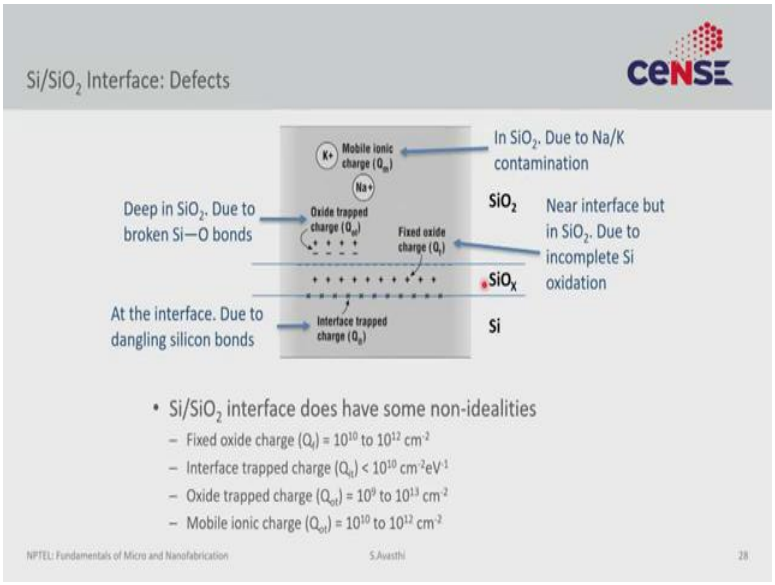
Fundamentals of Micro and Nanofabrication
Prof. Sushobhan Avasthi
Centre for Nano Science and Engineering
Indian Institute of Science, Bengaluru

Lecture - 14
Native Films: Defects at Si/SiO2 interface

This lecture in the module of additive processing is the final part of the three-part series looking at native oxides. In this lecture, we will look at the silicon-silicon dioxide interface; silicon-silicon dioxide interface is of remarkably high quality, so, it is worth looking into it a little more closely and try to understand the defects at that interface and the processing techniques we can use to reduce them.

(Refer Slide Time: 01:04)



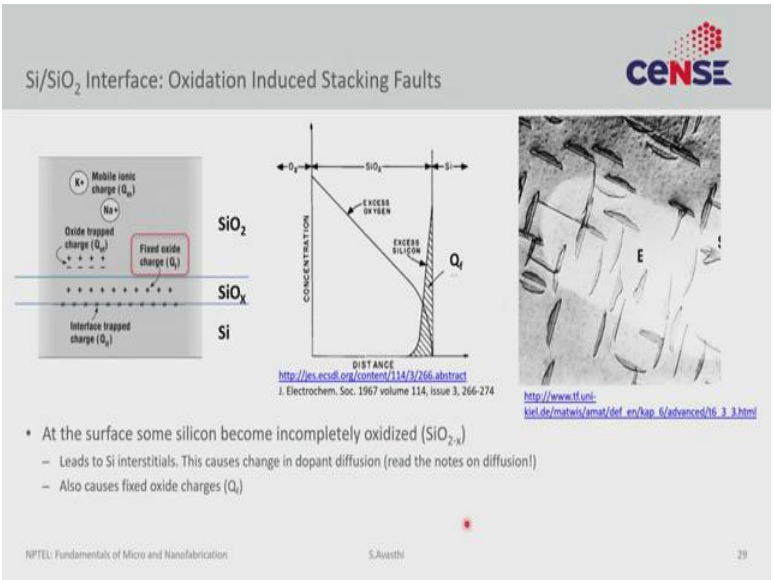
Primarily, there are four-five types of defects that can exist at the silicon-silicon dioxide interface. The silicon is crystalline and silicon dioxide is amorphous. So, there must be some transition region where the oxide is not stoichiometric, represented by SiOx. This lack of stoichiometry would imply some amounts of defects, typically called fixed oxide defects or Qf. They tend to happen at the interface but in silicon dioxide.

The second type of defects called interface traps can occur at the interface but in silicon. As silicon reacts with oxygen, there is no guarantee that all the silicon atoms will be bonded to oxygen. If a silicon atom is not bonded to an oxygen or to another silicon atom, it has a dangling bond or an unsatisfied valency that lead to defects called interface

trap defects or Qit. As you go away from the interface, you start seeing some of the bulk defects in silicon dioxide; this is not quartz or high-quality crystalline silicon dioxide but is amorphous silicon dioxide. So, there can be broken bonds of silicon-oxygen every now and then and these missing or broken or unsatisfied bonds would cause some amount of trap charges that can be positive or negative in the bulk of the silicon dioxide.

