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

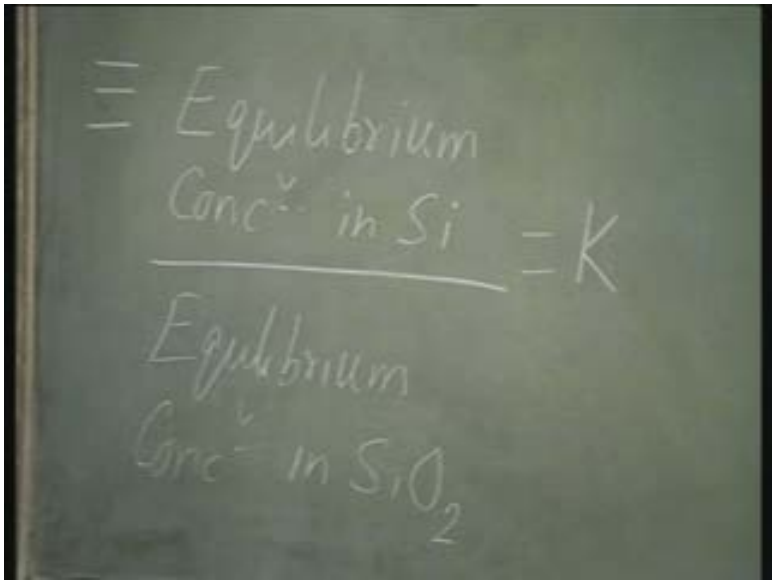
Lecture - 13
Oxidation III – Dopant Redistribution

So, today we are going to talk about dopant redistribution during oxidation. That is you know that the silicon is doped either p-type or n-type, moderately or heavily as the case may be and when we carry out the oxidation, what happens to the impurity? So, that is called the dopant redistribution effect. In this context, first of all we will define a term called the segregation coefficient. Segregation coefficient is something, I think once before also we have mentioned. In this case, we are going to talk about the segregation of the dopant impurity, the relative segregation in silicon and in silicon dioxide. Which way the dopant moves? Does it preferentially go into the oxide or does it preferentially stay back in silicon? That is going to vary from impurity to impurity.

(Refer Slide Time: 2:16)

So, in that context we will define a term called the segregation coefficient. This will be defined as the equilibrium concentration of dopant in silicon divided by the equilibrium concentration of the dopant in silicon dioxide. That is it is a ratio.

(Refer Slide Time: 2:47)



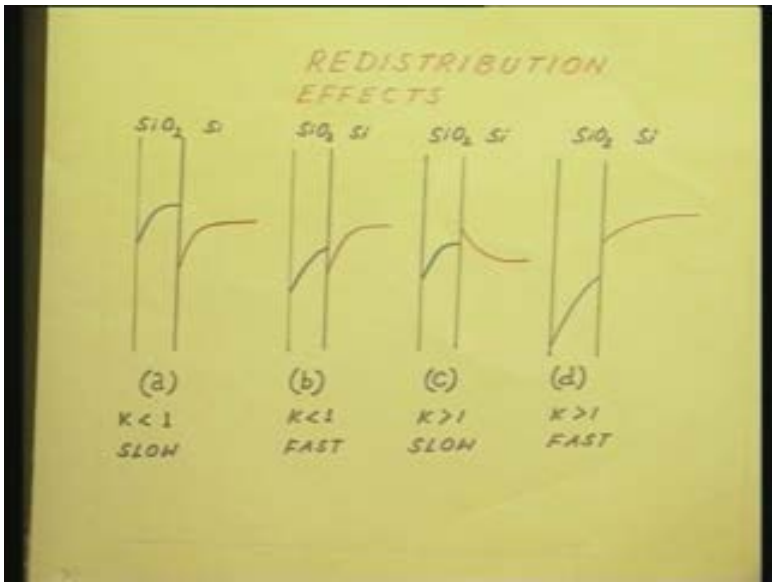
This is defined as the equilibrium concentration of the dopant, equilibrium concentration of the dopant in silicon, it is the ratio of the equilibrium concentration of the dopant in silicon to the equilibrium concentration of the dopant in silicon dioxide. This ratio is called, usually the symbol is K or M. As this is a ratio, you can understand that its value can be either greater than 1 or less than 1. What do I mean? When K for example say is less than 1, that means the equilibrium concentration of the dopant is less in silicon compared to in silicon dioxide. If K is greater than 1 that means the equilibrium concentration of the dopant in silicon is greater than in case of silicon dioxide.

Now, depending on these cases whether K is less than 1 or greater than 1, I may have different possibilities. In addition to this segregation coefficient, there is another factor to be considered. That is how fast the dopant can diffuse through oxide. Remember that we said one of the most desirable properties of silicon dioxide is that it can act as a mask against dopants. What do I mean? I mean that most of the normal dopants that we use for silicon, they do not diffuse through silicon dioxide as fast as they diffuse through silicon. That is the silicon dioxide blocks its movement, silicon dioxide acts as a mask. So, for these dopants we can consider that they diffuse slowly through the oxide and then there are some other impurity elements, for example gallium, which will diffuse very fast

through silicon dioxide. That is a reason why gallium in spite of being a Group III element is not normally used as a p-type dopant in silicon, simply because silicon dioxide cannot block its movement.

