JUDY HOYT: OK, so there are two handouts for today that are in the back there. There is the lecture notes and handout 8, and also, problem set number 2, or homework number 2 is going out today. Your homework 1s are graded, and I'll have them back next time.

So this handout is a fairly long discussion. There's two major areas I want to cover in today's lecture. I want to finish up chapter 4. There was a section in chapter 4 we didn't get to talk about yet, which is about characterizing. We talked all about impurities and how they can be issues and problems. I'm going to talk about how we characterize those impurities.

And then we're going to skip chapter 5. Remember, chapter 5 is on lithography. And we have a whole one-semester course here at MIT on lithography, Hank Smith's class. I'm going to skip that. You're welcome to read it, of course. But if you don't, you should go ahead and start reading chapter 6 on thermal oxidation. I'm going to start that.

In fact, the next three lectures are going to be on thermal oxidation. And I've listed here-- I won't go through them all. You can read them later-- some of the major topics from chapter 6 that we're going to discuss. So let's go on to talk about-- finish up the characterization on slide number 2.

This slide is supposed to give you an idea of the different surface analysis techniques and their detection sensitivity, how low of a level of an impurity can they detect? And I apologize right off the bat because not all of these acronyms are defined. And it's hard when you haven't seen these before to see all these ridiculous looking acronyms, Auger, [? OAES. ?] We're going to talk about Auger in this lecture. Raman, FTIR, we'll talk about a few of these.

These are techniques you can see just based on where they lie on this vertical bar. The vertical bar is atoms per cubic centimeter. So here, 100% would be 5 [INAUDIBLE] 22. That's 100% of silicon lattice atomic density. If you want to measure an impurity on the level of 1 to 0.1 atomic percent, that level, you would use Auger, or [? ESCA, or ?] one of these types of techniques.

Most of the time that's not going to be nearly adequate. It depends on what you're making. But we talked about needing to know impurity levels much, much lower than that in the 10 to the 10th or so atoms per square. This is if you integrate over some depth, or 10 to the 15th per cubic centimeter on the left here.

So the types of techniques people use for that are listed here in this lower shaded bar. X-ray fluorescence, which we'll talk about and surface SIMS, these are the two most common techniques for measuring impurities. And you can see they measure them down to the something like tens of parts per billion range.

So those are some of the options available to us. Next slide, on slide 3, gives you a rough estimate of the depth of analysis, the depth from the surface of some of these techniques. And I've circled the ones that will be of interest for this lecture.

So X-ray fluorescence, which is the acronym is TXRF, it's samples typically the near-surface region depending on how it's done. But it's in the near surface region, say, the top 100 angstroms or so, roughly. Maybe it's a little less than that. There's a special technique called surface SIMS, secondary ion mass spectrometry. We'll talk about that. That's designed to, again, sample just the top 100 angstroms, just the surface region.

And then there's a standard SIMS or depth-profiling SIMS. You can profile pretty much as deep as you need to. Typically, you profile here in the 1,000 to 10,000 angstroms or 0.1 micron to 1 micron. And this is a depth profiling type of SIMS. Whereas, this measures surface SIMS, as the name suggests, measures just what's in the very near-surface region. So those are some those three are three important techniques.

Slide 4 is meant to be a generic view-- and I took this out of your text-- of how surface contaminants are measured in silicon wafers. A lot of the time, what you're doing is you're taking a silicon wafer, which is shown schematically on the bottom, with some surface layer, however deep it might be, with some contamination level in it.

And we interact that surface with some kind of beam. It could be an incident electron beam, as shown on the left. And we look at what comes off. It could be an incident X-ray beam or an incident ion beam, as in Rutherford backscattering or SIMS.

So a lot of these-- these are beam techniques, where we interact. We shoot a beam into the surface. It has some energetic interaction. Something comes off. Could be an X-ray, could be an electron, could be an ion. And we analyze what comes off. That's the basic idea, very schematically.

Let's go on to slide 5, and slide 5 is about what happens when we bombard the surface of silicon with electrons. And what it shows very schematically is a plot on the vertical axis of the number of electrons coming off as a function of their energy. And you can see various humps here.

At very low energies, there are a large number of electrons coming off. These are called secondary electrons. Maybe they're in the 5 eV range. These are the electrons that are coming off that are forming images if you look at a scanning electron micrograph. I've shown a couple of scanning SEM micrographs. In fact, last time we had that gettering picture with all those little dots in it. And you can see throughout the wafer, see all those dots which were the oxygen precipitates. That was a scanning electron micrograph. So that was looking at these secondary electrons.

At the very highest energies are backscattered electrons. Basically, the incoming ones just bounce off the atoms and get reflected back with some energy similar to what you shot the electron in. They don't tell you much-- neither the lowest energies or the highest ones tell you much about what you bounced off of.

But there is a region of intermediate energies here called the Auger electrons, which is shown in here-- there aren't as many of them-- that give very specific information about the species that they hit or that they interacted with because they interact with the atoms in the substrate with the core electrons of those atoms that are more tightly bound to the nucleus. So here, at these intermediate energies is where we can get some information about what it was that the electron interacted with, what element.

So we go to slide 6. Again, this is a very schematic illustration of the energy levels involved in this type of process. So let's just go through the schematic. What's shown up here in the upper center dot is a primary electron, which is this black dot. It's coming in, and it's interacting-- in energy space, it's interacting with an atom, let's say, down here that has a certain energy level.

And what happens is basically it knocks-- these are the energy levels, this Ek, El1, El2, these are meant to represent the core energy levels of the electrons that are bound closer to the nucleus in the core. We've been talking all about valence electrons up here in the valence band and electrons in the conduction band. It doesn't interact with those, but it goes into the atomic core and interacts with the electrons that are there.

OK, and basically, it kicks out one of these core level electrons from this Ek energy, where this becomes a secondary electron. Now you have an open energy level. And one of the higher level electrons, this one at El, can drop down into that level. And the spacing of this level between the L1 and the K level is very much a characteristic spacing of the atom. Depends on the element, OK?

So you can imagine that this energy here, this transition energy is very specific to the atom. That transition energy is then, in this particular case, given off as kinetic energy to an ejected electron, which is called an Auger electron. That ejected electron then has an energy that's characteristic of this level spacing.

So by measuring its the energy spectrum of those outgoing Auger electrons, it tells me something about the atoms that were in the substrate, what those spacings were. And as long as you know you have a fingerprint for what those spacings should be, you can tell what atom the electron interacted with. So that's the Auger electron spectroscopy.

There's another way to do it where, instead of releasing an electron, an Auger electron, you may actually release an X-ray. This energy spacing between El1 and Ek can be given off to eject an X-ray. So you may also look for heavier elements at the X-rays coming off. And that will give you some species-specific information.

So you're get the qualitative idea here. I'm putting in some energetic particle. I watch transitions that happen in the core levels. And then I look at what comes off, either another electron or an X-ray. So that was an incident in that last slide, slide 6. The incident particle was an electron.

You don't have to use electrons, as it turns out. You just need something that can bring some energy into the atomic core. So here on slide 7, you can imagine the incident energy might be provided by an X-ray. Doesn't have to be an electron.