Another type of defect that can exist in the bulk of the silicon dioxide is mobile ions. We try very hard to clean the wafer, use the cleanroom, ultra-pure chemicals so that we never have sodium or potassium contamination. Practically, there is always a little bit of ionic contamination no matter how much we try and that ionic contamination would cause some amount of defects that we touched upon in the introduction. They can cause some transient effects, hysteresis, and drift in the characteristics of the devices. These are the four types of defects at the silicon-silicon dioxide interface. In general, the quality of this interface is extremely good, which means that most of these defects are in the low 10^{10}-10^{12}/cm^2. The number of silicon atoms is around 10^{15} /cm^2. So, these numbers represent one in 1000, 10000 or one in 100000 type of defect density, which is very low.

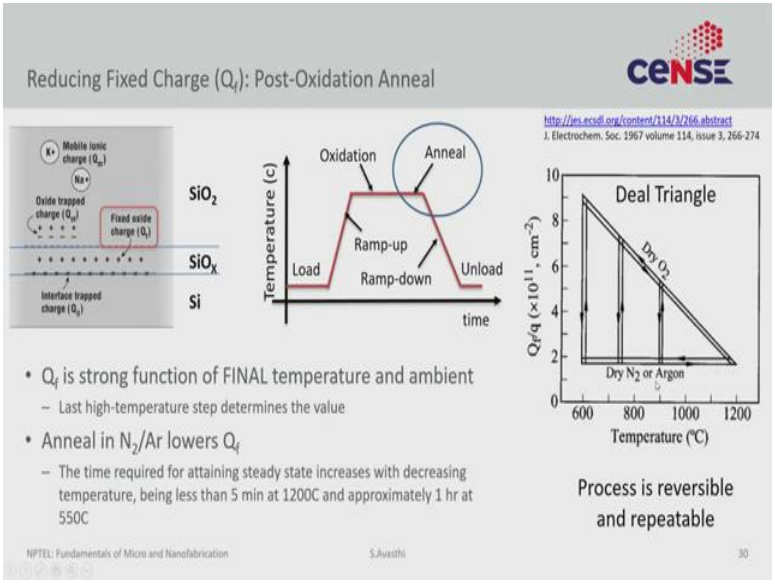
(Refer Slide Time: 04:12)



Let's start with the fixed oxide charges that exist in the silicon dioxide, but near the interface. They happen is because of the diffusion of silicon. We have already discussed the diffusion of oxygen from the gas phase through a layer of silicon dioxide under the concentration gradient. This oxygen reacts at the interface to form silicon dioxide. While

oxygen diffuses from left to right because the concentration of oxygen is higher on the left, silicon also diffuses from right to left because the concentration of silicon is higher on the right and lower on the left. This excess silicon that is diffusing from silicon into silicon dioxide causes some amount of defects in the silicon dioxide because it makes the silicon dioxide non-stoichiometric. As these silicon atoms leave the silicon substrate, they leave voids or stacking faults that are also a problem on the silicon surface. We shall talk about those defects a little later. This is the fundamental origin of Qf.

(Refer Slide Time: 05:39)



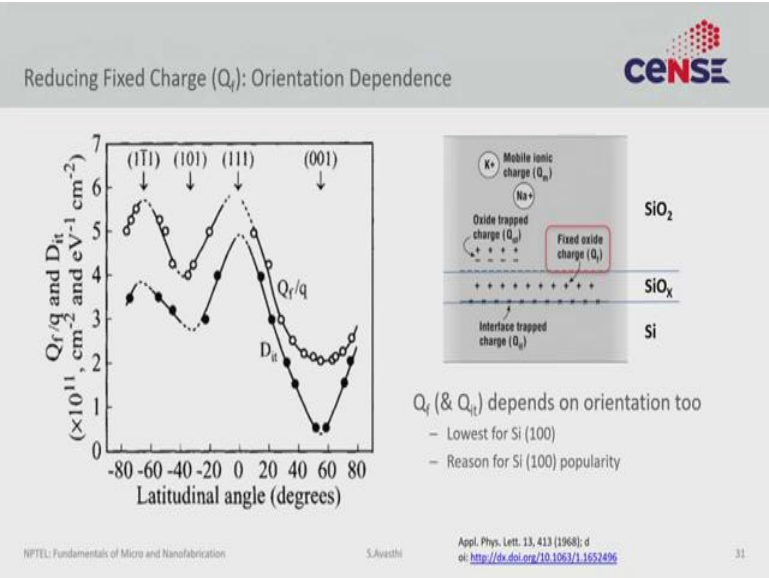
The way we can reduce this fixed oxide charge density is through post oxidation anneal. If you remember the last lecture, we saw a generic recipe for silicon oxidation, and in that recipe, we had nitrogen anneal step before ramping the temperature down. This is a result that was published in 1967. It is called a Deal triangle in honor of one of the author of the paper – Deal, and it is a very interesting and remarkably powerful result. The first thing they observed was that the amount of defect density at the interface, the fixed oxide defect density depends upon the ambient (oxygen vs nitrogen) and the temperature of the wafer. As you increase the temperature during the dry oxidation, the defect density falls. As much as possible, try to grow your oxide at higher temperatures. If your thermal budget allows, go as high as possible. The second thing to note is irrespective of the temperature, if you anneal the wafer in an inert atmosphere, for example, dry nitrogen or argon you always maintain a very low defect density.

