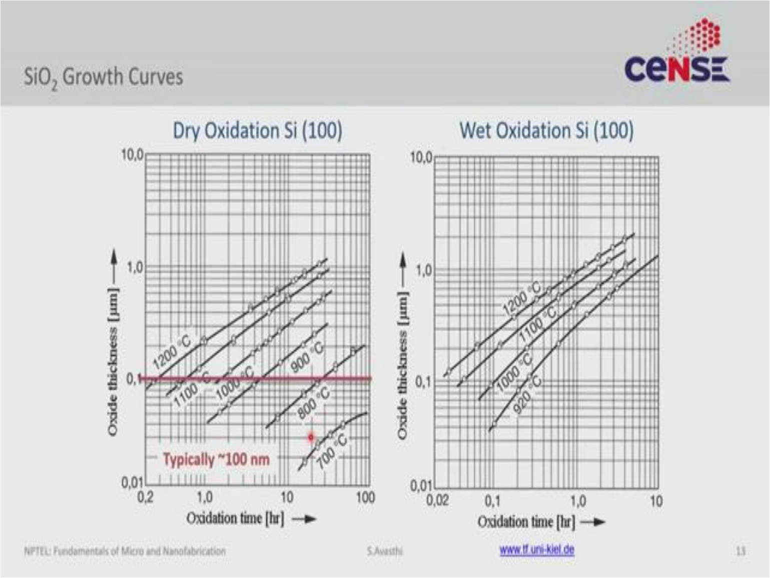
Fundamentals of Micro and Nanofabrication
Prof. Sushobhan Avasthi
Centre for Nano Science and Engineering
Indian Institute of Science, Bengaluru

Lecture - 13
Native Films: Advanced Concepts

This is a module on additive processing and we are specifically going to discuss further details of the native oxide growth. So, in the previous lecture, we went into quite some detail into the exact mechanism of how silicon dioxide is grown on the silicon wafers by just providing an oxidant, either dry (oxygen) or wet (water vapor). In both cases, because of the differences in diffusivity, solid solubility, etcetera the growth rate is very different.

(Refer Slide Time: 01:02)



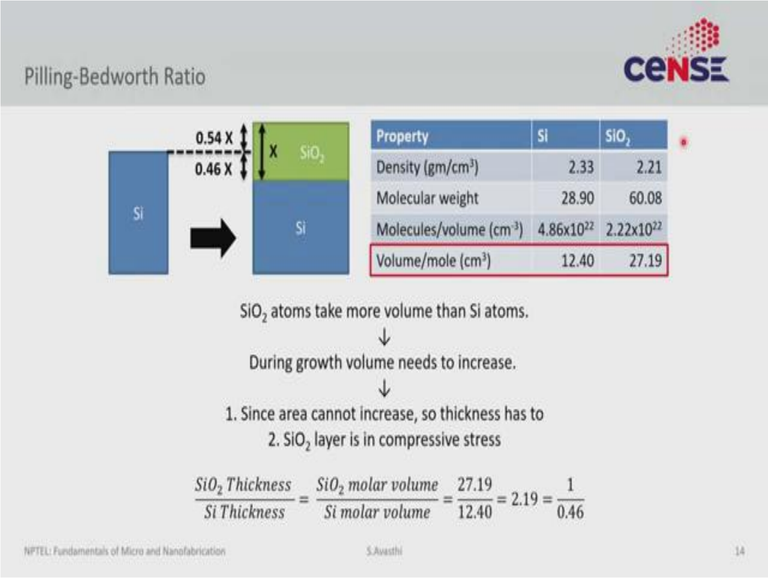
SiO_2 Growth Curves

NPTEL: Fundamentals of Micro and Nanofabrication S.Avasthi www.tf.uni-kiel.de 13

Wet oxidation is much faster than dry oxidation and, in general, the quality of the dry oxidation is much better than wet oxidation. If you need a high-quality and relatively thin dielectric, you do dry oxidation. Typical thickness is anywhere between 20 to 100 nm. If however, you want something thicker, then you use wet oxidation. Wet oxide can be as thick as up to 1-1.5 µm. For even thicker oxides, neither of these processes work; you will have to look at other types of deposition techniques.

One of the consequences of growing the oxide on top of the silicon wafer is that you have to consider the pilling bed worth ratio.

(Refer Slide Time: 01:56)



Pilling-Bedworth Ratio

Property	Si	SiO_2
Density (gm/cm^3)	2.33	2.21
Molecular weight	28.90	60.08
Molecules/volume (cm^{-3})	4.86×10^{22}	2.22×10^{22}
Volume/mole (cm^3)	12.40	27.19

SiO_2 atoms take more volume than Si atoms.

↓

During growth volume needs to increase.