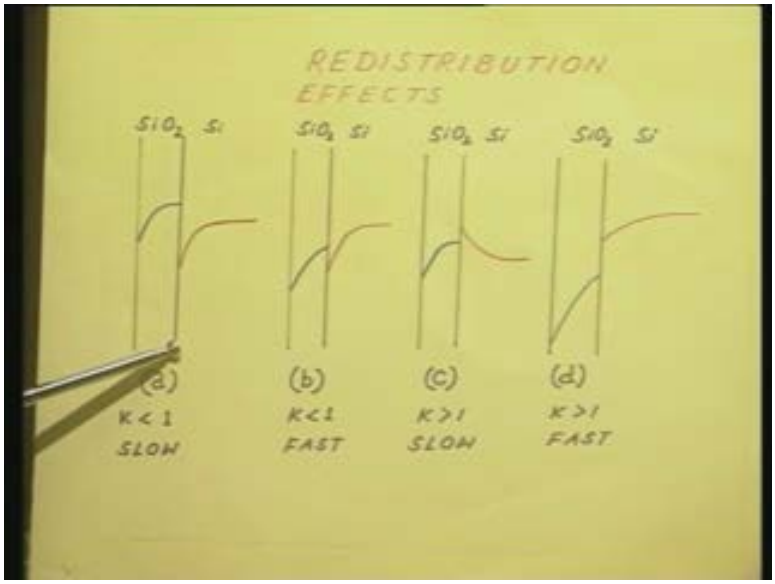
So, you see now I have two facts to concern, consider. One is the segregation coefficient and the other is whether the dopant diffuses fast or slow through silicon dioxide. Depending on that I can have a variety of situations.

(Refer Slide Time: 6:15)



For example concentrate on figure a. The legend says K is less than 1. K, you know, is defined by the equilibrium concentration of the dopant in silicon divided by the equilibrium concentration of the dopant in silicon dioxide. That is less than 1 and I have also added rather cryptically a term called slow. That means the diffusion of this impurity through silicon dioxide is slow. An example is boron, boron in silicon. Boron has a segregation coefficient less than 1 and boron is a slow diffuser through silicon dioxide. That is why we can use boron as a p-type dopant. So, what is the situation? Since K is less than 1 that means the dopant boron will preferentially go into silicon dioxide, because its equilibrium concentration is more in silicon dioxide than in silicon.

(Refer Slide Time: 7:30)

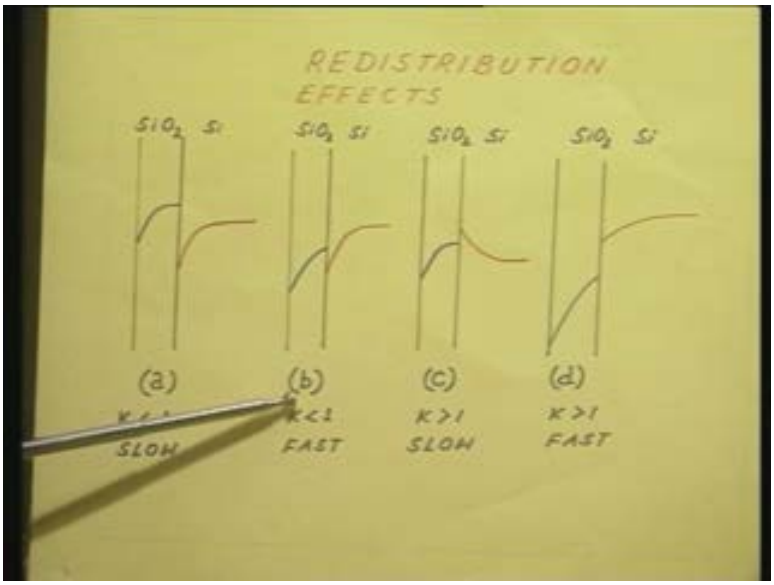


You see, this line is the interface between silicon dioxide and silicon. At the interface, on the silicon side the boron concentration is less than on the silicon dioxide side. Do you see? In other words, at the silicon-silicon dioxide interface, there is a depletion of boron. Near the interface the boron concentration is less. Do you see that? This is because of the fact K is less than 1. Now, the boron profile inside the silicon dioxide. At the interface, this concentration should be higher on the oxide side that is because K is less than 1, equilibrium concentration of boron in silicon dioxide is high and then it will remain more or less constant over quite a distance inside the oxide that is because it is a slow diffuser in oxide, it cannot move very fast. Only towards the surface it starts to fall. Do you understand? This is when the segregation coefficient is less than 1 and the diffusion of the dopant through the oxide is slow.

Two facts you have to be very careful about. That is at the interface, on the silicon side the concentration is less, signifying K is less than 1. On the silicon side, the concentration of boron is less as shown by the red line. On the silicon dioxide side it is more and this concentration remains more or less constant over a considerable distance that is because the diffusion is slow and then towards the surface it starts to fall. Let us look at a variation of this thing. Let us consider that the segregation coefficient is still less than 1,

but the dopant is a fast diffuser. It diffuses fast through the oxide. A practical example is again boron, but this time we will use a hydrogen ambient. If you use a hydrogen ambient, then boron can diffuse fast through silicon dioxide. That is why normal boron diffusion is never done in hydrogen ambient, but always in oxygen or nitrogen ambient. In hydrogen ambient, boron can diffuse fast through silicon dioxide. That is my case b.

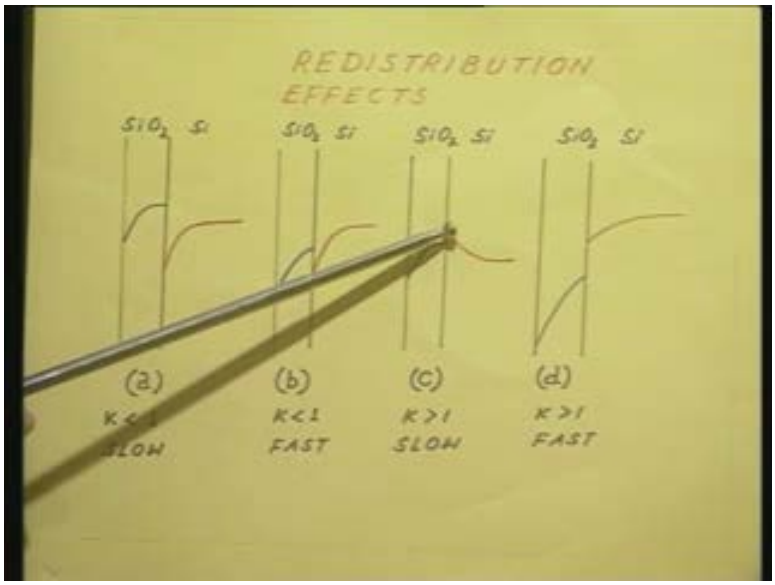
(Refer Slide Time: 10:20)



