JUDY HOYT: OK, maybe we'll go ahead and get started, at least with the announcements. Yeah, hopefully everybody's recovered from their Thanksgiving feast. And looks like people are sleeping in today. What I'm showing up here is the schedule to orient us. This is lecture 22. We'll talk about silicides, device contacts, and I've added in a new material this year on novel gate materials.

And part of this is covered-- the first two topics are covered in chapter 11. And then we have one more lecture, which is on strained silicon and silicon germanium growth and processing. And then next week we have two class periods scheduled. There'll be four speakers in each. And those will be the oral reports and the student reports given on Tuesday and Thursday. And in fact, I have a slightly better version of this schedule, I think, in this Excel file, and this is posted up on the web, by the way.

So on December 7th, we have these four speakers. We're going to hear short presentations, about 16 minutes each, from these four students on everything from high k going to diffusion and gallium arsenide. And then on the Thursday of the ninth, we have these four students scheduled. And I just want to remind people, if you're doing an oral report, you're expected to provide handouts, the same type of handout that I use during lecture.

And if you need help in making Xerox copies, contact my assistant. Her contact information is published on the web. Make sure you get to her early enough. Don't give her the job the same day, that morning, but maybe the day before would be good. And also, if you're doing a written report they are due on Thursday, December 9 in class. If you have any questions about the final project, just please email me or contact me after class.

OK, so that's the bookkeeping and to remind people about the final project. Let's go on and start-- we can start today's lecture. The notes for today's lecture are given in handout 36. And today we want to talk about a couple of devices, a couple of areas-- the formation of silicides. I've referred to silicides a number of times in the course, but we really haven't gotten a chance to talk about what they are. Now we'll find out.

How we make contact, how we make the structures and make good, electrical contacts and the device. And as I mentioned, I've added a new module this year on novel gate materials. And we started some of that last time. So far, we've talked about how to get these insulating and doped regions inside the device. We've talked about oxidation, implantation, diffusion, thin film deposition, and how we etch structures.

But at this point in the device fabrication of the seam laws flow, we need some way to make a good electrical contact. And I put good electrical contact in quotes because it's somewhat nebulous. But we'll see in this lecture what we mean by good electrical contact-- two doped regions. And so this lecture is kind of the first link to the backend technology course. The contact itself is still generally considered part of the frontend because you're contacting the silicon. And then beyond, that everything is in the 6.773, which is the backend technology.

This is just a schematic I've shown a couple of different times to point out to you what I mean by the contact. You're familiar with this by now. We have the silicon source and the silicon heavily doped drain. Here's the gate material. And I need some way of making a contact between this electrical connection here, this metal line, and the silicon itself. And you can see that can consist of several different materials. In this particular picture, these sort of dark regions contacting the silicon are going to be made of silicide. And we'll talk today about how you fabricate those silicides.

Let me also-- I want to talk about local interconnect. Sometimes the contacts themselves are also used for something called local interconnect. And I should point out here, this is an example of a local interconnect where I've got a contact region that is being used to interconnect, say, from one layer to the next, or locally from one part of the device to a neighboring device, but not across the whole chip. These are called global interconnects because these wires go potentially across the entire chip up here.

But this is a local interconnect. It only extends over a very small distance. In fact, on slide number two of your handout, I've got an example of a local interconnect and what it might consist of. This is just the circuit schematic. If you have taken electrical engineering courses, you know this is a bipolar transistor with a base and the emitter. This little arrow represents the emitter region. And you see in this circuit schematic that the emitter of this transistor is connected by a wire to a resistor, r.

Well, this is the circuit schematic. And below is the actual implementation of this circuit in silicon. And in fact, if you recognize that this is an NPN device, and this little n region right here is the emitter-- this is the P-type base as shown here in red, and this is the N-type collector. And what you see here is this emitter region has a metal contact to it. And it is connected locally to this P-type tube region, which is, in fact, a resistor.

This is a resistor because the current can flow from this metal contact here through the P-type region-- it has a certain sheet resistance, or a certain resistance-- and out this other end. So this is an example of using a very short metal wire. Might even end up being silicide or something. A very short metal wire locally on the chip to do interconnect from, say, one neighboring device-- in this case a transistor-- to another neighboring device, a resistor. So that's what we mean by local interconnect.

Talk a little bit about historically about how contacts were made. And this is shown on slide 3. In the early days, if this was the drain or the source of a transistor, a MOSFET, people simply used aluminum, which was the metal of choice, directly on silicon to make both the contacts and the interconnect material. So it was convenient. You could contact the silicon with aluminum, and then you could run wires along the chip. And that could interconnect the various devices.

The advantages of aluminum has a low resistivity. It's the second lowest of all the metal candidates, copper being the lowest. And it has very good adhesion to silicon and to silicon dioxide. It's a very stable metal, and it makes good electrical contact to heavily doped silicon, as long as the silicon is heavily doped.

Why does it do that? Well, aluminum tends to reduce any native oxide present on silicon. You know if you take a silicon wafer and you do an HF dip, you'll get rid of any native oxide. But if you let it sit for any period-- a few minutes, an hour or so-- you grow a thin, natural native oxide, SiO_2, at room temperature on silicon. May only be about 10 angstroms thick, but it's still there. The nice thing about aluminum is that it tends to reduce or eat up that oxide on the silicon, and it forms a very thin layer of Al_2O_3.

But this layer is quite thin, and the aluminum itself can diffuse right through it. So this enables you to form a very good electrical contact without having an intervening layer of oxide. Some metals you might put down and they won't eat through the silicon native oxide, the SiO_2. And then they'll just be sitting there, and you won't get a good electrical contact. Aluminum has this property that eats through that little native oxide so you get a good contact. So it was a convenient and a good material for people to use for many years.

Slide 4 just shows you some of the basic properties of interconnect materials just so you get oriented. Here is aluminum on the top. Again, one of the most popular for many years. The resistivity is listed in the second column, somewhere around 2.7 to 3 micro ohm centimeter for the resistivity. Again, the resistivity is a property of the material. And the last column shows the melting point. So it is a low melting point material.

The only other metal of consequence that has a lower resistivity is copper, which is shown here, the third one, the third row. That's about 1.7 to 2 micron centimeter. Copper has a much high melting point. Copper has other difficulties associated with its integration. Copper, if it gets into silicon, can be a deep level. So it can cause lifetime problems. Copper oxidizes very readily, even at room temperature. So you need to protect it, if you're all aware of that.

And copper is more difficult to deposit than aluminum. But copper is the material of choice for today's interconnect. I think I brought a wafer from Intel about halfway through the course, a 12-inch wafer, and showed it to you. And you noticed it was all pretty copper colored. It's because the modern interconnect material is copper. A few places are still using aluminum. Aluminum was historically what's used. Copper was introduced in research in the mid 80s and in production about 10 years past that.

On slide 4, I'm showing some basic physics that you may or may not be familiar with. But you need to understand the basics of it in order to understand what we mean by a good electrical. Contact their contacts can, to a semiconductor, can be sort of divided into roughly two classes. There's an ohmic contact and a Schottky contact.