So you may bombard the surface with X-rays instead. And there are two different categories here, XPS, X-ray photoelectron spectroscopy, where you bring an X-ray in that provides the energy and an electron comes out, or X-ray fluorescence, where you bring an X-ray in, and another X-ray is emitted with a characteristic wavelength.

And what's happening here in this XRF, this X-ray fluorescence, is pictured schematically here on the left for an element that's relatively light, has fewer shells in its electrons, in its core levels, and one that's relatively heavy on the right, which has lots of core levels.

So you can imagine an incident X-ray coming in. It's knocking an electron out. And you get a transition energy between two core levels. And what comes off is an electron with-- an X-ray with some energy N nu with a characteristic wavelength.

The problem with doing this with a light element is there are too many electron shells. So this process is not that favored, and it becomes difficult to do. For heavier elements, it's much more practical, heavier elements being the transition metals or things that are heavy, like silicon. There are lots of these core levels. You bring an X-ray in. It has a high probability of interacting, producing an X-ray coming out that has a characteristic spectrum. So you see, if you plot the X-ray density or intensity coming out versus energy, you see these peaks. And this spectrum becomes a fingerprint of what elements the X-ray has interacted with.

So if we go on to slide 8, this is a very common technique that's used in semiconductor analysis. I'm showing here a typical, what's called, total X-ray fluorescence, TXRF, the spectrum. And the plot shows on the y-axis the intensity of the X-rays coming off in counts per second, detected as a function of the X-ray energy, so that H nu or whatever you want to call it, that's coming off. And again, these energies are going to be characteristic of the core levels of the atom.

So next to each peak, the person who did the spectrum actually is familiar enough that they were able to identify the peaks. For instance, this peak right here, little above 2 keV, corresponds to chlorine. And you can integrate the area under that peak and get a rough estimate of the amount of chlorine atoms per square centimeter that exists on that surface.

Silicon, of course, is going to be large because it's a silicon surface. There's a lot of chromium on this sample-- look at the peak labeled chromium-- and a fair amount of iron. So the nice thing, it gives you a broad spectrum. You can just shoot the X-ray onto the surface in about a 1-centimeter area. And you can immediately survey in the transition metals what's on that surface. It's very useful and commonly-used technique.

So let's go on to slide 9. So this TXRF, what are some of the good things about it? Well, I just showed you, it's a survey technique to get a broad idea of what different elements are on the wafer. And it can be automated. It's reasonably quantitative. We can integrate those peaks and compare them to standards.

It doesn't destroy the wafer. You notice the X-ray just comes in, interacts, and goes out. It's not like I have to sputter off, or I don't even have to break the wafer. You can do it on an entire wafer. They have machines for 12-inch wafers you can put in there. And they'll tell you what impurities are on that wafer surface. So that's really nice.

A weakness, it doesn't detect low-z elements because the physics of-- there aren't enough core level shells. So aluminum, potassium, and sodium, both of which are very nasty, and silicon can't be detected. So that's a major limitation.

It doesn't tell you about how it's distributed in depth. It doesn't tell you anything about the depth distribution. And it's pretty big spot. About 10 millimeters is the smallest you can focus the beam down. So you're looking at large spots. That's OK. But if you want to look, let's say, inside an individual device and see what impurities were in a bad device, if you had a fail device, TXRF is not the way to go because the device isn't much bigger than-- it's much smaller than 10 millimeters. That's the whole die on that wafer being passed around.

And the surface has to be specular and polished. If the surface is patterned-- like on this wafer here, you see some patterns, it's not perfectly specular-- you're going to have a problem because you're not going to get the kind of reflectance. So it's typically done in an area that's relatively well polished.

OK, there is an alternative technique, which actually is quite complementary. Where TXRF has weaknesses, surface SIMS has some strengths. So people often use both of them. This is an example of an ion beam technique. So instead of bringing in X-rays, I'm bringing in ions. And the ions that come in are called the primary ions or the primary particles. The two typical ones to use are either cesium ions or oxygen ions.

So you have a primary beam of particles. Energy varies. This says 10 keV. In some cases, you can bring this energy all the way down to 1 keV. And there's an advantage in bringing the primary energy down lower. What you see is this primary beam comes in, and what does it do? Well, it interacts with the solid, the surface. And of course, the atoms, the ions penetrate in, and they get implanted some depth. They penetrate in some depth. That depth that they penetrate depends on their energy. So if I'm at 10 keV, I may go down to 100 angstroms.

Now, the ions that come off, the atoms that are sputtered off, some of them are ionized. And that's what we analyze in the machine. And we measure the atoms that come off. They only come off the top 10 angstroms or so because that's the escape depth of which they have enough energy. These other particles that come off can get out.

So that's good. We can sample the top 10 angstroms. But notice, I disturbed-- in doing the measurement, like Heisenberg, in doing the measurement, I disturbed the crystal. I shot a bunch of ions in here, and now they're down here at 100 angstroms deep. And they knocked things around in the substrate. So whatever atoms were down there, they disturbed them. They moved them around. They might have knocked them in deeper.

So if I can lower this primary particle energy down to one keV, they won't be knocked in so deep. And they won't disturb as much of the crystal. So you get better depth resolution in general as you bring the primary ion beam energy down.

Basic idea is you have a primary ion beam. It comes in. It sputters off a certain number of monolayers in a certain period. And you measure the ions that come off the substrate. And you count how many ions are of silicon, how many coming off are copper, how many are gold, whatever. And you get an estimate. You can get a spectrum of what's coming off. You can't count too many because it has to go through a spectrometer. And the spectrometer can only count so many at one time. But that's the basic idea of secondary ion mass spectrometry.

If we go on to slide 11, I actually took this off the website. One of the commercial places that does a lot of this analysis commercially is called Charles Evans and Associates. And if you go on to their website-- you just search on Google. You can get to their website. And they have a lot of information on these techniques.

I won't go into slide 11 in any great detail. I'll let you read it later. I just put it up there to show that companies have done a lot of work in order to quantitatively be able to measure in the top, say, 10 to 20 nanometers or 100 to 200 angstroms, quantitatively measure the contaminants in that surface region. It wasn't always possible to do that. But they have nice techniques to do that now. And if you're interested, you can do more detailed study on the web.

Let's go on to slide 12, which lists some of the key features of surface SIMS. Well, it's been approved by ASTM for measuring some of the elements and exactly some of the elements that the TXRF can't measure. So that's nice. It can quantitatively measure contamination for sodium, aluminum, potassium, iron, on into [INAUDIBLE] silicon. And it can measure lots of the elements. But it's actually been approved as being a standardized technique.

It's quite accurate. It can detect many elements and isotopes. But detection limits is quite low. Look at this. Look at these numbers, 10 to the eighth to 10 to the ninth atoms per square centimeter for most metals. Now, why do we care? Well, what was on the ITRS? When you read your ITRS in your homework, the number of atoms per square centimeter that people were interested in was what order of magnitude? 10 on the 10th, exactly. So you better be able to measure below that. Otherwise, you can't even tell what you have. So it's right in the range where we need to be measuring for the ITRS.

It also can give you information about the profile in depth. And that's kind of nice. If you go back, if we go back a couple slides to slide 10, what happens is you collect these ions as they come off. And you measure them as a function of time. In the first couple of seconds, you sputter off the first few monolayers, the second few seconds you sputter deeper.

