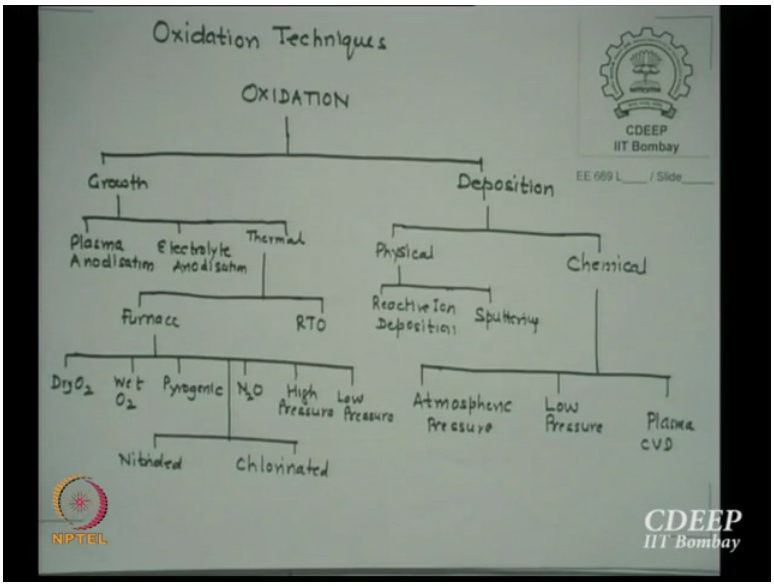
Fabrication of Silicon VLSI Circuits using the MOS technology
Professor. A N Chandorkar
Department of Electrical Engineering
Indian Institute of Technology Bombay
Lecture 13
Thermal Oxidation of Silicons

(Refer Slide Time: 0:17)



Okay, so we were looking for oxidation. We already modeled it. The growth models we have created already from Deal-Grove, we also have seen other kinds of models for oxidation, particularly for thin oxides. And this was done because the technology was scaling down and the oxide thickness was also scaling down. So we wanted to know what models one should use if thin oxide comes. However one now knows that SiO2 has already lost race with hafnium oxide and maybe lanthanum oxide for 28 nanometer down.

So there is no necessity of Grove-Deal model for this because hafnium oxide is not natural oxide of silicon, so obviously this growth model cannot be used. It is a deposition system. So we will see later when we look for depositions. How to deposit metals or oxides or any other thing and which will have a different property and different models. For silicon dioxide 65, some process some companies still have devices, particularly analog with silicon dioxide as insulators.

RF circuits are mostly with SiO2 but anything below 45 at least for digital or even 45 digital is hafnium oxide which is more standard Intel process. So we look into that when it comes but as of

now we assume that we should know oxidation technique anyway, so we did that. Of course even in the whatever known process you will do, you will have some field oxide as we shall see and that will be done by normal oxidation technique.

So some model is good enough there for creation of thicker oxide which is maybe 2000 to 3000 Armstrongs. Right now for 0.5 micron or above process the oxide thickness was 6000 Armstrong to a micron, now it will be less than 2000 Armstrong field. So there is a thickness variation but still it is thicker oxide enough for Grove-Deal model to be acceptable. For gate oxide yes, this may not be now useful sooner or later.

But they do not think it because there is research going on to look for some equivalence of silicon oxide now which grown some materials like, there is a material silicon nitride. We are trying to see whether we can grow out some thin oxides there. So some other semiconductors may use Grove-Deal model again. So let us look for oxidation techniques. There are two techniques. One of course is growth, that is you must have silicon to create growth out of it, silicon dioxide. The alternate method is deposition in the substrate is unimportant.

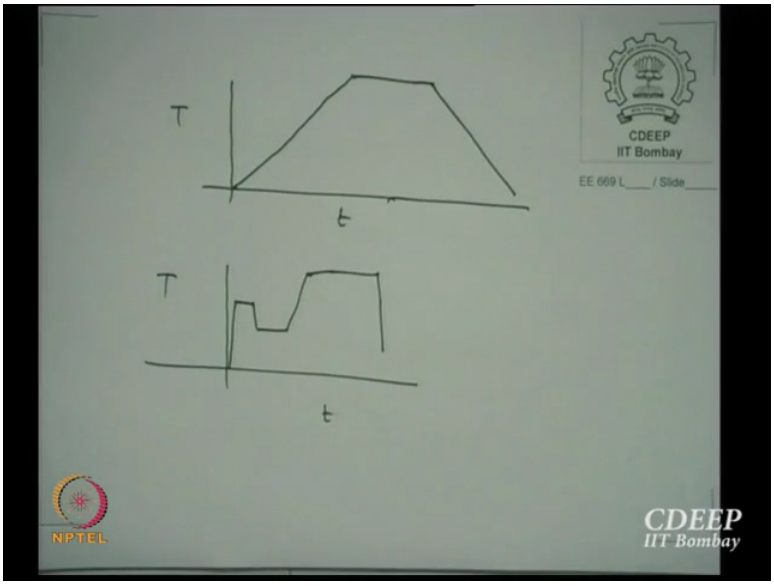
I can deposit silicon dioxide on any surface, it can be silicon as well but can be any surface. That is called depositions. The growths can be plasma anodisation, electrolyte anodisation and thermal. This is a process which we looked into way back in 80s. It did not find the acceptance from the industrial community because the in industry one of the major criteria is what we call throughput. What is throughput?

The number of devices once can create in a single run. So in a furnace if you cannot put 200 wafers, that will be very costly for them. So the throughput is plasma anodisation system, could be 4 wafers at a time. So how many times I will go and put 200 wafers, so that became very costly. But this wastage can grow even 10 Armstrongs of good oxide which we tried, two of my PhD student worked for this.

But somehow of course as I say it is not industrial process. You can also have sol gel method which means you can deposit in by electrolytic, electrolytes like a cell. You can deposit SiO_2. However the most common one as I say is thermal which has furnace, we will see today. And there is this process which is called rapid thermal oxidation, RTO, rapid thermal oxidation.

Actually RTP is then, actual name, rapid thermal processing. Anything can be done in RTP including oxidation. RTP is the process, rapid thermal processing.

(Refer Slide Time: 4:46)



For those who are first time hearing this sentence, normally if you see a furnace temperature rise, time versus temperature, you ramp this, actually you call ramp the furnace. You start with room temperature and go to the temperature of your choice. So typically it is like this and then becomes like this. Then when you ramp down, it becomes like this. Now this takes hell of heat and hell of time.

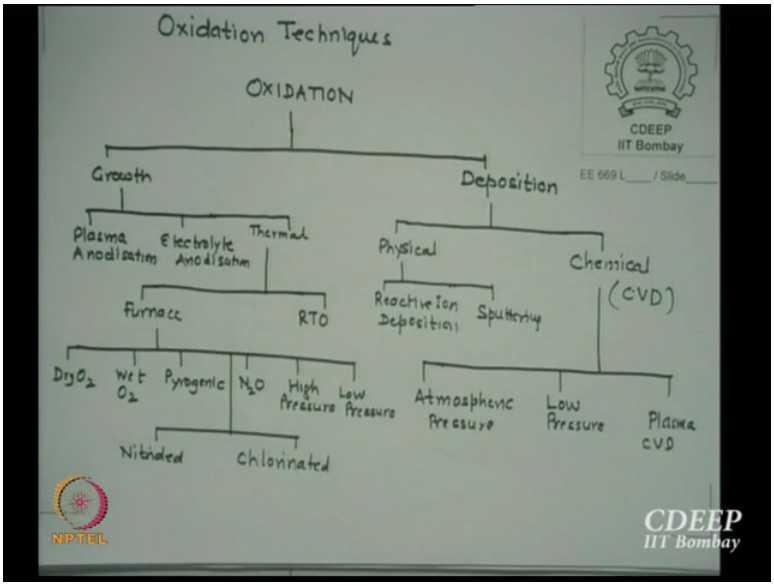
So what we want to do is if you want a very short duration oxidation going on, then one can do something like this. I may ramp up short time, maybe slightly reduced if you wish. I can further ramp up, very sharp ramps, longer, maybe ramp down. So any thermal profile in a very short duration can be attempt, can be attempted in rapid thermal processes. This word rapid itself is proof enough, very sharp rise. To improve the thermal conductivity of the system so that the rise is very fast, the rapid thermal processes have halogen lamps around in a parabolic system.

And their on-offs which is controlled by microcontroller. One of the small RTP system was created way back by my BTech student in 1990 or 92. He is currently looking into analog design in textile instrument, manager of that. So Kiran Godbole, so he made first RTP in India, ourselves. So we have a microcontroller design, we actually picked up, switched on and off. The

problem became cooling. So we have a jacket cooling, we have lot of problem, so making a system is a tough job but we made one. So this is available now in market.

So the problem in RTP is also same. Number of wafers you can do actually, are relatively smaller compared to any furnace kind of processes. So this RTP is still there for specific purposes, at least anneals we can do if not the process.

(Refer Slide Time: 6:51)



So there is a possibility, this is no different from furnace. All that you are doing is temperature is created in a different way. In furnace we can have dry oxidation, we can have wet oxidation and we can also have another one which I will show my results on that, pyrogenic. Then we can use oxide, N2O oxides. Then we can have high pressure oxidations, we can have low pressure oxidations.

And for this we can also, during the oxidation we can introduce nitrogen or we can introduce chlorine which is called nitrided oxides and chlorinated oxides. This nitrided oxide was very important for us because this was the way we actually made first India's radiation hard circuits, some other day. So these are possible thermal processes. If you look for depositions, you have two possibilities, either it can be physically deposited or chemically deposited.