A Schottky contact is shown here. And if you just look at the current, this little schematic shows the current that flows in the device, so through the device, as a function of the voltage that you apply. Schottky contact is a rectifier. It is when you apply a large voltage in one direction, the reverse direction, not much current flows. So that's not a very useful way to make a contact to a device.

If you apply a small voltage in the other direction, you get a little current, and then you apply above a certain threshold voltage, and you get a lot of current. It's basically a diode. And the current transport is by this sort of thermionic or thermal emission over this barrier. So this is not considered desirable for making a good contact. An ohmic contact on the other side, or a tunneling contact, occurs when you can actually tunnel right through this barrier. The barrier is so thin that you can actually tunnel right through it. The depletion region is very thin.

And if you do make high doping right near the metal, you can reduce this depletion layer width. And it enables this tunneling. And so if you look at the current voltage characteristics-- so current plotted on the y-axis, voltage on the x-axis-- you see it's very symmetrical, and you get a very, very high current for a very small voltage in either direction. So this is considered desirable. This is what we want. We want that ohmic contact. And it tends to happen, particularly with aluminum and most metals, if you make the surface doping very high.

A concept that we want to talk about that is shown here on slide 6 is called the specific contact resistivity. And we usually use the Greek letter rho sub C. The definition of it is given here. It's defined as being equivalent to the derivative of the voltage current density characteristic at a metal semiconductor contacts. So it's partial v, the voltage, by partial j, basically.

And so we usually assume a structure where the current density is uniform across some contact area. Then you can calculate the contact resistance R in ohms. So R would have the units of ohms from the specific contact resistivity by the following equation. The resistance of a given contact in ohms is just the voltage divided by the current flowing through that contact. And you can calculate it by rho sub C, which typically has units of ohm centimeters squared or ohm micron squared-- so it's got a units of resistance times an area-- and divided by the area of the contact.

So if someone tells you, oh I've made a contact. It has a certain specific contact resistivity, 10 to the -7 ohm centimeter squared, then you can design your contact area as you need to to get the right contact resistance in ohms. And basically what this tells you is-- and you notice you it's the total resistance of a contact in ohms is inversely proportional to area. So if I make a smaller contact area for a given specific contact resistivity, you're going to have a higher resistance.

And this is a problem if I'm scaling devices. If I'm packing more and more devices on the chip, it means I want to make the total area occupied by every device smaller. That means I need to make the contact area smaller. But if I leave the contact resistivity the same, I don't do anything different about how I make the contact, then as I'm scaling these devices, of course, the area of the contact is going down to get more of these devices on a chip. That means the resistance is going up as devices are scaled.

And we'll see some specific numbers on this. This is a problem as we scale, unless we scale roh sub c, unless we do something to make better contacts as we scale devices, just because of this geometric factor. As I make the area the current goes through smaller, the resistance of that contact goes up.

So the second type of contact-- again, this is one that's not generally considered desirable, but we need to understand a little bit of the physics of it-- is a Schottky contact. And the energy band diagram for that is shown here, where the metal is on the left and the semiconductor is on the right. And a Schottky contact is governed by thermionic emission. So it's a thermionic process emitting carriers over this barrier.

And we know from the Richardson equation we can write thermionic emission current density J as being proportional to the temperature squared, some constant a star. But most importantly, it goes exponentially like the barrier height. So it's e to the minus q phi B over kT where the barrier height is just the band bending. It's this barrier between the Fermi level in the semiconductor and the Fermi level in the metal.

So that's going to tend to be a property of the doping in the semiconductor and what kind of metal you put on it. So if we take this equation. this current equation, which we know, the current voltage characteristic is exponential in the applied voltage. And we just use the definition of roh sub C, you can calculate, as shown on the bottom here of slide 7, what the specific contact resistivity should look like for a Schottky barrier. And you see it goes exponentially, like the barrier height. So if we change that barrier height, we can change rho sub C.

So what we do in practice is we look at that equation and we try to get into a different regime. Instead of being in the regime where we're dominated by tunneling over this barrier by thermionic emission, rather, excuse me, over the barrier, if we make the barrier distance really narrow, really small, you know from your quantum mechanics classes that you can actually get quantum mechanical tunneling through the barrier.

So here on slide 8, what I've shown is a tunneling contact. It's a Schottky contact in the limit of very, very high doping. So we form these-- you see we still have this barrier, phi B. But the distance, the depletion layer thickness in the semiconductor, is very, very small. And that happens when you make the doping high. So in fact, for a tunneling contact, if you go back to your quantum mechanics, you know if you have a certain barrier height, phi B, and a certain barrier thickness XD, that the tunneling current is exponentially dependent on those two quantities. It looks something like this.

So how do I make the depletion layer thickness small in the semiconductor when I put a metal contact to it? Well, again, if you're going back to some of your basic electrostatics, XD is inversely proportional to the doping, or the square root of the doping. So all I need to do to make XD small is to pump up this doping very high, the doping in the semiconductor.

So when the doping is high enough, XD will become small. And then when XD gets less than about, say, 2 nanometers or so in this equation, you get a very large tunneling current, basically. So what people do in practice is you dope very heavily above, say, above mid 10 of the 19th, you get quantum mechanical tunneling. And that's really what's dominating. So if you put these numbers and in here and you look at the definition of the specific contact resistivity, you get an equation like what's shown at the bottom of page 8 here.

And you can see, again, it depends exponentially on the barrier height and inversely exponentially on the doping. So for low contact resistivity, to get this number down, I need a small barrier. So you would choose a particular metal that gives you a relatively small barrier height, phi B. And you need, most importantly, very high doping. So you need to pump up the doping as high as possible. And then you will get an IV characteristic that is essentially ohmic.

So here I'm just doing a simple calculation on slide 9. Here's an example if we use the expression for the specific contact resistivity at very high doping levels where we have tunneling dominate. So we can write that expression that I just showed on the prior page saying that rho C is rho C zero, some pre-exponential, times the exponential dependence on phi B and D, and where this exponential multiplier C1 is given by this number, 7 times 10 to the 10th.

So now let's say I assume I have a metal semiconductor contact where the barrier height is 0.6 electron volts. And rho C zero, this number, is about 10 to the minus 7 ohm-- that's a little hard to read, but that's ohm centimeter squared for an aluminum and silicon contact. So the question is if I change the doping from 1e19 to 1e20 by a factor of 10, how much does rho C go down? Just to give us an example of how much you benefit yourself. And again, this is an exponential relationship. So you expect a big change.

So in this case, the first case, I have a doping of 10 to the 19th. You plug in all these numbers. You put in the square root of 10 of 19 down here and you get a RHO C of about 6 times 10 to the -2 ohm centimeter squared. It's a pretty big number, as it turns out, if you plug in a standard size of a contact on a silicon chip. That's a 10 to the 19.

At 10 of a 20, you do the same sort of thing, and you do the math, and now you get about 6 times 10 to the -6 ohm centimeter squared. So that's about a factor of 9,000 lower. So it changed by almost four orders of magnitude, the contact resistivity, just by upping the doping by a factor of 10. She got four orders of magnitude. So that's dramatic difference.

