the oxide over the source and drain. So, that will look something like this.

(Refer Slide Time: 7:51)



MOSFET has, apart from source and drain, a third contact. That is the gate contact. So, I need to realize the gate. You know the MOSFET gate is the heart of the device. That is this is where we have the metal oxide semiconductor structure. In order to realize the gate, first of all what we have to do is to do the gate oxidation. So, here comes a question, I already have an oxide between the source and drain, can this oxide serve as the gate oxide of the MOSFET? The answer is no, it cannot.

There are various reasons why this oxide, the thick field oxide, which is already present here cannot serve as the gate oxide. The first reason being that the thickness of this field oxide is very high. You see, the purpose of this field oxide when we opened the windows for source and drain was to mask against source and drain doping. We wanted to selectively dope the silicon substrate and for that reason the field oxide thickness must be quite high. In fact, the thickness of this field oxide will be something around 8000 angstrom; 0.8 micron to 1 micron may be.

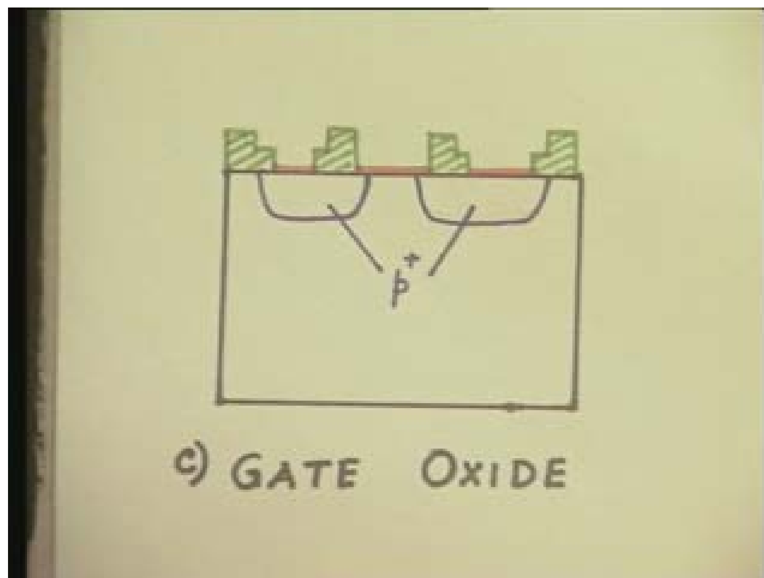
So, this field oxide thickness is very high and the gate oxide thickness of your MOSFET will be much less. The gate oxide thickness has a direct bearing on to the threshold voltage of the MOSFET. So, for modern day MOSFETs, when we talk about very small dimension, very small geometry MOSFETs, then we also talk about the gate oxides of thickness say, may be 100 angstrom or even lower; even lower than that. Even, where you have a channel length of 1 micron, even for a channel length of 1 micron, you would like to have a gate oxide may be of 200 angstrom thick, 200 angstrom thick and even for the older day MOSFET technologies, even for that the gate oxide thickness was never higher than 1000 angstrom.

Please note the difference in the order of magnitude. On one hand, your field oxide is very thick. It is about 8000 angstrom or may be even more than that. On the other hand, the gate oxide is approximately one tenth of that. That is one reason. The second reason is the quality of the oxide. The gate oxide of the MOSFET is taking part, taking an active part in the device performance. It is not merely used as a mask. The quality of that oxide and the oxide and underlying silicon interface is very, very crucial in determining the

performance of the MOSFET. So, the gate oxidation step in the MOSFET fabrication is an extremely critical step, very stringent quality control is required and when we discuss the oxidation as a unit process step, we would like to highlight how the quality of the oxide can be improved.

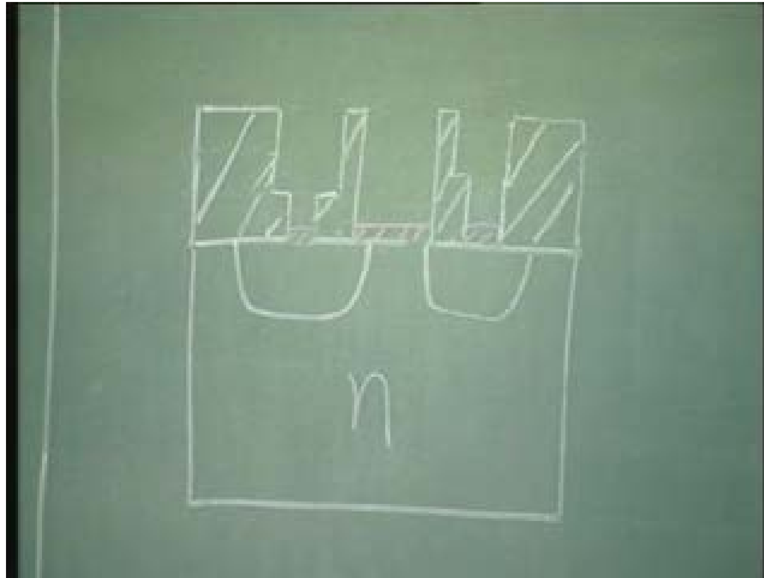
For the time being, let me only say that the quality of the gate oxide is an extremely important parameter, stringent quality control is needed and under no circumstances you can use the field oxide as the gate oxide. You cannot use the same processing step for the gate oxide. In fact, the quality of the gate oxide was also one reason why it took more time for the MOSFET to be realized. Even though its theory was already known, because of the non idealities in the gate oxide it took a long time for the MOSFET to be realized in practice.

(Refer Slide Time: 13:39)



So, the next step will be step number c in our card that is that of gate oxide. You can see the thin pink line. It shows the difference in the thickness between the field oxide which is green hatched in color in the card and the thin pink gate oxide over the, between the source and drain and also over certain regions in source and drain.

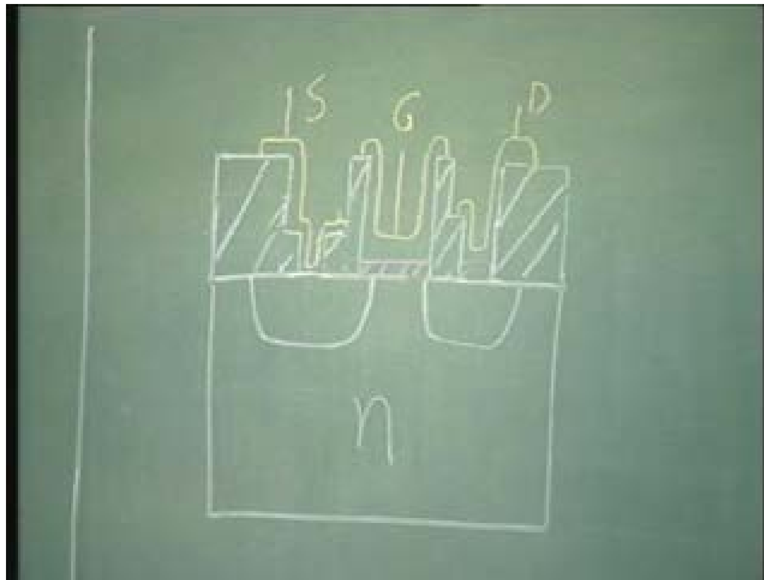
(Refer Slide Time: 14:12)



That is the figure actually looks something like this and then we have the gate oxide over here and also over here. That is we have opened windows in the gate region and also over the source and drain and then subjected it to a process of very stringent quality oxidation and grown a thin layer of oxide over this gate region and also over the source and drain and in the final step once you have realized your gate oxide, your job is almost done and in the final step all you have to do is to realize the metal contacts.