K is still less than 1, but now instead of slow, I am considering a fast diffuser through oxide. Again you see, if you look at the silicon-silicon dioxide interface, because K is less than 1, on the silicon side the boron concentration is falling near the interface as shown by the red line and at the interface on the silicon dioxide side, I have a higher boron concentration. Do you see? This is because segregation coefficient K is less than 1, but because of the presence of hydrogen, now boron can diffuse fast through the oxide. That is why you see it is, the concentration is falling very fast. It is diffusing out through the oxide. That is the difference between these two pictures. In the first case when the diffusion was slow, you have seen that the boron concentration remains more or less constant in the oxide, only towards the surface it starts falling. Here you see however that it is diffusing fast out, right.

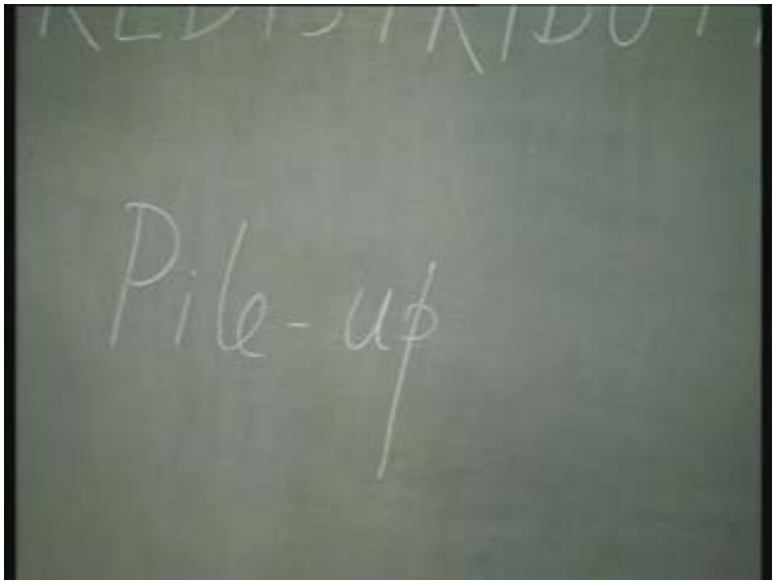
Let us now consider a case where the segregation coefficient is greater than 1. A case is phosphorus. Phosphorus has a segregation coefficient of K greater than 1 and phosphorus is also a slow diffuser through oxide. It stands to reason because, phosphorus is a very common n-type dopant. It can be used because silicon dioxide can mask phosphorus, its diffusion through silicon dioxide is slow. In this case what will happen?

(Refer Slide Time: 12:23)



Because K is greater than 1, you see on the silicon side, I see a pile up of phosphorus. That is at the interface the phosphorus is preferentially staying back in silicon. So, what you see here is that red line moving up. In these two figures what did you see? The red line was moving down. That is because K was less than 1. Now, it is moving up because K is greater than 1. It is commonly referred to as a pile up, pile up, phosphorus pile up.

(Refer Slide Time: 13:05)



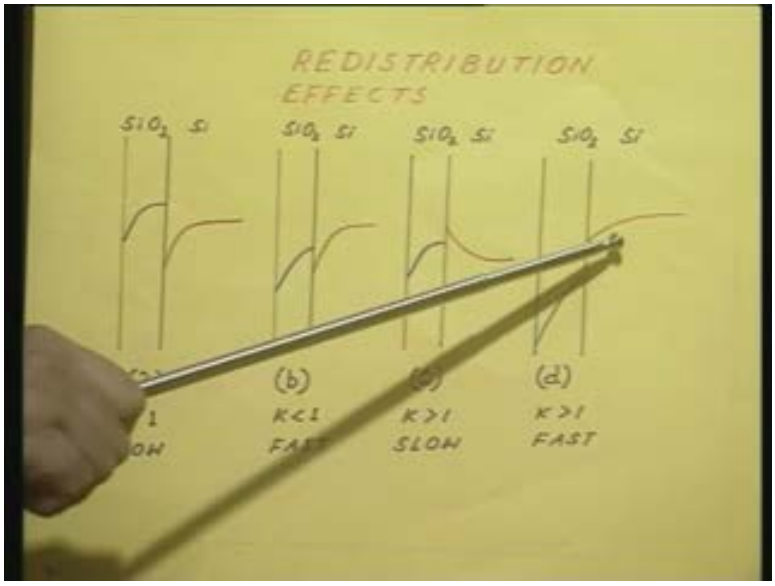
Phosphorus is piling up at the silicon-silicon dioxide interface. That is why on the silicon side I see the phosphorus pile up. On the silicon dioxide side, you notice that at the interface, the concentration of phosphorus is less on the silicon dioxide side. This black line starts at a level lower than the red line at the interface, right and phosphorus is a slow diffuser in the oxide. Therefore the profile of phosphorus in the oxide is also going to be somewhat similar to this. That is it stays more or less constant over a considerable distance and only towards the surface it starts falling.

The important difference between a and c is in the silicon-silicon dioxide interface. In one case, the boron concentration at the interface is low on the silicon side. The phosphorus concentration at the silicon-silicon dioxide interface is high on the silicon side. That is because of the difference in the values of K, whereas the profile inside the silicon dioxide is dominated simply by the diffusion factor and since both of them are slow diffuser, look at these two profiles, they look more or less same inside the silicon dioxide, right.

