JUDY HOYT: Where we were-- where we are. I've been out last week, I went to a conference. So you had two-- your first two lectures on dopant diffusion given by the TA. And here we are, Lecture Number 10, we're going to talk about some more advanced models for dopant diffusion. And last week, you also had your homeworks Number 3 handed out. So hopefully, you've all started on that. You're using SUPREM for your homework.

Get used to running SUPREM because that homework is due this Thursday. And then next Tuesday, you'll have another homework, homework Number 4 going out, which is also going to use the process SUPREM IV, and more sophisticated models. OK. So as I mentioned, today's lecture is essentially the third lecture out of a number that are on dopant diffusion and profile measurement.

Last couple of lectures, Maggie talked about the relatively, quote, unquote, "simple cases of constant diffusivity." Diffusivity does not change in space, basically, throughout the sample. That applies when you have a low dopant concentration. We'll talk today about what happens when that assumption is broken, such as, when you're diffusing the well of the CMOS. So it's relatively low dopant concentration, say, 10 to the 18th or less.

And we also saw-- you saw last time that you can design a diffuse layer based on a sheet resistance requirement. As a device engineer, you might be told you need a diffuse layer, such and such a thickness, and it should have so many ohms per square of sheet resistance. You saw how you could use some of the urban curves, or you can use numerical techniques to design a diffuse layer.

Last time, also on Thursday, Maggie introduced a relatively simple but very powerful method for doing numerical solutions. There's only about three or four cases for diffusion that you can solve analytically. And of course, we try to give you some of those in the homework. But most of the time, you have to do it numerically. And this is a very simple technique. It's called the finite difference algorithm.

It's slow, it's not actually what's used in SUPREM IV. SUPREM IV is the-- was more sophisticated than that. But if you were stuck on a desert island and all you had was a computer with you, you could write your own finite difference technique in about three or four minutes, and you could solve a diffusion problem like that numerically, even if you didn't have SUPREM IV.

Maggie also talked about that you need to modify Fick's law for something called electric field effects, so electric fields that are induced. These can enhance the effective diffusivity of a high concentration species. So you have a high concentration of arsenic, let's say diffusing, it can actually, in a low concentration of boron background, its diffusivity can be enhanced by up to a factor of 2 due to this electric field effect. But even more of an enhancement can be obtained, more than a factor of 2, very dramatic diffusion of those species of lower concentration.

So she showed an example with a low boron concentration that was constant at the beginning of the diffuse profile and then diffusing arsenic into it. And then after the diffusion, it actually hadn't moved, even though the profile was flat. That cannot be-- that's non Fickian diffusion. Fick's law would say, once the profile is flat, no diffusion. So that's due to the electric field effect. So that's kind of what we went over-- what you went over last week.

Today, I want to cover three items-- concentration dependent diffusion, because this is really prevalent in Silicon, especially, in making MOSFETs-- segregation interfacial dopant pileup, and starting to look at an atomic scale model of dopant diffusion. OK. Let's go on to slide number 2. And this plot introduces so-called Fermi level effects, or concentration dependent diffusion. And it's a cartoon plot, but it shows concentration on a log scale, on the y-axis, versus depth.

And there are a couple of solutions here. Look at this one starting at the surface concentration of 10 to the 18th going down here in red. This is a complementary error function, just like we learned a couple of lectures ago when you have a constant surface concentration. If you have a higher surface concentration, say, up in mid 10 of the 20, this red curve here, dotted curve, shows what a complementary error function would look like.

Well, what we actually find when we do diffusions of a number of dopants in Silicon, that's not what they look like in high concentrations. They don't look like this red curve. They look more like the blue curve, or the green one. They have a much flatter top, and they have a much more abrupt drop-off. And this tends to occur when the dopant concentration is greater than ni. So n much, much greater than ni. So here's n, the concentration of arsenic-- let's say this is arsenic-- is 2 10 to the 20 up here at this point. And ni is a few 10 to the 18th.

So orders of magnitude larger than ni, you see this that the complementary error function does not yield the correct profile, has more box-like than it would be. And by the way here, this little dashed line, horizontal line, shows you where ni is. So clearly over this portion of the profile, n is much, much greater than ni. So at high dopant concentrations we observe for a lot of the dopants in Silicon that the diffusivity appears to be larger than it is at low concentrations.

And so what this means is that fixed equations have to be solved numerically since, basically what we're finding, what we see, is that diffusivity is no longer a constant throughout the sample. In fact, it depends upon the local dopant concentration at each point. So at each point along this profile from here, here, here, here, and here, the diffusivity is a slightly different number. And in fact, according to this, for the blue curve is proportional to the ratio of n over ni.

And of course, n over ni is changing as I walk down this profile. So there's no way I can do an analytic solution. Every point in space I need to apply a different diffusion coefficient. That's what we mean by not equals to a constant. So in that case, remember this was Fick's second law, when I cannot pull the d out of the partial derivative. So partial d, partial t is equal to the first derivative with respect to x of this product. The diffusivity of the dopant, the effective diffusivity times partial c, partial x.

Where now da effective-- I should write here-- is actually a function of x in this high concentration case. So I cannot pull it out of the derivative. I have to solve it most likely numerically. OK. So that's an introduction to what people see. How do we explain this effect? Well, as you might imagine, people explain this dependence on the Fermi level. Basically, it's the dependence on dopant, or Fermi level, based on concentrations of point defects.

We saw several chapters ago if I move the Fermi level up and down the band gap, so if I increase the dopant above NI, that I change dramatically. In fact, you had a homework problem that was due last week. You saw when you went to a high concentration that the total concentration of vacancies went up because some of the charged vacancies went up. OK. More vacancies, if you need vacancies for diffusion, make sense the diffusivity would go up.