So this tells you, you need to have-- if you're going to get reasonable contacts in silicon technology, unless you want huge devices where the whole chip is dominated by contacts, which is ridiculous-- you won't be able to get enough devices on your chip-- you're going to have to get the doping in the source drain regions high, at least 10 to the 20 or higher.

This number itself is still considered a little bit on the too-high of a side for a contact resistance. But it's just interesting to look at these numbers. And in fact, I've taken on the next slide figure 11-7 from your text. And you can see that these numbers are in pretty good agreement with what people have measured.

Here on slide 10-- I took this from your text that shows the contact resistivity in ohm centimeter squared as a function of doping here. And you can see the doping up on this scale here for two different metal systems. One is the aluminum contact to N-type silicon, and the other is platinum silicide to N-type silicon.

And the data is the bullets that people have measured, like these open squares are for aluminum contacting N-type silicon. And the theoretical curve is shown here with the solid line. There you can see there's reasonable agreement. Notice here this is a log-log scale. So we're getting an exponential dependence in the heavily doped regime. So again, as doping is increasing to the left on this x-axis-- so here's a doping of 10 of a 19th. Here's a doping of 10 of the 20. And we see a number of orders of magnitude. Basically force of magnitude drop as they go from 10 to 19 to 10 to the 20.

So you say, OK, that's not a big deal. I just keep increasing the doping in my source drain regions over time. The problem is we know that there are electrical solubility limits. We cannot just arbitrarily keep increasing doping. We're trying to find ways to do it. But you know if you do an implant and you do a certain anneal, you get doping up to a certain value, maybe depending on the doping, maybe in the low 20s. Maybe 2 to 4 times 10 to the negative. Beyond that, it's very hard to go much further.

So the problem is that rho C is not really scaling because of the doping electrical solubility limit. So the contact passivity doesn't scale as we shrink technology. And this is a major problem. People are looking for new methods, new materials, whatever, some way of getting the doping up, or a new method of making contact. That's kind of a fundamental issue.

In fact, here on slide 11, I show that this is really one of the concerns. This is a table that I took, table 71a, from the 2003 international technology roadmap for semiconductors. I've shown this before, but we hadn't gotten to this contact stage. And just to remind you, in the upper right corner, this is what the MOSFET looks like. And these black regions are going to be silicides, either cobalt silicides or nickel silicide. We'll talk more about that.

But the point is they're a contact between a metal and a heavily doped semiconductor, which is silicon. And if you look at some of the requirements here on the ITRS roadmap, here we are in, let's say a year ago, 2003. Look at the maximum contact resistive. So this is the rho C. This right here is represented what we call rho C.

In 2003, they wanted to have about 2 times 10 to the -7th ohm centimeter square. And you notice it's sort of a light orange color, which means, according to our key, that there is an interim solution on how to get to this is known. But if you go to 2004, you want to lower it according to ITRS about 1.6 times 10 to -7. That's all in yellow, which means their manufacturable solutions are known. They haven't yet been integrated.

And when we get to 2008 in the red, to get down to about 0.8 times 10 to -7 ohm centimeter squared, there aren't any known solutions that are manufacturable. So we're coming up against the red brick wall here because the device scaling and the need to make the contact size smaller means that we need to drop rho C at this rate. And it's just nobody knows yet how to make a contact that has 0.8 times 10 to the -7 ohm centimeter squared for the specific resistivity.

So it's the contacts that are really hurting us. Look at the sheet resistance you might be concerned about. Well, how about the resistance of the metal itself? Resistance of the electrons flowing through the metal is not a real big problem. When you look at the contact silicide sheet resistance in ohms per square, it needs to be in this range of 6 to 10 ohms per square.

That's not a problem. None of that-- that's all in white, which means people know how to do that. The real problem is making these really good low resistivity contacts to the silicon. I just want to show an example of how-- so you get a feel for how the numbers work of how one does a contact resistance calculation. This is for a MOSFET. This is shown on slide 12 of your handouts.

Again, this is a cross-section view of a MOSFET. So on the left, you can imagine this region as being the heavily doped source. In the middle shown in orange is the gate. And this is the heavily doped drain. And these black regions are metal. Could be silicide, but there's a metal region. And I'm showing the dimension of the contact, so the region over which the metal is in contact with the silicon, in this dimension is 0.2 microns wide.

OK, now into the page, or into the board, we're assuming that it's 1 micron. And in fact, this is a cross-section view of the device. If you were to look down on the device, a top view, this is what you would see. This would be your source contact region. Here's your gate in the orange. And here's the drain. And again, the region over which we have to make the metal contact is assumed to be 0.2 microns in this direction in width and 1 micron in this direction.

So the area of this contact clearly, if the current is going to go flow through it, the area is going to be 0.2 times 1 square microns, or 0.2 micron squared. So that's the area. Assuming we go back to ITRS, assuming we are here roughly in 2002 and the contact resistivity is about times 10 to -7 ohm centimeter squared-- OK, so I'm assuming this as a specific contact resistivity-- then can calculate the resistance of just the drain contact itself.

The resistance is just that rho C divided by the area of that contact. And you get 100 ohms. So the resistance of the current just to flow through this contact is 100 ohms on the drain side. It's going to be 100 ohms on the source side, as well. So that's 200 ohms right there, just to give you a rough idea. And this is equivalent to the ITRS 2003 requirement. So we're talking about 200 ohms of the total resistance of the device just being due to the contacts because of the size.

And that's a reasonably large resistance. So we'll see that the contact resistance actually does dominate. Again, what you could say was, well we'll just make the contact wider, which you can do, but that means your chip ends up being larger. You can't put as many devices on the chip. So there's a fundamental trade-off here, unless we can get this rho C number down. And that's exactly why this number is dropping with time. People want it to drop because they want to be able to continue to scale the area of the device. It's a problem, though.

On slide 13, I'm showing you a structure. If you are studying contacts or if you ever make devices, you always want to know, what is my contact resistance? And this is a classic structure to actually measure contact resistance. It's called a Kelvin structure. I took this particular picture out of your textbook, figure 11-35. There's a reference to it in your textbook, a paper that an article that talks about it in much more detail on how it's actually made.

On the left-hand side is showing the mask layout in order to make one of these structures. And what you do is you have these dark regions here that are funny shaped are considered to be the N plus region that you want to contact. So that would be called the n plus diffusion, has this particular shape. The dashed lines represent the metal. So that would be the metal level. So this is going to take at least three masks to powder.

And these little square regions with the X's going through them are the contacts. So that's the region where one layer contacts the layer above it or below it. So this is sort of a planar view. If you want more of a bird's-eye view of actually how it looks, you can look on the right-hand side. And these L-shaped brackets are made of metal. So there's one here and one here.

And basically you perform a four-point kind of measurement. What you're doing is this little region here that is square in the center that has a dimension of L in one dimension by L in the other-- so this is an L by L squared-- the current then comes through on this leg. So the current flows through the diffusion. Then it flows up through the square, the square contact. And then it flows out through probe number two on the metal.

So you put an ammeter here. You put an ammeter between probes 2 and 3, and you measure that current that's flowing through that square contact. And then you put a volt meter between probes 1 and 4 and you measure the voltage drop across that face. And then you just divide V divided by I, whatever you measure-- the measured voltage between 1 and 4, divided by the current flowing between 2 and 3. And that is your resistance. And that's going to be equivalent to rho C divided by the area of the contact, L squared.