I now come to the final variation of the theme. That is K is greater than 1, but it is a fast diffuser. I have already given the example of gallium, gallium is an example of this.

Gallium has a segregation coefficient greater than 1 and gallium is an extremely fast diffuser in the oxide.

(Refer Slide Time: 15:17)

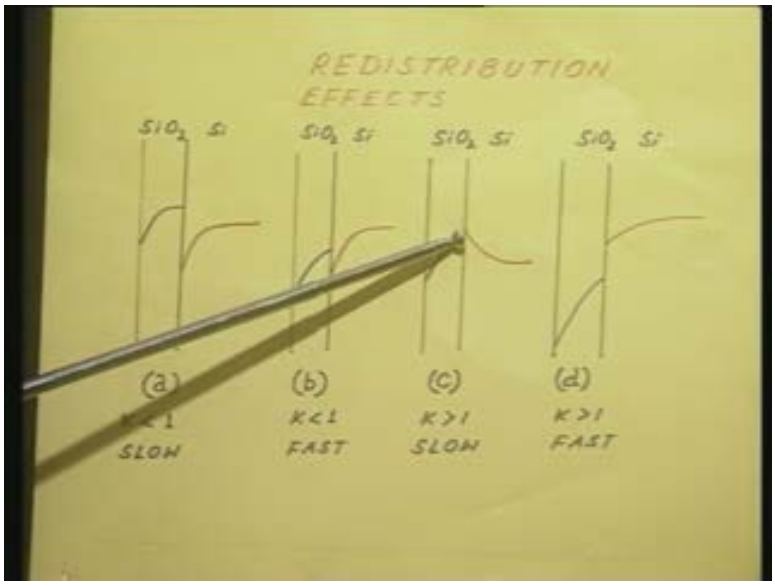


This figure however is a bit more confusing. You see, gallium has a segregation coefficient greater than 1. So, what should you have expected? You should have expected a pile up at the silicon-silicon dioxide interface, isn't it? In the phosphorus case, because K was greater than 1, you got a pile up. So, even in the case of gallium, it stands to reason that since K is greater than 1, I would expect a pile up of gallium at the silicon-silicon dioxide interface. Instead of that you see, if anything, it is falling. That is because, here the diffusion factor is going to dominate the phenomena.

Gallium is an extremely fast diffuser. There is no chance for gallium to pile up at the interface, it just diffuses out. It just cannot pile up at the interface, it is forced out through the oxide. That is why you see, inside the oxide it just falls drastically, so sharply it is falling and even this effect continues even at the silicon-silicon dioxide interface. It just draws the gallium out from this region, so that if anything, actually the gallium concentration is a bit less at the silicon-silicon dioxide interface than inside the bulk. It does not get a chance to pile up, to build up.

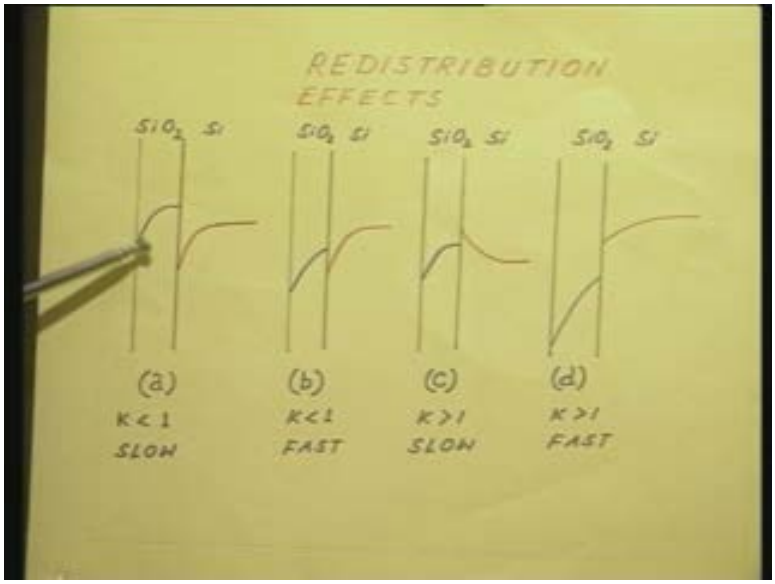
Now, in the light of these four cases, now let us see whether we can take a look once again to the oxide thickness problems on a heavily p-type and a heavily n-type silicon. That is to say, suppose I have a silicon wafer which is heavily phosphorus doped, n plus with phosphorus.

(Refer Slide Time: 17:26)



The situation that you are expecting is case c, right. This silicon is heavily doped with phosphorus. So, at the interface, what do you expect? At the silicon-silicon dioxide interface, you expect a pile up of phosphorus. Now you remember, thermal oxidation is taking place inwards. That is oxide is going to be formed in this direction. So, it is this heavily phosphorus piled up region which is getting oxidized, agreed. This affects the reaction rate. You are modifying the quality of this region. This is the region where the oxidizing species are reacting with the silicon, so you are affecting the reaction rate. So, so long as the oxidation process is governed by the reaction rate, you can expect a thicker oxide growth. That is why you see that when the silicon wafer is heavily phosphorus doped the thicker or the faster oxide growth phenomena occurs at relatively lower temperature, where the reaction rate is dominating the process.