So this-- basically, we saw that charge point defects obey the same statistics as shallow donors and acceptors. In fact, we wrote equations in Chapter 3, we wrote down these equations for the concentration of any charge point defect here. Here's the concentration of vacancies that are single negatively charged. We could write it in terms of the neutral vacancy concentration, which is only a function of temperature times this exponential of e Fermi minus ev minus, OK.

So it's the distance between the Fermi level here and this energy-- defect energy level ev minus, that distance, that whole thing to the kt. So as I move the Fermi level up and down, I can exponentially increase or decrease. And we can write the same kind of equations for interstitials. So as you can imagine then, we're going to use this idea to say that the diffusivity, which depends directly on these concentrations, if you're diffusing with point defects, must then depend on the Fermi level, and use that to explain the concentration dependence.

So I just want to derive now here on-- shown on slide 4 of your handouts an explicit expression for the charge defect concentration. This time in terms of carrier concentration. On the last slide, if we just go back one second, here it's implicit in here. I'm going to show you the carrier concentration is embedded in here. Right now, you don't see it explicitly, but I want to make an expression where we can see the dependence on n over ni directly. And we'll do this for a particular case and then we'll generalize it.

So this is a little bit hard to see, the font size is a little small. But what this is saying is from the last slide we just saw, that the concentration of cv plus is equal to the neutral concentration of vacancy. So the concentration of neutral vacancies times this exponential ev plus minus e Fermi over kt. So that's-- and this is our band diagram. Remember, e sub v here is the valence band position. ec is the conduction band. This-- unfortunately, the notation is a little bit unfortunate. It's awfully similar.

But ev plus is the energy level in the band gap of the single positively charged vacancy, OK, it's ev plus. It's not the valence band. The valence band energy is just E sub b for the valence band energy. The mid-gap position here is called ei, which is shown by this dashed line. The Fermi level in this particular example is given right here by ef. So those are all the relevant energies.

So I want to now take this expression, this ev plus minus e Fermi and expand it out in terms of some other relevant quantities. So this ev plus minus e Fermi, that's actually just some distance between here and here, from this point here to here. So it's equal to this vector, this a value. But you can also write a as just the sum of d plus c minus b. Just add up d plus c, subtract b, that's equal to the energy distance a.

And so you can take each one of these terms, d, c, and b, and write down quantities for them. For example, d, this distance right here, is just ei minus e Fermi, OK? So I can write d like that. d is just the valence band energy ev minus ei, so we've written that. And subtracting off b, b is just ev minus-- ev, the valence band energy minus ev plus. So if you want to look at it in terms of those distances, I think that helps. You can also recognize that mathematically all we really did in this equation is we added and subtracted quantities from the same side. So we haven't changed anything, but we've rearranged the terms in a way that's going to become useful.

Then you can just rearrange this here where you substituted in here an eg over 2 for ei. So I've just rearranged. So it's the same quantity as we talked about. So if I substitute this in, this expression right here, into the argument on the top of the numerator of this exponential, this is what we get. We get something that looks like this. cv plus is the neutral concentration of vacancies-- the concentration of neutral vacancies times this exponential, times another somewhat more complicated looking exponential. So we've just rearranged these energy quantities.

But I've done it in a particular way. And the reason we've done it, we factored out this ei minus ef is because that is directly-- that exponential of that is directly related to n over ni. That's why we wrote it in terms of this expression. Because, in fact, you know that the electron concentration divided by ni is just exponentially dependent on the distance between the Fermi level nei. That's something that we learned from Chapter 1. So that over kt, the exponential of that over kt, that's just n over ni.

So I have something in this equation that looks just like that. So in fact, if I invert this, ni over n is just the ei minus e Fermi over kt, just inverting that. So this expression right here, this exponential, I'm going to be able to put in a term ni over n, replace Boris. In fact, that's exactly what we do here. In this equation that's in the red box, we can write then the concentration of v plus as ni over n times the concentration of neutral vacancies times 1 more exponential factor, this e to the minus, and then this everything in curly brackets.

Well, let's look at the numerator here of what's in curly brackets. This eg over 2, if I expand this out, there's a negative sign in here. I've used this simple mathematical expression here, eg over 2, plus ev plus, minus the valence band energy. Well, just rearranging, that's ev plus minus this quantity here, the valence band energy plus eg over 2. Well, that is just ei, what's in parentheses, right? What's right here is just the definition of the mid-gap point ei, the intrinsic energy.

So this numerator up here in this exponential can be written as ev plus minus ei. Well, then the exponential of that whole thing over kt, that's just equal to, essentially, the concentration of-- intrinsic concentration of v plus, of positively charged vacancies. So basically what happens is, this expression in the rectangular red box shows me that I can write the concentration of v plus under extrinsic conditions. I can write it as the concentration of v plus under intrinsic conditions times ni over n.

So I immediately have factored out the dopant dependence. And so if I substitute in here-- well, it's simple. If I substitute in here n equals ni, this term just goes to 1, and then the concentration is just the intrinsic concentration. But if I were to pump ni way up, or n way up, let's say I make n very large, this number becomes very small and the concentration of v plus goes down. Alternatively, I can make n very small by going to very heavy p type material. So I make n's very small, that pumps up this number-- this whole quantity.

So in p type material, this concentration is going to be very high. It's the exact same equation we just learned that you did in your homework problem where you were manipulating the energy levels and subtracting all these energy differences in the band gap. The only difference is, now we have a more convenient way of remembering it and of writing it in terms of the ratio of the dopant concentration to the intrinsic concentration. And that's very convenient for thinking about these concentration dependent effects.