So this is a very common way to do it. You might say, well, why do you go to the effort of having separate probes? Why don't you just measure the current and the voltage on the same probe points? And the reason you do this is because you don't want to have extra contact resistance, say, of your probes going down. And there's always a voltage drop there.

So the probes through which you force the current are separate from the probes through which you measure the voltage. So that, therefore, you're only measuring, really, the voltage drop across this the contact face, the square face. So it's a very common structure, the Kelvin structure, you'll find on a lot of test masks and test circuits that people use to measure the contact resistivity.

Here's an example shown on slide 11 using a cross-bridge Kelvin structure. And you have a 1 micron by 1 micron opening. You find that you get a current of 10 microamps through the contact, so that's I23, when you measure a voltage drop of about 320 microvolts. What is the specific contact resistivity? So you just divide the voltage drop divided by the current. And that gives you a resistance number. In fact, that's 32 ohms.

And by definition, that's equal to rho C divided by the area of the contact. So you can solve for rho C here just by multiplying 32 ohms by the area of the contact. And you get 3.2 times 10 to -7 ohm centimeter squared. It's pretty close to the ITRS requirements. Still a little bit high. It's still about a factor of 2 too high. But it just gives you an idea. Typically you use several different dimensions. So you either use a 1 by 1, a 5 by 5, and a 10 by 10 square.

And you make sure you get the same specific contact resistivity number over all of those because sometimes you can have current crowding effects. When I drew it in this picture here on page 13, I said the current is uniformly going through that face. So you can imagine a flow of water uniformly flowing with the same flux all across the face. That may or may not be true, depending on the series resistance of this N plus diffusion and things like that. So you want to use different areas.

And for different size areas, you should get the same number. If you don't, then you probably have current crowding effects, and you need a more sophisticated two-dimensional model. So that's a typical way that people would measure their contact resistance. So that sort of introduces the whole idea of scaling and why we want to do this. I want to bring up some other requirements.

We said it's good, it's important, to make a low-resistance contact. But there are other requirements besides that. So clearly we need to have a high doping concentration at the interface that's going to allow the electrical tunneling to take place. You need to be-- the interface has to be free of contaminants, as I mentioned. You want to get rid of the native oxide. So you preferably use a metal that will kind of eat away, if there is any little native oxide, that will eat it away.

And you clearly don't want residues. You don't want a lot of excess carbon or other things at the interface because that's going to cause an increase in your contact resistance. You don't want nitride or oxide there. That's one thing to get low-contact resistance. The second thing you want is good thermal stability. You don't want the contact structure to change or degrade in some way during the process. After all, there will be subsequent thermal processing.

Once you make those contacts, you still have to make all that multi-level metal. And those multi-level metal schemes involve annealing steps or deposition steps that could be in the range of 400 to 500 degrees. So you don't want anything weird happening down in your contact when you heat the thing. And in particular, you're also very concerned about junction leakage. Remember, this electrical metal, this metal, is making electrical contact to a PN junction. This is an N plus P junction.

So you want it to be-- ultimately, this N plus P junction to have good junction characteristics and low leakage in the silicon underneath the contact. So here on slide 15, I'm showing an example of a figure I took out of a text where something very bad has happened, which is called spiking of the metal. And you can see what's happened.

This is the metal. Aluminum is shown here. And this is my N plus drain or source. And this chip has been heated up. And the aluminum has actually spiked through the junction. It's actually touching now the P-type silicon. So it's really shorted out. It's essentially shorted out this PN junction. It's created a lot of voids. And this is a big problem.

And that's because aluminum has a finite solubility for silicon. The silicon actually gets sucked into the aluminum and causes this spiking effect. And this is one reason why, if you have a very deep junction that's many microns deep-- well, it would never spike because it never gets that deep. But now the junctions are very shallow, they're 0.1 micron or less, people never put aluminum directly in contact with silicon.

So in the old days, this is the way people made contacts. Today if you do this with any kind of reasonably shallow junction, when you go to heat the thing up to do the final forming gas anneal, you're going to spike the junction and you're going to destroy the device. Here's an example on slide 16 of what happens in some extreme cases of junction spiking. What we're actually looking at-- what I showed you here in the prior slide, this is a cross-section view. You can see the way the aluminum has spiked into the silicon.

In fact, if you look in plan view at a tilt angle of about 45 degrees, this is a contact region. And around it is oxide. These are 5 by 5 micron holes. So they're fairly large. These were annealed by rapid thermal processing. And then the aluminum layer was removed just to see what happened. And in fact, this was annealed for 10 minutes at 425. So this was aluminum directly in contact with silicon at 10 minutes at 425.

You can see it's created all these voids because there's a solubility of silicon in aluminum. Silicon from the substrate has actually gone up into the aluminum. And it leaves behind voids, which end up being filled by the overlying aluminum. And when you etch the aluminum off, you see all these holes. And if you do RTP at a lower temperature here, 350 for ten seconds, you don't see nearly as many spikes, nearly as many holes.

But 350 for ten seconds is really too low for any backend processing. So this is why you never put aluminum directly in contact with silicon unless you're making a really deep junction, like a micron-deep junction, you have some kind of device. But in MOSFETs today, the typical source drain junctions are 0.1 microns. So this is sort of a classic case of junction spiking.

There are a couple of different solutions people have come up with over the years. One of them is shown here on slide 17, although it's not perfect. People had the idea of, well, don't use pure aluminum. Use an alloy of aluminum with 1% to 2% silicon. And since we said that the silicon is soluble in the aluminum, if you put silicon in the metal itself, so when you sputter, you don't sputter pure aluminum. You sputter an alloy of aluminum and silicon.

Then you think, well, OK, it won't suck up any silicon from the substrate because it's already in equilibrium. It's already got all the silicon it needs, the aluminum eutectic. So it's not a problem. But it's still a bit of an issue is that the silicon itself can actually precipitate out of the aluminum as an example. It's been heated, and then they remove the aluminum film. And you get these little silica precipitates, which can increase the specific contact resistance, especially in N-type silicon.

So it's making an aluminum 1% silicon contact to silicon, which was, again, after pure aluminum, people did the aluminum 1% silicon solution. Was a solution maybe in the 1980s or so, but again, it's still not something that people typically do because you're going to get a higher contact resistance. And you can still get spiking. It's not a perfect solution. But that was one solution that people did use at one point.

The way people go today is to form what they call a barrier layer. And a barrier layer does exactly what the name says. It forms a barrier between this aluminum and the silicon. The barrier has a very low specific contact resistivity. So it still makes a good contact, but it doesn't allow the aluminum or the silicon to talk to each other. So you cannot get this void formation. You can't get the spiking.

So typical barrier layers might be a thin layer of titanium, say 1,000 angstroms of titanium is often used, or use sputter tie tungsten or maybe tie silicide. So in between the heavily doped silicon, you now have an inter layer of this material that forms a barrier. So this prevents the chemical interdiffusion between the silicon and the aluminum. Generally has low stress. You pick a material that has good adhesion so it doesn't peel off.