↓

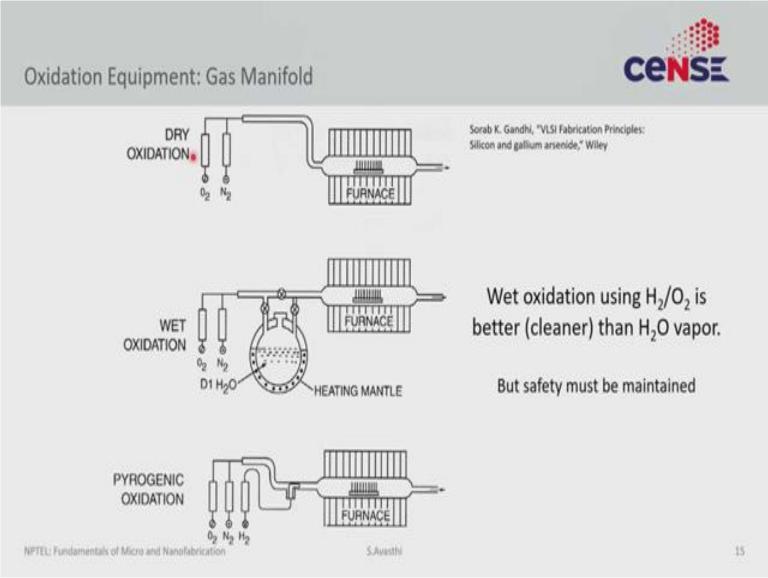
1. Since area cannot increase, so thickness has to
2. SiO_2 layer is in compressive stress

$$\frac{SiO_2\ Thickness}{Si\ Thickness} = \frac{SiO_2\ molar\ volume}{Si\ molar\ volume} = \frac{27.19}{12.40} = 2.19 = \frac{1}{0.46}$$

The density of silicon and silicon dioxide is not the same. As atoms cannot be created or destroyed, the number of atoms must be constant. So, if the silicon dioxide atoms take more volume than silicon, the expansion must happen somewhere. The film area, of course, is constrained because you are doing the oxide growth on the wafer so the thickness increases. You start with a silicon wafer of a certain thickness t and grow a certain x thickness of silicon dioxide. The silicon dioxide thickness will be about 50 % larger than what you expect as the silicon dioxide atoms take more volume than silicon. Since the area cannot change, the thickness does.

The exact values can actually be calculated by looking at the molar volumes. The density and the molecular weight of silicon and silicon dioxide are given. Molar volume (the volume occupied by one mole of molecules) is the ratio of the molecular weight and the density, which is around 12.4 cm^3/mol for silicon and 27.19 cm^3/mol for silicon dioxide. There is almost a factor of 2 difference that reflects in the factor of 2 increase in the effective thickness. In microfabrication, sometimes if you do not take into account this vertical increase in thickness, you can’t understand the cross-section that you get.

(Refer Slide Time: 03:48)


Oxidation Equipment: Gas Manifold

Sorab K. Gandhi, "VLSI Fabrication Principles: Silicon and gallium arsenide," Wiley

Wet oxidation using H_2/O_2 is better (cleaner) than H_2O vapor.

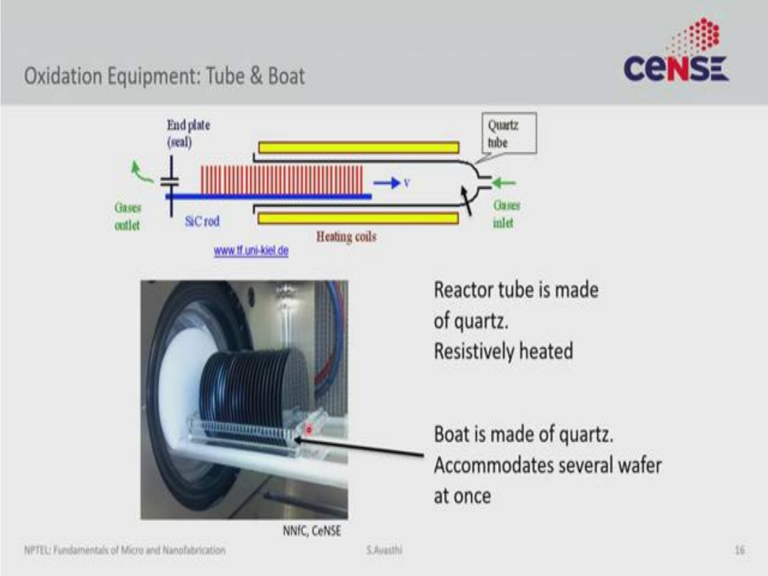
But safety must be maintained

NPTEL: Fundamentals of Micro and Nanofabrication S.Avasthi 15

Next, let’s look at some of the equipment used to grow silicon dioxide. The silicon dioxide growth is one of the most important and challenging aspects of microfabrication, at least on silicon. The silicon-silicon dioxide interface is one of the best in the world of any other material system. Its quality is what allows CMOS to be as good a technology as it is. How do you achieve that remarkable quality interface? we do that by maintaining our equipment. This line diagram is a considerably simplified representation of the tool. You have a tube inside the furnace and inside the tube, you mount the wafers vertically and flow the oxidant with nitrogen. The role of nitrogen will actually be a little more clear down the line, but simply speaking, it is the nonreactive component that allows you to control the recipe a little better and reduce defects.