So if we go on to slide number 5, then basically what I've just written down is that we can write cv plus as ni over n times the concentration under intrinsic conditions. Or if you want to invert this, it's a little bit easier to think it either in terms of p over ni, or n over ni, you can just write ni over n. Well, you know what n is, right? pn product is always equal to ni squared. So you can write n as being equal to ni squared over p.

So if we want to write it in terms of the majority carrier given this expression, then I can substitute for this lower n in the denominator, ni squared over p. And what I end up with is p over ni times that concentration. So clearly, the concentration-- cv plus goes up in heavily p type material. And that's kind of what we know is saying mathematically what we knew intuitively. ev plus is down here. When I make the material very heavily p type, I bring the Fermi level down towards that. And as that distance decreases, I'm going to pump up the concentration of v plus vacancies.

And in fact, it's directly proportional to p over ni. Similarly, for the double positively charged, I can write it as cv plus plus is equal to p over ni squared. It turns out there's a square quantity in there. cv minus is what we just derived is n over ni times the intrinsic concentration of cv minus, and cv double minus, depends similarly with the n over ni squared.

So in general, you could write a general using this derivation, sort of a general rule that the concentration of any vacancy in any charged state r, r could be zero for neutral, it could be plus 1 minus 1, that concentration under extrinsic conditions is just n over ni to the minus r times the concentration of that species under intrinsic conditions.

So all this is saying, that's a generalized term, is that the concentrations of charge point defects and, of course, the total point defect populations increase or decrease directly proportional to n over ni. So with all this mathematical manipulation, what it boils down to physically, if the dopants diffuse using these point defects, the vacant charge vacancies are interstitial. Then the diffusivity of the pair that is of the dopant and the charged vacancy, or interstitial, is proportional to the point defect concentration. Then the total diffusivity will follow these same trends.

So if I-- as the concentration of vacancies goes up because I'm moving the Fermi level up very high, so I get-- let's say I get a lot more of these cv minuses, total concentration of vacancies goes up. If I have a diffuser like antimony that diffuses with vacancies, then you expect its diffusivity to be enhanced because it has more vacancies around it to diffuse with. And it should be enhanced according to this n over ni expression. So the higher I make n over ni, the more of these vacancies, the more the diffusion coefficient should go up. That's the general argument.

So what evidence do we have of this? Let me show you some experimental data which indicates, although it doesn't tell you that there are vacancies or interstitials involved, but at least indicates these dependencies make some sense-- the Fermi level dependence, that is. And here's some experiments that are called ISO concentration experiments-- and you'll see in a moment why we call them ISO concentration-- indicating the dependence of the diffusion coefficient on the concentration.

So you might have boron 10 diffusing in a boron 11 background, for example. And so just-- let's take a look at what we mean by this. This is an ISO concentration diffusion experiment. So if we were to plot concentration of a species as a function of depth, what we do is we put two species in the sample. The first one is a background species, which I've shown here by this orange box, or this constant concentration profile in this region.

So this could be my B11, so it's boron 11. It sets the Fermi level to a constant value in this one region, right, because it's a high concentration of dopant that's p-type, so it defines p over ni. And then underneath that at a smaller concentration, I could put boron 11, which is the dopant that I want to study. I want to study its diffusion coefficient. So I do a diffusion experiment where I start with an initial profile, a Gaussian profile of boron, I let it diffuse. All of it in a background concentration of high concentration boron 10.

OK. I measured the diffusivity. Now, I go again and I take another sample. And this time I put a higher concentration of B10, again, constant in space but higher than it was in the previous sample. And what you see is that the diffusion of the dopant of low concentration dopant is more enhanced now. And you keep doing this ISO concentration, you keep putting in this box like profile as the background at higher and higher concentrations, and you measure each sample and you measure its diffusivity.

So the nice thing about that is the profiles remain Gaussian because in space, the diffusion of the dopant in any one of these regions is a constant now, as long as you stay inside the orange box. It's just higher than it would be in the absence of the background dopant. So it makes analysis of the profiles a lot easier. You can still get Gaussian diffusion. So people have done these type of background concentration studies where they move the Fermi level up and down with another species.

In fact, here is some data I took from a literature on slide number 7 from Marc Law's work published back in 1993-- so it's about 10 years old now-- where they did just that. They had a background concentration of anti-dopant that they created with, say, arsenic. And then they looked at diffusion of phosphorus underneath it. The Fermi level was constant in each one of these samples. They got this triangle, and they got, say, the electron concentration was 10 of the 19th here, going all the way up to low 10 of the 20.

And then they extracted the diffusivity of the phosphorus as a function of the background electron concentration. And in fact, you see this is the experimental values, diffusivity going up very rapidly. When you get above about, oh, say, mid 10 to the 18th, this was at 1,000 degrees. Well, what is ni equal to at 1,000? Maybe you remember from the homework roughly? About mid 10 to the 18th, in that range.

So you see, that's exactly what's happening, the passivity is a constant. This was the calculated. This is the calculated theoretical line where they use this type of expression. So the effect of diffusivity was given by a constant, d0 plus d minus times n over ni, plus d double minus times n over ni squared. So there's three terms here.

And what this equation tells you is, well, if I plug in n equals ni, then it becomes a constant, OK? At ni or below, this whole thing goes to a constant number. And so if you go below ni, indeed, the diffusivity is approaching a constant here. It just looks like it's about 1 and 1/2 times 10 to the minus 14th. As I crank up n over ni, this time right here starts to take over and over ni. And you see an increase. And I crank up n over and ni even higher, see here I get a factor of 10 or 20. This n over ni squared kicks in and, indeed, you see this square dependence.