The physical deposition is essentially called using a reactive ion deposition system or sputtering. They are similar but not same. In the chemical you can have for CVD which is called chemical

vapor depositions, CVDs. Maybe I should write, the chemical vapor depositions can be done at atmospheric pressure at very low pressures and the most likely these days we will do is plasma CVD. So let us look at, as I say this deposition part we will take care when we start CVD system. So there we will show you how to make.

So any material can be deposited in fact including oxides. Right now we have looked into furnace and maybe RT is same. So we have looked into dry O2, wet O2 models and maybe today we will show you other processes quickly. There is no difference between this, N2O plus silicon is SiO2 plus nitrogen is released. So that is no different. High pressures, high pressure is called a bomb. So you actually put water wafers, water solutions and seal the tube and heat, so the pressure breaks.

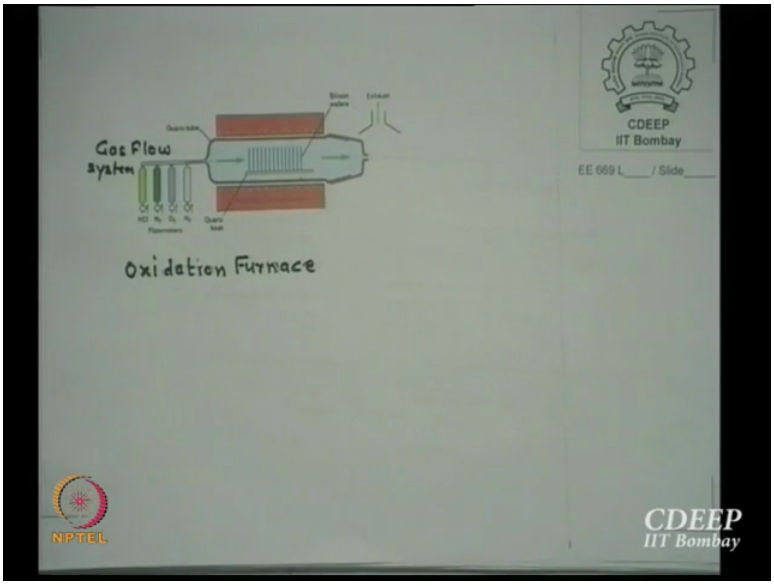
So you have to guarantee that the pressure is not so excessive that they blast out. So there is a game there and of course the advantage of high pressure oxidation is I already said. At high pressure the rate is very high. So very short time you can grow thicker oxides. This is essentially used in some old processes but just to make a point complete I say. Then there can be a low pressure. I can evacuate the furnace to a certain vacuum and then pass oxygen to maintain certain amount of gas pressure inside.

So I can change partial pressures there and this also has an advantage or disadvantage as we may show you one of them. During oxidation as I say I can add ammonia so that I can have nitrogen inside in silicon dioxide lattice, is called SiON bond which can be formed there and it is very important in radiation hard circuits. Many times we do put chlorine. Now it is almost given up but there was a time when everyone was trying chlorinated SiO2. The reason was that chlorine is a first group element and as we shall see soon, there is what we call dangling bonds at the Si-SiO2 interface.

So chlorine can actually remove those bonds by attaching to them. So it may reduce what we call interface state density. But chlorine oxide has a problem, it is a larger atom, then it has affinity to silicon, so it may form silicon chloride and evaporate and it gives pitting. That is why it is called pitting. So we stopped using chlorinated oxides many years back. So we replaced from HCL oxidation to TC oxidation, we tried in my time 70s, we have gone through all kinds of

chlorinated and nitrided processes. Nowadays except for Red Hats nitrogen oxides are also not used. Okay, so this is typically the techniques available.

(Refer Slide Time: 11:08)



A typical furnace, taken from Google looks something like this. You have a Quartz furnace. This is the resistive heaters, these are resistive heaters, this is called the central zone, this is called early zone, this is called the end zone. So we can pass the gases through rotameters or through mass flow meters. You have a cell possibility, nitrogen, oxygen, hydrogen, whatever you wish. You can control the flow rate here and pass the gases inside.

Whenever we start oxidation, the first thing we do is to soak the wafers to a given temperature. What is word soak means? When you introduce the wafer from this side, they are at room temperature. So when they enter central zone, there will be time they will require to actually reach the furnace temperature, that is called soaking. So during soaking I do not want oxidation to proceed. So what should I do? Pass something inert which does not oxidize it, so we pass enough nitrogen. It can be as high as 100 liters per minute oxygen, nitrogen flow just to allow soaking in the wafers to that temperature.

So there is at least three to five minutes are required for soaking. So you introduce the wafer from this side, start nitrogen huge amount so that slowly temperature rises to the wafers, quartz, rack, everything and then this is universal, uniform temperature everywhere. Once that happens, I first introduce, I am doing dry oxidation, so I pass oxygen initially whatever amount of

oxidation which can fill this and start reducing nitrogen and close the nitrogen, so that whenever I finish nitrogen, that is the time when oxidation starts.

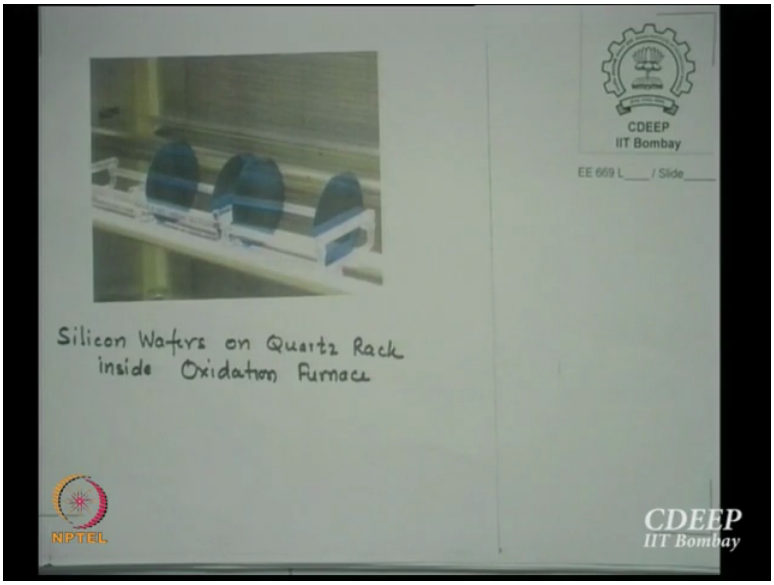
Because oxygen is already inside, nitrogen was not allowing that oxygen to react but as I close the nitrogen, oxygen, oxidant is available for silicon to oxidize the silicon wafers. And one can see from here that I can oxidize any number of wafers if I have a central zone thick enough, wide enough. If you have a large zone, what does that mean? Bigger size furnace, huge thermal heating, so huge power loss but that is the way you can make.

You also have as I said coils on the early zones and the end zones, this is to maintain the gradient because this will lose the energy, thermal energy. So you want to maintain the furnace to uniform central zone furnace, so you must heat both ends. So there is a separate coil, separate monitor and also there are enough thermo couples to monitor the central zone temperatures. So the oxidation when it finishes, before oxygen is switched off, you again start nitrogen, large amount.

Then no oxidation takes place and then switch off the oxygen. So all processes starts with nitrogen and ends with nitrogen and in-between is the time for which a particular process is maintained. If you want HCL along with oxidation time, so during oxygen you can add 5 percent HCL wafers, diluted HCL through a bubbler can be in, you can pass HCL inside and you can have chlorinated oxide as we can say.

Or if you want the next process, you will see if we can create something using H2, so we can also do hydrogen insertion if we need. Earlier almost everywhere we had these rotameters or flow meters. For more accuracies now we use mass flow meter for a given gas.

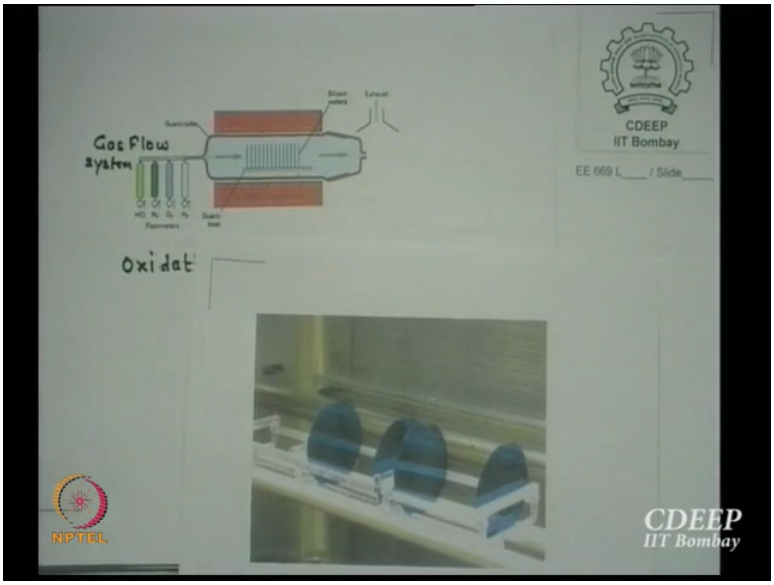
(Refer Slide Time: 14:59)



Typically if you see a rack, these silicon wafers are mounted in a quartz rack, you can see they go vertically down and they fit it into the slots. This is how the rack there, so larger the rack size, larger is the number of wafers one can introduce inside. So any number of wafers you can load depending on the temperature, central temperature zone you have wider or smaller and the rack size you have.