And of course, good electrical conductivity and low contact resistance silicon aluminum. So that's to satisfy all of these things. And these materials, titanium and tungsten, have reasonable contact resistivity. They are still not low enough to meet a lot of the ITRS requirements in the future, but today they're good enough. Oh, this is just-- on slide 19, I took from a different textbook. This is a figure from Mayer and Lau. I had mentioned Mayer--Lau textbook at the very beginning of this course. So you have a reference to it.

And this is just an example of in his book how he's saying titanium may work as a sacrificial barrier. May or may not exactly work this way, but it's one potential methodology. You can imagine having a thin layer of titanium with aluminum on top. When you heat the structure, you may get formation of TiAl3 for some time and portion. And then you heat it a little bit longer and all the titanium is consumed.

And then here at this point in D, you can start to get the aluminum incursion. So the idea is you put down enough titanium that you're not going to end up with a spiking problem. And typically, if you want to anneal a typical forming gas anneal, 450 or so, 1,000 angstroms of titanium seems to be adequate barrier layer. But again, it depends on your backend thermal budget.

Slide 20 is just another example saying that depending on the barrier layer that you choose, you have to be careful. A lot of these materials for barrier layers. Ti-tungsten or Ti-nitride are very often-- they're polycrystalline, basically. And what can happen if you're not careful is the aluminum can actually diffuse through the long grain boundaries and can still make its way to the silicon and end up causing problems.

So people often, when they deposit these material, they sometimes sputter them in an ambient that has some impurity. Could be nitrogen, is one of the most common. Sometimes people actually do sputter Ti-tungsten, and then they expose it to air for a period, a few minutes or half an hour. Then they put the aluminum on top. The idea is that these either nitrogen or oxygen impurities are going to end up being very high concentration in these grain boundaries. This is called a stuffed barrier.

By stuffing the grain boundaries, it helps prevent the aluminum from diffusing down in and getting to the silicon. So again, we're going to do a heat treatment. You're going to put aluminum on top here. You need to try to prevent the incursion of the aluminum. So how do you do the sputtering of the barrier layer is actually very important. What ambient you use will determine whether it's a good barrier layer or a poor barrier layer and how long it's going to stand up. So a lot of it's quasi empirical, just testing of what works.

On this page 21, I'm showing a classic process that was developed a number of years ago called the salicide process. And this is something you need to be familiar with. Salicide stands for-- the sal comes from self-aligned silicide process. And you'll see when we go through it what we mean by self-aligned. So we start here with our naked device ready to be contacted. We don't have any contacts. We just have N plus source and drain. And you have a heavily doped polysilicon layer.

We then form our oxide spacers. So you know how to form sidewall spacers now. You deposit a conformal layer of SiO2, say low temperature oxide, and then you etch it back anisotropically. So you end up with these little stringers on the edges, which we call sidewalls. OK, those are critical. So they're going to form sort of a blocking region, which will be critical in the salicide process.

We then deposit metal, M, over everything. So you can do some kind of PBD deposition. And here's your titanium layer that goes everywhere, obviously across the whole wafer. And then you do a magic anneal at the right temperature. And this anneal is such that wherever the metal, the titanium, is contacting the silicon, you form a metal silicide, say TiSi2. So it only forms where it's in contact with silicon, which is right here, this dark region, and on top of the gate.

So on the source, gate, and drains where it's in contact with silicon-- remember, the polysilicon gate will also react-- you have titanium disilicide. Everywhere else you have still remaining some unreacted metal. So it's still sitting there. Now, a key to the processes that you dip it in some solution-- sometimes it's HF, sometimes it's sulfuric-- that removes unreacted metal but that does not etch the silicide itself.

So we've created a material, a titanium di silicide or a metal disilicide, that because of the virtue of its chemical structure, it now stands up to the etch, which will remove the unreacted metal. So the reason it's called self-aligned, you notice I did not have to do any photolithography to pattern this metal. The metal was deposited over the entire chip. It was reacted with the silicon, and then it was just etched off in a blanket etch.

So there's no photoresist step here to pattern this metal. That's why it's self-aligned because wherever there's exposed silicon, you will end up with a metal contact, a line to that exposed silicon, without having to do another lithography step. So this was a really great invention. You would save yourself an alignment tolerance. So it's very good alignment.

You form a low resistance contact to the source drain and to the gate simultaneously. So this kind of a salicide process was a big breakthrough in CMOS technology, say around the 80s or so. Slide 22, I just took from a different text. It has a slightly different type of drawing. Maybe it's easier to understand. They have a little different notation. It's the exact same process. It's the basic salicide process. You have an N plus source and drain. You form your spacers. You react the metal.

It does not react. Hopefully it does not creep up the sidewall, which is a problem. You then selectively remove the unreacted metal form, put your dielectric down everywhere else, and put in your metal contacts. So that's a basic metal contact scheme that people use today. What do people use? Which materials are commonly used? Well, some of them are shown here on slide 23. The first one that was used historically was Ti-silicide. Ti-silicide was very good because it has a low resistivity phase.

And we're talking about maybe a 15 micro ohm centimeter for a particular phase that's formed on a particular type of annealing. There's a little difficulty, though. As people make the polysilicon gate length shorter, it's harder and harder to get this particular phase to form. This is called the narrow line effect or the narrow width effect. Usually the narrow line effect. Ti-silicide also has a tendency to agglomerate when it's very thin at higher annealing temperature. So it's not that thermally stable.

So for a number of reasons, particularly the narrow line effect, industry moved a number of years ago primarily from Ti-silicide, although some people may still use it, to cobalt bisilicide. It does not have that narrow line effect, but it gives you a slightly higher resistivity. It's a little more sensitive to surface contaminants. The beauty of Ti is, again, the titanium eats its way through things, through oxides, sort of reduce oxides.

So cobalt, you have to have a really clean interface. Cobalt has a little less lateral encroachment over the oxide spacer. And I'll show pictures of that. So cobalt was used and is still used in some processes. The latest silicide that people are exploring in research and development-- and at some point will probably be in production-- is nickel silicide.

Nickel silicide has the lowest silicon consumption. In order to do this, you need to react and you need to consume some of the silicon. And you don't want to consume very much of that shallow junction. So nickel is good because of that. It can be formed at very low temperatures. And it doesn't have very bad narrow line effect for silicide in the gate.

Big problem with nickel is you have to be careful of your thermal budget. And watch out for nickel. Nickel is a very fast diffuser, remember, in silicon. Nickel is also a deep level. So nickel contamination of equipment and of the wafers is an issue. And people need to deal with this. But nickel silicide is becoming more and more prevalent in research and development.

I mentioned there was a problem with this encroachment over the spacer. And this is illustrated on slide 24. On the left-hand side, I'm showing the case of a cobalt disilicide formation where it's the metal that diffuses. So you can imagine that you have this metal that's deposited everywhere. And you're going to form a reaction between the metal and the silicon. The question is, how does the reaction occur? Does the metal diffuse in and meet the silicon down at this interface or vice versa?