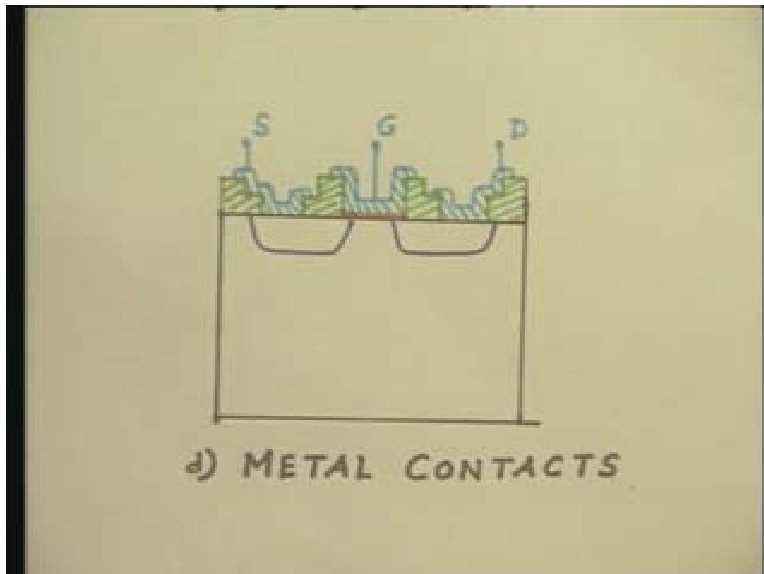
The metal must be on top of your gate oxide. That will be the gate contact and it should be in contact with silicon in source and drain region.

(Refer Slide Time: 16:18)



So for that, we need to open a window here, open a window here and then put the metal which is aluminum. This is the source contact, this is the gate contact and this is the drain contact.

(Refer Slide Time: 17:01)

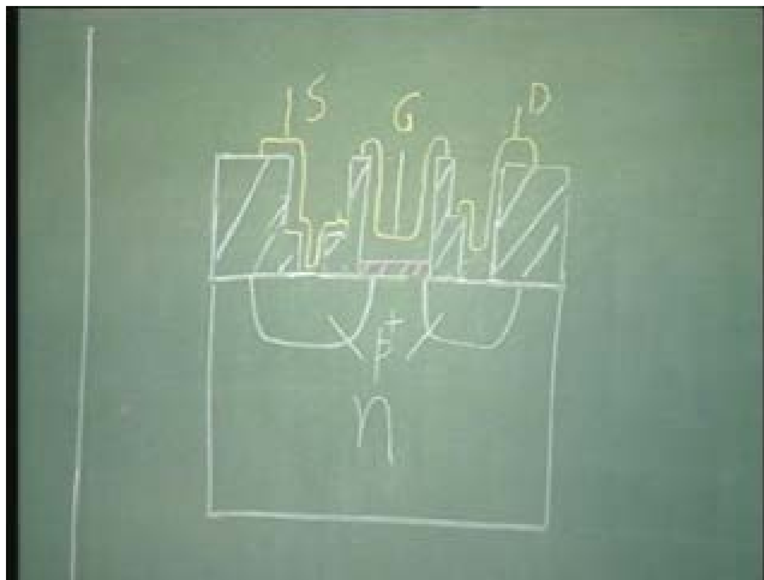


So, you have realized your MOSFET. Compared to the fabrication of a bipolar junction transistor technology, realizing a MOSFET is really very simple. Essentially we have just

four processing steps - that of field oxidation, that of source and drain doping, that of gate oxidation and that of contact metallization. Yes. But as I said, with these four simple process steps it is very difficult to control the quality of the MOSFET, to control the performance of the MOSFET. Why is it so?

Let us, let us look at the performance of the MOSFET, performance figure of the MOSFET. What are the factors which determine the performance of the MOSFET?

(Refer Slide Time: 18:18)



See, in this particular set of technology steps, I have chosen the substrate to be n-type. That means my source and drain doping was p plus and in other words, this MOSFET is a p MOSFET or a p-channel MOSFET. Now of course, the question that comes first to the mind is why did I choose a p MOSFET and not an n MOSFET. As you all know, MOSFET is a uni polar device. That is the majority carriers are the only carriers here. So, it makes more sense to have a n MOSFET, because electrons are the carriers there. Electrons have higher mobility, therefore the current drive of the n MOSFET will be much higher. Why then did I choose a p MOSFET? Let me tell you, it was not an arbitrary choice. I did it deliberately simply because it was the p MOSFET which was

fabricated first in the integrated circuit technology, even though everybody was aware that an n MOSFET will be potentially a better performer.

Why was it? You know, a MOSFET can be an enhancement type MOSFET or a depletion type MOSFET.

(Refer Slide Time: 19:43)

Enhancement mode MOSFET are normally OFF MOSFET, depletion type MOSFET are normally ON MOSFET. What do I mean by these terms? I mean that if it is an enhancement mode that is normally OFF, which means when no gate voltage is applied this MOSFET is not conducting, this MOSFET is OFF. I have to apply a gate voltage in order to make the MOSFET conduct. In contrast, a depletion mode MOSFET is conducting when no gate voltage is applied and I have to apply a gate voltage in order to switch it OFF.

MOSFETs are particularly useful in digital circuits and there you will want the device to be normally OFF. That is where no gate signal is coming, the device should be normally OFF. Otherwise if you are using a depletion mode type MOSFET, you are wasting a lot of power. The device when it is idling, it is conducting. Therefore, you are wasting a lot

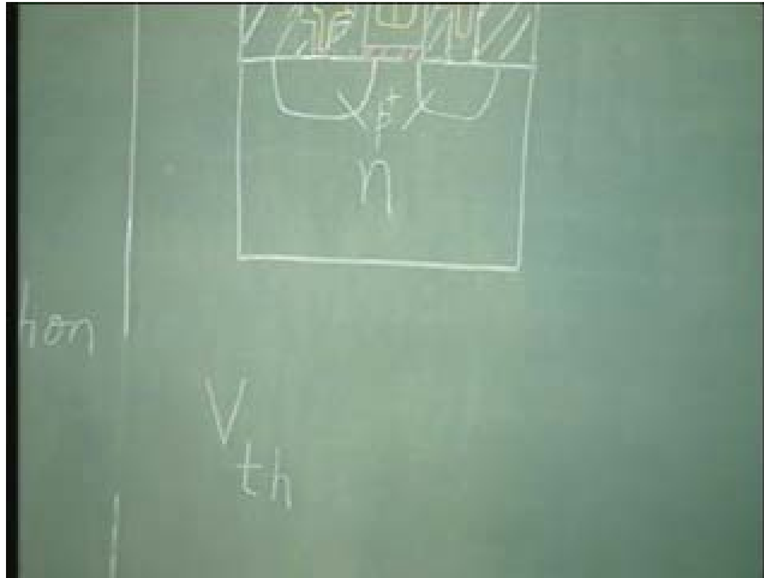
of power. So, for digital applications primarily, we prefer to have enhancement mode type MOSFET.