The defect density is only high when there is oxygen in the ambient and it does not change with the temperature if the ambient is inert. Of course, inert ambient does not allow you to grow the oxide, but during annealing, you do not change the interface defect density. Furthermore, this triangle is reversible which means that if you start oxidizing a wafer at ~ 900^{o}C, you would get a defect density of ~ 5×10^{11}/cm^{3} but if you anneal the same wafer under nitrogen for half an hour before unloading (maintaining the same temperature), the defect density would fall to 2×10^{11}/cm^{3}.

That is why this step is added to the recipe because it reduces the defect density. More interestingly, this is a completely reversible diagram. If you take that wafer with the low defect density and reduce the temperature in nitrogen, the defect density will not change, but if you flow oxygen on the same wafer before unloading, the defect density will increase and reach the number that you would expect had you done the oxidation at that temperature (say 700^{o}C) in the first place.

This is a state diagram and the only thing the wafer remembers is the last temperature and ambient you leave it in and it maintains that level of defect density. As long as you add an inert anneal at the end of your oxidation recipe, you can always guarantee a low defect density irrespective of the oxidation temperature.

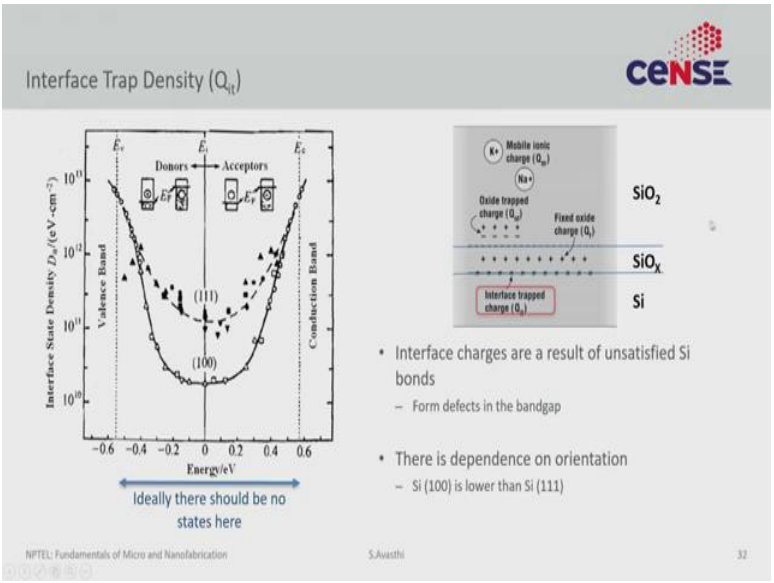
(Refer Slide Time: 09:32)



This fixed oxide charge also has an orientation dependence. So, (100), (111) and (110) surfaces do not all have the same density. In general, the (100) surface tends to give the

lowest defect densities and this is one of the reasons silicon (100) wafers gained popularity when CMOS came along. For BJTs, it does not matter because the silicon-silicon dioxide interface is not that important, but in the COMS technology, where the field-effect transistors are used, the silicon-silicon dioxide interface is extremely critical. Hence, they switched from (111) orientation to the (100) orientation.