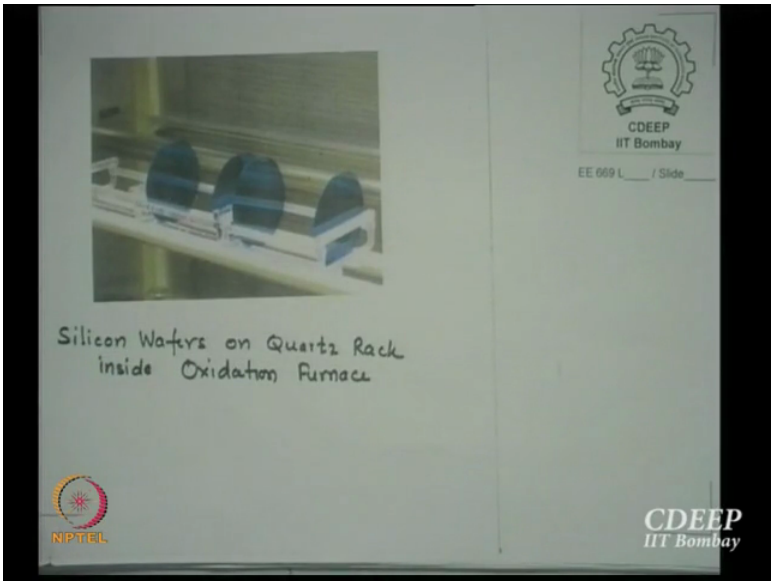
The rack thickness is, rack slots are essentially for the given wafer size because thicker, larger the dia, thicker will be wafer, so slots will be of thicker this. In 3 inch wide wafers may have say 500 microns of thickness, so slot will be 550, so just fits in. So depend on the wafer size, the slot size is also different, so is the rack size, so is the furnace.

(Refer Slide Time: 15:56)



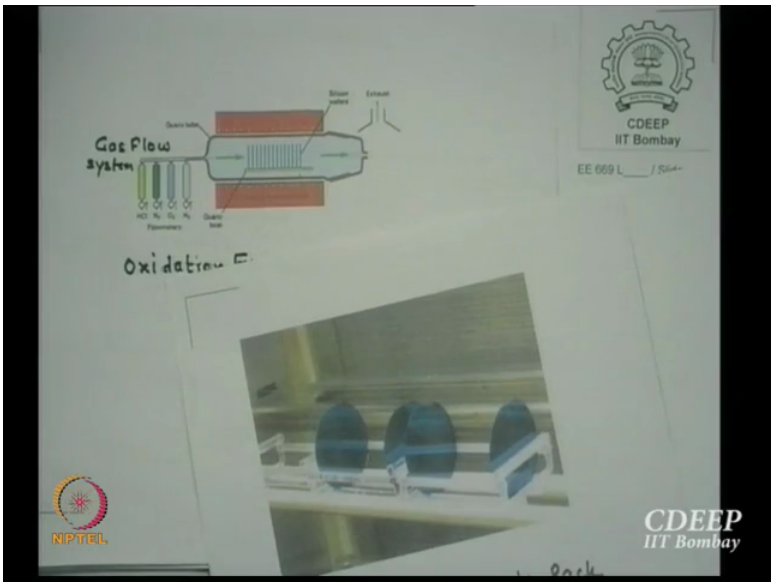
For example, this tube size you can think of it, I am having 8 inch wafer or 12 inch wafer now, so for 12 inch wafer the extra part at least should be 4 more inches. So 16 inch tube you can see how big it will be. And holding those rack coming out, coming in, it is a game. When you start, then auto may decrease. Do not, there is everything is motor controlled. So power electronics is heavily used to do this what is called as at what rate I should push in, what rate I should take out, that is possible. In other times, we had only 4 wafers, we cannot afford more than 4 wafers at a time, too much cost for us, 8 dollars a wafer, so inch. So with available money we could at best put 4 wafers. In normal any process what we call, called dummy wafers.

(Refer Slide Time: 16:41)



What we do is some wafers are always kept say 1, last couple of them and in center also couple of them and they are never taken out. They are always there. So oxidation keeps, that actually is adjusting the flow as we say, some other day. So there are some dummy wafers which are never taken out, they are just sitting there all times. So of course these are not dummy wafers but I am just saying. Even in diffusion, every process we do, there are certain dummies which are, and some of dummies we can actually take out and monitor also. But normally dummies are at least 4 percent in the whole system. Yes.

(Refer Slide Time: 17:48)



“Professor-student conversation starts.”

Student: Sir, from the nitride, an oxygen is already present so that the oxidation will start but sir, do not we pass oxygen from outside into the furnace?

Professor: Yeah, I mean the tube I have already shown, both from the rotameter I will start both gases. Oxygen is in now but it is not nitrogen being such a heavy amount, it does not allow oxidation to proceed.

Student: Okay. Then sir, when you also said that we stop like when we want to stop oxidation, we start the nitride…..

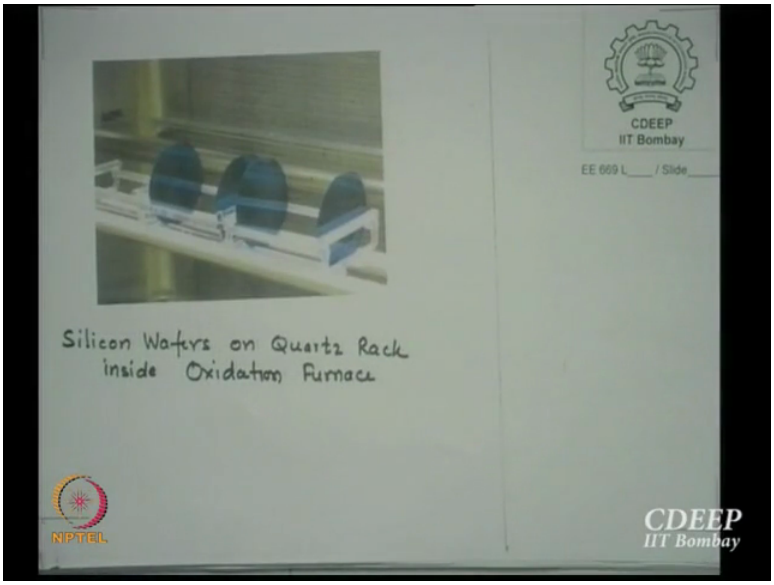
Professor: Nitride.

Student: Can we just stop oxygen and….?

Professor: Yeah, there you have the point. If we do this, the problem with just no gas system is this silicon dioxide which has the oxygen there, actually may dissociate itself at that temperature where there is nothing else to come out, stop that process. So if you stop nitrogen, that additional oxidation process will stop. So it is essentially to maintain the growth oxide to its own whatever thickness we did. Also there is a possibility that what is outgassing from this furnace should not go into the furnace. So we actually push nitrogen, so everything comes out. You are right.

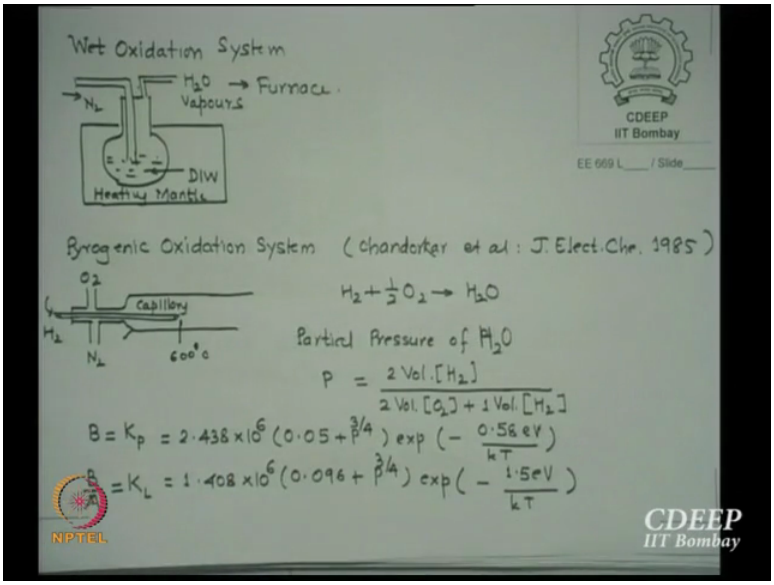
“Professor-student conversation ends.”

(Refer Slide Time: 18:49)



And that nitrogen flow is very high, so it does not allow anything to happen in fact. We can also use argons, any inert gas. Argon is costly and not easily, very pure argons are available. Of course they are available but much cost, higher cost.

(Refer Slide Time: 19:09)



Then there is a wet oxidation system. Wherever that H2O2 we are saying there is another tube we can add, another flow meter through this. And we can have a heating metal inside which quartz this has been put, quartz bubbler. We put deionized water here, pass nitrogen slow amount maybe 30 cc to 100 cc. And since it is heated at 95 degree centigrade, water vapors are picked up

by this very small amount of nitrogen and pushed in through the furnace. This is called wet oxidation. The partial pressure of nitrogen has to be very small compared to water vapors.

So we pass, then why do you bubble? Because to come out at a given pressure, water vapors, you need something to push. So this gas, nitrogen allows you to actually push the gas out, water vapors out. That is why it is called bubbler. The nitrogen actually bubbles, it goes inside and bubbles. This wet oxidation is H2O as I say, already we have seen the model for H2O, same as dry oxide. Then there was an issue many years ago, since these water vapor bubblers are outside the diffusion furnace system, you can see this is outside the furnace.

So what happens that one cannot guaranteedly say the water vapor is how much inside and whether it is the purest form because you do not change water every now and then because this is deionized closed water system. And in those days, we figured out which is even now it is figured out that the water quality is not as good, water vapor quality is not good as if would have directly passed the water vapors inside.