So this is based on experimental data, and it fits this type of empirical fit quite well. Now, what are these numbers? And what is the meaning of d zero, d minus, or d double minus? Well, we don't really know the meaning. But what we assign the meaning to is, we presume that this coefficient d minus has something to do with the diffusivity of the pair of the dopant with the single negatively charged point defect, whatever it should be. So it could be that the pair of the phosphorus pairing with d minus.

So that diffusivity value has to do with that pair diffusivity. This term here has to do with the pairing of the dopant with v double minus. Or if you believe in interstitials, the I double minus, whichever. These experiments don't tell you whether it's a dop-- whether it's a vacancy or interstitial enhancing it. All it says is that at some point defect it has-- the concentration of that point defect depends on the Fermi level because it's charged.

So as I move the Fermi level up by increasing n over ni, the diffusivity overall goes up according to n over ni. That's basically what it tells us. So those are the types-- one type of experiment people have used to observe these so-called Fermi level effects. So what we do based on this experiment showing on slide number 8, we write the diffusivity in terms of the local carrier concentration at each point in space in the sample.

So here's an example. Equation 2 shows you the case for an n type dopant, a convenient way to write it as the effective diffusivity of a. a could be the arsenic, or phosphorous, or whatever, is a constant d0, plus d minus times n over ni, plus d double minus n over ni squared. So we write that type of expression. And we write a very much analogous type of expression for p type dopants. But now, the dependency is on p over ni. So presumably, this term here, this d plus, refers to the diffusion of the dopant, maybe boron, with a single positively charged vacancy, for example.

So as I increase p over ni, that v concentration goes way up and so does the effective diffusivity. So again, we talked about what each one of these d's corresponds to. So again, if I'm under intrinsic conditions and I have an n type dopant where intrinsic means p equals n equals ni, that's the definition of intrinsic. But I just substitute in here for n, ni, and you see that the diffusivity is indeed a constant. It's the sum of three numbers, but it's a constant number. It just depends-- doesn't depend on the local concentration when you're intrinsic.

And each individual diffusivity, this d to the r power, or if you want to call it-- it's not really to a power, it's just a symbol, a superscripted symbol that tells you the charge of the point defect, each one of these diffusivities is exponentially activated. You can write it in Arrhenius type fashion. So d is equal to d dot zero. This is using the SUPREM-- you should learn this because you'll use it in your homework-- the SUPREM sort of notation. Any diffusivity is d dot zero. That's the pre-exponential times e to the minus d dot e. That's the activation energy divided by kt. So each one of these terms has this exponentially activated.

So sometimes this equation, equation number two here, is rewritten in a slightly different fashion. People like to talk about these parameters beta and gamma. Beta is defined as the ratio of d minus to d zero. So it's that ratio. And gamma is defined as the ratio of d double minus to d zero. So when I make that definition, then I can rewrite equation 2 in this fashion.

The effective diffusivity is da star, where this is under intrinsic conditions, and you factor that out, and it's 1 plus beta n over ni plus gamma n over ni squared divided by 1 plus beta plus gamma. So it's just another mathematical formulation. People sometimes like to think about these beta and gamma terms, the ratios of the diffusivities, rather than the absolute numbers.

For a p type dopant in this equation, you would just replace the n by p, basically, for a p type dopant. And then the beta and gamma are redefined according to the appropriate positively charged diffusivities. So again, this is the, quote, unquote, "Fermi model." When you run SUPREM IV, as you'll do for your next homework set, and you use a Fermi model, you're invoking this type of concentration dependence on the diffusivity for the dopant.

So if we go on to slide 11, just to show you, I've taken from Tables 7-5 in the text. These are the quantities. This is the way we write the diffusivity, and these are the quantities that are in this equation. There's the pre-exponential factor, d dot zero, and then there's the activation energy. So just here are some examples. The first two rows refer to the d zero term, both the pre-exponential and the activation energy. And the last row, a couple of rows, refer to the double minus term.

So just from looking at this chart, let's say we take the case of arsenic. And these are the numbers I took out of your text, and some of these are also used in SUPREM. Based on looking at this chart, people have fit data to arsenic. And what can you say about the different types of point defects that arsenic supposedly diffuses with? What are they based on-- if there's no number in the chart, then there's nothing been observed for that. So arsenic diffuses with what types of point defects?

Single negatively charged and neutral, just because there's nothing else filled in in the chart. So people have observed for arsenic, primarily, an n over ni dependence. For phosphorus, how about phosphorus? What is it diffused with based on this chart? Got neutral, single, and double negatively charged. So there's three terms. And I showed you that when we saw Mark Law's data. He had those three terms. He fit the arsenic-- the phosphorus data back on slide 7. He fit this to a three term type of model. So these three terms come from experiments like the one that he did.

OK. So these are the Fermi models for extrinsic diffusion. So when n is greater than ni, or when p is greater than ni, we use these types of equations. And this is what's used in your process simulator SUPREM IV. OK. Let's go on to page number 10. So on slide number 10, I just want to go over a little example just to give you a feel for how these numbers work. An example asks you to calculate the effective diffusion coefficient at 1,000 degrees for two different box-shaped arsenic profiles. So they're going to give you an easy profile. It'll be a box like thing.

And there are two different ones. One is doped at 10 to the 18th, and the other profile is doped at 10 to the 20. So the first thing you need to do when you're calculating any of these is figure out for the temperature you're at, what is ni, because everything varies. If you're less than ni, ni or below, then you have a constant diffusivity and you just plug in for n, n equals ni, right? If you're above ni, then the n is equal to the-- essentially, whatever the dopant concentration is because it's being controlled extrinsically by the dopant concentration. The carriers come primarily from the dopant.