So you're constantly sputtering the surface of at a constant rate, and you're measuring what comes off as a function of time. So you know by collecting this as a function of time pretty much exactly what depth those ions came from. So if you want to sputter a little deeper, you can get some information about the profile.

And finally, going back to slide 12, it has a relatively small detection area, say, 50 by 50 microns. Still not that small compared to the size of a device. But it's better than 10 millimeters. So you can get in-- you can put special structures on the mask with this kind of size area that are dedicated for device characterization if you're doing research and you want to be able to measure on the chip, measure the characterization, the contamination.

Let's go on to slide 13. What I'm showing here, again, I took this off of the Charles Evans and Associates website. And it's a little bit dated. It's a few years old, but it gives you a rough idea of comparing total X-ray fluorescence, TXRF, in these columns on the left, to surface SIMS.

In terms of its detection limit, so the detection limit is the lowest concentration, which either technique can measure. And so you can get a rough idea. Let's say just for comparison, let's look at potassium here. TXRF doesn't measure it very well. The lowest it can measure here is about 20 times 10 to the 10th, or 210 to the 11th, not all that sensitive. At least a few years ago, that was the limit. Maybe they've improved it.

But look at the detection limit using surface SIMS. It's 0.01 times 10 to the 10th or 10 to the eighth per square, much more sensitive. So it gives you an idea of what the detection limit is. So if I know that I'm concerned about potassium, I would shoot for SIMS, not TSRF. So it's a nice quantitative comparison for the different elements.

So I went through that relatively quickly, I admit. But you can read in your textbook on the characterization techniques. In chapter 4, they're covered. And I would also encourage you, if you're doing research in this area and you need to know more, go to the Charles Evans website. They are a commercial vendor. They do sell characterizations for a business. But they're the world's experts as a result on that. So it's a nice place to get information.

So I want to move on now starting with slide 14 and talk about oxides and the IC industry. And I think we hinted at this in the first lecture or two. But really, it's silicon-silicon-dioxide interface. And it's perfect properties. That is the number one reason why silicon dominates integrated circuits today compared to germanium, compared to gallium arsenide or any of the other semiconductors.

And these are some properties here listed in this bullet list of silicon dioxide that are quite desirable. It's easy to selectively etch silicon dioxide and to pattern it. It also masks the diffusion of a lot of common impurities, which is a really great thing. You can put down SiO_2 and pattern it, and some impurities will not diffuse through it.

It has great electrical properties. It's a very good insulator, at least when it's thick enough. It has a high breakdown field. It tends to passivate junctions very well. It has a very stable and reproducible interface. For any of you who have ever tried to make oxides of this quality on other semiconductors, like germanium or gallium arsenide, you immediately appreciate the properties of silicon dioxide because on those other semiconductors it's 10 PhD thesis to try to get a decent passivation of those semiconductors. So silicon is really-- we were very, very lucky coming upon silicon-silicon-dioxide interface. It's really unique.

On slide 15, this is a little bit dated, but this just shows some of the uses of silicon dioxide in IC technology. And there are two columns here. On the left, I'm showing thermally grown oxides, and we'll talk about that process in the next few lectures.

On the right are deposited oxide layers. These are put on the wafer, rather than growing and consuming the wafer, by a process called chemical vapor deposition. And we'll talk about these a little bit later in the class, in the course later on. And these deposited oxides are different. They're usually not used for layers below about 10 nanometers. It's harder to control. And the properties of the silicon silicon dioxide interface for a deposited as opposed to a grown oxide are not nearly as good.

But you still need them. You need them in the back end between metal layers, and you need them to do masking. So they're very important. But we're going to focus now on thermal oxides. And this just shows thickness and the type of oxides that are used. The thickest oxides are used in the field. The field surrounds the active device region. So it isolates individual devices.

Thinner oxides, like 100 angstroms, are used for gates or pad oxides. In fact, I put a little arrow here and showed it going down. In fact, gate oxides today in the highest performance devices are quite a lot thinner than when this textbook was written. So they're actually more like 20 angstroms. So they're actually down in this range of tunneling oxides. And then finally, the very thinnest oxides are those chemical oxides that result from RCA clean, as we talked about last time, that are grown chemically rather than in a hot furnace.

Let's go on to slide 16. This is just-- the quality didn't come out all that well. But this is a high-resolution cross-section transmission electron micrograph. And what we're looking at here on the top is polysilicon gate. This very thin layer right here, labeled gate oxide, is 1.2 nanometers thick, or 12 angstroms. That's the gate oxide that's been thermally grown. And at the bottom is the silicon.

And in fact, it's the formation of these gate insulators that really is probably the most critical application of the process of silicon thermal oxidation in ICs today. We'll talk about thermal oxidation to grow thicker oxides, field oxide, et cetera, which is, of course, of interest. But in terms of the most critical application of thermal oxidation, it's really to try to form the gate insulator.

However, remember we said that silicon dioxide is a perfect insulator if we go on to slide 17. There's some basic physics, though, that we can't get around, which has nothing to do with silicon dioxide. It's just the fact that there's something called quantum mechanical tunneling. If you have two regions, one on the left and one on the right, in the middle there's an insulator.

There's a barrier, an energy barrier. If you've taken quantum mechanics, you know that when that energy barrier becomes thin enough, depending on its height, but when it becomes thin enough, there's a finite probability that electrons will find their way through that barrier. They will tunnel through. They don't go over. They go through. And this isn't a problem when you're thicker.

But when you get down to this regime, for example, below about 30 angstroms, you can have direct tunneling right through from the gate electrode into the channel. And this constitutes a gate current, which, again, the whole idea of CMOS is that there's no current through the gate insulator. That's not true anymore. There is current flowing through these gate insulators because they're so thin.

And this is just a diagram I took from an older paper now. But on the vertical axis is a log scale, shows the gate current density. So that's how much current is going through that insulator as a function of gate voltage. As I increase the gate voltage, of course, that goes up. And the different parameters shown here are the thickness of the SiO2 layer.

So here in the bottom, it's showing for 36 angstroms, you see. And you can see it goes up exponentially. Below about 30 angstroms, we're going up exponentially. Here, I'm only going from 29 to 25 to 20 angstroms. And we're going up five orders of magnitude in just five angstroms in the amount of current that flows through that oxide.

So it's an exponential process. It's very, very highly dependent on [INAUDIBLE]. As we go thinner, as we go here to, say, to 15 angstroms, really getting quite a bit of current tunneling through that. And that's a major issue in modern technology.

So as a result, if we go on to slide 18, what are people doing? Well, you have a couple of options. One is don't scale the oxide. Don't make it so thin. Stop scaling it. Well, it turns out that's an issue because we need to get higher current drives. And one way we do that is to make the oxide thinner. So if we want higher performance, we want to make the oxide thinner. On the other hand, the gate is no longer a perfect insulator, the gate insulator.

So there's another idea. To achieve higher current drive, well, we need to increase the capacitance. Usually, we just decrease t-ox in this formula shown on the slide. Well, the other thing you can do is increase the dielectric constant, k-ox of the insulator. So that's the name of the game people have been looking at and doing research for the last five or 10 years. What other insulators can we make besides SiO2 that have all the great properties, but they have a somewhat higher dielectric constant, k nu, which is shown in this formula here.