Well, it depends on the particular type of silicide. In the case of cobalt disilicide, the metal diffuses through the silicide into this interface here and then reacts. So you get silicide reaction or formation at the bottom here. So you're less likely to have creep up. In the case of titanium disilicide, in fact, what happens is the silicon is what's the fast diffuser through the Ti-silicide. It diffuses up and meets the metal.

So if you do it long enough and if you have a short enough spacer, you can see the disilicide can creep and grow up from the source and drain down from the gate. And eventually you may actually bridge. And then you have a big problem because then your gate is electrically shorted to your source and drain. And then your device is dead.

So titanium disilicide has a little more tendency to do this than the cobalt disilicide. This is part of the reason people went to cobalt despite, the fact, it has a slightly higher resistivity. Slide 25 is just kind of an illustration of some of the kinetics of what's happening when you have titanium on silicon that's reacting.

Here we have a certain thickness of Ti-silicide that's already formed. People do model this. There are silicide models in SUPREM-IV. Not necessarily perfectly accurate, but what people model is the diffusion of the silicon from the bulk through the titanium disilicide up here to this top interface and form a new layer of titanium disilicide. So this is what might happen in an inert ambient.

Now, if you do the anneal in a nitriding ambient, so in an ambient that has nitrogen, at the same time, you're forming Ti-silicide at this interface, you can be reacting. The titanium is fairly reactive. You can be reacting in form Ti-nitrite up here at the top interface. So this would be in a nitrogen ambient, you can get simultaneous formation of Ti-silicide down here and then Ti-nitride on top of the titanium. Titanium is very reactive, so it's not unusual to try to form a Ti-nitride to prevent it from oxidizing.

Slide 26 shows you some of the different uses of silicide in silicon technology. And I've actually updated it. I added a fourth one. So silicide are used, as you can see, to strap the poly. And the reason you do that is you're trying to reduce the resistance of the gate. So you form a little silicide during the salicide process. Strap the junctions. Well, what does that mean? Again, you're going to reduce the sheet resistance of this junction by forming a thin layer. It also forms a barrier layer, as we mentioned.

It can be used as a local interconnect. Here the silicide has actually been formed above this oxide. So some silicon must have been deposited prior to that. And finally, it can be used as a gate material. And at the end, I've added a few slides that show you that people are actually using silicides as metal gates. This is a table here shown on slide 26 of some of the common silicides and some of their important properties.

So here you notice for Ti-silicide there are two phases. There is a phase called the C49 phase. That forms at low temperatures. So if you were anneal the titanium and react it with silicon, say between 500 to 700, say around 600-- you get a resistivity of about 60 to 70 micro ohm centimeter. That's too high for the film resistivity for most applications.

So people typically do a low temperature anneal to form the C49 phase. They then etch off all the unreacted titanium. They put it back in the RTA and they pop it up to 800 or 900 to do a high temperature anneal to form the C54 phase, which has a much lower sheet resistance. Nice thing about Ti-silicide, it's reasonably stable, maybe up to about 800, 900 degrees, something like that.

It does consume a fair amount of silicon. So it consumes 2.2 or 2.3 or so nanometers of silicon per every nanometer of metal that's consumed. But depending on how thick of a silicide you're trying to form, you may eat into your junction. If you have a very shallow junction, you have to watch out. You don't want to eat in too far. So cobalt disilicide is another example here. You can see it's got a somewhat higher resistivity. This is the temperature at which it's formed. It consumes a little bit less than the case of titanium.

The modern silicide that I mentioned here is this nickel silicide, NiSi. It's formed at very low temperatures, 450 or 500. And it has a low consumption, only 1.8 nanometers of silicon consumed per nanometer of metal. One thing about this, though, if you look at the stability temperature-- look at this column here called stable on silicon. Be very careful. People have listed nickel silicide as being stable up to about 650. It may not even be stable that high. Maybe 600 or so.

So unlike some of these others, which you can heat up Ti-silicide or platinum, you can heat up to 800 or so. You cannot do that with nickel silicide. So if you're using a nickel silicide process, you're going to be limited to a lower backend temperature. But you can get reasonably low sheet resistance. Slide 27, I just want to show you an example. And we'll go through these calculations at the end of the lecture. But I want to go on and do the gate material work first.

But what this is is a cross section of a MOSFET. And we've seen this cross-section a number of times now in our class. We now have enough ammunition that we can actually go sit down and calculate all these resistances or estimate them given a geometry of a particular device. So the resistance that we really, when we scale the channel length, what we're really trying to scale is the channel resistance, R chan.

OK, that's fine. We make the channel shorter, we can reduce that resistance. But the problem is, what happens to-- if we don't do something to all these other parasitic resistances, the net resistance of the device is really not going to go down by very much. So there are three resistances we really want to be able to think about. One is this little resistor right here, which is the resistance associated with the source drain extension. It has a certain sheet resistance. It's doped to a certain level. It has a certain junction depth. It's usually very shallow.

So this resistance of the source drain extension is one resistance we have to calculate. The second one is the resistance of this region here, R, that-- and generally, this region has been silicided. You notice the source drain extension is under the spacer by definition, so it's not silicided. So the current has to flow strictly through the silicon. So it will have a certain higher resistivity.

In this case, this region is blue. It's been silicided. So it's going to have a lower resistivity. It's got a metal on top of it. So that's the second resistor we need to calculate. The third one is the resistance of the contact itself. So it's the contact resistance from the current flowing through the silicon up into the metal. And that would calculate if we know the area and the specific contact resistivity.

So all three of these-- the source drain extension resistance, the silicide resistance, and the contact resistance-- all three are going to add up and give us different contributions depending on the geometry. So I want to go through a calculation of this. I'm going to hold off for now because I want to make sure we cover the novel gate material aspects. But we'll come back to this. And you should be able to sit down at this point and calculate all three of these things and see what they look like for modern technology.

Let's go on for now to slide-- there's one more-- by the way, I sort of didn't tell you the whole truth. I said there are three resistors, this one, this one, and this one. It's not quite that simple. These back-of-the-envelope calculations are terrific because you can do them in five minutes in class or in 10 minutes in your office or whatever. There is another resistance called the spreading resistance. And that's actually pictured-- it cannot be calculated by hand, but it's actually pictured on page 28 here, just to give you an idea.

This first was discussed by Ming and Lynch back in the late 80s. But what it is is you have this channel region that's very, very thin. Typically the channel where the electrons are the holes are traversing across and they're going across the gate length, that's maybe only 30 angstroms thick. So that's where the current is all flowing in the channel. It's a very, very high density of carriers.

The current then spreads out as it goes into the source drain extension. This is such an old paper. It's before they had source drain extensions. But you can imagine it then spreads out here into this region over certain distance. And how far it spreads out and exactly how it spreads out depends on the doping and the geometry of the structure. It's a very two-dimensional problem. You can't calculate it by hand.

So this that's called the RSP here in this little diagram where he wrote down the different resistances. So our spreading is one that generally has to be computed either by a two-dimensional simulator or you have to get it out of test devices or some kind of test structures. It's not something you can simply calculate by hand. But it is a major contributor-- it can be a major contribution to the total series resistance. So it's something we need to be concerned about.

But I just want to point out, so those three resistances we can calculate by hand, that gives you the minimum series resistance of the device. There will always be a little extra, which is spreading resistance, which you have to simulate in a two-dimensional simulator. That's how we make contacts.