Now, let us look at it from an n MOSFET and from a p MOSFET point of view. You know, if I have an n channel MOSFET, n MOSFET, if it is an enhancement type MOSFET that means where no gate voltage is applied, gate voltage is zero, it is not conducting and when I apply a positive gate voltage, it starts conducting and we call that gate voltage to be its threshold voltage at which it starts conducting. So, ideally speaking, an n MOSFET, if it is an enhancement type MOSFET, it should have a positive threshold voltage. If it is a depletion mode type MOSFET, then it is conducting when no gate voltage is applied. That is at gate voltage equal to zero then MOSFET conducts and in order to switch OFF, I have to apply a negative voltage to the gate; agreed.

In other words, the depletion mode n MOSFET should have a negative threshold voltage. Till a negative, that particular voltage is applied on the gate the MOSFET conducts. When we apply a gate voltage more negative than the threshold voltage, it shorts out; agreed and when we come to a p MOSFET, you see, for a p MOSFET the gate voltage supply is negative. So, an enhancement type MOSFET in order to make it conduct, you have to apply a gate voltage, a negative gate voltage in this case and if it is depletion mode type, we have to apply a positive gate voltage. So far so good.

Now, the problem is this. For, particularly for digital circuit applications, you would like to have enhancement type devices. That means if you want n MOSFET, you would want a positive threshold voltage. If it is a p type MOSFET, you will want a negative threshold voltage for enhancement mode type devices.

(Refer Slide Time: 24:10)

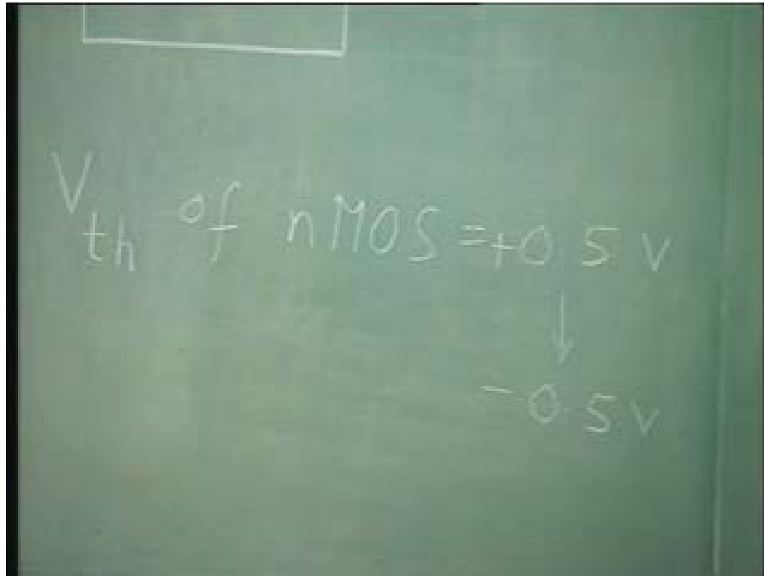


Now, the threshold voltage of a MOSFET is going to depend on various factors. It is going to depend for example, on the gate oxide thickness, on the substrate doping concentration, on what kind of metal you put on the gate, so many things. These things are to some extent in our control. But, there are certain things which are not in our control, process related things, which are very difficult to control and these are unwanted charges in the oxide. Ideally speaking, the gate oxide should be a dielectric, but during the oxidation of silicon, there are, there are possibilities that some charges will be introduced in the oxide.

It may be due to incomplete oxidation, it may be due to presence of mobile ions like sodium, potassium. The net effect of these charges in the oxide, usually all these charges are positive, that you have a surplus of positive charges in the oxide which ideally should not be there, ideally it should be zero, but now you have some positive charges sitting in the oxide. In order to counter balance this and bring it to the ideal condition, you have to therefore apply negative gate voltage. So, the net effect of having positive charges in the oxide is to push the threshold voltage towards the negative direction, because you have to apply a negative voltage at the gate.

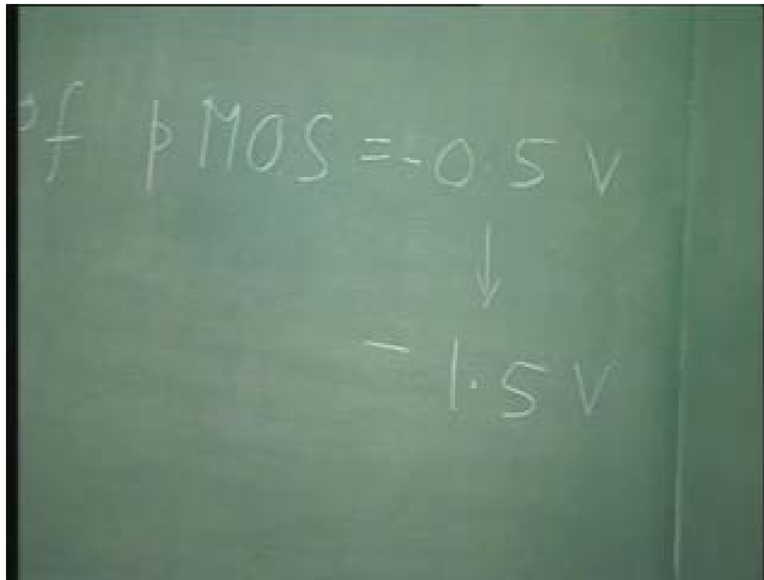
In the light of this, now let us see what is going to happen to my enhancement mode n MOSFET. Ideally speaking, it should have had a positive threshold voltage.

(Refer Slide Time: 26:37)



Let us say hypothetically that the threshold voltage of my n MOSFET, it is a hypothetical value, let us say it is 0.5 volts, hypothetically speaking. Now, if you have a large density of positive charges in the gate oxide, the effect is to push the threshold voltage towards the negative direction, so that it can change its sign and become, let us say, minus 0.5 volt, if you have pushed it by 1 volt towards the negative direction. What do you have? You opted for an enhancement mode type device, you have now got a MOSFET with a negative threshold voltage that means it has become a depletion type device. All your circuit goes haywire.