So to avoid this, we suggested a new process which we call as pyrogenic oxidation system, way back in 1985. In those days patenting was very costly, so we could not patent. This was our own design and own process but since as I said in 1985, actually work was done in Ti in 1982 but published in 1985. So what we did is the following:

We have this furnace tube and we introduced, there is an enveloping, this is a bigger tube and through which we can pass either nitrogen or oxygen. And at the end from there we seal a small capillary inside. From the end side I could put a capillary inside which has 1 millimeter dia and it is pushed till it reaches temperature on the end zone or early zone or front zone as it is called to 600 degree. So we actually push the seal like this that whenever that capillary reaches, it has 600 degree to see.

Now from this capillary I introduced hydrogen and I introduced oxygen from the rotameter. This of course also has a flow meter, both have flow meters. So I can pass some amount of volume of gas of hydrogen, some amount of volume of oxygen through this. Oxygen is, so when oxygen enters, it forms envelope around the hydrogen. You can see oxygen is coming from outside, capillary is inside, so when the hydrogen released, it finds oxygen surrounding it.

So all around is oxygen flow and hydrogen is released in and at this since at 600 degree even at 400 but at least say for 600, H2 plus half O2 is H2O. So we create, since the gases available to us are ultra-high pure, so the water which I can create inside a furnace water vapors are relatively far superior than any of the externally added water vapors. This has many advantages. Once of course that you get highly pure water wafers inside. The advantage was that if you do ideal gas laws, maybe I leave it to. In case you cannot, then see my paper.

We have a partial pressure of water in total gas system is given by two times the volume of hydrogen, this use gas laws. So many molecules react with so many to give so many. We use P is equal to NRT and KT formulas. So I can get a partial pressure of two volumes of hydrogen upon two volumes of oxygen plus one volume of hydrogen is called the partial pressure of water wafers in the net gas stream. One can see from here if I chose not to have H2, then the partial pressure is 0, so there is no steam obviously, so partial pressure is 0. There is no water there.

So only oxidation, dry oxidation can take place. If complete oxidation takes place and P is we call it 1. So you find out that complete hydrogen is oxidized by oxygen, so we get full steam, so we say wet oxidation. So by adjusting the oxygen to hydrogen ratio, I can adjust from dry oxidation to wet oxidation and in between I can vary from 0.1, 0.2, 0.5, 0.8 different water vapors concentrations. Is that clear? This system has big advantage because if the water vapors are smaller, the oxidation rate will be smaller. If water vapors are thicker, larger, then there will be thicker oxides.

Now I can adjust my suitable time, I want to do, that means I do not want thermal process to go more than this time for diffusion not to change. Then I adjust my partial pressure for a given time and adjust so that in given time I get oxide of my choice. This was the biggest advantage we got, that I can make in a given time desired oxide thickness which neither wet or dry can do, because if you want certain thickness, dry will take at that temperature this much time or wet will take this much time for this, that temperature.

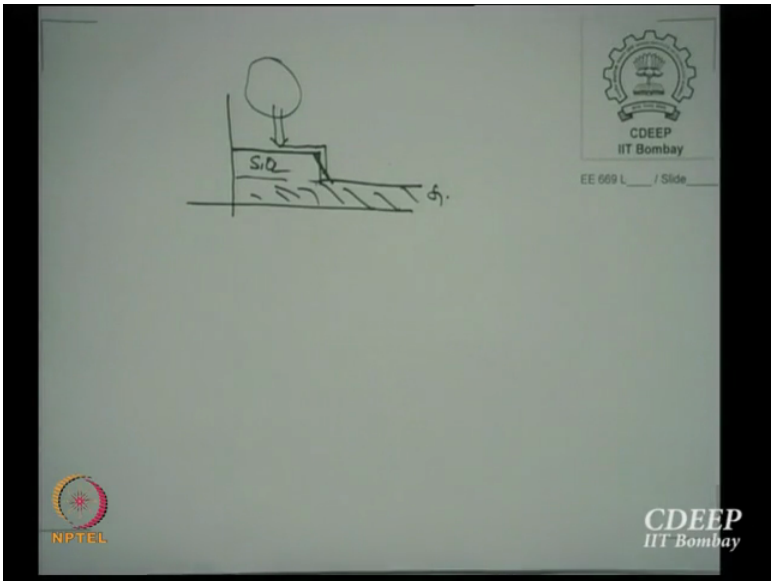
Temperature we do not want to change because furnace ramping will take enough time then. So we do not want, unless you have RTP, we will not do that. So this process was first time done 1982 when I was working at Tata Institute in Mumbai. And we actually measured the oxide thickness at various partial pressures at various temperatures and whatever possible flows and we

fitted the curves again because we thought everyone will like to use the parabolic rate constant and linear rate constant.

So we equivalently fitted into the curves and got an expression which looks very funny, 2.438 into 10 to power 6 into 0.05 plus P to the, P is partial pressure to the power 3 by 4, exponential and this fitted to all possible partial pressures, 0 to 1. And I can see from here, someone were asking. Now this energy and this pre-exponent is my choice, I give some variation, I allow this only vary it here and allow only this much to vary.

So I say okay, it should not be more than 1ev, should not be less than 0.525ev. Now the fit function will try to adjust the pre-exponent with term so that it fits. So do not put any physics on it, there is not physics on it, this is just a fit function. But for a simulator how does it matter? As long as give them P, as long as I give them T, I will get B and B by A. And if I get B and B by A, I know my growths anyway. So for me as a simulant person, I do not care whether the physics is followed here but the advantage I see in technology is that I have now same thickness of oxide, at different times I can adjust by adjusting the partial pressures. And that is something very important in actual thermal processes. Is that point clear? So this is as I say is hours and nothing very great about it.

(Refer Slide Time: 27:59)



Before we go, of course we have to monitor thicknesses. One of the easiest method of finding the oxide thickness is to create a slot or a step. Of course I will come to that little latter again but just

to show you how do we do it. I will do many methods but right now, for example, on a wafer say let us say this is oxide and this is my silicon. This is silicon and this is oxide. I create a step by etching, then I run something called profiler, surface profiler which has a transducer kind of, acoustic transducers which is fitted with all kinds of circuitry.

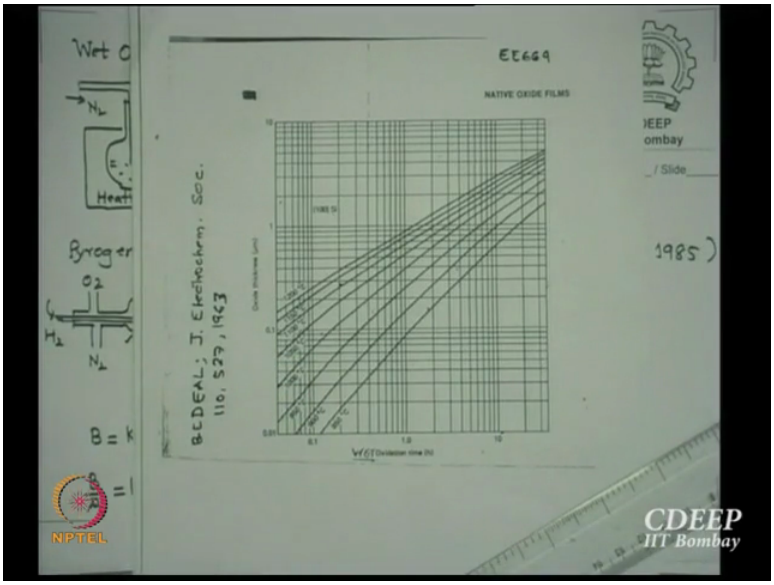
And it has a tiding system, so it goes there, jumps down. So as soon as that transducer jumps down, the acoustic sensor finds out that it has traveled vertical down, so displays how much is the thickness. However the accuracy of profiler is, unless this oxide is sufficiently thick, one cannot say it is accurate because it gives at least 20 percent errors in actual monitoring. Firstly you cannot create a step. The way oxide will be formed, will be some kind of angle. So when the profiler goes, it assumes as if it is going vertically but actually it is climbing down. So the error is in built and there is nothing much but this is very important.

The other method of this is very interesting, when the white light shines on the oxides, you can see depending on the thickness you will see a color on the oxide. You may see tan, you may see, for example, different thickness of oxide will show you different colors. So example, there is seven, one cycle will be repeated again, blue, green. White light, all blue to red, it will keep repeating every now and then as n times, n lambda in optics, n times. So first cycle is 1, second cycle is 2, so the colors will repeat every now and then.

But if you know roughly what time you did, you know which zone you are. So if I see the color is tan, I say it is 0.05 microns. If I see a blue, I know it is 0.12 micron and this is simple optics. So by just seeing the wafer, I roughly know, like 6,000 Armstrong used to be garnish and pink. So I know as soon as I see a wafer, oh, it is okay, 6,000. Because I know I did one hour or something, so it will be around 6,000.

So if I see a good color of garnish and pink, I say yeah, I have reached 6,000. So this first guess we do, how much oxide we have grown. These are only for the overall thing but the actual thickness has to be monitored in many cases. And we will see little later what is ellipsometry does that. So profiler, color monitoring, color monitoring I repeat that the intensity of light changes as the thickness of this because light is absorbed there in silicon dioxide. So as the color changes, I know how much thickness I have.

(Refer Slide Time: 31:17)



So now having shown you this oxidation this, something which you have to learn to do in this exam as well, you may need that, probably I am not sure but possibly. Here are the two graphs which I have provided to you. This is 100 wafer silicon, please remember four graphs have been given. Two for 100 and two for 111. If the orientation is not specified, it is 100, otherwise it will be specified as 111.