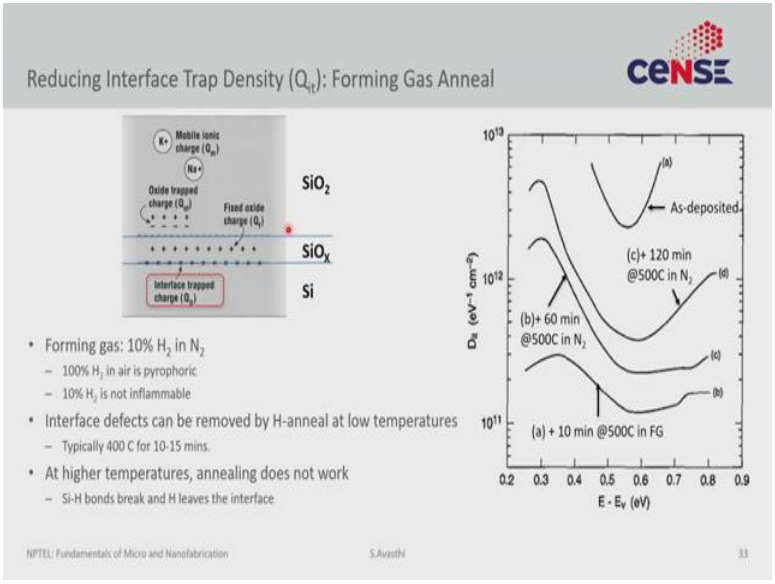
(Refer Slide Time: 10:18)



Let’s now discuss a little bit about the interface trap density. As we discussed, these are the defects that exist in silicon, but near the interface and they exist because silicon has dangling bonds or unsatisfied valencies and because of those unsatisfied valencies, you have mid-gap defects. One way to represent these defects is given in the figure on the left, where you have defect density as a function of energy within the gap.

Silicon is a semiconductor. Ideally, it should have a density of states in the valence band, conduction band and no states in the bandgap. Practically, that does not happen and you have a finite number of defects in the forbidding gap because of these dangling bonds at the interface. This graph shows the density of these defects vs energy in the bandgap. There are two things to notice: there is an orientation dependence just like previously, (100) oxidized surface tends to have lower defect density than (111). The second thing is this U shaped curve, so, a lower density of defects in the middle, higher density of defects at the edges.

(Refer Slide Time: 12:04)



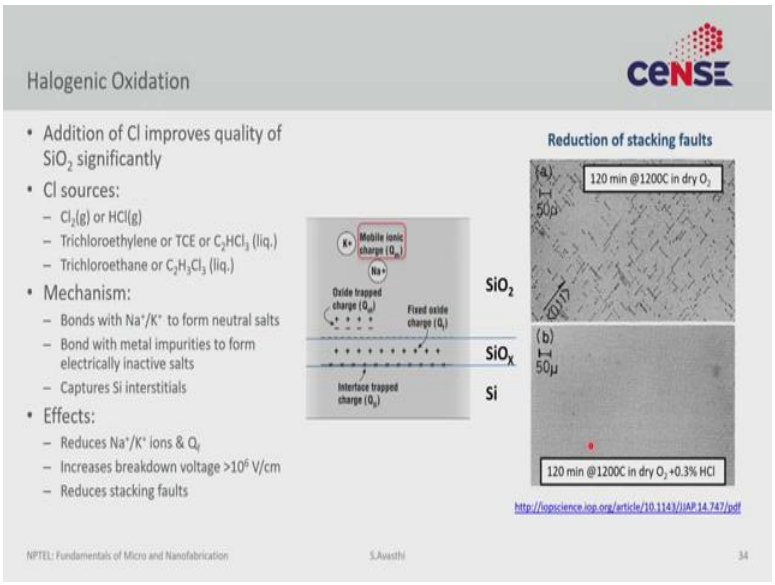
How do we fix this? the easiest way is to use forming gas anneal (FGA). Forming gas is a mixture of hydrogen and nitrogen and the reason we use forming gas and not pure hydrogen is purely for safety. A 100 % hydrogen is a very dangerous gas as it can explode on contact with oxygen. However, 10 % of hydrogen in nitrogen is very safe and you can use it in an ambient atmosphere without any danger of explosions. That is why the industry likes forming gas more than pure hydrogen.

What does forming gas do? the active component here is hydrogen. Hydrogen diffuses through the oxide and reacts with these silicon dangling bonds at the interface, forming a silicon-hydrogen bond that is strong and passivates the unsatisfied valency and hence, removes the defect. The figure shows how effective this process can be. You start with an as-deposited defect density in the same U-shaped curve that we saw earlier. Now, if you anneal this wafer in forming gas ambient at 500^{o}C for 10 minutes, you see more than an order of magnitude change in the defect density.