So I try to increase the dielectric constant by-- one way is by adding a little nitrogen to the insulator. And people use oxide nitrides, or you might go to a whole new material. So what people refer to is for the same gate capacitance, what we typically do is we define an equivalent gate oxide thickness called t-ox equivalent according to the formula on the bottom.

So what is t-ox equivalent? t-ox equivalent is equal to just k, the dielectric constant of silicon dioxide, which is at 3.9, divided by the dielectric constant of my new gate insulator, my high k. So let's say you went from 4 here to 8 times the thickness. So what that means is, basically, you can actually increase the thickness, go to a slightly higher thickness of the high k and get the same capacitance.

So I'm upping k here, k nu, so I can increase thickness, which means your quantum mechanical tunneling is going to go down. You'll still get the high capacitance and the high gate drive. So it sounds like a perfect, from a mathematical point of view, physics will look at this, say, perfect. Increase k, increase t, everybody's happy.

The problem is it's hard to come up with a material that has a higher dielectric constant that has all the perfect and ideal properties of the silicon silicon dioxide interface. In fact, such a material has not been discovered yet despite millions and millions of dollars that's been spent in trying to comb the periodic chart for such a material.

So if we go on to slide 19, let's look at what some of the future projections are for the scaling of the gate insulator. And I should say gate insulator in this slide because we're not sure-- it won't be SiO2. It'll be an oxynitride with a higher k or maybe a new material that is yet to be figured out.

And what I'm showing here, each column corresponds to a different year, just like we usually do on the ITRS. And what I wanted to focus on is a couple of interesting rows. This row right here, if I highlight this row, that says equivalent physical oxide thickness for microprocessor units. So this would be for high performance. So this is t-ox equivalent.

In 2004, for high-performance devices, it's supposed to be about 1.2 nanometers, or 12 angstroms. So the high-performance technology is supposed to be about 90 nanometer technology, 12 angstroms. That's how thin we are.

And look at the current. The gate dielectric leakage, the maximum current tolerable in such a device is about 170 nanoamps per micron. So it's pretty high current. So that's for the highest performance device. And if we go out in that same row and we work to higher years, look what happens.

By 2006, we end up in yellow. Yellow means people have an idea how to do it but maybe not totally manufacturable. But there's some ideas. And if we go beyond that, 2007, it's all red. So we want to achieve equivalent oxide thicknesses of 0.9, 0.8, and 0.8 by 2008. People just don't know-- don't really have the right material at this point. So there's a lot of research in this area.

So what people do instead, we haven't figured that out yet, is they make different types to reduce the total amount of power that's burned on the chip. They make different types of devices. You have a very high-performance device that's quite fast. But it has reasonably thin physical gate oxide thickness. But then you can also, on the same chip, going down here in this region that I've circled down here where it says equivalent physical oxide thickness for low operating power.

So this is a device burns less power. Notice, the oxide is thicker. It's 1.5 nanometers instead of 1.2. And the maximum tolerable gate leakage is on the order of 1, 1 nanoamp per micron instead of 170. So this device is supposed to be lower operating power. You make it a little thicker, and you change the VT to achieve that.

Finally, if you go down to this column for the low standby power device, it has even thicker oxide. So on the same chip you might have a 2 nanometer, or 2.1-- 21 angstrom oxide. And the leakage current much lower through the gate for this device, about three picoamps per micron. So three orders of magnitude lower than for the other device and five orders of magnitude lower than for this. So quite a bit different on the same chip. This is the solution people are coming up with.

Only the fastest devices have to have the very thinnest oxides. Nevertheless, in a couple of years, we will be hitting the red brick wall. So there's a lot of work to be done on alternative gate insulators. Let's go to the next slide. I've also taken this from the 2003 ITRS. It's a little bit complicated and not the most pretty plot in the world. But what it does show is a couple of interesting things.

It's a plot of the leakage current density, JG, through the gate. So this is the current going through the gate oxide as a function of year on the left axis. And you can see that this red line represents the JG limit. This is what device designers would like. No higher than this amount of gate current is what they can tolerate for a high-performance device.

This black curve is actually what people think, a dielectric called oxynitride. So this is not SiO2, but it's SiO2, SiO, where a certain amount of nitrogen has been added. And you can see the gate current on the left that is being projected. It looks something like this black curve. It's increasing exponentially.

And somewhere around 2006, 2007, these curves cross. So beyond that oxynitride, you cannot stay on the same scaling curve. It won't satisfy the gate leakage current. It will leak too much through the gate. So beyond 2007, it's not clear if people can use oxynitride. So they have to come up with something else.

This brown curve up here that's marked EOT, the equivalent oxide thickness, it's referenced to the right-hand vertical axis, which shows in angstroms the equivalent oxide thickness. So this is just exactly the data I showed you going down. So starting in 2003, we have an EOT of 13 angstroms on the right.

And it's going down and down so that by 2007 or so here, the EOT is nominally 9 angstroms. So this is the way people would like to scale. But clearly, in the next few years, oxynitride is not going to work. You're going to need some other material, like hafnium oxide or one of the silicates that people are studying.

And so one more-- go to the next slide, on slide 21, one more thing I just wanted to point out about alternative gate dielectrics is-- this I took from *IEDM, International Electron Devices Meeting* publication from about two years ago. There's been a lot of data. I don't mean that people have searched the entire periodic chart, but there's a lot of work that's been done by different people to study different materials. And this is just a way of summarizing them.

One parameter people look at is the current density that leaks through the gate as a function of the EOT, equivalent oxide thickness. And I won't go through it in detail. But just the point is that a number of different materials are under consideration. I think right now the most popular are hafnium oxides and hafnium silicates, HfSiON.

But this is an example of a topic, if you're interested in doing some of the library research for your final report, of a topic you might want to consider looking at is high k dielectrics. What some of their properties are and how they're formed is an interesting thing to study.

So given that background information, I want to go on to page 22, slide 22 here. And even though I've just finished telling you that in the future pure SiO2 probably won't be used, it's still the most important-- there's nothing that's ever beaten it, so it's the standard. So we're going to study the process of thermal oxidation in this course.

And this shows schematically, very schematically the basic process. We have an oxide that's growing at a high temperature on a silicon wafer. And what happens is an oxidant, such as oxygen or water, diffuses through that oxide, has to react at the surface. And new oxide is formed right at this interface right here.

So we have a chemical reaction taking place at the interface. These are examples of the chemical reactions, silicon plus oxygen going to SiO2, or silicon plus moisture going to SiO2. These are the reactions that take place at that interface. And we'll spend, not this lecture, but probably next lecture talking in detail about mathematical models for that process.

But rather, before we get into all the mathematical detail models and the Deal-Grove model, there are a few basic properties about the process we should think about, that, in fact, when the silicon surface is oxidized, there is a volume expansion associated with that. And this upper schematic picture shown here in the upper right is something I took from chapter 6, figure 6-39.

What it represents is a cross-section in depth of the crystal, where these open circles here are meant to represent the silicon atoms in a lattice. This interface right here is meant to represent the silicon silicon dioxide interface. And this upper region here that has both black atoms, which are oxygen, and the white open circles, which are silicon in it, that's the SiO2. So here's the interface.