Please remember in mass technology everyone will use 100 but otherwise if specified then you will have to use 111 if specified. So I am only using as of right now 100 wafers, oriented wafers. So this is for dry case, this is a dry oxidation time and it has been shown at 800, 900, 1000 and actually you can of course not accurately but you can even extrapolate them because log-log, so you can extrapolate down to find at different times the thicknesses.

You can see from here the x axis is in hours, please remember x axis is 0.1 hour, 1 hours and 10 hour. And on this side is the oxidation thickness which is in microns. This is 0.01 micron which is 100 Armstrongs, 1,000 Armstrongs and 10,000 Armstrongs. Unfortunately the current trend is to talk about nanometers but in our time we used to talk Armstrongs, so I am more conversant. 10 nanometer is 1 Armstrongs. So 1 nanometer is my 0.1, 10 Armstrongs is 0.1 nanometer, so it 10 to power additional term has nano to 10, it has gone.

Let us say I do a dry oxide for some time, maybe I have a scale. For the heck of simplicity maybe I do it 1,000 degree centigrade for 2 hours. This is 1, please remember this is log scale, so

this is 2 hours. Any amount you can do but I am just taking some example, I have not done earlier. So it is 0.09 is the oxide thickness at 2 hours, maybe even assume 0.1 but at least it is 0.09 microns. So 900 Armstrongs.

You can see it is slightly lower than 0.1 and by log it is 0.9, so it is 900 Armstrongs of oxide is grown. If I do oxidation, dry oxidation for 2 hours, I get 900 Armstrongs of oxide. But I follow it by a wet oxidation for say 30 minutes at same temperature, any other temperature. So what I do is I did it this oxidation at 1,000 and I am doing wet oxidation at say 1,100 for 1 hour. So what I now do is to find out the new oxide thickness grown, already there is an oxide and you are going further. So I use this wet graph and first I do it at the wet oxidation temperature.

I find how much time it would have required to grow 0.09 microns of oxide, that time I calculate. That is called my initial time, I have not done with wet at that time but equivalently saying at 1,100 I would, if I would have done it to grow 0.09 microns, how much time I would have taken. So let us say this is 0.09 somewhere here, so it is, you can see it is roughly 0.09, 0.05 hours. So 10 minutes or something, 12 minutes of oxidation was required to grow at 1,100, 0.09.

Is that point clear to you? What is I did, it transferred the earlier oxide thickness into the new oxide stream equivalent time I calculate. From the oxide thickness I know and the new temperature which I am doing if it is 1,000, then I have to take this graph and then say 0.09 will be somewhere here, it will be 0.08 or 0.09 hours. So it depends on what temperature you are doing next process.

You first translate the first process into the new temperature time, this, find that time. So in this case let us say it is 0.01, then add 30 minutes to it from here ahead. Is that clear? 0.09 is equivalently I got it plus the new 30 minutes at that time which you are doing wet oxidation to add this time to this. This is small change here occurred because dry oxides are very thin comparatively.

But in two wet cycles if I go through, it may happen. I have done a wet cycle at 1,000 and then at 1,100. So I have enough thickness initially and then I have to translate on the other temperature. What time I would have done to do this and add the new time again. So if I add 30 minutes to this which is roughly half an hour, so from here to here it is 0.5 here, 0.5 here. So at 1,100 I

actually 0.1, 0.2, 0.4, 0.5 okay. Say at 0.5 at 1,100 degree centigrade somewhere here this is 1,100. So this is the oxide thickness, I just check.

So it is very close to 0.28 or something or 0.29 microns. So this fact has to be understood that when I translate one temperature cycle to the other, to the new, for the newer cycle, I must get prevalue converted to this one. Let us say again I do something, that oxide should go back equivalent time in the newer temperature regime and start the new timing from there. Is that clear to you? This is something, that means you may have to move from one graph to other if more than one oxidation cycle is going.

Many times dry cycles are very short durations, so their oxide times, thickness is so small, so one can assume they are negligible compared to wet oxide times. But there can be two wet oxide cycles in which case the first wet oxide thickness may be sufficiently large which at the newer regime we have to first transfer and then add the additional time to this and find the new oxide. Is that clear to you? This is a very important thing, reading the graphs.

"Professor-student conversation starts."

Student: (())(38:05)

Professor: Let us say I do two wets, one at 1,000 and then I did at 1,100. So let us say at 1,000 degree I did it for 30 minutes somewhere here 5. So this is around let us say 0.2, little less but 0.2. And then I did it at 1,100 next for 30 minutes. So for 0.2 Armstrongs, okay you can go back on this side, at 1,100 degree this time is around 0.3. It is not 0.5, it is 0.3. At this 0.3 hour to next 0.5, that means 0.8 hours, so you go somewhere here, go to 1,100 and find the new thickness. Is that clear?

"Professor-student conversation ends."

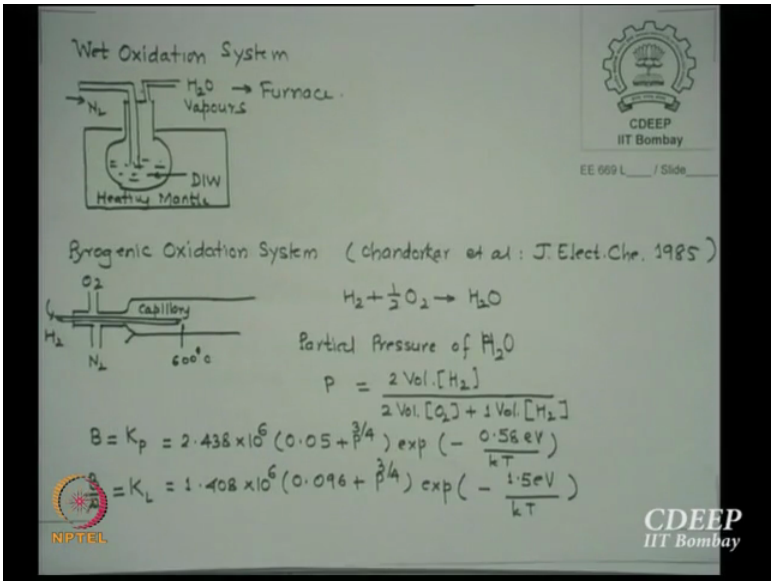
So any temperature cycle, earlier one should be translated to the new temperature cycle, find the time, equivalent time there, add the new time whichever the process is asking and find the final result. So if you go through n number of cycles, n times you have to move from temperature to temperature or dry to wet. So this is very important because otherwise your thickness monitoring may be absolutely wrong.

There as I keep saying most cases dry cycles are so short time, so their equivalent wet cycle time you cannot even, it may be less than 0.01, so you say where I did, okay we at 0 and add just wet cycle. But in many cases if it does not happen that, you must figure out where is the oxide, initial oxide time for the, if I change the temperature. Please remember I wet, for the same wet I showed, I found at 1,100 or even at let us say at 950.

So for 30 minutes of 950, I would have done, okay this point something this. Now at, say at 1,000 this is the time I have done. If I do next time 1,000, I know how much initial time I would have spent to get 1,000 degree same thickness. So you actually go horizontally, for the new temperature find what is the time, add the new time now and keep finding. As many cycles you go through as many times you will have to come on graph and remonitor it. Is that point clear to you? This is very important because as I say if you make errors, dry wets, I normally do not make much errors because wet cycle is much stronger than the dry cycles.

So this wet time cannot be modified too much but in cases where two wet cycles go, your dry cycle is long enough in which case these transfers are important. So is that point clear? So reading from the graph, equivalent this, is that clear? Equivalent to this, from here to here, here to here. We will have to keep doing if it is more than 1 cycle of different temperatures and different types of oxidations. So please remember these two graphs are there for required, just to monitor the exact thickness which is not very exact because at the end from the graph you will read some mischief.

(Refer Slide Time: 41:17)



I keep repeating the log graph is initially spaced, long spaced and as you go towards 1, it becomes compressed. That is the log numbers, log 30 is 0.47, log 40 is 0.6, log 50 is 0.7, so obviously when you reach higher values, they come smaller and smaller. So the log graphs are on the top becomes condensed whereas below they are spaced. So 0.2 is here but 0.8 may be, 0.5 may be far away but 0.5 to 0.6 is very close, 50 to 60 is not very different. So this log reading I am again, every year I tell.

“Professor-student conversation starts.”

Student: How you gave expression?

Professor: No, do not go B, B by A. This is already given, actually data given to you. This is what we did for different temperature, actual oxide was measured. These graphs are shown to you, time versus thickness. But I am traveling, if I give you B and B by A, then use those expression there but if I will not give you B, B by A anyway. If I am using normal process, I will say okay, oxidation was, dry oxidation was done for 1 hour at 1,000 degree, followed by wet oxidation cycle of 950 degree for 1 hour.

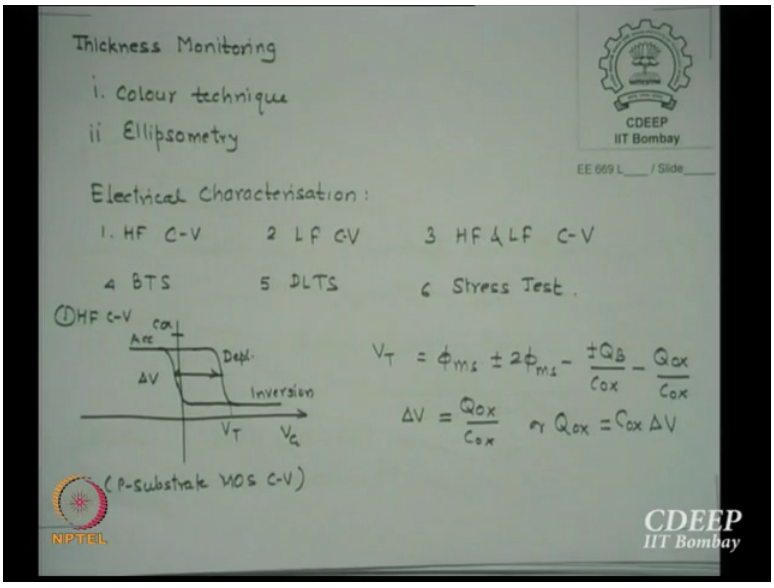
“Professor-student conversation ends.”