So in this example, we calculate ni to be about 710 to the 18th. So if I have a dopant profile that's at 1 E18, well, that's much less than ni. Then what is n? n is equal to ni, right, the intrinsic carrier concentration. Because that's just due to the thermal activation of the breaking of bonds. So if you just substitute in here d0, which we got from the prior-- this is the numbers we get from the prior table on the prior slide, and d minus. And again, n over ni is 1, so this term is just being multiplied by 1. There's a 1 here in front of it. Add those up and you get a number that's about four-- 1.4 times 10 to the minus 15th centimeters squared per second. Again, we should be aware of the units.

So you say, OK, when I use the table and I use the two term model, this is what I get. Now, a sanity check on that value might be when we were talking earlier in Chapter 7, we talked about intrinsic diffusion, and we didn't give a two term model. We just gave a constant diffusivity number, which was obtained by fitting a single activation energy to those expressions. So the question is, when I use this two term model, how does my number compare to the case when I use the single term model that's shown in Table 7-3?

And in Table 7-3, you were given this number here, this exponential dependence. The activation energy, if we want to average out these activation energies, you were given was 3.99, and the pre-exponential was 9.17. When you calculate that out at 1,000, you get about 1.48 times 10 to the minus 15th, very, very close. So again, at a sanity check, this two term model is not screwing up your intrinsic diffusion coefficient. You're getting exactly what you would have gotten had you gone back to the simpler table and the simpler calculation in Table 7-3.

But now if we do the next part, how about for the case where the profile is still 1 E20. Well, at 1 E20, what do you know? Well, you know n is much, much greater than ni. So n over ni is large, that means the concentration of single negatively charged point defect, cv minus goes way up. So the diffusivity should be enhanced. So now solved, right, at that same equation, here's the first term. The d zero term doesn't change, right, because it's not multiplied by anything.

This is the d minus term. But what it's multiplied by here is n over ni, which is a big number, 10 to the 20 over 710 to the 18th. That's more than an order of magnitude. That's two orders of magnitude. So you're really pumping up this term. So now what do you get? 1.6 times 10 to the minus 14th. So the highly doped layer has a 10-fold higher diffusivity diffusion coefficient in the extrinsic material. So it'll be diffusing with a diffusivity that's 10 times greater.

And that's exactly how these calculations go. Now, if you have the concentration changing in space at every single point along the profile, the computer has to keep track of the diffusivity value at every single point and apply that correct diffusivity value when it's doing the calculation of a diffusion profile. OK. So let's go on to slide 12. So basically, the consequence of this from a practical point of view is that the profiles are very steep. They have-- they're flat topped and they fall off rapidly.

And why is that? Well, if I'm working-- walking my way along this profile, first of all, I'm much, much greater than ni by over a factor of 10. And the diffusivity here is-- it's a large number. So that means it's going to-- when you're high, when you're above and over ni, you're going to be diffusing very fast. So these guys get over to the right very quickly. But as I start to go to-- my concentration starts to drop at the diffusion front, when I get close to n equals ni, the diffusivity is dropping like a rocket, right? Because n over ni is going down as I move down this profile. In fact, n over ni is only 1 here.

So what's happening is the diffusivity is falling off very rapidly at the diffusion front. So it tends to make a very sharp, abrupt profile there, because as you walk down here, your diffusion coefficient is going way down, it's slowing down. So you get these box-like profiles. A box-like profile originating from-- and again, this assumption was that you have a constant source surface concentration of arsenic, or whatever so that the surface concentration is about 5, 10 to the 20. Instead of getting the complementary error function, which is here, you get a more box-like profile. And that's exactly because of the concentration dependence to the diffusivity.

And people observe-- you observe these all the time when you do arsenic source strains. It's never-- it never looks like a complementary error function. It's always very box like, and that's why. Besides source drains, actually, you can see this-- we always talk about Moss beds. But if you're doing bipolar technology making NPN structures, you see the exact-- the same issues.

And here's an example. On the left, I'm showing a starting structure for an NPN bipolar transistor. So this is a plot on a semi-log scale of concentration as a function of distance into the device. And what it is is that the collector down here is lightly doped around 10 of the 16th n-type with phosphorus. Here's the base. We're assuming the base was grown by epitaxial crystal growth. And we'll talk later in the class what that is in the course.

But it has a very abrupt, constant doping profile. Like this is about 1,000 angstroms. And then you deposit a layer of n plus polysilicon on top, and that gives you this high arsenic concentration. Now, what you're interested in is annealing this to drive in the arsenic a little ways into the epi, into the base for 1,000 degrees for 30 minutes. And you want to see what the profile looks like. Well, on the right, I'm showing the SUPREM simulation that includes both the electric field effects that Maggie talked about last time, as well as this concentration, or Fermi level effect.

And you can see what that looks like. Here is, again, concentration versus distance. And look at the arsenic profile. Here it is in the poly relatively constant. It is indeed very box like. It doesn't look like a complementary error function. It's almost constant, and then shuts off very quickly right here. That's due to the concentration dependent diffusivity. Here's the boron profile. It kind of Gaussian-ish, sort of, but it has this little dip in it right here, this divot, right where the arsenic-- right where the pn junction, the metallurgical junction happens.

And in fact, this little divot is due to the electric field effect, which significantly impacts the profile, the boron profile, near the junction. Remember, the electric field here is being generated by-- mostly by this rapidly decreasing arsenic concentration, and the boron feels that and gets-- its diffusion gets modified. There's no way you would calculate this by hand, you would come up with anything like that. But this is exactly what is simulated in the numerical simulator. And this is what people measure by SIMS and things.