However, you have to be careful. If you take this same wafer, turn off the forming gas and introduce nitrogen which is inert, you make the defect density worse. The hydrogen from the forming gas would come off or desorb from the interface, increasing the interface defect density. The longer you keep the wafer in this nitrogen atmosphere, the worse it will become. The defect density became very good in curve b, a little worse in curve c, and all the way to back to where it started in as you keep on annealing it without

hydrogen. Practically, the forming gas anneal (FGA) must be the last high-temperature step of your system because after that, if you introduce any high-temperature step, the hydrogen that you have incorporated at the interface would desorb and you make the interface worse. Remember that forming gas anneal does not work at high temperatures (> 500^{o}C). You do not need to do it for longer than 10-15 minutes. After FGA, do not take your wafer again to high temperatures or you will lose all the passivation effect.

(Refer Slide Time: 14:57)



Let us talk a little bit about this mobile ionic charge and how to remove that. A part of the solution is to maintain cleanliness. We have already discussed surface cleaning techniques: using COMS grade chemicals, using cleanrooms, never touching the wafers with hands etcetera. A trace amount of the ions can be removed by the introduction of chlorine during the oxidation process. That is called halogenic oxidation. So, this can be achieved by using either chlorine gas or HCl.

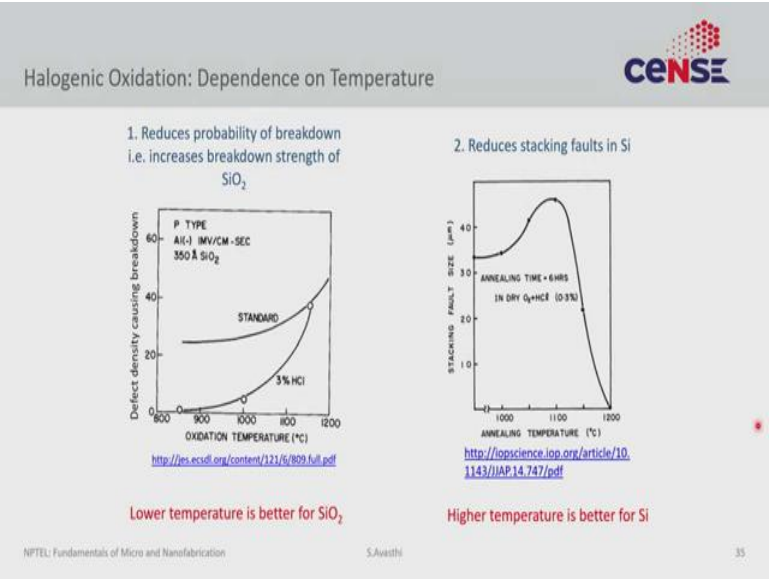
Both of them are a little dangerous, making it challenging from a safety perspective. Both of these gases are very corrosive. Even if you have a very high-quality stainless steel tubing or housing, small amounts of leaks would cause a lot of corrosion in those steel lines. As this is a high-temperature furnace, it is not practical to have everything made of plastic and you need metal at some point. However, the industry has processes to control for that and hence, it is a popular method to create very high-quality gate

oxides. If you do not like these two gases for safety or corrosion reasons, you can also use trichloroethane or trichloroethylene which are liquid sources of chlorine.

The mechanism of the halogenic oxidation is simple; sodium and potassium both form very stable (and neutral) salts with chlorine that have a lower diffusivity in silicon dioxide than sodium ion alone. As an added effect, it also captures the silicon interstitials existing at the silicon-silicon dioxide interface. This reduces the oxide trap charges and enhances your breakdown voltage. A higher breakdown voltage allows you to put more voltage on top of your gate, increasing the (field-effect) mobility.

What we have not touched upon till now is that during the oxidation process, you cause some stacking faults in silicon. Here is an example of a silicon wafer stripped of its thermal oxide. At the interface, now the silicon surface, you can see these straight lines which are stacking faults that form on the surface during the oxide growth. This is acute when you are doing the oxidation at high temperatures. It turns out that this halogenic oxidation reduces the stacking faults as it captures the silicon interstitial. If stacking faults are a problem, halogenic oxidation might be the answer.