So there is now B to no or B by A, so only you have to go on the graph, get the values. Please remember the first one should be translated into the second one before you get the actual oxides.

Or of course I can do means I may give you final and then you will have to return back and say what first I would have done. The mischief could be like this, I may give you final and say okay, then what time I would have done to get this, so you may have to find the dry oxide time.

Inverse problem, okay if you can do this problem, I can ask you inverse problem. Before we go to many other thing, I hope, Professor Vasi is already through with you to mass capacitors, I trust. Okay, before that I will just show you what characterization I will do.

(Refer Slide Time: 43:25)



I will talk about ellipsometry, I will talk about, this is all thickness monitoring, either by color technique or profiler or by ellipsometry, either of the three. There can be high frequency capacitance voltage characteristics, there can be low frequency capacitance voltage characteristics. In general we do both HF, LF together, is called HF LF CV technique. Then we will do BTS which is bias temperature stability, CV measurements.

We may do also DLTS which is called deep level transient spectroscopy, DLTS. Deep level transient spectroscopy, DLTS. Of course I will not give detail of anything because I think I may make mischief for doing, I may teach devices course here, so I do not want to. And maybe I will do some stress test which is called TDDB, time dependent dielectric breakdown. I may also show you some secondary current measurements with isochronal timings, that is time is same but temperature varies, temperature is same, time varies. We may also do many noise measurement, one is called post-office noise which is very important.

“Professor-student conversation starts.”

Professor: We also may monitor flicker noise, you know what is flicker noise due to?

Student: Dangling bond.

Professor: Dangling bond, someone said. That is what the, that is why this LF CV is required. I want to know how many dangling bonds I have. It is very funny if I say someone who is working on circuits, since someone said it is only low frequency noise below 100 hertz, so in RF system why we are so much worried about flicker noise? We should not, only thermal noise should have it. But in RF we really worry about flickers. So why? Think of it. Those who are circuit oriented things, they know much more. Think of it.

Flicker noise below 100 hertz, it is 1 upon F noise. As the frequency goes down, increases, the noise keeps going down. Where is the issue? Vertically think of a system which is called transmitting or receiving system. Why in RR transmitter receiving system we are worried also for flickers? Something if you are working, Vishal Gupta or others, in RF area Madam Schijoy or someone, think of it why flickers are so crucial.

Anyway we will only look into that okay, this NIT or DIT is essentially flicker related. So these are some measurements, so I will come back to it little later. Before that since he has not done it, I actually thought he will, we would not had done it.

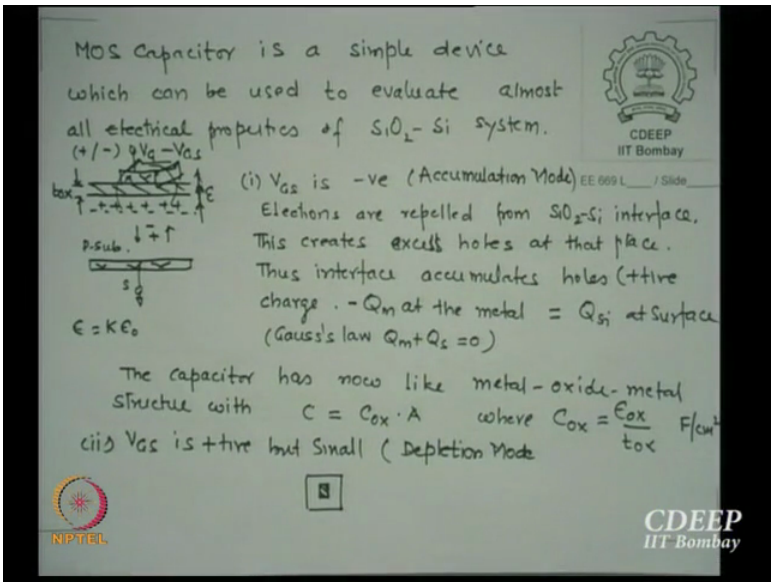
Student: (())(46:10)

Professor: Stress is time dependent dielectric breakdown. So I apply constant voltage for longer time at a given temperature, time dependent. So I increase the time and I find after some time the dielectric strength whatever it is, at that current starts shooting. So it is called TDDB, time dependent dielectric breakdown, is one of the major measurements in thin oxides, TDDBs. You do not want to read other papers generally but in case anyone is working with in the MOS area, even our group, Professor Vasi, myself and Professor Lal have published at least 100 papers in or maybe more including Ram Gopal Rao was one when he was our student.

“Professor-student conversation ends.”

We have published enough papers of reality what we did in the lab. So please see them. They are not Deal-Grove or as our models, then our measurements, our everything. Now there is a tendency in, they do not even refer the last year your own paper in next paper as if something wrong. The figures you see, it is called curtailing, I know this lady and these two people different places. He will send all in reference all my papers. He will also, I will also put every paper. So my citation index really varies. That is how the promotions are given, tenureship gauge how much citation index you have. This is why citation index has (())(47:43).

(Refer Slide Time: 47:46)



Here the mass capacitor which is a very simple device which we use to characterize oxides. Most electrical properties can be found through this different kinds of CVs. So what is mass capacitor theory? So let us start looking at it. I have a silicon wafer, it can be either p-kind or n-kind and then there is a thickness, some oxide thickness of tox and there is a metal plate which may have area of A. Please remember this is 2-dimensional figure, the area term will appear because of this W into L for example.

If this is L and this is W, the area is W into L. So there is a metal plate area which is W into L which is sitting on silicon dioxide or if you look at the plan, this is, this metal plate, plan. This is cross-section. These days drawing is not compulsory part in IIT, no one learns drawing. So cross-section end-view abhi sabko pata nahi, side view kya hota hai, pata nahi. Unless isotropic 3-dimensional figure is shown no one understands.

Toh mein thoda bataiye plan hai, yeh cross-section hai. If I apply first thing, I can have either initially 0 voltage and afterwards I may apply minus. The substrate which is p-type or n-type, right now I have chosen p, the substrate is not source, it is S is named because the VGS is named common in MOS transistor. So I kept S as substrate but in their case S is the source. But their substrate is normally connected to source. So it is okay, so it is VGS but not necessarily in most what we call as dynamic threshold, devices, circuits, we do have bulk separate.

So disadvantage with me is that I have done so much circuits, I know where, I confuse sometimes whether I should talk about circuits, devices, materials, so thoda jumble ho jata hai. Okay, B substrate I apply minus VGS, this is a metal plate and this minus VGS creates a fixed negative charge, electronic charge on the metal plate. If I apply negative potential, you have by law of Gauss, there will be a negative charge sitting on the metal plate which I say which has the density per unit area as we call coulombs per unit area, is Qm which is minus.

Since I apply minus charges at the metal and silicon dioxide, right now I assume is a very good dielectric and it has no charge inside. My assumptions, it is a good dielectric and it has no charges inside. So if I apply minus Qm according to the Gauss's law, the net charge around the system must be 0. So Qm plus Qs must be equal to 0. Since Qm is minus, so minus Qm must be equal to silicon charge. So if I minus Qm, I must get my induction or what Gauss's law, positive charge at the surface.

If this positive charge has to occur, how this positive charges can occur. I want, I have put minus charge. Equivalently surface should get positive charges because electric field lines must start from positive terminal and go to the negative. So equivalent charges must be created. Each starting point must have the end point. Now this essentially means that the charges, opposite charge will, identical equivalent charge will be and again I appeal there are no charges in the oxide as of now.

If I do that, then I find which is the way I can create positive charge. This material was p-type, it has huge number of holes anyway. Enough but please remember the rest of the part other than surface shown here is charge neutral. What does that mean? P plus Nd is equal to N plus Na. This charge neutrality holds, there is no electric field here. So charge neutrality holds, this is called charge neutral region.

However the number of holes are enough there, so some of these holes travels to the surface, why? Because since it is minus VGS, the electric field is bottom to upward, both across silicon as well as across oxide. Minus potential, so electric field is vertically up, going up. Now these holes travel in the direction of electric field. So holes from the p substrate actually go into the surface. Now question is then, the substrate loses holes? No, the battery will supply those many. So the neutrality still will hold, do not ask me from where, that is enough charge is available to you.

So this excess holes which are coming in the p substrate near the surface, one can say there were holes initially itself whatever doping you have, plus additionally you have created more holes there, so it is called accumulation. So we are accumulating holes, larger the VGS or minus VGS I apply, more and more holes will appear there and how many exactly the charge I put on the metal, that net charge system must be 0 across the loop.

This is very popular for at least dual degree student will watch, or many of you might have, this is a charge problem is everywhere. This is a very important, some circuits they show or some Q1 plus Q2 plus Q3 must be 0. So some fun is always done in all physics papers. Now since this is charge neutral, of course it has small resistance. I will not say there is no resistance because after all it is a silicon. But assuming that resistance is 0 or very small, this large positive charge here compared to what you apply here, you can say, what is metal definition of a metal?