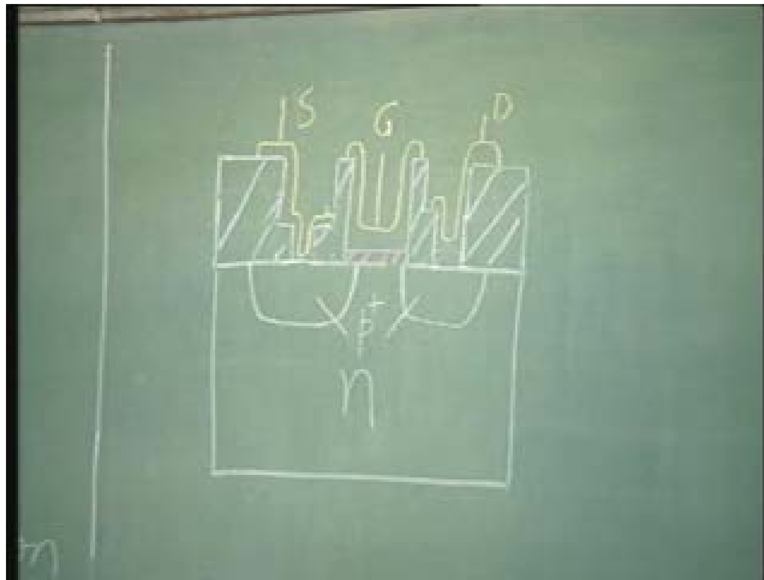
(Refer Slide Time: 27:47)



In contrast, let us think about the threshold voltage of a p MOSFET, ideal hypothetical enhancement mode p MOSFET. It should have a negative threshold voltage and again as before, I have a large density of positive charges in the oxide and therefore the threshold voltage got shifted more towards the negative direction and it has probably got changed to by the same amount, minus 1 volt. But now, I have the value of the threshold voltage as minus 1.5 volt. Your circuit will probably still perform, because you still have an enhancement mode device. The only difference is your threshold voltage had got shifted.

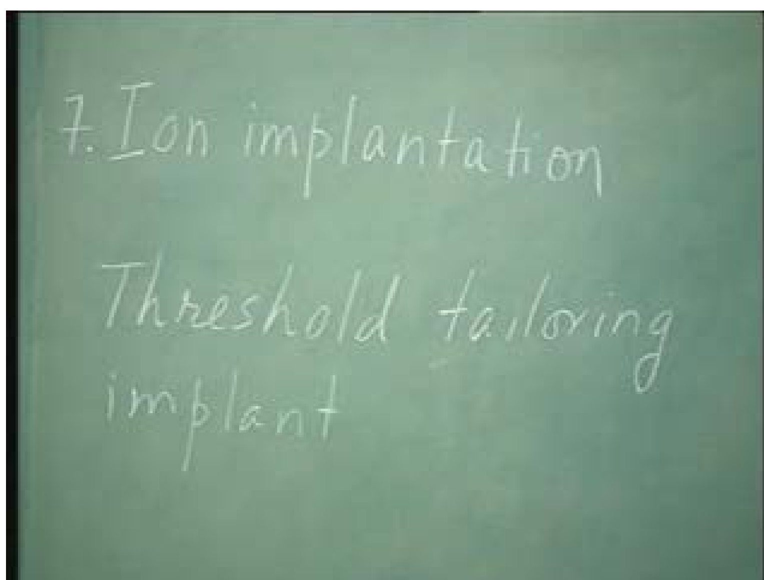
In the earlier days, when the technology was not quite so advanced and one did not really have stringent control over the quality of oxide, people could not control this positive charges in the oxide.

(Refer Slide Time: 29:19)



It was a safer idea, safer bet to go for an enhancement mode p MOSFET and that is why the first MOSFET, cross sectional diagram I have shown you, is on an n type substrate with p plus source and drain. That is it is a p MOSFET. With the advance of technology it became easier to control the oxide charges and with that came the n MOSFET. Not only that, we have in our hands now a sophisticated doping technique called ion implantation.

(Refer Slide Time: 30:02)

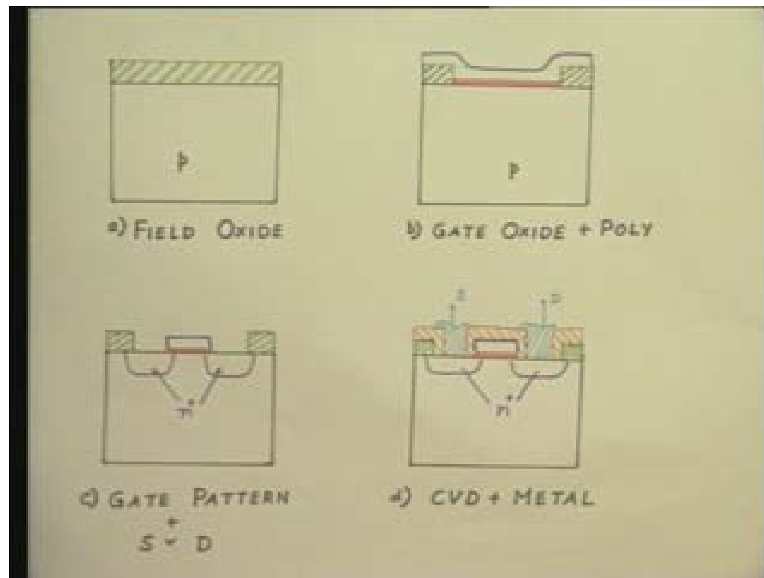


So, ion implantation actually plays a very important role in today's MOSFET technology. Like I told you that the threshold voltage of a MOSFET depends on various factors and one of them is the substrate doping concentration. Ion implantation is a very sophisticated doping technique by which you can locally alter the doping concentration. So now-a-days, it is a common practice; it is in fact one common step in MOSFET fabrication to have what is known as a threshold tailoring implant, a threshold tailoring implant. This implantation is done only in the channel region of the MOSFET; no where else, only in the channel region of the MOSFET very close to the surface, just in order to tailor threshold voltage, so that you can realize the threshold voltage of your choice.

This would not have been possible with the other less sophisticated doping techniques like diffusion. It is possible with ion implantation. So, that brings us to our processing step number 7 that is ion implantation. You see, even though it looks on the face of it that the MOSFET technology is simpler, it is in fact not so. It is true that you need lesser number of steps, but these steps have to be more carefully controlled and there are hidden depths in this. Like for example, the threshold tailoring implant. You can fabricate a bipolar junction transistor totally without ion implantation technology, it is possible. But, it is almost impossible to have a MOSFET technology without ion implantation. Now, even though the device is called a MOSFET - metal oxide semiconductor field effect transistor, nowadays in most cases, the metal that is the gate metal has been replaced with a poly silicon layer.