So you really need numerical simulation to accurately model modern devices bipolar, be they MOSFETs or any other type of device because of these high concentration effects. OK. So let me go on to talk-- so we've talked about the electric field effect and the high concentration and what happens. There's another effect when we get to an interface that we're going to need to take into account that's going to be important in determining practical profiles. And this is called segregation.

We know that dopants have different solubilities in different materials. So let's say I have a dopant and it's coming up against an oxide layer. It's in Silicon and there's an oxide layer right next door, next to it. It could have two layers right next door, could be silicon and nitride, or whatever-- silicon nitride. But they have different solubilities and so they tend to redistribute across an interface between two materials until something called the chemical potential is the same.

And basically, the ratio of the equilibrium dopant concentration on each side of the interface is the segregation coefficient. So we've already seen segregation coefficients. We talked about it in Chapter 3 on crystal growth. We defined a segregation coefficient k0 with respect to crystal growth to be the concentration in the solid of the dopant, say, the boron or the arsenic, divided by the concentration in the liquid phase. So there we had an interface. One was the same material, it was silicon. It's just that one case it was liquid, the other case it wasn't solid.

We can do this exact same thing, we can define a segregation coefficient, in general, between two materials, material A and material B in this segregation coefficient k and k0, or it may be-- the subscript may indicate which two materials may be from silicon to silicon dioxide, whatever, is in general-- is the concentration in material B divided by the concentration of material A, just the ratio. So given long enough time, if you were to put some dopant in this material and let it-- and heat it up and let it move around, it's going to arrange itself so that the ratio of the concentrations on either side of the interface is exactly k, k0.

If k0 is 1, then it will arrange itself to have the same concentration at that interface on either side. If k0 is 10, it's going to want to have a factor of 10 difference in concentration going from material B up to material A. So there is-- this different solubility causes this equilibrium segregation. OK. When I'm calculating in SUPREM, when I'm calculating the interface flux, what's the boundary condition at the interface? Well, you saw-- if you're inside a given material, you saw last time in the finite difference case, if you're inside silicon one given material, you could use your neighboring concentrations to figure out a flux to either side.

It's a little bit different if you have an interface flux. In fact, what you write is that right at this interface between A and B, the flux F to the right is equal to-- now, we don't use a diffusion coefficient, it's equal to a transport, interface transport coefficient h, which has units of length per unit time in, for example, centimeters per second. It gives you an idea how fast this thing is going to reach its equilibrium concentration. It's h times this ratio times this quantity, ca minus cb over k naught. Where again, k0 is defined according to this.

So if you just look at this flux equation for a moment, what is it saying? Well, it says if, let's say, h is some reasonable value, if ca is much, much different, the concentration on this side, then cb over k0, this difference is going to be a large number. There's going to be a big flux. It's going to allow flux to flow, basically, until ca approaches cb over k0. And then the flux goes to zero. So it's going to force then the concentration profile across that interface to be pegged to have a difference, a ratio, that's equal to k0, and how rapidly it approaches that equilibrium.

Well, that will be given to a certain extent by what h value you use for that. Because at each time step, you have a flux that's equal to ca minus cb over k0 times the h value. If h is very small, it'll take a long time to reach the equilibrium segregated condition. If h is very rapid-- high, then it reaches that very quickly. So h is a measure of how easily the species is transported across the interface.

So a very common thing that you will see happening when you anneal wafers, or when you grow oxides, is segregation at an interface between silicon and silicon dioxide. And so let's look at that case. By the way, I should say that segregation coefficients, the ratio of some species on one side of an interface to another material, the other side of the interface, it sounds trivial, but it's-- actually, in practice, it's very hard.

You might say, well, why don't you just use SIMS? Just profile through the sample and measure the concentration on this side and the concentration on that side. Well, a lot of techniques like SIMS near an interface, they don't perform very well. Because as you get close to the interface, you're changing the material and the sputter rate starts to change. All the assumptions that you need to make in SIMS tend to be degraded, to a certain extent. So you would think we have perfect numbers for this but we don't. We have rough numbers. And these are the rough numbers that are in SUPREM.

They are-- the k values, of course, are adjustable, and you can adjust them to fit your-- whatever your experiments show. But k0, so the ratio of the concentration of the dopant in silicon-- to that in silicon dioxide. For boron, it's less than 1, it's 0.3. So what that means is that boron wants to go into the oxide layers, concentration in the silicon will be lower. So boron tends to be depleted in the silicon and higher in the oxide.

Arsenic, the n-type dopants are just the opposite. They tend to segregate into the silicon. So they pile up at the silicon, and they're lower in the oxide. And this is giving them a ratio of 10. Of course, it also depends on temperature. So you need to take that into account. So in fact, if we go to slide 16, we can see an example of this. These are some SUPREM simulations. And what's been done here in the upper left, you're seeing the simulation of the oxidation of a uniformly doped substrate. So initially, it was uniformly doped.

So the concentration of boron at the surface was all the same, say, 10 to the 18th, throughout the entire wafer from the surface all the way through. And you take that and you put it in a furnace and you oxidize it. Well, you can see where the oxide was grown. All of a sudden, the boron has a profile to it. And it's a little hard to tell the profile from these contours. Each contour, each color represents a different boron concentration.

But if you want to take a cut right through the center, this is what the boron looks like. There's a certain concentration of boron in the oxide here, around 6 or 7 times 10 to the 17th. Then there's a drop. There's a factor of 3 drop because, again, we said the segregation coefficient was about 0.3. And it's depleted somewhat in the silicon. So it's come down in the silicon.