In wet oxidation, things would be a little more complicated; you have a bubbler filled with ultrapure water, hopefully, DI water. You bubble the inert nitrogen through this water and as the nitrogen goes through the water, it picks up some of this water vapor that gets transported into the furnace where it does the oxidation. A more modern method of doing the same thing is pyrogenic oxidation where instead of having a bubbler filled with water which can get contaminated with time, you flow hydrogen and oxygen in an ultra-pure gaseous form. Hydrogen and oxygen internally react inside the furnace to form water vapor which does the oxidation. An important thing to notice here is when you flow hydrogen and oxygen together, you have to be extremely careful because you are essentially creating a mini-explosion. So, the electronics and the safety structure of this pyrogenic furnace need to be carefully designed.

(Refer Slide Time: 06:05)



Oxidation Equipment: Tube & Boat

www.tf.uni-kiel.de

Reactor tube is made of quartz.
Resistively heated

Boat is made of quartz.
Accommodates several wafer at once

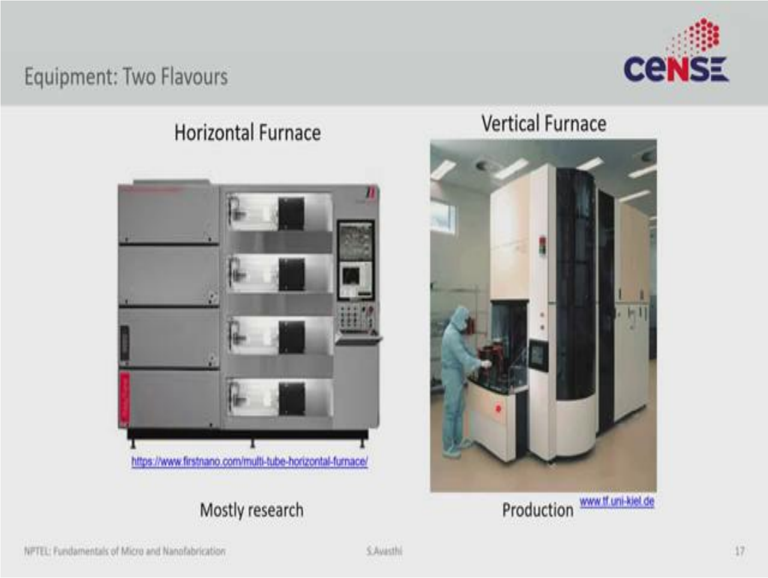
NNfC, CeNSE

NPTEL: Fundamentals of Micro and Nanofabrication S.Avasthi 16

Because you are so sensitive to purity, this is the most sensitive step. You want your gate oxidation to be absolutely pure. So, you try very hard to remove any contamination or a source of irreproducibility. Anything that is uncontrolled must either be removed or controlled. For example, you may notice that the wafers are loaded into a quartz port. You cannot have, for example, a metal port or a or an impure glass port. It has to be an ultra-pure quartz port made from the same ultra-pure quartz as the rest of the furnace.

The baffle here allows you to reduce the heat loss, thereby maintain the temperature in the furnace more accurately, which is very important if you want to control the thickness and the quality of the oxide very accurately. Typically at the other end, you have a face seal that seals the tube so the contamination in the atmosphere cannot get in. However, you still need to flow the gas, so, there probably is a little outlet connected to the exhaust. So, there is a continuous flow of fresh oxygen. other than that, it is a relatively simple system with complex electronics.

(Refer Slide Time: 07:36)



Equipment: Two Flavours

Horizontal Furnace

https://www.firstnano.com/multi-tube-horizontal-furnace/

Mostly research

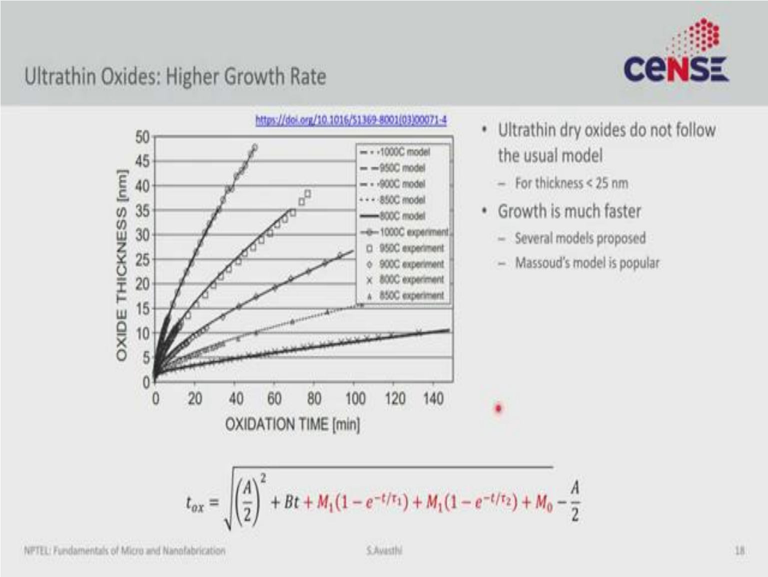
Vertical Furnace

Production www.tf.uni-kiel.de

NPTEL: Fundamentals of Micro and Nanofabrication S.Avasthi 17

There are several vendors to will sell you a high-quality oxidation furnace. Older technology or small wafer models tend to be the horizontal reactor. Production tools and large area wafers tend to have a vertical tube. The reason for vertical versus horizontal is purely for a space constraint. The cleanroom is a very expensive real estate, so, you do not want very large equipment and one way to save space is to have a long tube lying vertically instead horizontally as it takes less space.