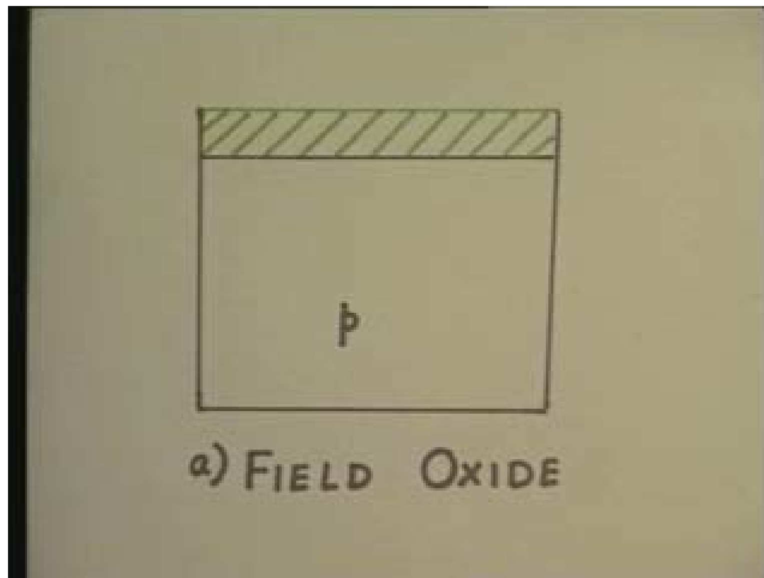
We will now look at the poly silicon gate MOSFET technology which is our card number 2.

(Refer Slide Time: 33:12)



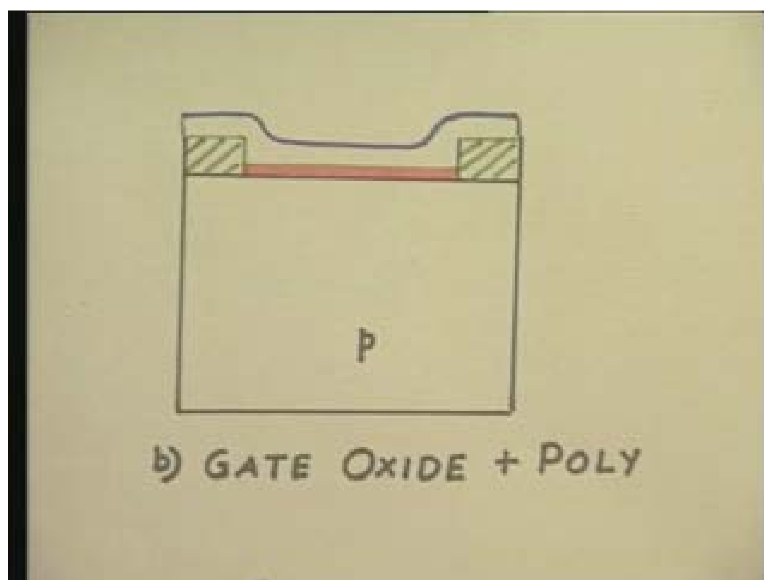
Yes, threshold tailoring implantation it is done only in the channel region, in order to realize a particular threshold voltage. We will discuss more about the threshold tailoring implant when we discuss ion implantation technology. For the time being, the only point I want to stress is that ion implantation has become a very useful processing step in MOSFET technology. You can, you can perhaps do without ion implantation in a bipolar junction transistor, it is possible. It is not impossible to realize a bipolar junction transistor without ion implantation, but it is, it is impossible to do it in a MOSFET technology. That is the only point I wanted to make here.

(Refer Slide Time: 34:41)



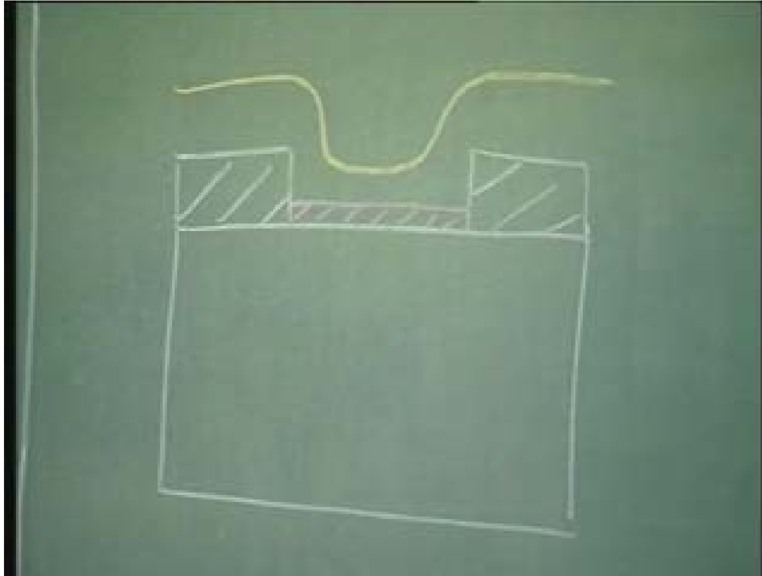
Now, if you look at the poly silicon gate MOSFET technology, looking at card number 2, you can see that the first step is same; that is the field oxidation. Notice here I have changed the substrate from n to p-type that is from metal gate to poly silicon gate the technology has advanced so far as to allow you to fabricate n MOSFET with proper threshold control. That is the significance. So, first step is the field oxide.

(Refer Slide Time: 35:22)



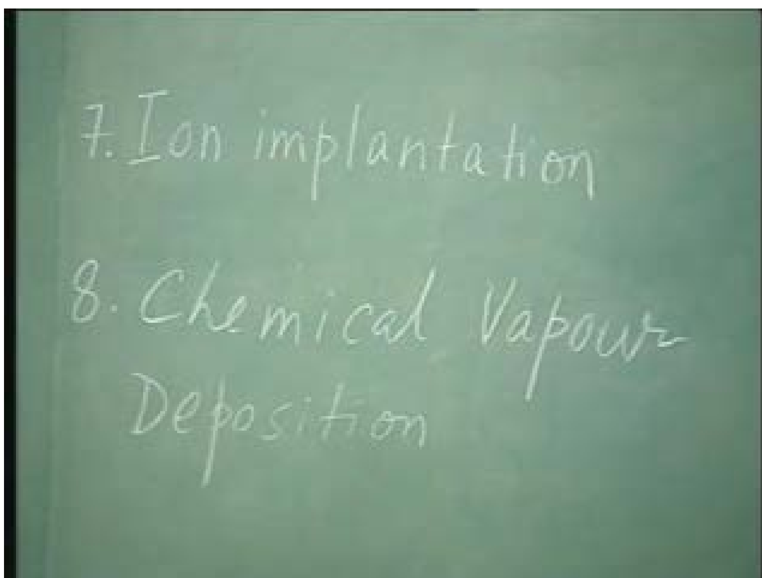
In the second step, that is in the step number b, we have actually opened a window in this field oxide, opened a window in this field oxide and then subjected it to the gate oxidation and on top of this, we have deposited a poly silicon layer.