So what it shows schematically is that, in order to oxidize this surface, what has to happen is the silicon bonds have to be broken because the silicon is bonded to each other. They have to be broken. Oxygen, these little black atoms, have to be inserted in between the silicon to form SiO bonds, such as shown here. And so the volume, that whole area has to expand to a certain extent because of the room taken up by the oxygen atoms. So there is a volume expansion associated with it. And the lower cartoons show schematically what that expansion looks like.

So let's say I take a unit volume of a substrate that's a cube, that's 1 by 1 by 1 cube in all three dimensions, shown schematically here on the left. And just take that square that cube and expand it 30% in all three directions. So this is a bigger cube. This blue cube in the center is bigger by 30%. So this would be if it were unconstrained. There were no constraints, just in free space.

But actually, the substrate restricts the expansion to one dimension, right? I can't just take the cube and expand it in the xy plane because I've got the substrate holding on to it. So in fact, the expansion only takes place in the vertical direction.

So for every one unit consumed-- silicon consumed in the vertical direction, we have basically a height here of 2.2 units when it actually grows the SiO2. So the substrate restricts this to one dimension.

So what does it look like? What does an oxide look like grown-- especially an oxide grown in a complicated two-dimensional fashion? We're going to spend a lecture talking about this. But this is a pretty picture. It's a scanning electron micrograph of something called LOCOS. I think the first lecture we talk about local oxidation of silicon, where we would put a nitride mask down on the chip where we don't want the oxide to grow. And everywhere outside the nitride mask is where the field oxidation takes place.

So this is an actual example of field oxidation that's taking place and what it looks like. This rectangular bar marked the location of the silicon nitride mask, that's the region on the chip where the nitride was there protecting the surface against oxidation. The gray region down here is the silicon substrate.

This region marked SiO2, you can see the volume expansion. It's expanded compared to-- the original silicon surface is shown by this white dashed line because we know oxidation involves the volume expansion by about 2.2x. So this occupies about twice, 2.2 times, the volume of the silicon from here to here that was actually consumed. So you're consuming the substrate when you're doing this.

Interestingly, though, in two dimensions and these three-dimensional structures, stress plays an important dominant role. We're going to talk about that in this course. So look at this little bird's beak. They call this a bird's beak that's squeezing out like this because it looks like a bird's beak. And sometimes they call this the bird's head. You have to have a good imagination in order to understand that.

But the bird's beak shape is actually due to lateral oxidation, some of the oxidant diffusing in here under the nitride mask. And also, its shape is controlled by the nitride mask pushing down on it. So the oxide is trying to grow it. Has to push up on the nitride. That takes some force. So this whole shape and the amount of lateral encroachment is very much dependent on the stress effects. And we'll talk about that in the next couple of lectures.

So complex shapes can be formed by local oxidation. We'll go on to slide 25. This is a very high-resolution transmission electron micrograph. And again, on your handouts, it probably doesn't look very good. But all those little dots, you see them in rows, correspond to pairs of atoms in the silicon lattice. So these are planes here. And they're all very regular because silicon is a crystal material, right, single crystal, perfect structure.

SiO2, however, is amorphous. It has no long-range order. It has some short-range order but no long-range order even though it's growing on a single crystal. So this is an amorphous material even though it's growing on a single crystal substrate. It's a glass.

If we go on to slide 26, there are no crystalline forms of SiO2 that match the lattice size of silicon. There is a little bit of short-range order in thermally-grown SiO2. And in fact, if you look at this schematic on the left, this shows that there's a tetrahedral type of structure on the left. You see a silicon atom in the center and four oxygen atoms bonded to it nearby forming some sort of tetrahedron.

These tetrahedra, though, are not actually bound in any real structure. They may form a ring structure a little bit by sharing oxygen atoms. But you notice, this lattice does not repeat itself throughout space to form some single crystal. So there is short-range order in the tetrahedra. But it is an open network, and it doesn't have long-range order.

So let's go on to slide 27 and just talk for a few minutes about stress and oxide. We're going to spend a lot more time talking about it in the next lecture but just to make the point that when oxide layers grow on silicon, they're under compressive stress. And it's for a couple of different reasons.

But basically, because the growing oxide is constrained by the interface that it's growing on it's constrained by the substrate this interface right down here. So it can only expand upward And this gives you an idea, 5 10 the ninth dynes per square centimeter for the amount of compressive stress that this oxide is feeling when it's grown on the wafer.

Now, at high enough temperatures, say, if you put it in the furnace to oxidize above 1,000 or anneal it, the oxide can relieve some of the stress. The oxide can actually flow a little bit by a viscous flow. But that's only at a high enough temperature where the glass starts to soften a little bit. At lower temperatures, it's just too viscous in order for it to flow. So the stress stays built in.

So there's these intrinsic stresses just because of the volume expansion. There's also differences in the thermal expansion coefficients of SiO2 and silicon. And that leads to a certain amount of intrinsic stress. Both of these effects are going to put the silicon in tension. So the silicon wafer at the surface is sort of being pulled by the oxide, which is being compressed.

So if you were to grow an oxide on both sides of a wafer and strip it off the back, the wafer would actually be slightly bowed, very so slightly, nothing that you could see by eye. But you can actually measure it in a special laser apparatus called a wafer curvature measurement. So there are ways of actually detecting the amount of stress just by measuring the curvature of the wafer when you remove the oxide from one side and leave it on the other side. That's how people determine some of the stress.

Let's go on to slide 28. This is a very schematic, and I apologize for this cartoon. If you've ever actually been in a fab clean room and you know what an oxidation furnace looks like, it doesn't look exactly like this. But it's an artist's conception. The equipment is actually relatively simple in concept. It's essentially a quartz tube, which is shown here by this black line. So it's a cylinder. It's a long quartz tube into which one can push a boat of a quartz carrier with wafers standing up in it.

So all these wafers are put vertically in the boat in little slots, and you have a whole bunch of them sitting right next to each other, within a few millimeter of each other. You have the whole stack of wafers. So you can oxidize 25 or 50 or 100 wafers in a furnace at a time.

You essentially just flow in oxidants, like oxygen and hydrogen, to create water vapor and moisture. You heat the whole thing by resistive [? heater. ?] So the quartz gets very hot, and the wafers get very hot. And then the oxidant goes out the back side. So it's relatively simple.

In practice, though, there are some modern furnaces or oxidation systems that are a little bit different. They also have vertical furnaces these days, where the wafers sit vertically like this to help avoid warpage and things like that. There are rapid thermal oxidation systems and fast-ramped furnaces.

In fact, I think on the next slide, slide 29, there is an example nothing like an oxidation furnace, but a system in which people do grow SiO2 and oxynitrides by a process called rapid thermal processing or rapid thermal oxidation. This is a photograph. I took it off the Applied Materials website. It's a little bit fuzzy because I blew it up.

But this is a particular system called the Centura rapid thermal processing system. And it's a single wafer system. And you can see it's like a clamshell. The clamshell has been opened up. A wafer goes in, sits here. In fact, you can see where a wafer might sit. The clamshell would be down, and the infrared lamps-- you can see them glowing-- the infrared lamps heat up very rapidly, heat up the wafer within a few seconds.