That's because there's been an interface flux. The concentration originally in the oxide was 0. The concentration in here was about 10 to the 18th in the bulk. And there's been a flux from the silicon to the oxide because in equilibrium, it wants to set this ratio to be equal to 3. And so you see this depleted boron concentration in the silicon.

How about n-type dopants? Well, arsenic and phosphorus we said their segregation coefficient is 10. So it's a positive number. So they actually tend to pile up at the interface, and they're low concentration in the oxide. So they tend to segregate into the silicon. This arsenic profile looks a little steeper than phosphorus because it has a slower diffusivity. So it's not getting to the interface. It's not able to diffuse across the interface, or transport across it as rapidly because the delivery of the arsenic from the bulk is a little bit slower.

So there's an example of segregation that happens during an oxidation process. So that's something that's incorporated in SUPREM and is very commonly observed. The next interfacial effect that I'll talk about, interfacial dopant pileup. Be careful, this is not equilibrium segregation. This gets a little tricky because we just talked about cross an interface, the doping concentration being different. This is actually piling up right at the interface. So it's a little bit different. It's a monolayer type effect.

And particularly as junctions become shallow, it's observed that some of the dopants pile up in this very narrow interfacial layer at the interface between oxide and silicon. This pileup is separate from and is larger than the normal equilibrium segregation that might occur during annealing. So just as an example, if this is the oxide to the left, the silicon to the right, normal segregation would give you a little height difference like this. This interfacial pileup gives you a height difference that's even larger, and it's just at the interface, and it integrates to some dose.

It may be as thin as a monolayer, but it can actually-- the interface essentially acting as a sink for the dopants. And it can trap up to about 10 to the 15th per square at that interface. That's a pretty big number. If I consider that I might only be implanting 2 times 10 to the 15th arsenic, and I implant it right near the surface, a lot of that arsenic gets sucked up right into that interface and never become electrically active.

So when people were implanting the arsenic really deep into the substrate, nobody noticed it because the concentration near the surface was low. But now that we have shallow implants, a lot of people are finding that their dopants are sucked up into that interface and so they lose a lot of effective dose that they would have expected to have. So people-- SUPREM IV now includes an interfacial segregation module to model this effect.

In fact, on slide 18 shows some data of what happens. Here on the left, this is from Cazenave, IEDM, 1998. This is concentration of arsenic as a function of depth, and the "as implanted" is shown here in blue with the boxes. And this is annealed, the rapid thermal anneal, 1,050 for 30 seconds. There's a thin oxide layer, by the way, that was on the surface. He was using as a cap when he did this anneal. And then he stripped the oxide before doing the SIMS. And this is the red layer, the red profile. If you integrate it, its area, or its dose, is only-- is 6 times 10 to the 14th. So 30% of the arsenic was lost.

Now, he had an oxide cap there, hoping to keep the arsenic in the sample, hoping to keep the arsenic from evaporating. It didn't evaporate, it got stuck at the interface and then was stripped off with the HF. So he prevented evaporation, but it is a problem. And in fact, if you go on the SIMS [INAUDIBLE] at the right, he tried to understand what was going on. What he saw, if he did not do the HF dip after the rapid thermal anneal, he saw this red profile.

So indeed, in the oxide right near the interface-- and again, SIMS doesn't resolve the interface very well-- he saw a pileup in that interfacial region. And that's where all that extra dose went. Then when you strip the oxide, of course, everything piled up, but the interface is gone and you end up with a much lower dose. So it's not very well understood. But you can imagine, this is very important for source drains. If we go on to slide number 19, in fact, if we look at right near the channel, we implant these very shallow junctions called the source and drain extensions, sometimes called the tips. These junctions today are on the order of 500 angstroms or less, maybe 300 angstroms. So they're very, very shallow. So you need a very shallow implant right near the surface.

And lo and behold, what's above that implant, of course, is an oxide. So you have an interface between silicon and oxide, and a lot of dopant right underneath there. So you will go ahead, you will design your source train extension to give you the right sheet resistance, and then you go measure it and you find out it's half-- or it's twice what the sheet resistance you would design. Because half your dose has segregated to that interface where it's not electrically active.

So it sounds like a subtle effect, but it's actually extremely important, particularly, when you need low sheet resistance contacts to the channel. It's very annoying that half of your dopant gets sucked up by that oxide. So it's just a fact of life and something you need to take into account. SUPREM does have empirical models. Of course, you have to adjust all the coefficients to fit your particular data.

OK. So those are some special effects that I wanted to include and when we talk about dopant diffusion. And now, I want to go on and talk about atomic scale model. Everything we've talked about so far has been-- except for several lectures ago-- it's been, really, at the macroscopic diffusion. We defined Fick's law, you added electric field effects and Fermi level effects. But a lot of effects, especially, those like OED, oxidation enhanced diffusion and TED, transient enhanced diffusion, they're action at a distance.

They're very important experimentally, but they cannot be explained by these simple macroscopic models. So we really want to look at dopant diffusion as best as we can at the level of the atomic scale. And here is a way of doing that, showing you on slide 21. There are two different mechanisms pictured on this slide. One that we've talked about, we haven't really gone through the specifics of it. But we've hinted at this vacancy assisted mechanism.

So imagine this chemical equation. I have a dopant A, it could be arsenic, boron, whatever. And it gets-- it's paired with a vacancy right nearby. And it goes into a pair, an av pair. av pair, what I mean by pair, well, they maintain within a lattice constant are two of each other. There's this pair and they move as a pair throughout the lattice. So how can that happen? Well, you can imagine this-- let's say this is my first time step.

Here's a vacancy right here. Here's a dopant atom pictured in the light. And the silicon atoms are all the dark black. So this dopant here may exchange site with the vacancy, OK? The dopant moves to this vacancy site, the vacancy moves up there. All right, fine. Well, if they just switch back, they haven't moved anywhere. They're just switching back and forth. That doesn't do you any good. But imagine I move the dopant to this point and the vacancy moves up there.