(Refer Slide Time: 35:32)



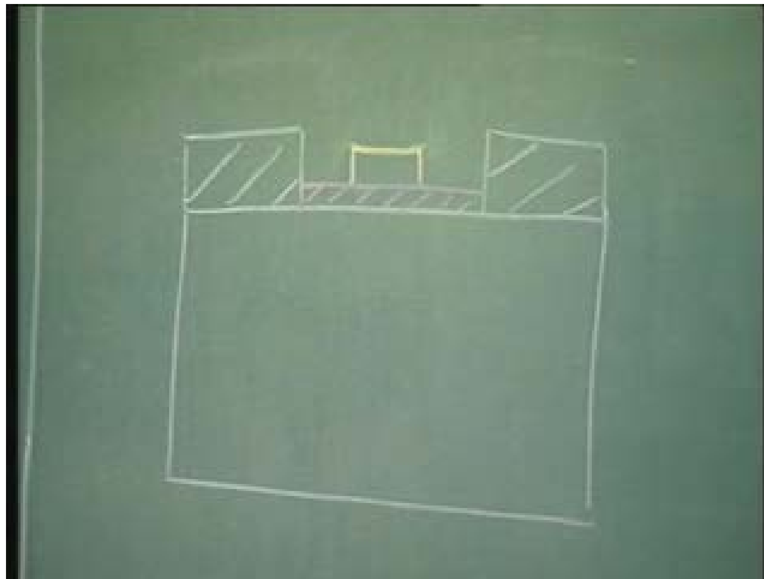
This brings me to our processing step number 8 that is chemical vapor deposition.

(Refer Slide Time: 36:22)

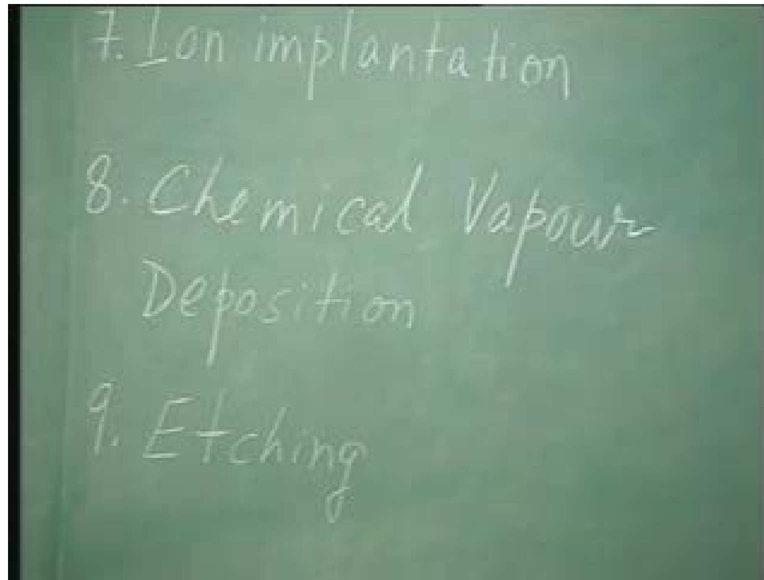


We will discuss why we choose to go from a metal gate technology to a poly silicon gate technology, but first let us look at the processing steps. So, you have by chemical vapor deposition or which is commonly called the CVD, you have deposited a layer of poly silicon on top of the silicon. Now, the next step is going to be patterning of the poly. You want to retain the poly silicon only over the gate and want to remove it from the rest of the region. We do that by means of etching.

(Refer Slide Time: 37:42)

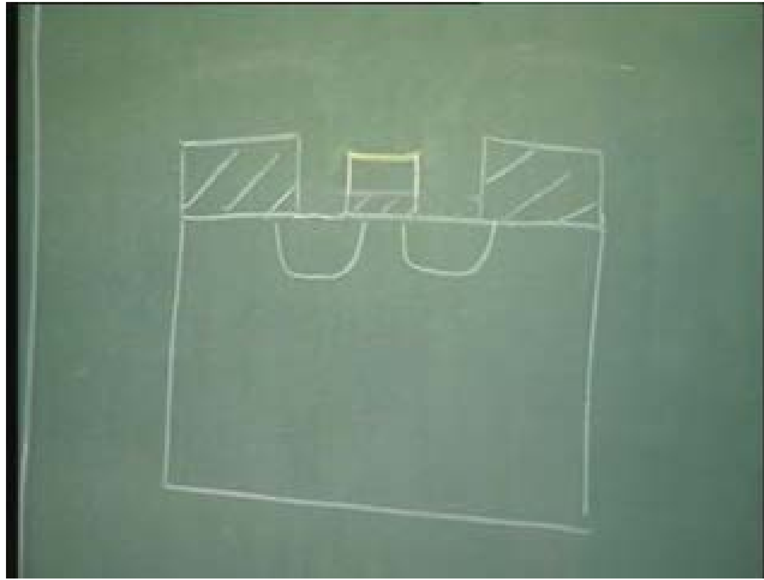


(Refer Slide Time: 37:53)



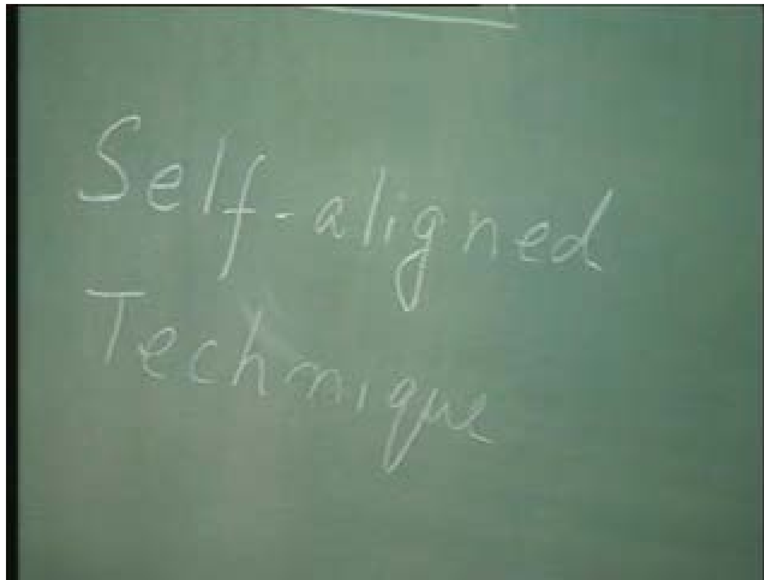
So, I have patterned the poly and retained it only over the gate top. Now I have, notice the important difference between the metal gate technology and the poly silicon gate technology. In the metal gate technology, we had first doped the source and drain region and gate oxidation was the pre final step. After that, immediately afterwards, we just deposited the metal and that was the end of the process. Now, however the gate oxidation is done much earlier to the source and drain doping. In fact, this is one of the earlier steps ...... to do the gate oxidation and immediately follow it up with poly silicon deposition. The advantages we will discuss later.