(Refer Slide Time: 08:13)



Ultrathin Oxides: Higher Growth Rate

https://doi.org/10.1016/S1369-8001(03)00071-4

- Ultrathin dry oxides do not follow the usual model
  - For thickness < 25 nm
- Growth is much faster
  - Several models proposed
  - Massoud's model is popular

$$t_{ox} = \sqrt{\left(\frac{A}{2}\right)^2 + Bt + M_1(1 - e^{-t/\tau_1}) + M_1(1 - e^{-t/\tau_2}) + M_0} - \frac{A}{2}$$

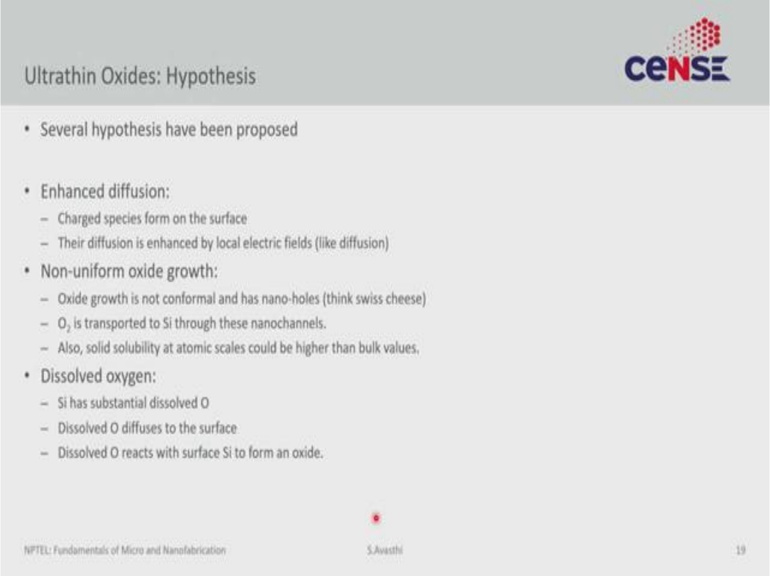
NPTEL: Fundamentals of Micro and Nanofabrication S.Avasthi 18

The next thing we will discuss is the growth of extremely thin oxide. We went into quite a detail into the Deal-Grove model where we looked at how the oxide grows on the silicon. It is a very elegant model that accurately explains reality but breaks down when you are looking at very thin oxides or the growth kinetics at the very start of the process.

For very small times or at low temperatures, you will grow oxides that are very thin, on the order of 5-10 nm. The growth rate at these very small thicknesses deviates from the Deal-Grove model, which typically underestimates the growth rate at these very small thicknesses. This problem can be solved in 2 ways; one is to understand the mechanism and the other is to apply a whole bunch of fitting parameters that at least accurately model the process without necessarily understanding the underlying physics. The Massoud model takes the latter approach and is pretty popular.

You don’t need to remember the formula. The point here is to sensitize you to the fact that while trying to grow ultra-thin oxides, you have to be a little more careful in using the Deal-Grove model. The second point is that the time dependence is very strong; the growth rate exponentially depends on the run time. Very small changes in time can make your thickness go up and down very quickly. This makes it harder to accurately grow a very thin oxide layer (say 2 nm) repeatably; you have to be very careful in your recipe.

(Refer Slide Time: 10:14)



Ultrathin Oxides: Hypothesis

- Several hypothesis have been proposed

- Enhanced diffusion:
  - Charged species form on the surface
  - Their diffusion is enhanced by local electric fields (like diffusion)
- Non-uniform oxide growth:
  - Oxide growth is not conformal and has nano-holes (think swiss cheese)
  - O_2 is transported to Si through these nanochannels.
  - Also, solid solubility at atomic scales could be higher than bulk values.
- Dissolved oxygen:
  - Si has substantial dissolved O
  - Dissolved O diffuses to the surface
  - Dissolved O reacts with surface Si to form an oxide.

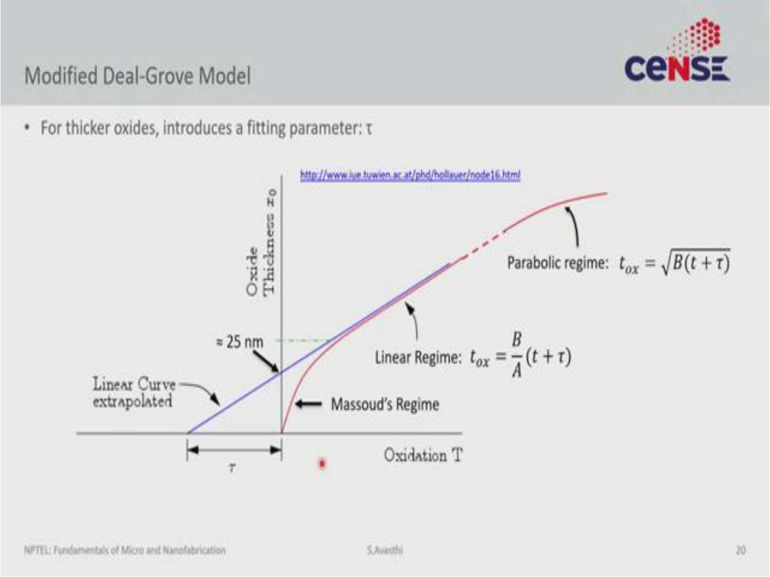
NPTEL: Fundamentals of Micro and Nanofabrication S.Avasthi 19

There are several hypotheses on a physical level of why actually this happens. I do not think there is consensus or clarity on what exactly is happening in these ultra-small thicknesses. Here are some of the possibilities that could be happening, but this is still an open question in my knowledge.