It's in a little quartz chamber where there's flowing oxygen or whatever. Heats it up. The reaction takes place and then cools it down. So this is a so-called single wafer rapid thermal processing, a method of making oxide. So you don't have to necessarily do gate oxides in an old fashioned furnace anymore. You can also do it in these new pieces of equipment.

So I want to go on to slide 30 now. And I want to talk a little bit now about a little more theoretical things, which is to talk about the silicon silicon dioxide interface. This is important from the point of view of understanding electrically the quality of what you've produced. And it does feed into how we actually do the processes to grow thermal oxides.

So this is a cross section here in pink. On the bottom is supposed to be the silicon, and the white region is the SiO2. And back in the 1980s, Bruce Thiel suggested that this picture of electrical defects that might exist, the evidence of which had been observed experimentally. And so there are different types of charges here that we're going to talk about, Qit, Qf, Qot, and Qm. On the next slide, I'll define those.

One point to make, though, is that to first order the interface is perfect. We're going to focus on defects because we're all neurotic, and we're interested in defects. But just to stand back for a moment and think, it's perfect in that only one in about 10 to the fifth atoms has a defect.

So if you were to go onto that surface and walk around at the interface and just count the number of atoms that where there was an unsatisfied or broken bond, you would count 1 for every 10 to the fifth. So that's pretty darn perfect. It's still not perfect. It's not as perfect as we'd like to make it. We always want better. But it's pretty darn good compared to any other semiconductor insulator system.

So as a result, the defect densities at the interface we measure are in this range per square centimeter, 10 to the ninth to 10 to the 11th defects on a per square centimeter basis. Remember, the surface atom density is about 10 to the 15th. So again, it's about 1 in 10 to the 5, something like that.

So let's go on to slide 31 and talk about these different charges. So first, there's basically these four types of charges associated with the insulator itself or the semiconductor insulator interface. The first one you may have heard of called Qf, or Q sub f, it's the fixed oxide charge. It is represented here by these little positive plus signs. It's a sheet of positive charge, and it's supposed to be within about 2 nanometers of the interface.

So you might ask me, well, what happens when you make an oxide that's less than 2 nanometers? Well, that's a good question. What happens to the fixed charge? So it's not it's not exactly clear. But the fixed charge, it has a very unique characteristic to it, and it's always positive. And we'll talk about why people think that is. And you can reduce its concentration by heat treating the oxide after it's grown, heat treating it at a certain temperature.

So we put into our recipes after we grow an oxide very often an anneal to reduce the fixed charge to as low as it can go. That's Qf. It's always positive. And it's a fixed number. It's fixed in the sense of, after you've done the process, you have the chip, depending on how you bias the chip, it doesn't change.

That's not true of Qit. Interface trap charge, Qit, can be either, as pictured here by the little x, x meaning we don't know its sign, it can be either plus, it can be neutral or negative. And it may change during normal device operation as you change the bias because electrons or holes can be captured.

So it has a behavior similar to bulk deep levels, the way iron and nickel can trap electrons and holes, the way that's discussed in chapters 1 and 4. So Qit is not fixed. It changes with time depending on how we bias. And we'll talk about how it changes with bias.

Third type, QM, mobile charge-- well, as the name suggests-- these ions are mobile. That is, they can move in the oxide from the top surface to the bottom surface, the bottom interface. They can move up and down upon application of electric field. They're ionized. So if you heat the oxide, say, to 100, 200, 300 degrees and you put electric field on it, you can actually cause these sodium and potassium to move from here to here and back.

Mobile ions are less important in modern manufacturing because clean rooms are so clean. People are never allowed to touch anything. Unless you're in a research lab where people are sloppy or something like that, you don't really-- you might worry about mobile ions. We do worry about them, of course, because here at MIT we're in a research fab. We have a lot of students. But in a real fab, people, in manufacturing sense, they're very, very careful about mobile ions. So usually, the mobile ion constant charge is relatively small unless you've made a mistake.

Fourth type, Qot, oxide trap charge, this is represented right here. It's in the bulk of the oxide. Notice, it's drawn up here. It's not at the interface. It may be created by other processes that the oxide was subjected to after it was grown. The oxide might have been subjected to plasma etching. The plasma could put traps in because the plasma has energetic ions going into this thing. Or it may capture electrons that are holes that are injected into the oxide during device operation.

So the energetic electrons may be coming along the surface and, boop, get popped up by some scattering event and create oxide trap charge. So it's in the bulk. And it may depend, not only on how the oxide was grown, but what happened to the oxide afterwards. So those are the four characteristic types, and we can see their signature by doing certain types of electrical measurements.

So let's take a look at slide 32. There are a lot of different measurement techniques-- and your book talks about them-- to understand the properties of oxide. You can do physical measurements. A very simple physical measurement any of you can do in the lab, take an oxide and etch it with hydrofluoric acid, and measure its etch rate.

And it turns out, the etch rate in Hf, so how quickly it gets removed by Hf is a function of the density and the index of refraction of the density oxide. So it gives you an idea of the density. You can do scanning electron micrograph or AFM, atomic force, to see how rough the surface is.

A very common measurement is called ellipsometry. This is a technique where you put a laser beam into the oxide, and you look at its reflection. And you measure-- actually, it's a polarized beam. You measure the shift in the polarization. And it's a very nice method of measuring accurately the thickness of the oxide and its index of refraction, which tells you about the density.

So almost every fab, and even the research labs, have ellipsometers, where you can take your wafer. You'd be able to take this wafer here from Intel, stick it on the ellipsometer. It would come back and say 5,000 angstroms or something, or a 532 angstroms. It can tell you the oxide thickness.

What we're going to talk about a little bit here is electrical measurements. And some of you may be familiar with these, but there's a measurement called a capacitance voltage technique. It's probably the most powerful measurement technique that you can subject an oxide to.

If you look in the text, section 6.4.3, there's two or three pages on it. I suggest you want to read through that, especially if you're not familiar with CV because it has a lot more detail than what we can do in the lecture.

So let's just go through some of these CV measurements. If you've had basic courses in electronics, you'll be quite familiar with this. And it'll be boring and a review. If you haven't, it'll probably seem a little bit mysterious. But bear with us. And again, if you go back and you read through that section, it hopefully will help.

The main point it will make, though, hopefully is that, by doing a simple measurement, which doesn't cost that much time or money, you can learn a tremendous amount about how that oxide was grown and what happened to it in its life. So I'm going to talk about a technique called high frequency capacitance voltage.

And in order to do this measurement you do need to make a device, a very simple one. In fact, it's the simplest device probably that you can ever make on silicon. What you have is you take a silicon wafer. In this case, it's n-type. It's doped with donors. And you grow an oxide.

You take it into the clean room, put it in at 900 degrees in an oxygen ambient for half an hour. You grow a certain oxide thickness. You put metal down, just evaporate aluminum and pattern into little dots. And then you make a contact to the back of the wafer, and you put a probe down the front of the wafer. So relatively simple measurement. We do this here at MIT in a class called 6.152J. Very simple. You make an MOS capacitor in the lab.