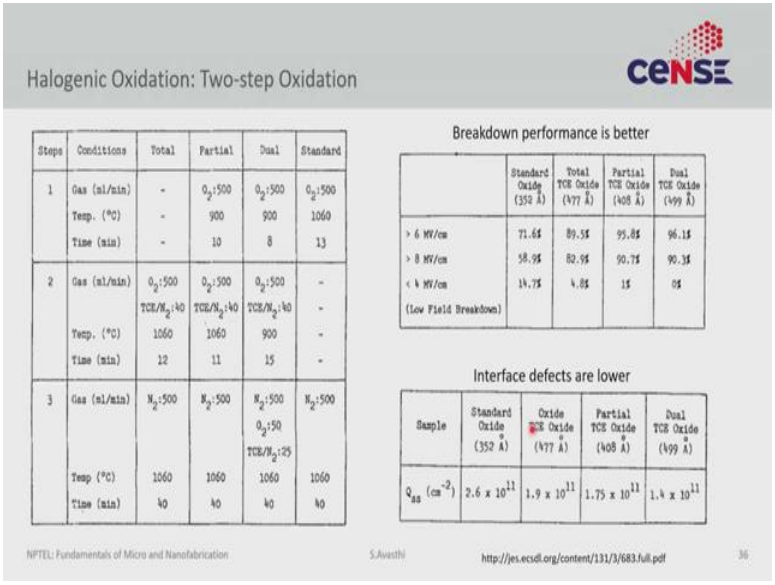
(Refer Slide Time: 18:06)



The temperature at which you do the halogenic oxidation depends on whether you want high-quality silicon dioxide or whether you want high-quality silicon. There is a trade-off. In general, the higher the temperature, the defects that cause a breakdown in silicon dioxide increase, making the breakdown voltage worse (lower). If you want a higher

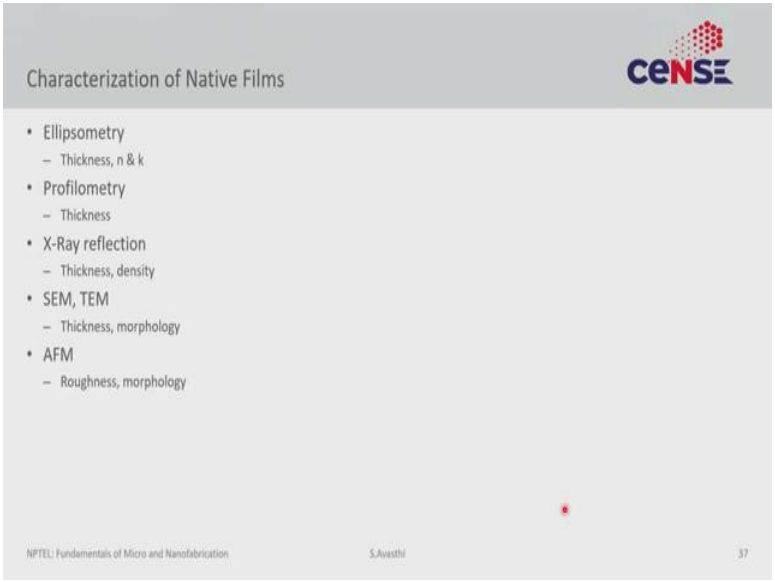
quality silicon dioxide, you may want to do it at a slightly lower temperature; however, the stacking faults are reduced in silicon only at high temperatures. So, if you care about the quality of silicon more, you should oxidize it at a higher temperature.

(Refer Slide Time: 19:02)



The industry often does two-step oxidation where there is both a low temperature and high-temperature process to achieve both the goals or at least reach a better compromise. This discussion is too advanced for this course. I will just leave this slide here for you to look at and appreciate why depending upon whether you want a higher breakdown voltage or lower interface defects, you can select one recipe over the other.

(Refer Slide Time: 19:34)



With that, we end the discussion on the growth of native films and look at one slide to discuss characterization. This is primarily a course on fabrication, but we will still look at some ideas on how to characterize the film that you grow. The easiest thing to do is to just measure the thickness, often using ellipsometry. From Ellipsometry, you also get some information about the refractive index that often correlates with density. Regarding the quality of the oxide, you can get some information by just looking at the refractive index and making sure it is closer to the bulk value.

Thickness can also be measured using simple profilometry with a simple stylus that goes over a step and measures the height. For very small thicknesses, people have also used X-ray reflection which is very accurate at giving you surface roughness and density measurements. TEM and SEM for looking at the cross-section, the morphology, the stacking faults, dislocations and the interface in great detail. AFM is used for roughness and morphology.

In the next lecture, we will start tackling the more complicated subject of chemical vapor deposition.