(Refer Slide Time: 19:21)



In contrast, if I have a heavily boron doped p plus wafer, then the situation I expect is case a, right, oxidizing ambient. It is a slow diffuser. Remember, boron will be a fast diffuser only in hydrogen ambient. We are carrying out the oxidation, so you are working in oxidizing ambient. So, if the case is a, in this case, what do you see? Boron has preferentially come inside the oxide. So, this increased presence of boron in silicon dioxide is going to affect the diffusion of the oxidizing species through the oxide. It is going to aid in the diffusion of the oxidizing species, because now you have a silicon dioxide, which is heavily boron doped. That is why for a p plus silicon wafer, the faster oxide growth rate will continue at all temperature, because you are affecting the diffusion, right. This is the two cases for a p plus doped silicon and for a n plus doped silicon. In both cases, the growth rate will be affected, but for different reasons. When it is phosphorus doped, it is affected because the reaction rate is affected. When it is boron doped, it is affected because the diffusion is affected, right.

So, now that we have talked about the oxidation, thermal oxidation of silicon in general, let us take a look at the practical oxidation systems. How do we oxidize a silicon wafer?
As I have already told you, oxidizing a silicon wafer, thermally oxidizing a silicon wafer is a very simple process. What you do is you put the silicon wafers inside a furnace.

Remember, contamination is always a key problem in all VLSI processing. Therefore, to avoid contamination, the furnaces, the tubes, the furnace tubes must be made of quartz. Quartz is silicon dioxide, so silicon dioxide will have the minimum source of contamination and even better idea will be to use pure silicon tubes, but that is very, very expensive. Very few places who really are bothered about the stringent quality of the oxide, they use the silicon furnaces, actually silicon furnaces. The furnaces are also made, furnace tubes are also made of silicon. But most commonly the furnace tubes are made of fused silica, quartz.

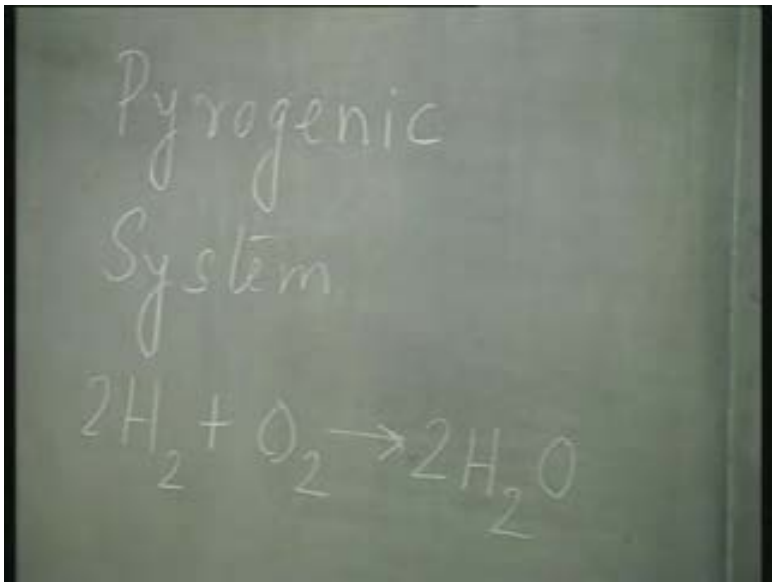
So, you have, you see, two options. One is you pass pure oxygen through it. Oxidation is called the dry oxidation process or you can pass steam. A very common way of passing steam is you have a water bubbler. That is you have a quartz container containing deionized water. Remember, for VLSI processing, we always use deionized water to minimize contamination. So, you put deionized water and this container is placed in a heater, so that its temperature is just below the boiling point of water, just at or below the boiling point of water and you pass a stream of gas. That gas can be oxygen, it can also be nitrogen, it does not matter. That is because, your oxidizing species are going to be steamed, not oxygen and since the steam oxidation rate is much faster than the dry oxidation rate, it does not matter whether you are passing oxygen or nitrogen, because the oxidation rate will be determined by the steam oxidation rate alone.

So, you bubble the gas through this container, so that the outlet is gas saturated with water vapour and that is now passed inside the furnace tube, in a quartz boat. We call it a boat that is a carrier which contains the wafers. So, inside this boat you place the wafers and put it inside the furnace where the requisite temperature is maintained. So, in presence of oxygen or in presence of this gas saturated with water vapour, oxide is formed on silicon. Usually, if you are going to grow a thick oxide, that is for masking purposes, thick oxide that is say, 5000 Angstrom or 6000 Angstrom or thicker than that, then we use a dry wet dry sequence. That is first 10 minutes you carry out dry oxidation,

then you carry out wet oxidation, for whatever time duration, then you follow it up again with a dry oxidation. The idea is to have the interface of a good quality. That is why the final dry oxidation. Then, you get your requisite thickness and you take it out.

There is another possibility. In some systems instead of using the water bubbler arrangement, you directly introduce hydrogen and oxygen in the system where they form steam.

(Refer Slide Time: 26:23)

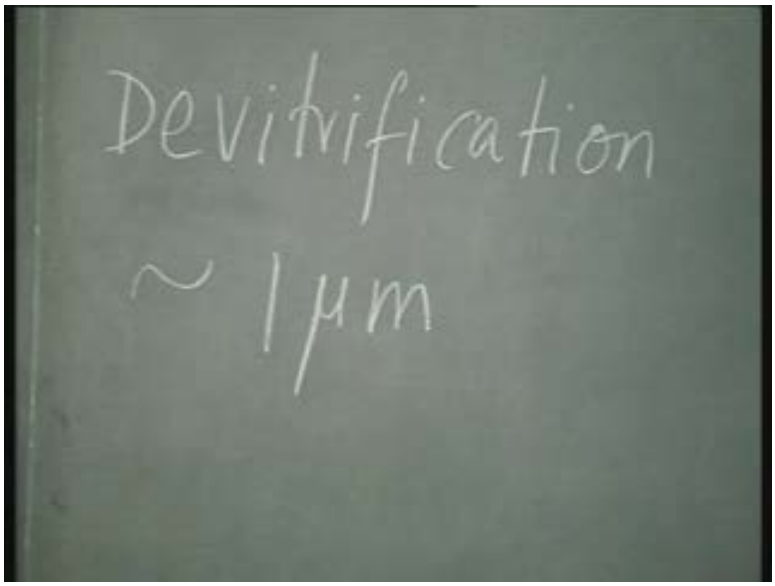


This is called pyrogenic oxidation system, pyrogenic water system or pyrogenic oxidation system, where hydrogen and oxygen - a mixture is introduced and it forms water inside the system and this reacts with the wafer to form the oxide. As I have already mentioned, dry oxide has greater density and better quality. What exactly I mean by better quality, we will come to, a little later. Wet oxidation on the other hand is thicker oxide and is used for masking or isolation. Now, there are a few variations of the theme.