One of the proposed hypotheses is enhanced diffusion; charged species form on the surface and are much more mobile for small thicknesses than what the Deal-Grove model predicts. This enhanced diffusion leads to an enhanced growth rate. Next is non-uniform oxide growth. The Deal-Grove model assumes that the oxide grows layer by layer in a very careful fashion, which may not be the case initially. So, instead of having a uniform single thickness which allows you to use a 1D model, you may have a non-uniform thickness, where some places are thinner or thicker, making a more complicated model necessary. Most of the diffusion might actually be happening through the thinner regions, increasing the growth rates than the expectation. Also, solid solubility is a bulk concept and does not scale well. If you have only a few atoms on the surface, the solid solubility can be higher than the bulk, increasing the growth rate.

Some mechanisms argue that the silicon substrate itself has some dissolved oxygen. If you remember our discussion on the substrate growth, the oxygen concentration in the silicon substrate is ~ 10^{16}-10^{17}/cm^3 and that oxygen might play a role in the growth of oxide at these very small thicknesses.

(Refer Slide Time: 12:11)

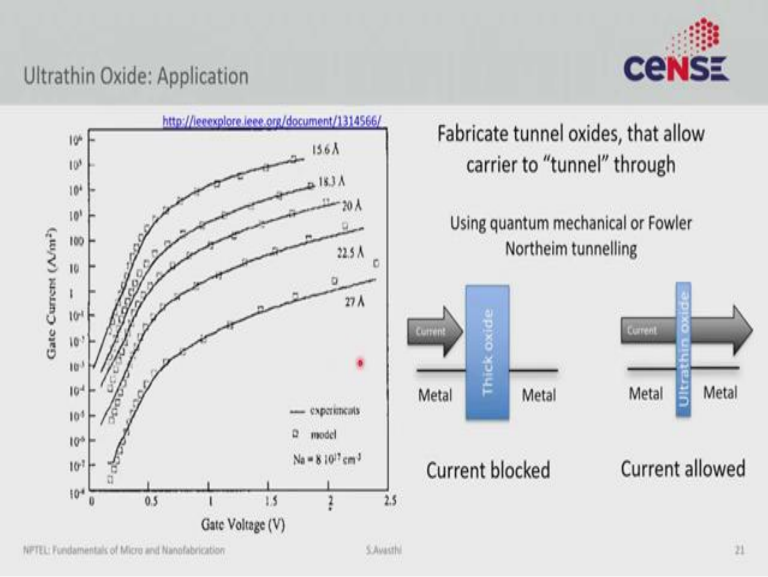


Modified Deal-Grove Model

- For thicker oxides, introduces a fitting parameter: τ

Practically, for higher thickness, this problem is handled by modifying the Deal-Grove model by adding a simple fitting parameter. This red line is the data. You have a linear and a parabolic regime predicted by the Deal-Grove model, but if you extrapolate the linear region (the blue line) down to lower times or thicknesses, it starts deviating and this is where you see the Massoud's regime or the ultra-thin oxide growth regime. However, if you are not interested in modeling accurately in this regime, you can just extrapolate this red line. This fitting parameter τ can then be used to fit the actual data without changing the formula that you already know. So, instead of having $t_{ox} = \frac{B}{A}t$, as predicted by a simple Deal-Grove model, you have $\frac{B}{A}(t + \tau)$, where τ is simply a fitting parameter. Using this fitting parameter, you can accurately get the oxide thickness above this green line, where the blue and the red curves are the same. This is what a lot of the oxide growth models use to predict the thickness.

(Refer Slide Time: 13:31)