(Refer Slide Time: 38:59)



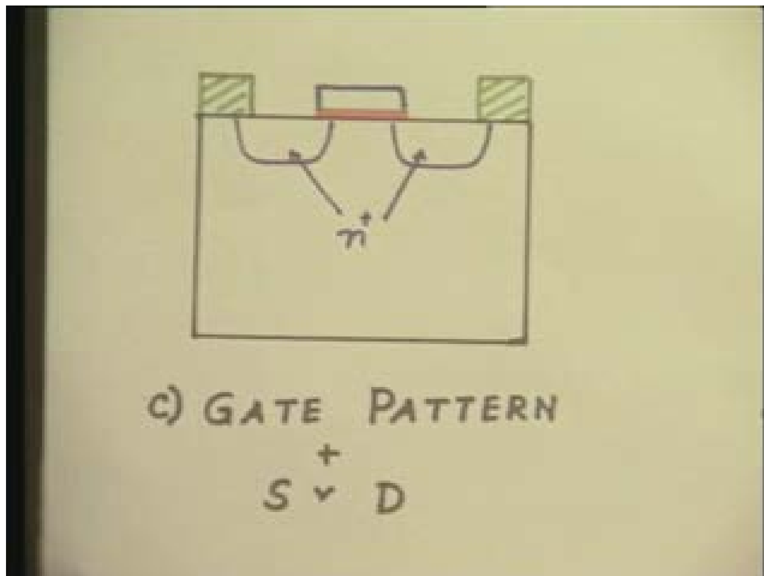
So, all you have to do now is to realize the source and drain. So, we open windows. It is very easy to open windows now. You do not even need a masking step, because you see your field oxide is much thicker than your gate oxide. The gate oxide in this region is protected by the poly silicon region. So, just a quick dip in hydrofluoric acid will remove the thin oxide from this regions, while it will not materially affect the thicker field oxide. Thicker field oxide will not get etched so easily, while the thinner oxide over this region will get etched quite fast and now you do not need any mask. Let me repeat it again and again, you do not need any mask now in order to realize your source and drain.

(Refer Slide Time: 39:56)



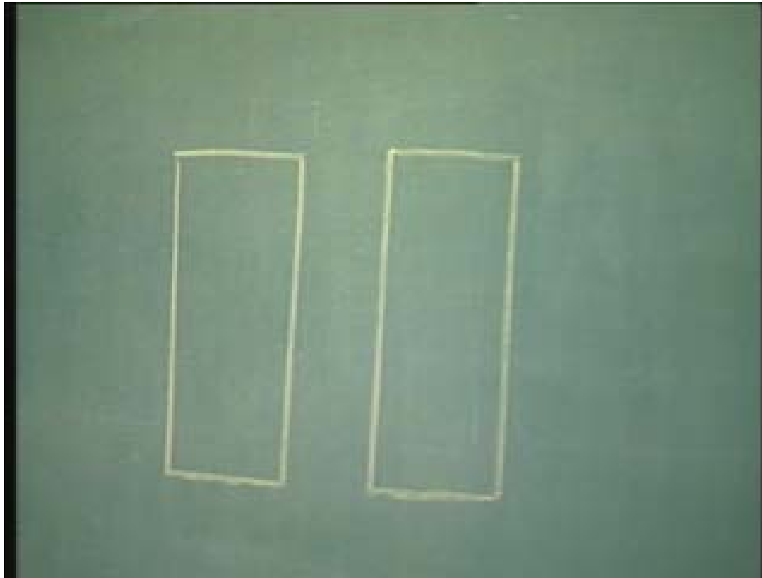
This is called a self aligned technique where the source and drain are automatically aligned to the channel region of the MOSFET. This will become clearer when you look the card number 3.

(Refer Slide Time: 40:25)



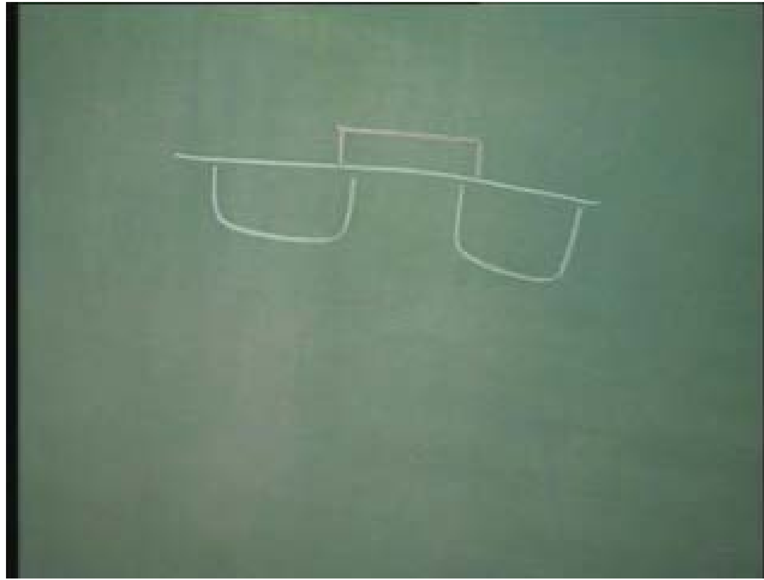
You see, in the metal gate process, your first mask was the source and drain diffusion or implantation mask. The first step was realizing the source and drain region.

(Refer Slide Time: 40:49)



So, the masks looked like this. For each transistor, you have one source and one drain. Rest of the portions are all covered by the thick field oxide. You have opened only two windows in that oxide through which you are going to dope your source and drain. After you have doped this source and drain, your next masking step was the gate mask. It is very important that the channel extends all the way from source to drain. So, what does it mean? That the gate oxide must overlap the source and drain region. In other words, what I am trying to say is something like this.

(Refer Slide Time: 41:47)

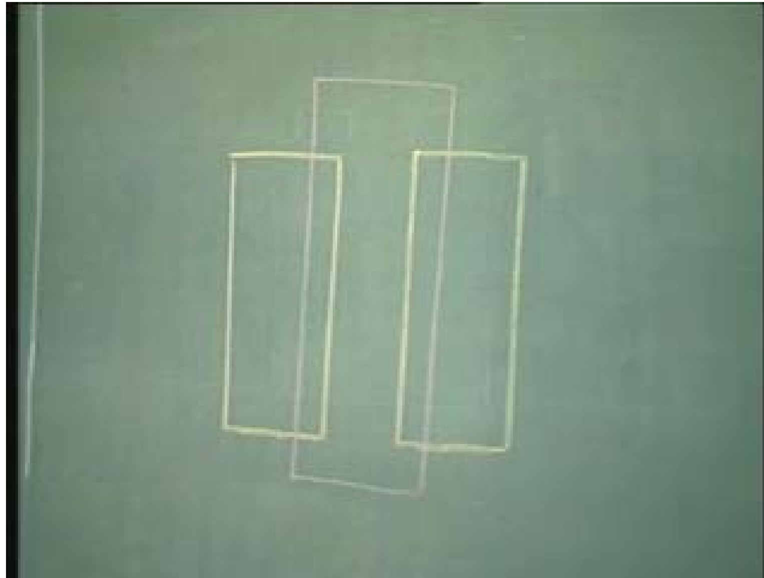


If I have my source here and drain here, then the gate oxide must extend from here to here. It should not be shorter than this. For example, it should not be like this.