This is physically what it looks like. It's got silicon, oxide, and metal. This little symbol in the middle is meant to represent from the circuit engine or the electrical engineers or physics point of view what it looks like. It's a capacitor. It's two electrodes with contacts on either side. And what I'm showing in this plot here is the capacitance on the y-axis as measured as a function of the gate voltage. So the amount of voltage put on the gate is referred to as that little aluminum dot on the surface of the wafer.

And we've biased this particular wafer in such a way that it's got a positive bias on the aluminum dot. In that case, electrons from the substrate will be attracted to the surface. And we do what we call-- we accumulate the substrate, the surface. The primary majority carrier in the silicon is electrons anyway. But I get more of them right at that silicon-silicon-dioxide interface. I accumulate them.

And so if I were to measure the capacitance of this thing, it looks just like a parallel plate from basic physics. I have aluminum. I have a dielectric. I have another sheet of electrons. Acts sort of metallic. That's called a capacitor. The capacitance you measure is the dielectric capacitance related to the dielectric constant and the thickness, c-ox. So as long as I bias it was sufficiently positive voltage, I get a constant. The capacitance is independent of the gate bias. So that's accumulation.

Now we start taking the gate bias and start making it negative. So this vg now goes to 0 and then to a negative number. Well, when that happens, we get something actually calling depletion. And depletion, as the name suggests, means that the surface region is now, instead of accumulated, it's depleted of free electrons.

The free carriers are now pushed away from the surface, shown sort of schematically by this arrow, the electrons going away. And we form a region here of a certain thickness, x sub d, where there are no electrons. There are no majority carriers. It's depleted of free carriers. And this x sub d region is called the depletion region. It has a certain thickness.

And in fact, the thickness of that depleted region, the thickness of xd depends on the bias. As I make it more negative, xd grows. In fact, what we have electrically, if you want to look at the circuit schematic, we have two capacitors in series. We have an oxide capacitor from here to here. And we have a depletion capacitance in the semiconductor, which is a variable. And it depends on the voltage. So in fact, if you measure the capacitance, it starts to go down. And you're sweeping out a curve as you make it more negative.

So if you go on to slide 34 and you continue on making it more negative, you finally hit a point called v threshold. And right at that point, you're at something called inversion. And in fact, what's happened, you've put so much negative charge on the gate that you actually start getting positive holes attracted to that interface, to that top interface. So you've now formed an inversion layer. If you're a circuit design person, right at VT, that's where the device turns on for a MOSFET. That's where you've just formed a conducting channel. This bright red region here is the holes in the inversion layer.

So in an n type semiconductor, in inversion, you have a p type in a layer of holes at the interface. This xd, which I said was expanding, it stops expanding. It reaches its maximum. And all of a sudden, you get a constant capacitance. You notice the capacitance curve now is flattened out, which is just a series capacitance of this c-ox and this c depletion region, where the depletion region is maxed out.

So you trace out a very simple looking curve, which we'll see how the properties of this oxide affect this curve. But the basic idea is that all regions, no matter where you're operating this device, the amount of charge that's on the gate here, on the aluminum, has to be balanced by the charges in the substrate. And the charges in the substrate will be either the depletion charge consisting of the charge associated with the donors in the depletion region, or the inversion charge, which is also positive, which are these holes in this region. So that's kind of a basic charge neutrality overall that has to hold.

So let's go on to slide 36 briefly. This is a little more detailed solid state physics than most you need to get into. If you're familiar with it, again, it'll be a reminder. If not, you have to take it as a truth. The question is, once I've formed this inversion layer and I have this Q sub I and I add additional charge to the gate, it's balanced by more inversion charge.

You might say, well, why doesn't xd just keep growing? People always wonder about that. Why do we always get more inversion charge instead? Well, it turns out it's just a lot easier at that point to create inversion charge, to create these holes at the interface. In fact, if you're taking a solid state physics class, you know that that hole density looks something like this. In fact, it has an exponential dependence on that voltage, on how we move that potential up and down.

So exponential means we only have to move the potential just a little bit, and we can get exponentially more carriers. So it's a lot easier to create this inversion charge at this point than it is to create more depletion charge. So that's kind of a reason why that happens. If you're familiar with it, that'll just remind you how that works.

So let's go on to slide 36. And now, I've traced out-- again, this is a very schematic sort of view. But I've traced out a plot of capacitance on the vertical axis. And the x-axis is DC gate voltage. And we just traced out for you this black curve, where we started at c-ox, a constant number. As I sweep down through 0, I go in depletion. And it goes down, and then the capacitance reaches a minimum. And that's what an ideal high-frequency capacitance curve would look like.

Now, basically, so what happens when we go to a lower frequency? Well, I have another curve here where we swept through this region, through depletion. And then all of a sudden, the curve, right after threshold in inversion at low frequencies, the capacitance goes back up to c-ox.

Well, it's because if I have a signal that's going at low enough frequencies, the inversion layer carriers can actually follow that signal. So instead of just getting this minimum capacitance of these two in capacitors in series, it actually goes back up to look like a normal parallel plate capacitor with capacitance c-ox. In fact, if we go to the next slide. I think it explains it a little bit more physically on slide 37, the different regions.

So in accumulation, remember, we said we just had a regular parallel plate capacitor, capacitance c-ox. I wiggle charge here on the metal gate, and you wiggle a little bit equal amount of charge in the substrate. In depletion, what's happening? You're wiggling charge on the gate. And instead of accumulating, you're actually depleting. And you're wiggling charge at the back end of the space charge region, which is sweeping out. And that's why the capacitance goes down.

Now, if I'm at low frequency, and I'm inverted, I've created this inversion layer here of positive charge. If I wiggle charge on the gate, it's low enough frequency that I can actually wiggle charge in the inversion layer. However, if I'm at a bias where I'm in inversion but I'm doing it a very high frequency, say, megahertz or 100 kilohertz, the inversion layer can't respond that fast. But the depletion region actually can.

So you actually get the series capacitance of both the inversion layer and the depletion layer. So this is all reasonably well understood if you want to read in [? Peres ?] book, it has a little more detailed explanation of these different regimes or regions of the curve.

So there's another region that tells us something of the CV curve, which is going to end up telling us something about the quality of the oxide. And it's called deep depletion. And I've marked that region here on a high-frequency curve, where it actually, instead of being flat going straight across, it actually starts to go down.

And what we've done in deep depletion-- remember, in this regime we have two capacitors in series, the oxide and the depletion region-- is that we sweep the DC voltage very fast so that the inversion layer carries can't follow it. So xd ends up expanding a little bit. And in fact, if you see deep depletion, it's a sign that you have a very high minority carrier lifetime. You have very few traps. You don't have much iron and things like that.

So people will often take capacitors and try to intentionally sweep the CV quickly, see if they can get this lowering, this deep depletion. The faster you sweep it, the further down it goes. If you don't see that lowering, you don't have very good minority carrier lifetime. You probably have a lot of traps or things somewhere at that interface. So this is a way of getting a rough idea of what kind of quality interface have you produced? Can you deep deplete the capacitor or not?

So basically, from looking at these capacitance voltage measurements, we can extract a lot of quantitative information. In fact, we can get the oxide thickness. Remember, we had c-ox. C-ox just depends on the dielectric constant and the oxide thickness.