Ultrathin Oxide: Application

http://ieeexplore.ieee.org/document/1314566/

Fabricate tunnel oxides, that allow carrier to "tunnel" through

Using quantum mechanical or Fowler Northeim tunnelling

Current blocked

Current allowed

NPTEL: Fundamentals of Micro and Nanofabrication

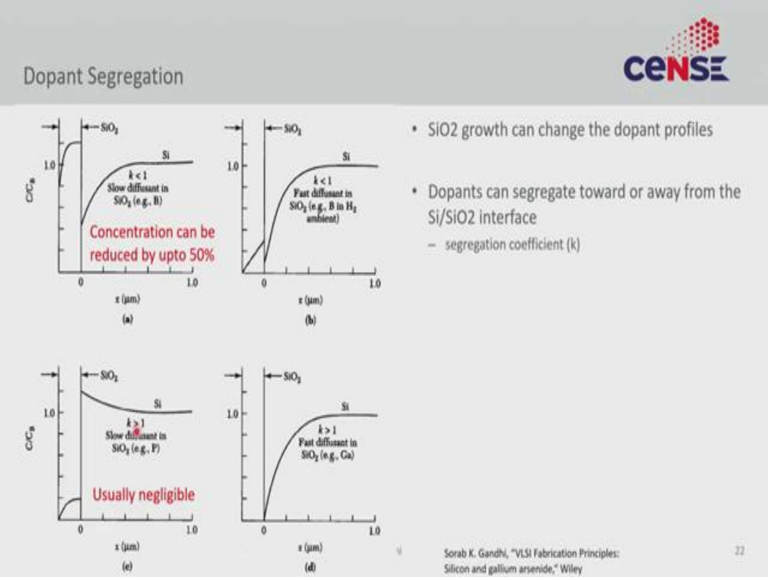
S.Avasthi

21

However, ultra-thin oxides are important for certain applications and one is to use them as tunneling contacts. We will not go into details of tunneling. Quantum mechanical tunneling predicts that if the barrier is very thin, the carriers can make their way through. In this example, if the oxide between the two metals is very thick, the current cannot flow through the oxide because the thick oxide acts as an insulator. However, if the insulator is very thin, then there is a finite probability for the electrons to make their way from left to right and that would lead to a current. These tunneling oxides are very useful if you want to achieve passivation while maintaining the current carrying capacity. If this does not make sense, just take my word that the ultra-thin oxides are useful.

In order to grow these ultra-thin oxides between 1-2 nm, you need to have a very precise recipe and for that, the Massoud and associated models become extremely important. Here is an example of what you can expect in terms of tunneling current. The Gate current vs voltage for various oxides thicknesses is plotted. For $t_{ox} <$ 15.6 Å or 1.5 nm, you can pull in a large current (10^3 A/cm^2).

(Refer Slide Time: 15:02)



Dopant Segregation

- SiO2 growth can change the dopant profiles
- Dopants can segregate toward or away from the Si/SiO2 interface
  - segregation coefficient (k)

Sorab K. Gandhi, "VLSI Fabrication Principles: Silicon and gallium arsenide," Wiley

Another advanced concept worth thinking is dopant segregation. If you naively think about thermal oxidation of silicon, you would expect that the dopant concentration would be maintained in silicon and the oxide. However, practically, the details are a little complicated. We will the concept of the distribution coefficient k and the diffusivity D. As discussed during crystal growth, k tells you whether an impurity would like to remain in the material A or B. The diffusivity D tells you whether the material whether a certain impurity will diffuse fast or slow in a certain medium. Depending upon the values of k and D, we have 4 cases.

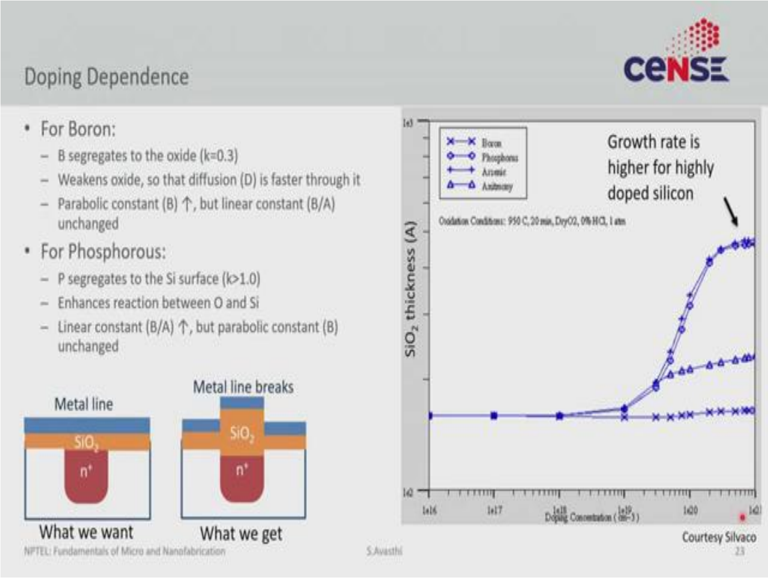
Case 1 ($k < 1$, small D): the distribution coefficient $k < 1$ means the impurity would like to remain in silicon dioxide vs silicon. And because of small D in the silicon dioxide, the impurity does not diffuse much in the silicon dioxide. As you oxidize more and more silicon, the impurity (Boron here) tries to remain in the oxide preferentially, and not in silicon. So, the silicon at the interface gets depleted of boron concentration as boron accumulates in silicon dioxide. However it does not diffuse in silicon dioxide very fast, so, it keeps accumulating in silicon dioxide and stays there.

Case 2 ($k < 1$, Large D): The impurity would like to remain in silicon dioxide vs silicon; however, the diffusion coefficient is now fast. Unlike boron in silicon dioxide, which does not diffuse, imagine some impurity, for example, boron in hydrogen ambient where it would diffuse through the silicon dioxide. Because of the distribution coefficient, you still get a depletion of dopants at the silicon surface, but those dopants do not remain in silicon dioxide; they out diffuse from the silicon dioxide into the gaseous medium. Here, boron goes into silicon dioxide because of $k < 1$ and then into the gas from silicon dioxide because of a large D. This leads to a severe depletion of the dopants at the interface.