I just wanted to go on and spend most of the remaining time of the lecture and talk about something else that's come up in the last three or four years, which is related directly to the contacts because it turns out we can also use silicides not only for the contacts, but we can also fully silicide the gate and use it for the gate.

And I think I showed you this slide last time, page 29. It was taken directly from the last lecture. I just wanted to remind you that there are some new gate materials that are coming along. This is the more classical older technology, the traditional technology. This is a photo from an Intel device. It's a somewhat older photo now. This was published back in about five years ago in 2000.

And what it shows is what we've just exactly been talking about. Here's the polysilicon gate. These are the sidewall spacers. You know how to form them. And this is the silicide in the source. You can see it looks very dark because it's been silicide. And the silicide in the drain. And the silicide that's been formed on the gate. So that was the salicide processed just by the process exactly what we just talked about. That's the classical type of structure.

We talked last time, though, that polysilicon can only be doped so high. Maybe you can get it to 10 to the 20 or mid 10 to the 20. But that's not enough carriers to prevent depletion of the polysilicon at very high gate biases. And you get a polysilicon depletion effect. So people are concerned. They want to get rid of poly as the gate material. It doesn't have enough carriers. Polysilicon may only have maybe 5 times 10 to the 20 electrons per cubic centimeter. They would like to replace the poly with a gate, with a metal gate.

A metal has 10 to the 23rd. So it's three orders of magnitude higher carriers. So it's not going to get depleted. So they want to remove the semiconductor from the gate. But poly is an extremely easy material to integrate. People know how to etch it. It's not reactive with SiO_2. That's not the case for metals. So here's an example.

Last time, in fact, I went through a process called the replacement gate process where we showed they actually used poly. They dug it out at the very end of the process. They etched out the poly, and they replaced it with a high-K material like hafnium dioxide and a Ti-nitride gate. So this was an example of something published about six months ago. So people are thinking of replacing poly, but it's got a lot of complex process integration.

An easier process integration that people are considering is rather than digging the poly out-- and whenever you etch out the poly, you always expose the very sensitive oxide silicon interface. So that's typically, from an integration point of view, digging out the poly is not that desirable. They said, well, why don't we just take the polysilicon and fully react it with a metal and convert it from polysilicon into a silicide? Put enough metal there, instead of, in this case, they only put a very thin amount of metal, so it only reacts to form maybe several hundred angstroms of silicide.

People are saying, all right, we'll put enough metal on top of that, make it really thick metal, so that it reacts-- and it reacted at a high enough temperature for long enough time that the metal reaction takes place throughout the entire poly. And you get a silicide all the way down to the gate interface. And here's an example of nickel silicide, fully silicided gate that was published about six months ago by annealing at 450 degrees and using a thick enough layer of nickel.

So not only are silicides being used in the source and drain, but they're also being used fully in the gate. Now, this is different from the salicide process. One reason would be is the gate's usually reasonably tall. The gate may be 1,000 angstroms tall, something like that. So you need to put enough metal down that you can silicide all the way through 1,000 angstroms.

Now, on the source and drain, you don't want to be siliciding down 1,000 angstroms because I said the junction depth is about 1,000 angstroms. So if you're going to do a FUSI process for the gate, you actually need to have two different siliciding steps. You need to have a step when you would silicide with a very thin amount of metal with silicide the source and drain. You need to cover them and protect them then, and then open up the gate and use a thick amount of metal to try to fully silicide the gate.

So it's a little-- it's still tricky. It's not as simple as using the old fashioned salicide process where the thickness that you silicide on the gate was the same as the thickness you went in the source and drain. It's different from that. But it's not quite as complicated as completely replacing the gate in the etch-out process. In fact, I've got here-- showing here you on slide 30 some fairly recent results from the last couple of years of how people have made nickel silicide FUSI. FUSI is the acronym for fully silicided gate.

How they've made this-- and this particular article came out two years ago from AMD. And what they're showing here is a method to form this. And so they do the standard transistor fabrication or flow here. So this is their poly gate. And here they've actually formed a little bit of silicide in the source and drain and a little bit on top of the gate. So they did the standard salicide process at this point.

Now they have to do a plan arising step. So to get from here to here, what they had to do is they had to put a metal and maybe some other harder material-- no, sorry. They had to put, yeah, contact material, and then maybe a dielectric over the entire thing, and then CNP it down so it's planarized because you know they're going to get deposition everywhere.

So they planarized it. And they expose, then-- in the plane rising process, they expose the polysilicon gate. And they can then put down metal here, and a thick layer of metal, and react it all the way to the interface. So here they have this very thick region where the entire gate has been silicided. And you notice during this last process here on the right, there's no siliciding happening down in the source and drain because that's all been protected during this process.

So it's certainly easier than the replacement gate scheme that showed last time. And it also avoids the PVD damage to the gate oxide. If I'm going to etch this gate out and I have to deposit a metal by PVD, your PVD processes can be very energetic. You often use some kind of sputtering where you have ions. And you can cause damage to the gate oxide. This doesn't have that at all because when you deposit the nickel, the nickel is going on up here. And then it diffuses in a siliciding process to get down to the bottom.

Here's an example of an electron micrograph of one of those devices. This gate has been fully silicided. So this is all nickel silicide. This very thin white layer is the gate oxide, very thin. And this layer here is the silicon channel. This is an SOI device. So this is a silicon channel. And this is the buried insulator, the buried oxide. And these are sidewall spacer layers.

This is just another image here on slide 31, another image of that gate. They've made a very short channel device, only something between 35 and 40 nanometers. It's a very ultra-thin body device, or reasonably thin body device. This is an SOI, it's only 25 nanometers thick. And they have oxide and nitride spacers on either side. The point they were trying to make with this was that the nickel did not diffuse in to the silicon.

Of course, it's a little hard to tell from this. The nice thing about it, if you zoom in at this point right here and you zoom in on the gate oxide-- so this is the gate dielectric, which is shown here by this amorphous material. It's about 2.1 nanometers thick. That's their gate dielectric. This is the metal gate now it's nickel silicide. It went down and it stopped at the gate stopped, reacting at the gate dielectric. And this is the silicon channel region.

It's reasonably smooth. People were concerned that the nickel might diffuse in and react with the oxide. But according to this particular temperature and time that they did, they get a reasonably smooth interface. Remember, you can care about that because it's going to be the channel. The carriers are going to be flowing right in the silicon underneath this. So you don't want to get any nickel into that channel. This is, on slide 32, that same paper.

They also did some Auger analysis. Remember, we talked about Auger spectroscopy as being a means of measuring the composition by going through a device. So this is a vertical scan through the device. So what they're actually doing is going back here, they're the Auger experiment, and they're sputtering right through here. And through the center of the gate, they're looking at what comes off, what they see, what Auger electrons as they sputter.

So here we see atomic concentration in percent as a function of depth, essentially. So this is the surface over here. And you can see that the red line is the nickel. The blue is the silicon. This is the thickness of the gate from the surface here to this point here, to the gate oxide. And what you see is there's quite a bit of silicidation. This is mostly nickel silicide.