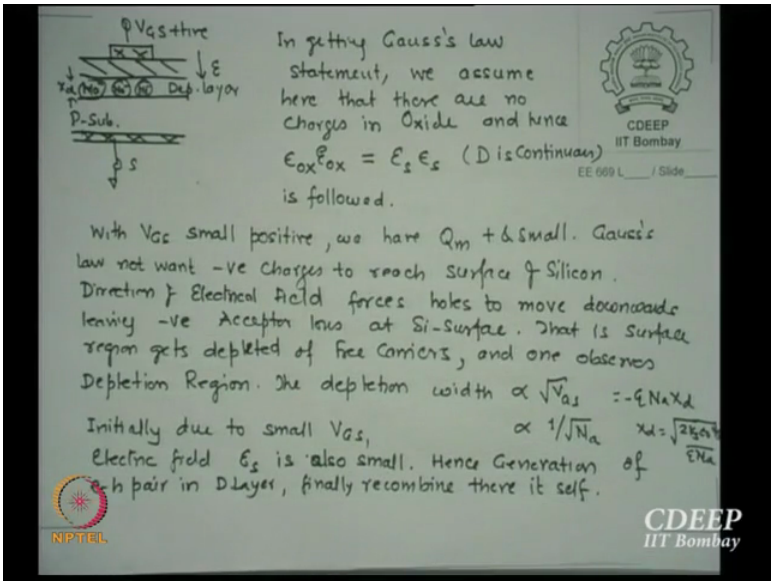
It has large free carriers, that is what we say. Either electron or holes, both free carriers. Large free carriers, so as larger the accumulation, more metallic it becomes. So this behaves like a metal, insulator metal system. And the capacitance of this metal insulator metal is the oxide capacitance associated with this oxide thickness which essentially is the way all MOS people define is Cox is per unit area. That is because we do not want to talk about area initially. When the device finally comes, I will multiply with area.

So the C, net oxide is Cox into A whereas Cox is epsilon ox by tox. Epsilon ox is the dielectric or rather permittivity of silicon dioxide which is ks times epsilon. Please remember epsilon is k times epsilon 0. Epsilon 0 is 8.857 into 10 to power minus 14 centimeter, per centimeter. k is the dielectric constant. So epsilon given means you should write kox which is 3.9. For silicon dioxide k is 3.9 and epsilon 0 as I just now said, so that is the epsilon S, or epsilon ox.

So this since I know tox, I know the material, so I know the capacitance per unit area. If I multiply it by the plate area, then I get the net capacitance. So this is standard capacitances, call it Cox dash if you wish, per area is multiplied. So the capacitance of a negative VGS system is tox, that is it, oxide capacitance. This fact has to be understood because if I have a ceramic capacitor which we use in our circuit lab, if I change the polarity of ceramic capacitor, it does not matter.

I can put like this, I can put like this, it is the same capacitor. But in MOS it is not so. If the bias changes, the C value is going to change and that is different from normal standard capacitors. Now the next case is I increase the voltage towards positive but increase not big, small VGS positive I apply. So till 0 one can say roughly the oxide capacitance is seen. As VGS becomes positive, that is this, maybe if you have written down or you understood, fine. If you understood, it is better than what you write down. Professor Vasi will spend enough time on this, so more clarification will come there.

(Refer Slide Time: 57:15)



Some theories which I propound here, maybe he will prove later wrong and then you can come to me. I say no, no, he is specific if you are right, I am also right, all funs. If I apply plus VGS, then I am applying positive charge on the metal. And by Gauss's law negative charges must appear at the surface interface. Now however accurate negative charges, now the electric field since VGS is positive is downwards.

So holes at the surface actually find the electric field, so first thing is holes move away from the surface downwards to the ground and leave. As the holes leave semiconductors, the acceptors are ionized. So in this they are ions, accepted ions which are not mobile. This area is called depletion layer, there are no free charges in this region. The charge in this must be exactly equal to charge at the metal because Gauss's law is always sacrosanct. There is another term which we use and if you have done your electrostatics well, we also see since silicon and silicon dioxide are two materials, the electric field on the two may not be same, they will be related to their permittivities.

And this is called D vector is continuous. D vector is continuous, this is our assumption. Sometimes if there are charges in oxide, this is slightly modified but otherwise you can assume it. And the relation we say is epsilon ox into E, this E is fields. So epsilon ox into electric field in the oxide must be equal to electric field in semiconductor into permittivity of semiconductor. This is sacrosanct as long as D is continuous. This is electrostatic, 10^{th} standard or 12^{th} or whatever it is.

Electrical engineering is only this much, aage kuch nahi hain. Yeh sab naam kahi hai sab. So if I have positive charge Qm, negative charges are due to the depletion layer and if you know the depletion layer, the charges and let us say this is my depletion layer thickness called Xd. So what is the charge density there? Q is the charge associated with carrier, negative this. Na, whatever is the number available into Xd because why, what will be the unit of this?

Q Na Xd, number, coulombs per unit area, so it is called charge density. So the depletion charge density is Q Na Xd. And if Qm is the charge density on metal, then Qm must be equal to opposite of Q Na Xd but Na are minus. So automatically minus charges are appearing to compensate for positive charge at the metal, induction, law of induction. One can see since the charge density in depletion layer, maybe you can write, Q Na Xd minus because it is Nas are minus.

Xd, we will derive this or at least show you later, k into (())(60:38) upon Q Na. (())(60:41) is called surface potential, maybe weight for it. So as larger the VGS I apply, I will find more charges in the depletion layer. Larger the voltage I apply, larger will be depletion their thickness.

And one can see that depletion width is proportional to surface potential and therefore the VGS and is inversely proportional to root of Na.

So larger the doping, smaller will be Xd. Smaller the doping, larger will be the Xds. But the charge density is Q Na Xd and this must be equal to the metal charge which you have applied. Since there are no free charges at the surface, we call this mode as depletion mode of a MOS capacitor. This is called depletion mode. Now there is something which we will discuss littler later also in DIT point but just quickly I may say what is happening.

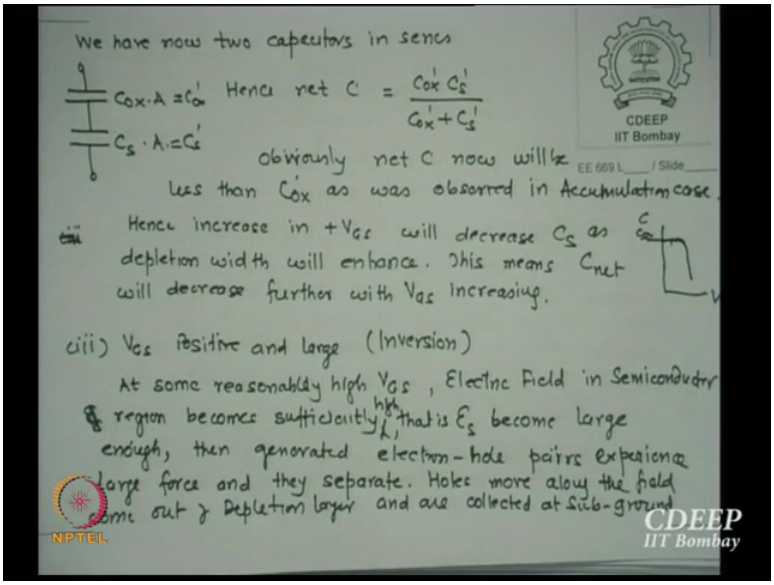
If you have done your PN junction somewhere well, I hope so, if there is a depletion layer, there is a electric field across by Poisson's equation, d by dx is rho by epsilon. So if there is a space charge here, there will be electric field there. Since semiconductor electric field is, will enhance with increase of Xd because rho will increase, larger charges, charge density means larger fields. So if I increase VGS, I will have larger electric fields in the semiconductors. Is that clear?

Going down, now this electric field, what is the purpose of electric field? If you have an electron hole sitting here and if I apply electric field, holes will travel in the direction of electric field and electrons will try opposite to the electric field. However when the initial VGS is positive, small positive, the electric field in semiconductor is very small because depletion layer is very small. Because it is very small, the hole electron, please remember in a semiconductor whether it is depletion region or neutral region, hole electrons are constantly generated, their pairs are formed thermally and recombines.

This is their generation, recombination is constant, all thermal equilibrium processes. But so is in the depletion layer also hole electrons are generated. Here of course there is no electric field, please remember below the depletion layer, charge neutral, no electric field there. Only in the depletion region, there is a electric field. So hole electrons which are generated in the depletion layer can recombine if this electric field is small because that time constant associates such that hole electron recombine.

So only depletion region is seen, nothing else happens. Hole electrons are generated but they recombine, so their depletion layer remains constant. You keep increasing VGS, depletion layer keep on increasing. But then if VGS becomes higher, that is the next step, so in the case of, before that field becomes higher, what is the capacitance I see now?

(Refer Slide Time: 64:01)



Here is the figure. There is oxide capacitance and in series to that there is a depletion layer capacitance which is called semiconductor capacitance. Is that okay? Oxide capacitance and semiconductor capacitance. So if I see that in series, they are in series combination of Cs and Cox, so then what is Cs? The semiconductor current. How much it will be? The epsilon S by Xd, whatever the depletion layer, that is the capacitance across the semiconductor.

However Xd is a function of VGS, is that point clear? So larger the Xd, capacitance will be smaller in semiconductors. In a series capacitance more, what is the way actually we write? 1 upon C total is equal to 1 upon C semiconductor plus 1 upon C oxide. If C semiconductor is much smaller, then it will dominate. If C semiconductor is very large, Cox will dominate, is that correct?

So if C semiconductor is very large means what? That means the VGS is so small, depletion layer is very small, so we say most of the time oxide capacitance is good enough even in small VGS. But as I start increasing VGS, Cs starts decreasing because Xd start increasing and the net capacitance will start then decreasing because Cs, 1 upon Cs plus 1 upon Cox is 1 upon C. Cs becoming smaller, will actually reduce the net capacitance. So initially if I have, I will show you this later again, it may be Cox but somewhere down.