Case 3 & 4 ($k > 1$): When the distribution coefficient $k > 1$, the dopants would preferably lie in silicon and not in silicon dioxide. Again, you have these sub-cases where the diffusivity D is low, for example, in phosphorus or high, for example, in gallium. Because of $k > 1$, the dopants accumulate in the silicon and the concentration increases at the interface. However, some amount of phosphorus does get incorporated into the oxide, so, you see some concentration, it will not out diffuse because the diffusion coefficient is very low; it just accumulates in the oxide layer. For fast diffusion, for example, in the case of gallium, all the dopants in the oxide keep out diffusing into the gas. So, the concentration in the oxide can never be maintained. Even though $k > 1$, because there is no concentration of the dopant in the oxide, it keeps pulling the dopant from the silicon. So, unlike in case 3, where you see an accumulation, here, you see a depletion of the dopants. So, in those 3 cases, you see a depletion of impurities and accumulation in one.

Now, why is this relevant? It is relevant because if you start with a uniformly doped wafer and do an oxidation step on it, at the interface, the doping concentration changes and the severity of that change depends upon the dopant and the processing conditions. If you are not careful about these effects you can under or overestimate the amount of dopant at the interface which can change your device characteristics.

(Refer Slide Time: 20:50)



Doping Dependence

- For Boron:
  - B segregates to the oxide (k=0.3)
  - Weakens oxide, so that diffusion (D) is faster through it
  - Parabolic constant (B) ↑, but linear constant (B/A) unchanged
- For Phosphorous:
  - P segregates to the Si surface (k>1.0)
  - Enhances reaction between O and Si
  - Linear constant (B/A) ↑, but parabolic constant (B) unchanged

Growth rate is higher for highly doped silicon

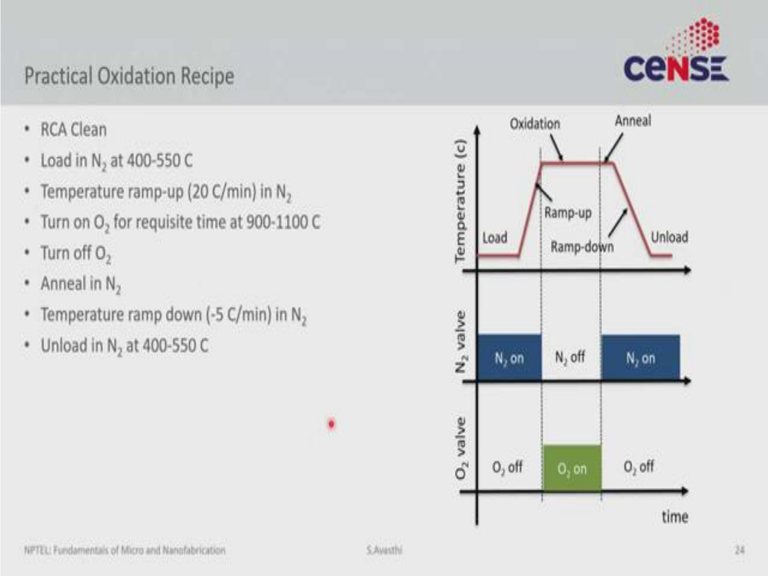
Courtesy Silvaco

Another complication is the effect of doping. So, based on the simple Deal-Grove model, you would not think that dopants would actually change anything. Unfortunately, that is not the case; at high enough concentrations, dopant tends to increase the oxide growth rate. At a very high doping concentration, 10^{20}-10^{21}/cm^3, you would get a thicker oxide than the simple Deal-Grove model would suggest. This enhancement is most severe for n-type impurity such as phosphorus as compared to boron. The curves for phosphorus and arsenic are significantly In the previous slide, we saw that boron has a $k < 1$, ~ 0.3.

So, boron that is in the oxide weakens it and increases the diffusion through that oxide. Faster diffusivity increases the parabolic rate constant, but the linear rate constant would remain unchanged and that higher parabolic rate constant causes the enhancement in the oxide growth rate. For phosphorus, things are slightly different, as it segregates primarily into silicon because the distribution coefficient $k > 1$. Because it accumulates in the silicon at the interface, the phosphorus concentration goes up. This enhances or catalyzes the reaction between oxygen and silicon. So, the linear constant will increase and the parabolic constant will remain the same. As the linear rate increases, you get an increase in the growth rate. In the linear regime, phosphorus has a stronger effect than boron.

There is a reason why these details are important. If you oxidize a silicon surface with an n^+ layer, the silicon dioxide layer will have a nonuniform thickness because the rate of growth at the n^+ region will be larger than the n region. If this difference is severe, then the metal line you deposit on the top will be discontinuous or broken. This can cause your fabrication process to fail. In some of these niche areas when you start having these weird problems, it is worth going back at the basics and question all the assumptions that you have made in modeling or in the process.