(Refer Slide Time: 42:10)

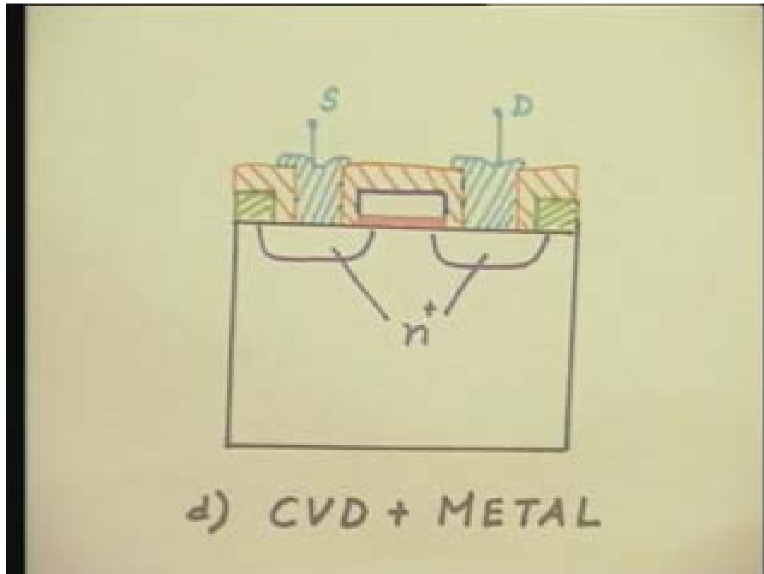
Then, I will not have a continuity of the channel. It is imperative that the channel extends all the way from source and drain.

(Refer Slide Time: 42:47)



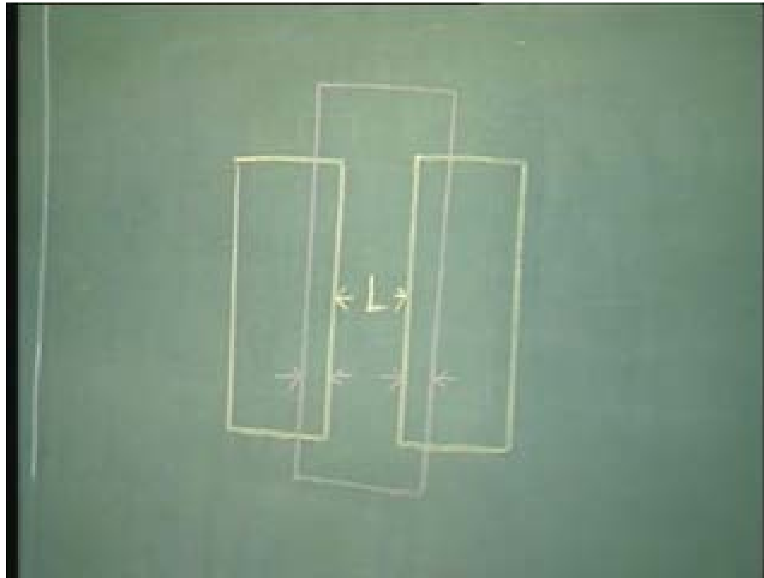
So, in order to make sure that the channel extends all the from source to drain the mask for the gate oxide should be something like this. Is it not? It must overlap the source and drain region.

(Refer Slide Time: 43:01)



For the metal gate technology, I must make sure that the gate oxide overlaps the source and drain region. But, this has its own associated problem.

(Refer Slide Time: 43:26)

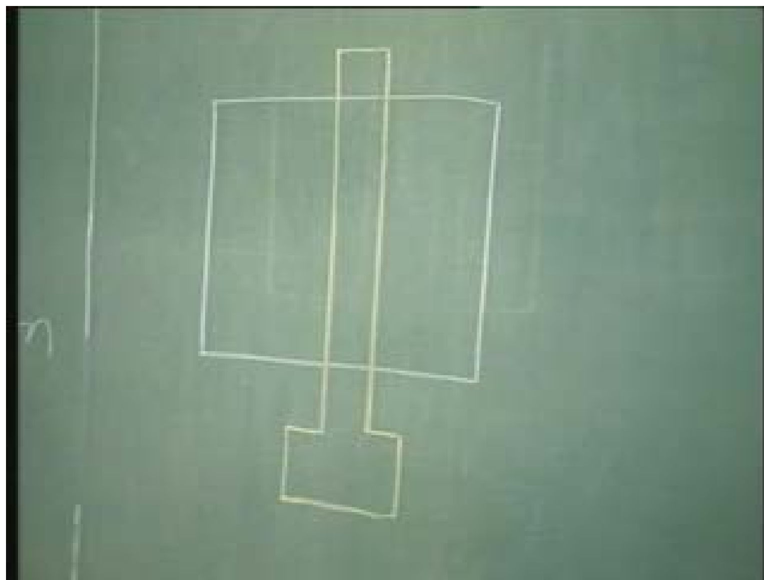


First problem is as we shrink the device dimensions, you see, this is the actual channel length; this is the actual channel length. But, I must have an overlap here and also an overlap here. That means that the second mask, the pink mask, the mask for the gate must be precisely aligned to the first set of masks. You already have the first mask pattern etched on silicon. So, the second mask, before doing the gate oxidation, must be critically aligned to the first pattern, so that I have this overlap in both the directions. For example, if by mistake your pink mask gets shifted towards right or left, you will not have a continuity and you can appreciate that as the device dimension becomes smaller and smaller, the overlap also becomes smaller and smaller and it becomes more and more difficult to critically align this mask. So, alignment becomes a very crucial factor particularly when you are going to devices with a channel length of 1 micron or less.

In a metal gate, this problem comes about because, we have to do the source and drain doping before the gate oxidation. The gate oxidation has to be the pre final step immediately followed by metallization. This is because we use aluminum for all the

contact metallization and you know aluminum has a very low melting point. So, after you have put aluminum, you cannot do any high temperature processing and unfortunately all the, almost all the processes in integrated circuit technology needs high temperature. So for metal gate technology, gate oxidation has to be one of the last steps. In contrast, in the poly silicon gate technology, because now instead of metal, poly silicon is used and silicon has a very high melting point, you can do all the processing steps afterwards. That is why we can have, we can start with defining the gate.

(Refer Slide Time: 46:12)



That is the first mask is where you open the window in the field oxide - step a and then you grow the gate oxide, deposit the poly silicon and then pattern the poly silicon using a mask like this. Poly silicon itself is going to act as the gate contact. So, underneath this poly silicon line you have the gate oxide; everywhere else outside this white square it is all field oxide. So, you have just these two halves of the window. You do not need another mask for your source and drain. You can dope straight through this chosen window. Outside, it is all protected by the field oxide and here it is protected by the poly silicon. So, your source and drain are automatically aligned to the gate, no need to take this into account in the mask. So, you do not have any critical alignment. This is automatically aligned; source and drain are automatically aligned to the gate.

Self aligned technology, self aligned technology is a key word in today’s integrated circuit technology. As the device dimension becomes smaller, it is very useful to go for the self aligned technology where by the criticality of the photolithography can be reduced.