First of all, let us say I want to grow a very, very thick oxide. That means I have to either increase the temperature or increase the time of oxidation, agreed. Now, whenever you are doing either of these two, normally the oxidation temperatures are say, from 900

degree centigrade to 1200 degree centigrade, depending on what you want. These are the normal range of temperature for oxidation. It can be a little less, it can be 800 also, but normal range is 900 to 1200 degree centigrade. Now, if you want to grow very thick oxide that means the wafers have to be subjected to very high temperature for a prolonged duration. This causes a lot of stress on the wafer. Also, the oxide that is formed, it is amorphous. This is what we want, amorphous silicon dioxide.

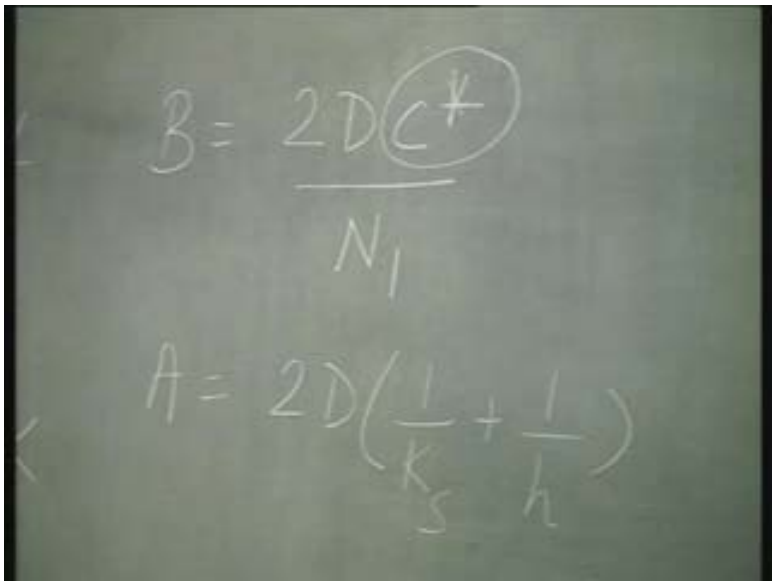
(Refer Slide Time: 28:56)



On prolonged exposure to high temperature, this amorphous oxide can get devitrified, devitrification or devitrified. Whenever we want to grow very thick oxide that means we have to carry out wet oxidation at very high temperature, for very long time and then devitrification occurs. What is devitrification? Instead of this amorphous silicon dioxide, we get poor quality crystalline silicon dioxide and that is no good. That will not serve either for masking or for isolation. It will just be poor quality crystalline silicon dioxide with a lot of cracks, lot of defects and that will not serve either as a good mask or for isolation. So, what is the way out? How do we grow good, thick oxide? By thick oxide I mean more than 1 micron. Normal wet oxidation, the limit is approximately at 1micron, beyond that it tends to devitrify.

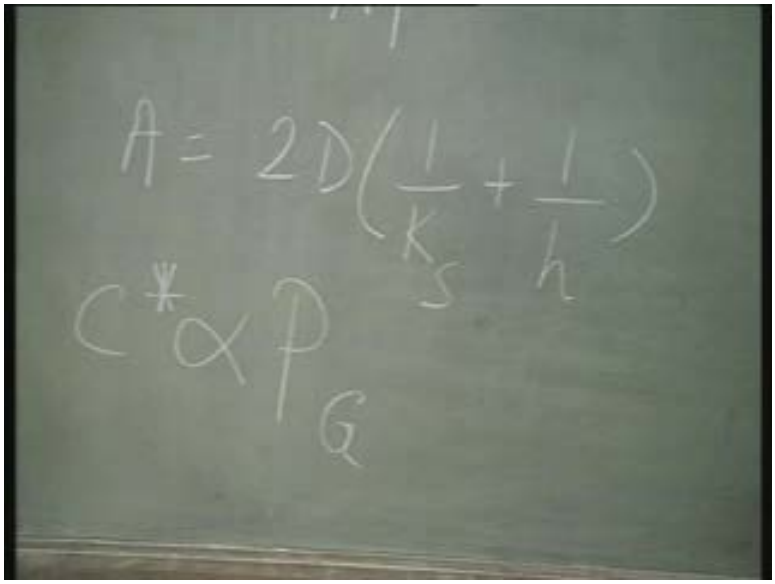
How do we grow good quality thicker oxide? That is one of the challenges which was posed to the VLSI processing technologists. I have to ask you to take another look at the oxidation rate constants. We have defined two rate constants, remember. The linear rate constant and the parabolic rate constant.

(Refer Slide Time: 31:07)



The parabolic rate constant B was 2 D C star by N 1 and A was ....., isn't it. So, you know, B by A will be C star by N 1 into 1 by K S plus 1 upon h, right. This C star that is the equilibrium concentration of the oxidizing species in the oxide, this is proportional to pressure.