It looks like here there's a little bit of silicon-- or here you have the silicon in the SiO_2. And then this is the oxide beneath. So you just get some idea. This looks like it's fairly stoichiometric, almost 1 to 1 silicon to nickel. Maybe a little bit nickel-rich. Again, using the types of techniques we talked about in our characterization lectures.

This is a different paper. This is a paper-- that last paper was from IEDM 2002 applied by AMD. IBM the same year also published at the same conference nickel silicide FUSI gates. This time they did them not only on fully depleted SOI. I just showed you a silicon insulator device. They also did it on a device, which is called a FinFET. We haven't had any time, really, to go into these new types of devices, but there are silicon MOSFETs these days that are not being made in a planar structure.

There's a device that emits very non-planar called the FinFET, when in fact, people form a fin. And the channel of the device is actually into the page. So it's into the board. And the electrons actually flow along the side walls. Electrons actually flow right along here. And you have a gate on the right side and a gate on the left side. In fact, the gate kind of wraps around.

So this whole thing ends up being a channel. And the source and drain are into the board. This is a device that has some advantages, although it's a little tricky to make. It's a double gate device. You end up getting two channels in this thin silicon film. So they may have FinFET and they demonstrated they could make a nickel silicide gate going all the way around that fin. In fact, this is the gate dielectric shown here, this amorphous-looking region with nickel silicide on the outside.

Again, using a FUSI process where they put poly everywhere, put down the appropriate amount of nickel, and then reacted it. And the reaction stopped right when it hit the gate dielectric. An interesting thing that was also done in this paper, if you end up wanting to do research in this area, is shown here on slide 34. This is something called gate work function engineering. We hadn't really talked about it, but the work function between the metal-- I think we may have talked about threshold voltage control.

The work function, which is a property of the metal material, to a certain extent. And the silicon, that determines the threshold voltage of the transistor, the voltage at which the transistor turns on. So that's a very important property. People have been using N plus poly and P plus poly on N FETs and P FETs for years because it has the right work function.

The problem with poly, as we've mentioned, is it doesn't have enough carries. It tends to deplete. So people want to use metals. The problem with metals is, what work function do they have? Well, they have the work function, typically, of the metal, which is a property of the metal. And there are only so many metals in the world. There are only so many silicides in the world.

But what IBM has shown in this particular paper is that they can adjust slightly the effective work function of that gate stack, so the effect of work function in between this metal gate and the silicon material. They said they can adjust it to a certain extent by how much they dope the polysilicon prior to the silicide reaction. And this is just a table that I took right out of that paper.

For the case of nickel silicide, what they found is, depending exactly on how much doping they put into the gate to begin with-- say if they don't dope it, or if they put in 1e20, 2e20, or 4e20, the effect of work function that they measure with respect to the silicon conduction band is shown here. It can be varied, maybe by about 100 millivolts. Maybe a little more. Something like that. I guess the nickel silicide was 0. So it's a little more, maybe a couple hundred millivolts, it can be varied by whether you put doping in or not and by how much doping.

And this is important because you need this flexibility in being able to adjust your threshold voltage. So FUSI gates are also interesting, not just because it's a little easier to integrate the processing. The frontend processing is what we're all familiar with. Everyone knows how to etch a poly gate and make a good poly gate, and you can convert it to silicide. But depending on how you dope it, you may be able to control the PT or adjust the PT a little bit. It's still a very tricky process.

The reaction temperature or the thermal stability temperature is reasonably low. So you cannot take these and then take them to a backend process that is too hot. So they're very much a research. It was only published by IBM a couple of years ago. It's certainly not ready necessarily for manufacturing right now. Perhaps in the near future, but just to give you an idea of the types of things that people are concerned about with siliciding.

Before I go through the summary, let's take a few minutes, because we have a couple of minutes right now. I just want to go through with you very briefly, if we back up a little bit here onto slide number 27. And you can sit down and do this back-of-the-envelope calculation yourself over the next couple of days. It's not a homework assignment, but I think it's something you should do-- how to calculate these three resistances. And then we can talk about it and go through it next time.

But what I want you to calculate is you're given everything you need to calculate the contact resistance. You're given the specific contact resistivity rho C. And you're given the area of the contact. So you know exactly how to calculate that. You just do a simple division. And you can calculate this resistor, the contact resistance. You're given the sheet resistance of the silicide, this blue region. You have its sheet resistance, or it's resistivity, I'm sorry-- 15 times 10 to the -6 ohm centimeter. And you know its thickness.

So resistivity divided by thickness, you can get sheet resistance. And then you can figure out how many squares the current has to flow through. So you should be able to calculate this resistance of the silicided regions. And then the little resistor here of the source drain extensions, we are given their resistivity, and again, its thickness.

So we should be able to calculate a sheet resistance, and from that figure out-- so you get the ohms square and the number of squares. And you get a resistor. So one, two, and three resistors corresponding to this contact resistance, the resistance of the current flowing through the sheet, and the resistance flowing through the source drain extension. You'll get three numbers, and you just get an idea of the order of magnitude of those numbers and how they compare.

So if you have time between now and next lecture, do that. We'll go through it at the beginning of the last lecture next time. But just an interesting comparison between those three numbers. OK, so let me finish up by kind of summarizing on this topic. What we talked about today, we need good device contacts. They have certain requirements-- a low specific contact resistance, rho C. Very good thermal stability so when you heat them up, they don't spike or do anything unusual.

In general, you want to get a high doping concentration of a silicon right at the interface. That will tend to lower the contact resistance by inducing quantum mechanical tunneling. However, there is a fundamental limit on how much doping we can activate in silicon. This puts a limit on the contact resistance and on the sheet resistance and on the parasitic resistance. In fact, the contact resistance is really a big problem. If you look in the ITRS, there's a lot of concern about how to lower that.

The so-called self-aligned silicide process, or also abbreviated salicide, has been the mainstay of technology for many, many years now. It has a lot of advantages. It's self-aligned. It reduces the sheet resistance of the deep source drain region. It reduces the gate sheet resistance because you put silicide on the gate. And it also provides a local interconnect layer. So it's been a very good workhorse.

Silicides can also function as a barrier layer to prevent spiking, potentially. Silicides do consume silicon. That's one problem with them. And as you move the metal closer to the junction depletion region, it's a big problem because you tend to get more leakage. In fact, this is why people use deep source drain regions to allow a thick enough region in between the depletion region and the point where the silicide is formed. So you can get a reasonable silicide with low junction leakage.

Because junctions are scaling to be thinner and thinner, especially in fully depleted SOI, people are moving towards nickel silicide because it consumes a lot less silicon per thickness of silicide than, say, Ti-silicide or cobalt silicide. So for fully depleted SOI, nickel silicide has a lot of advantages. And the last thing, which isn't in the summary we just talked about, is that silicides are even being considered to be fully silicide purpose for the gate. And that's something that's in research right now, you may see in production over the next three or four years or so.

So that's all I have for this lecture. Take a look at that simple calculation example by hand. And then next time we'll meet for the final lecture. And again, I announced it at the beginning. The people who are going to be speaking in the order, it's also posted on the website. But if you have any questions about your oral report or whatever, please get back to me. OK, thanks.