(Refer Slide Time: 24:03)

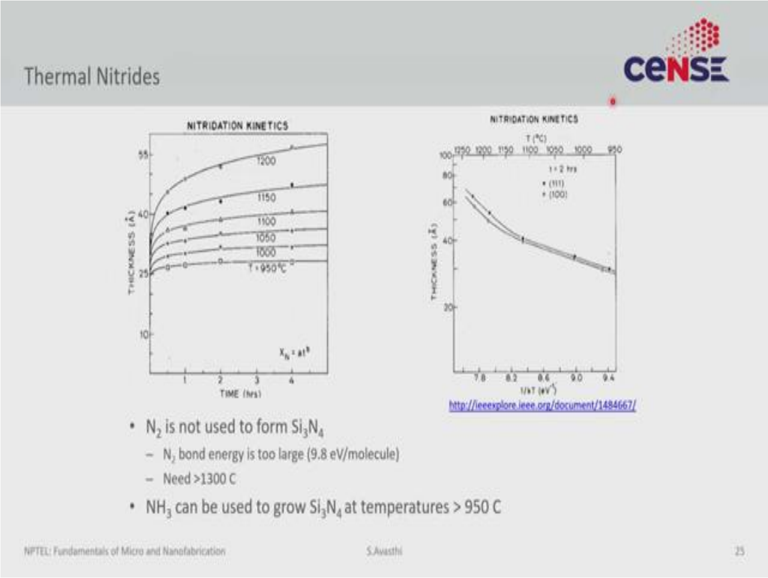


Practical Oxidation Recipe

- RCA Clean
- Load in N_2 at 400-550 C
- Temperature ramp-up (20 C/min) in N_2
- Turn on O_2 for requisite time at 900-1100 C
- Turn off O_2
- Anneal in N_2
- Temperature ramp down (-5 C/min) in N_2
- Unload in N_2 at 400-550 C

Practically this is what a typical oxidations recipe looks like. You take a wafer and do an RCA clean, then load it in nitrogen at 400-550^{o}C, ramp up the temperature in nitrogen, then introduce oxygen for the required amount of time and temperature that you want the oxidation to be done at. Turn off the oxygen, anneal for a bit in nitrogen, ramp the temperature down and then unload. We will discuss the reason for this anneal later.

(Refer Slide Time: 24:42)



Thermal Nitrides

http://ieeexplore.ieee.org/document/1484667/

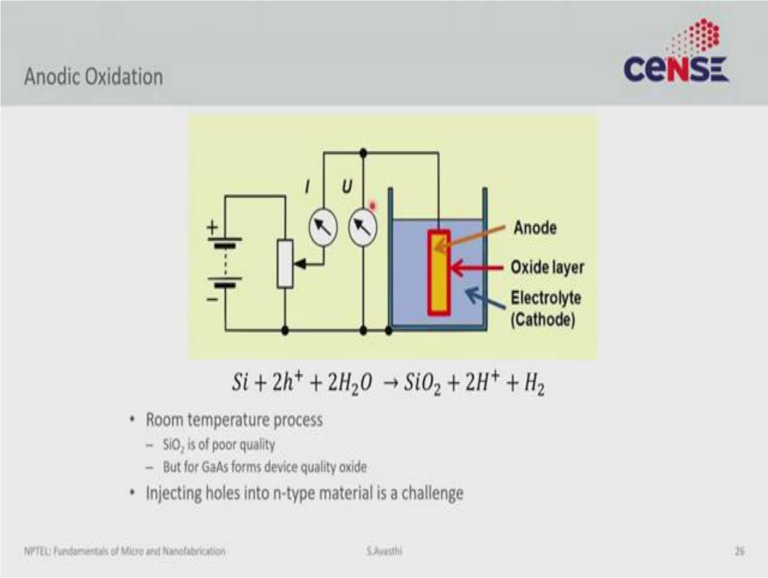
- N_2 is not used to form Si_3N_4
  - N_2 bond energy is too large (9.8 eV/molecule)
  - Need >1300 C
- NH_3 can be used to grow Si_3N_4 at temperatures > 950 C

NPTEL: Fundamentals of Micro and Nanofabrication S.Avasthi 25

We have only discussed oxidation but theoretically, you can even do native nitridation. Growing silicon nitride is not possible with nitrogen because it is a very inert molecule and has a very strong bond. You require remarkably high temperatures (> 1300°C) to break that bond and form silicon nitride, which is not practical. Instead, you can introduce ammonia, which is a more reactive form of nitrogen that can form nitride at temperatures ~ 950-1000°C.

If you look at the paper that explored this, the thicknesses that you grow are very small because silicon nitride does not allow nitrogen diffusion very well. Practically, this is never done; if you have to grow silicon nitride there are better ways of doing so.

(Refer Slide Time: 25:39)



Anodic Oxidation

$$Si + 2h^{+} + 2H_2O \rightarrow SiO_2 + 2H^{+} + H_2$$

- Room temperature process
  - SiO_2 is of poor quality
  - But for GaAs forms device quality oxide
- Injecting holes into n-type material is a challenge

You can also do anodic oxidation that is very popular for a lot of metals; however, in silicon, the oxide that you grow with anodic oxidation, which is an electrochemical way of growing the oxide is not very good. So, it is rarely used in microfabrication. However, if you want a very thick layer, you may consider it. For gallium arsenide, you tend to form slightly better quality oxide than silicon so, you might consider this method for it.

With that, we will end this lecture, and in the next lecture, we will go into little more details of the defects that can actually form at the silicon-silicon dioxide interface.

Thank you.