(Refer Slide Time: 32:15)



This varies with pressure, gas pressure. So, if you raise the pressure, if you carry out oxidation under high pressure, you are going to increase both B as well as B by A. Both the linear and parabolic rate constants are going to increase. So, this is something which is done. It has a number of interesting fallouts. First of all, your problem of growing thicker oxide is now solved, right. You have increased both the linear and the parabolic rate constants. Therefore, at the same temperature and in comparable time, compared to the atmospheric pressure oxidation, you can now grow a much thicker oxide without the fear of devitrification, because you are not subjecting for any longer duration or at any higher temperature. That is one possibility.

The other possibility is suppose I do not have to grow such thick oxide, it will do to grow even 5000 Angstrom, I have now one more parameter to play with. By increasing the pressure, I can cut down either on temperature or on time and grow comparable thicknesses. Do you understand? Previously, I had only two parameters to play with, temperature and time. Now, I have a third parameter, pressure. By increasing the pressure, I can now afford to cut down either the temperature or the time. What are the advantages? If I can cut down the time, my throughput increases, right. For the same

oxidation which previously needed 1 hour, I can probably finish it in half an hour. That means instead of 1 run, I can now have 2 runs, throughput is doubled.

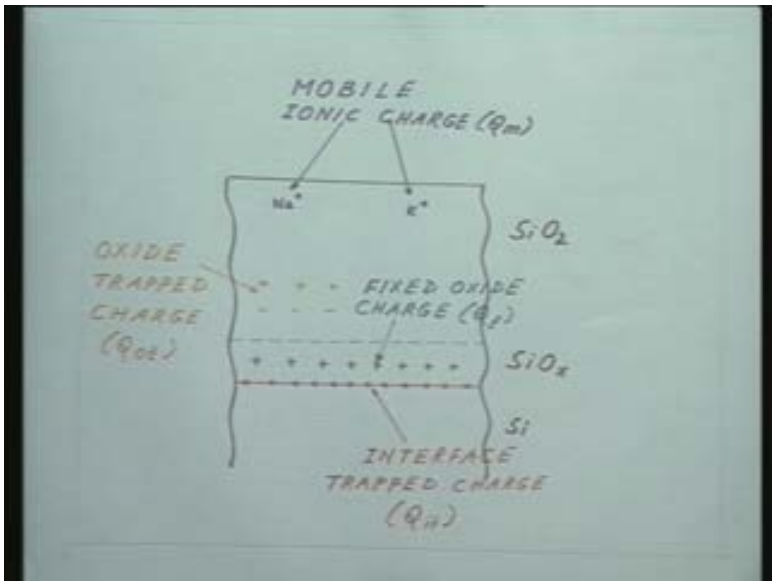
Alternately, let us keep the time same, I have cut down the temperature; so, my thermal budget is reduced. It costs money to run the furnaces, to raise the furnace temperature. They are all electrically heated. Electricity costs a lot of money and also your wafers are now not being subjected to as much thermal stress. The thermal stress is becoming less and you take it from me, the thermal stress problem is getting more and more serious as we increase the diameter of the wafer, because you see the wafer thickness is not really increasing at a proportional rate. For 2 inch wafers, the thickness used to be, may be 300 odd microns. Now even for a six inch wafer, it is say, about 500 microns. So, it is not really increasing at a proportional rate. That means the wafer now is more prone to thermal warping.

Warping means it bends, it buckles at high temperature. So, by cutting down the temperature, you can let your wafers stay more healthy. They are not subjected to as much thermal stress. At the same time, you are cutting down your thermal budget. So, you are saving a lot of money and believe me, engineering is all about saving money. Your product must be cheaper than others product, only then you can sell. So finally, that is the bottom line, the cost is the bottom line.

So, high pressure oxidation in both senses, in more sense than one, it offers interesting possibilities. You can grow thicker oxide, you can also grow normal oxide at a reduced thermal budget or at a reduced time. So, high pressure oxidation is a very interesting proposition today and in our lab, we have one high pressure oxidation system. It is most probably the only one in India and this is indigenously developed. It is developed in our lab, by our people in our lab and high pressure oxidation system is available. So, sometime I will take you to the lab and let you have a look at this.

I said I will talk about the quality of the oxide. How to, I said that dry oxidation will give you better quality oxide compared to wet oxidation. Now, exactly what do I mean by the quality of oxide?

(Refer Slide Time: 37:37)

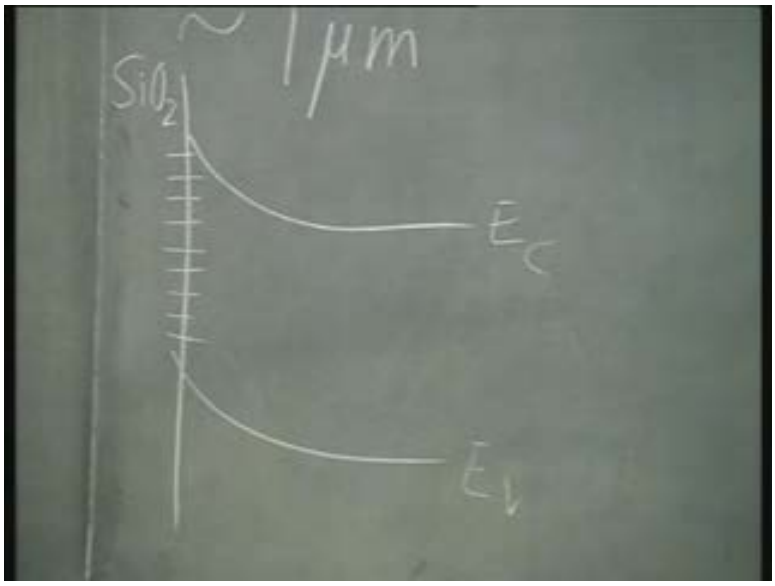


I have a card here, which outlines some of the non idealities inside the oxide. See, the silicon dioxide, ideally it should be an insulator. There should be no charge in this, it should be an insulator with a perfect interface between the silicon and silicon dioxide. Although silicon dioxide does fulfill a lot of these requirements, there is still some non idealities and the most important way of visualizing what the oxide quality is, is by identifying the charges that are present in the oxide, because they represent the non idealities. So, in this card I have schematically represented the various types of charges and where these charges are located.