The net capacitance, let us say I apply V versus VGS. So initially it may be Cox, smaller VGS and then it will start dropping to a lower value depending on how much VGS I apply. If I apply

VGS sufficient enough, it may go towards very small net capacitance by this simple series formula. And obviously it will be much smaller than Cox because net capacitance is decreasing. Initially it was only oxide capacitance, Cs is now in series to that, so it start dropping.

However as I just now said if I increase VGS further, if I increase VGS further, then the thickness of the depletion layer in semiconductor is large enough, so is the electric field is very large relatively. Now the hole electrons which were generated in the depletion regions which thermally were generated actually experience a force due to electric field. What is the force?

“Professor-student conversation starts.”

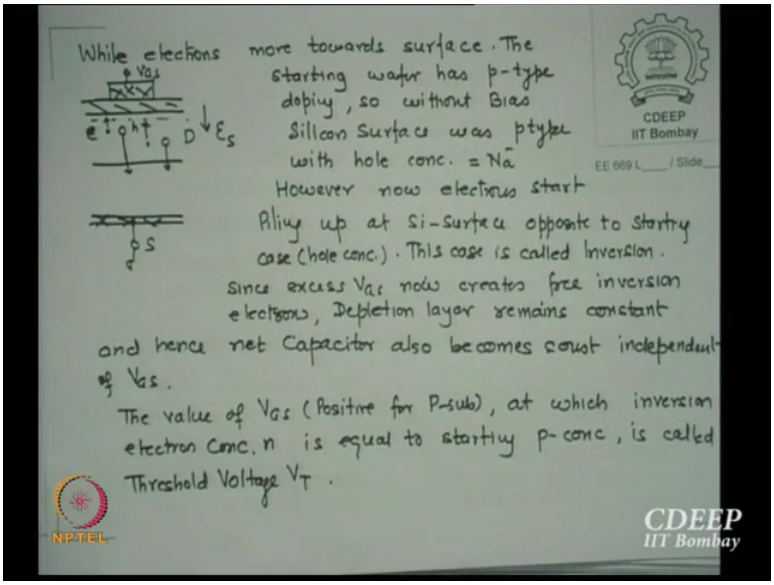
Professor: If E is the electric, Es is the electric field, how much is the force on carriers?

Student: Q times.

Professor: Q times the electric field is the force. And we know electrons travel opposite to the direction of electric field and holes in direction. So what we say, the additional field now separates hole electrons and they are not allowed to recombine. So what happens? You write down this later.

“Professor-student conversation ends.”

(Refer Slide Time: 67:29)

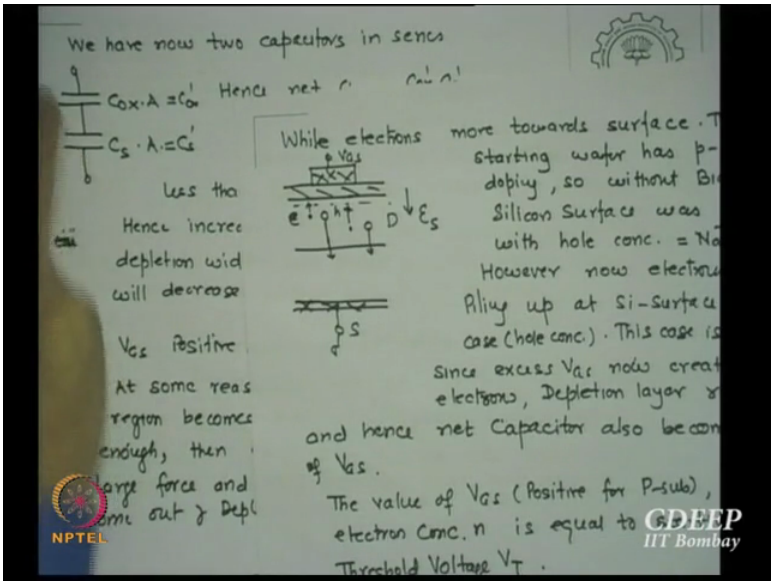


The hole electrons in the depletion layer, holes actually are separated from electrons and holes will travel downwards because electric field supports it. Electrons will go towards surface and since this is the insulator, nothing can pass through it, electron start piling up at the interface. One can see from here that there will be a depletion layer already reached maximum, why? The reason is once these electrons start separating, whatever extra voltage I apply, will be there to create this electron separation.

So the depletion layer then becomes constant. Though it is assumption but can assume it is constant. More and more electron reach to the surface, so we say we started p substrate at the surface, p doping and now it has become n kind. So it is inverted. We started with holes there and now it has electrons reach there. So the surface is now inverted and the region is called inversion channel. That is the principle of MOS transistor.

Since there is excess electron channel here at this creation, larger the VGS, larger will be the electron concentration there. So we find since the depletion layer is fixed, so of course there is a capacitance associated with it. Channel also but that we will see later. We always say the depletion layer capacitance is constant, oxide capacitance is also constant.

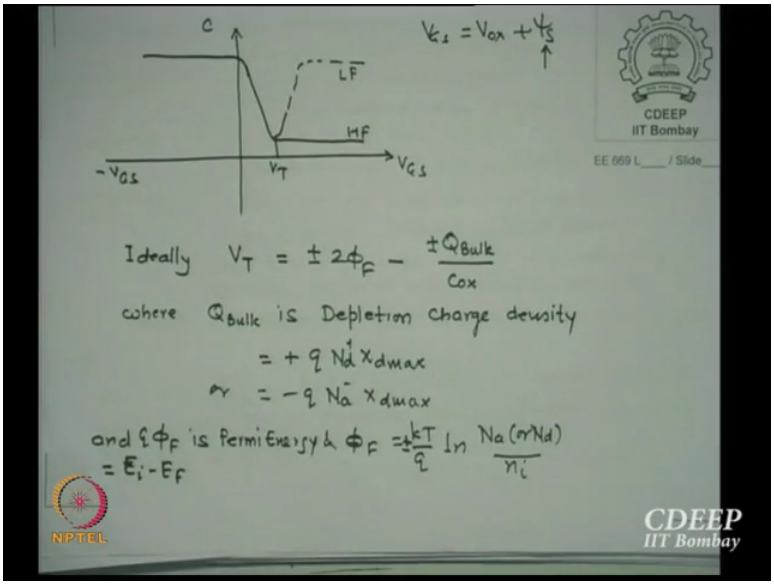
(Refer Slide Time: 69:10)



So by this figure which I just now did, if this constant, constant means how much is Xd here, Xd max. Before it starts separating whatever is the maximum depletion layer width, Q Na Xd max is

the maximum bulk charge or depletion charge we have before inversion sets in. Since this is constant now, this is constant now, so the capacitance become, net capacitance become constant.

(Refer Slide Time: 69:42)



And here is the figure which shows something. Initially minus VGS oxide capacitance is Cox, small positive VGS it starts dropping down. As the depletion layer enhances, the capacitance start reducing of semiconductor and therefore the net capacitance start dropping. Somewhere here the VT as the word shown here, the electric field is sufficiently high so that hole electrons can be separated.

The Cs becomes constant and therefore the net capacitance become constant. This is called high frequency CV. Now why high and low, we will see this later. So in high frequency CV, the way we monitor that the initially for p substrate or n channel device, the Cox for all minus VGS is constant, even it is constant little bit inside, positive VGS and then it start dropping. And at VGS where inversion starts, we call that voltage as threshold voltage for the MOS capacitor and later as we say in the transistors.

So the threshold voltage is the threshold starting inversions, threshold starting inversion is called threshold voltage. This is not a sharp point the way I have shown, Professor Vasi will give you many other reason why it is not so sharp. We also have shown another graph but later we will discuss that again. If the measurement frequency are, if the frequency with which VGS is changing is very low, the capacitance comes back to Cox. That is called low frequency CV, for

high frequency CV is this. But low frequency after inversion sets in, the oxide, the capacitance returns back to oxide capacitance and this is called low frequency CV.

So what is VT definition? When the, no, there is a definition we made which is mischievous. We say once the electrons start coming, inversions have come. But VT is not defined on that. We say when the number of, or electron concentration becomes exactly equal to the hole concentration of the substrate, then only we say it is inversion which is essentially called strong inversion. So the definition is when the, at least same number of electrons must be created as many as holes were.

But actually inversion starts even much before that, once the this crosses and VGS sufficient, it will start inverting. So from inversion, weak inversion to strong inversion, you have an inversion but VT definition says it should be as much as electrons as it was from the initial concentration of holes. That is the only deficit. There is mischief, we should actually define when the inversion starts. But we do not define. We say how much, as much as this additionally. When I show band banding next time, this will be proved very much, why we say inversion, a strong inversion is needed.

So what we say at inversion, please we will come back again, the furnace surface potential, why the surface potential we say? If I apply VGS, part of the potential goes to the oxide like a resistive network, plus part goes to semiconductor. Whatever VGS I apply, part goes to semiconductor and part goes to oxide. This is a sacrosanct. Potential divider, ek resistor aur dusra resistor, there is nothing great.

But rear side I do not know, and that is what I will find. But I will say whenever this surface potential becomes two times the furnace potential, then the inversion sets in. So we say VGS is equal to VT when (())(73:39) is this. And this Vox can be written as q b by Cox. Plus minus because it can be p substrate or n substrate, Q Nd or minus Q Na. So this is the threshold voltage available to you. However assumptions made are the metal work function is same as semiconductor and also there are no oxide charges. These will come in and there will be two additional terms may appear. But you need to know, so next time.