So once I get that number, I know the area, I can calculate the oxide thickness. I can get the substrate doping from the depletion capacitance. And you can get all these different interface charges, Qf, Qit, the mobile, and the oxide trap charge, all of those just by doing CV measurements in their different configurations and frequencies. We can get this information, which tells you about the electrical quality of the device.

So in fact, on slide 40, what I'm showing are some quote unquote "realistic" capacitance voltage curves, much more like you would measure in the lab. What we've shown so far on a [? cv ?] plot, it looks like this. So we have this curve here. Here in accumulation we're at c-ox. It comes down and sweeps out like this. So that's the ideal high-frequency curve.

In fact, that's not what you would measure in a real device. You would measure that curve but shifted over to the left. So if you look at this curve right here, you see it's shifted. And in fact, it's the high-frequency curve but shifted by two terms. And the terms look like this. There's one term that goes like the coulombic charge, little q, times Qf, the fixed charge, divided by c-ox.

So this is the Qf term. This is just-- remember, Qf is positive. So it just literally takes the threshold voltage and shifts it by Qf over c-ox. So you see this shift because of this positive fixed charge. The bigger the fixed charge, the bigger that shift. And in fact, it's the shift from the ideal curve that people use. That's the way people use to measure Qf.

Phi [? M S ?] is the metal semiconductor work function difference, which is just sort of a property of the metal that you use, whether you use aluminum or what you use. So there's also a shift associated with that. And that's a fixed number depending on the polysilicon gate that you used.

So with this positive fixed charge shifts the CV to the left, and you get this curve. That's a good sign. Now, most of you would actually, if you were to measure your devices, you'd get a distorted curve. Instead of looking just like with this, with this abrupt drop off, you'd see it looks a little bit distorted from what you calculated in the ideal case.

And in fact, this distortion is due to the fact that there is variable interface trap charge. Remember, we talked about Qit? And we said the amount of charge and Qit depends on the bias voltage? So as I'm sweeping from here from right to left, I'm uncovering or causing extra-- my ability to see, to charge an uncharge Qit. And that ends up shifting the curve by variable amounts. So it gets distorted. So the amount of distortion is a way of telling how much Qit you have.

And in fact, if you look on slide 41, this is pictured for you. If you've had solid state physics class and you know about energy band diagrams and states, then this is a way of thinking about that. If you haven't, again, it's the type of thing you'll have to take with a grain of salt or study up on it.

So what we have pictorially here up in this upper band diagram is the conduction band here shown of the silicon. And here's the valence band. So here's the band gap between those two. This is the semiconductor of a silicon, and on the left is the oxide. And notice, right at the interface, there's a whole bunch of little levels. Each little bar here represents an energy level that electron can occupy.

And you notice that there are energy levels within the entire band gap. There are certain density of states throughout the band gap. And we're going to assume they're donor type, and they're distributed somehow. Their donor type, meaning that they're positive when they're empty. So if there's no electron in there, this has a positive charge. If it captures an electron, it's neutral, OK?

So now we look at this energy band diagram in the center set of curves for different bias conditions. So we can look at it under inversion. So when the surface is inverted, it turns out that Qit is large, and it's positive. It's mostly empty. We have the bands bent. None of these have any electrons in them. So I have now a lot of positive charge at that interface under that bias condition, OK?

So here I am at inversion, region C on this bottom curve. And lo and behold, I have a lot of positive charge. That means I am shifted from the ideal case quite a bit to the left. And you see why in region C of this realistic curve I'm shifted quite far over. So I'm distorted.

So now let's go to region B of that curve. So that's here under depletion in the center. When I'm depleted, the bias voltage, again, is a little bit closer to 0. I don't have as much band bending. And in fact, some of these density of interface traps is now filled with electrons, maybe about half of them. So the amount of Qit that's uncovered at that interface now in this bias condition is less. So I have less interface trap charge. So indeed, if you look at the ideal curve and you look at the distorted curve, at point B, there's less of a shift than there was at point C because there's less charge at that interface, less positive charge.

And now, let's finally go to accumulation. I've bent things-- the voltage now is such that all of these traps are occupied by electrons. So Qit has its minimum value. They're all occupied. So Qit is low. And so in region A, where we're in accumulation, in fact, the shift from the ideal is very, very small. It's a small shift.

So as I sweep here, from inversion, I get the curve. A lot of shift to depletion, not so much shift, accumulation very little shift. So the distortion of the curve has to do with these traps that depend on the bias level and the fact that it varies with bias.

So how do people measure Qit? Well, there's one way of doing it. There are a couple of different ones, but what people do is they measure a curve under high frequency, and they choose the frequency such that the traps, the Qit's can't respond. So they get kind of an ideal like curve, shown here.

Or you can calculate the ideal curve. If you're a more theoretical person, you can calculate it. And then they do a low-frequency curve, and they choose low enough so that the traps can respond. And in fact, they see a difference. In a certain bias regime, you can see a difference in the capacitance. And that difference in capacitance can be related to the density of the traps that are in a certain region of the bandgap because where you are in the bandgap corresponds to where you are in bias.

So you could give someone this data who is familiar with it, and they could plot then for you the density of traps in the bandgap at that interface between silicon silicon dioxide as a function of the energy in the bandgap. And you can see it has this familiar u-shaped profile. In the center, it has some number. Mid-gap it 's typically 5 10 to the 10th traps per square centimeter. As you go closer to the edges, it goes up dramatically. But this is a way by CV of measuring Dit or Qit.

And in fact, we want to keep it as low as possible so our threshold voltage is moving all over the place and we're getting ideal characteristics. And what we do is, after we do an oxidation, the last high temperature step is typically some kind of forming gas anneal. And after annealing, this Dit goes down by about two orders of magnitude.

So you notice the last step in a MOSFET or an MOS flow, if you've done it, is to do an annealing and form a gas, which has hydrogen in it, at about 450 or 500. And that takes this Dit from up here all the way down to here to a level that's tolerable for the device to operate.

So I know I went through a lot of that CV stuff fairly rapidly, and I apologize for that. But I wanted to give you more than the details, give you a flavor for it. If you've had courses in electrical or solid state physics, you'll recognize some of it. If you haven't, you can go through in the text in more detail.

But just to summarize on the oxidation, it's really probably one of the most critical processes for CMOS. Thermal oxides that are thermally grown are used for a lot of different things, tunnel insulators, masking oxide, field oxides. If nothing else, you should remember that the electrical properties of silicon silicon dioxide are the best, and they're superior to all other cases. And that's why it's taking so many years for us to find another insulator, another high k. It really is tough to beat, or it's tough even to equal it.

There are charges, though. It's not perfect, remember? Nobody's perfect. 1 and then 10 to the fifth is a problem at the interface. But there are charges that exist and things like fixed charge, which is positive, Dit or Qit, mobile charge-- you don't have mobile charge if you're really clean-- and Qot.

These charges do exist, but they can be characterized by relatively simple techniques, like MOSCV. So we'll go on next time and talk about the physics or the kinetics of how this process of thermal oxidation takes place. But please go ahead and start reading chapter 6.

Also, your handout, homework number 2 was handed out. So you can go ahead and start working on that. That's due next Thursday. Thanks. it.