First of all, note this red line. Red is danger, red also means very important, right. So, this is my very important line, demarcation line that is the line between silicon and silicon dioxide. It is the interface, silicon and silicon dioxide interface. Now, this interface is going to dictate a lot of the things, particularly in relation to a MOSFET. There are non idealities present in the interface called the interface trapped charges. These interface

trapped charges may come about because of various things. It can be because of some mechanical damages already present in the wafer, it can be because of some broken bonds, broken silicon bonds, it can be because of various things. So, this is one major source of non idealities and it adversely affects the MOS device characteristics.

(Refer Slide Time: 40:09)



As the name itself suggests, it is present right at the silicon-silicon dioxide interface and is generally represented, if you look at the energy band diagram, this is your silicon dioxide. It is generally represented as states present all over the forbidden energy gap. These are available energy states present all over the forbidden energy gap. They can trap and detrap electrons that is it takes part in the electronic activity. So, this is one major non ideality. The other major non ideality is a little distance away from the interface in the oxide. This distance is about 100 Angstrom or so, marked with green and you will notice I have put only positive signs there. That is because in silicon thermal oxidation, these charges are mostly always only positive charges and they come about because of incomplete oxidation. Again remember, thermal oxidation is taking place inwards. That is the interface is the region which is oxidized the last.

Now think of a practical process. Towards the end of the oxidation you are going to switch off the gas supply. So, chances are that is where the oxidation process is going to be incomplete. You are reducing the oxidant species supply. So, chances are this is where the oxidation is going to be incomplete. That is why here the composition of the oxide will not be exactly SiO 2, but it will be SiO x, x being less than 2. I have less oxygen in silicon dioxide. This automatically means that I have more positive charges, isn't it. That is why this fixed oxide charges are almost always, they are only positive in nature. The term fixed oxide charge obviously signifies that these charges are fixed. They cannot move about, they cannot be decharged or discharged.

So, this is the second major non ideality and this fixed oxide charge also affects the threshold voltage of a MOS device. Because they are positive, they will push the threshold voltage of a MOS device into the negative detection. That is now you first have to apply a gate voltage in order to neutralize these positive charges, so as to get back the ideal condition, so that much extra negative voltage is added to the actual threshold voltage of the MOS device. That is why we say that if the fixed oxide charge is present, it pushes the threshold voltage to the negative detection and then, all over the oxide I may have this orange marks which are called oxide trapped charges. That is these are the charges trapped in the oxide.

How did these charges come about? There are a variety of sources. One source is suppose your device is exposed to radiation environment, then hole, electron pairs will be generated. Holes are relatively slow, therefore they will get trapped in the oxide. Other possibility is avalanche injection. Suppose you have a MOSFET with a very short channel. That means as the channel length is shorter, the electric field inside the channel is that much higher. So now, your electrons moving in the channel may have acquired sufficient energy to surmount the oxide silicon barrier and get injected into the oxide. Incidentally it is also called hot electron effect. The electrons are hot, they have acquired a lot of energy. Energy is equivalent to temperature, right. That means the electrons are hot, so they can have enough energy to surmount the barrier and get injected into the silicon dioxide. So, there are a variety of reasons, variety of sources by which these oxide

trapped charges may occur. Most important are these two which I just outlined. One is if you are exposing your device to a radiation environment, the other is if avalanche injection is taking place and then finally I have what are called the mobile ionic charges.

I have mentioned the two most important mobile ions, sodium and potassium and you know what are the greatest sources of this sodium and potassium. We humans; you have to handle your wafer and if you are not careful, from human body, the sodium and potassium ions can get into the oxide. Its another major source is water. Unless your water is properly deionized, they contain a lot of dissolved sodium and potassium salts. That means a lot of sodium and potassium ions and they are mobile. They play havoc with the system. They can move all the way from the silicon dioxide surface to the silicon-silicon dioxide interface, depending on the detection of the electric field.

See, they are positively charged, sodium and potassium. So, suppose on top of this you have your metal gate that will be a MOS structure, right, metal oxide and semi conductor. Now suppose you apply a gate voltage, positive gate voltage. All these mobile ions will get pushed to the interface and suppose we apply a negative gate voltage, all these mobile ions will move up towards the surface. They are free to move. This check, whether mobile ions are present or not, it can be done very simply by doing a proper CV analysis, capacitance voltage analysis. It is a routine check in all fab labs, which make MOS devices.

If you have sodium ions or potassium ions, might as well close down the lab, get rid of your sodium and potassium ions, because if they are present, none of your devices are going to work. So, it is a routine check. You must first get rid of sodium and potassium ions. Well, there are various ways and means of getting rid of this sodium and potassium ions. First cardinal rule is your system must be clean. If there are sodium or potassium ions, mobile ions present in your system, it says very badly about the cleanliness of your system. It says that your system is not clean enough. So, first of all you have to tighten the cleanliness of the system and there are also various other ways of getting rid of these mobile ions.

One of them is called chlorine oxidation. In the next class, we will talk about ways and means of getting rid of these unwanted oxide charges, mobile ionic charges as well as the fixed oxide charges and the interface charges. The oxide trap charges, most often they are not really related to processing. They are most often encountered when your device is fabricated and after that it is being exposed either to radiation environment or your MOS device is being operated under very high channel electric field. So, they are not really so much related to the processing conditions, but the other three are - mobile ionic charges, fixed oxide charges and interface charges. So, in the next class we will talk about how to get rid of these charges. That is very important when you want to talk about the gate oxide in a MOS device.