Now, all of a sudden, this vacancy is sitting here. It can move independently for a second of the dopant and exchange with its neighboring silicon atom. So now, I have a vacancy here and a silicon atom here. And the vacancy can move over here. So all of a sudden now, the vacancy is sitting over here now, again, next to that dopant atom. Now, it can exchange with it.

So just by moving the vacancy around in a circular motion around just open atom and exchanging sites with it, the vacancy and the silicon can essentially move as a pair gradually through the lattice. And it's much easier to do that than if you imagine there was no vacancy there, and every time the dopant had to move, it had to break the bonds. You don't need to do that when you have a vacancy.

So this pairing of the vacancies is believed to be a very efficient mechanism for dopants to move in silicon. You can do something similar with interstitials, or interstitial c assisted mechanisms. So here I have a chemical equation, dopant a, plus an interstitial forms, an ai pair. And it will help in this assist diffusion. Well, here again, my pink atom here is the dopant. Here is an interstitial.

It can come along and kick out the dopant off a lattice site, make it interstitial, and help it get moving that way. Or here's an interstitial c, what is that? Here is a silicon atom sharing a lattice site with another silicon atom. It can then start to share with the dopant atom. And then the dopant atom can move along and share with the next bonded silicon atom. And so it can move along, perhaps, the bond direction as an interstitial c as two objects sort of sharing the same lattice site. So either interstitials excess hanging around, or vacancies hanging around, either one, these point defects can assist with the motion of the dopant in the lattice.

So we're going to make some inferences about mechanisms. I think we talked about this a little bit last time, or a couple of lectures ago when we talked about stacking faults and oxidation. This is a picture of local oxidation. So over on the right, I'm having oxidation take place. On the left, I'm underneath a nitride so there's no oxidation. And what people see is that deep in the substrate underneath where you're oxidizing, you see that oxidation induced stacking faults. Remember, we said they grow underneath the region where there's oxidizing underneath. The region where there's no oxidation, they don't, they stay constant.

So it's believed that oxidation injects from the surface, injects interstitials into the bulk, which aids the growth of stacking faults, and also can enhance the diffusion of dopants like boron. Now similarly, it's also been found, if I took this starting wafer and instead of putting in an oxidizing furnace and growing an oxide, I put it in a furnace with ammonia. And I grow over here in this region on the right, I would grow silicon nitride. So I'm nitriding. I'm thermally nitriding. I'm reacting silicon with ammonia.

People have found that this actually has the opposite effect, that boron diffusion is retarded, and stacking faults actually shrink. So people believe that thermal nitration injects vacancies so-- by these inferences by observation of stacking faults and things. So the nice thing now is, if I do an experiment, I can inject interstitials by oxidizing in one region. I can tend to take the wafer, or another wafer, and put it in a furnace with ammonia and inject vacancies. And I can see what happens to the dopant profiles under these different injection conditions. And then, therefore, decide does the dopant diffuse faster with interstitials? Well, then, it must tend to diffuse with interstitial pairs. So there's a way to make an inference about the diffusion mechanism using oxidation and nitridation.

So there's been a number of experiments that have been done over the years, the last 20 years, and some of them are shown here on the results on slide 23. And what people have seen over the years is that boron and phosphorus, and to a little bit, a certain extent arsenic, they have enhanced diffusion coefficients under the influence of thermal oxidation. So during a thermal oxidation, the boron, the phosphorus diffusivity tend to go up. Antimony is just the opposite. It slows down compared to inert when you're in an oxidizing condition.

So here's just an example of a plot. On the left axis is concentration versus depth. And so here's an example of antimony. Antimony was diffused, and this profile here without any dots on it is the inert case. And if you look at the junction depth for the case where it was diffusing under oxidation, it's actually shallower. So there was less diffusion of the antimony when you diffused it in the furnace-- the same temperature, but under oxidizing ambient.

Boron is just the opposite. Look at boron. This is the inert case for boron. The junction depth here is about 0.4 microns. When you diffuse it with oxidation going on above it, again, it's not touching the boron, the oxidation is taking place up high in the sample. The boron has a junction depth about 0.8. So it's dramatically enhanced. The idea is that oxidation increases the concentration of interstitials, silicon interstitials, ci.

Now, it decreases cv from their equilibrium values. So from this, I would conclude that boron diffuses with i, with interstitials, because I put more i in the sample, it goes faster. And antimony diffuses probably more with vacancies. Because when I decrease the vacancy population by injecting a lot of excess interstitials, antimony slows down. So antimony must favor diffusion with vacancies.

So it's by these types of experiments, injecting interstitials with oxide, oxidation, or thermal nitridation to inject vacancies, that people try to figure out what is the mechanism of the dopant diffusion. So let's go on to slide 24. And in fact, the interesting thing-- the injected interstitial level depends on the generation rate at the very interface between the oxide and the silicon, and the recombination rate at that interface. So there's certain generation rate of these interstitials, and certain number combine-- recombine. Those adult go into the bulk.

And in fact, the concentration of excess of interstitials depends on the oxidation rate. That's what people find, which is interesting. So if I oxidize faster at a given temperature, I'm going to get a higher oxidation, or interstitial concentration. So for example, if I plot the interstitial supersaturation ratio-- and this is the ratio of ci divided by ci star, where ci star means the concentration of interstitials in equilibrium. So that's in a neutral ambient without any perturbing due to oxidation.

And I look at this ratio, the dashed line is for wet 002, and the solid is for dry. Well, we know the oxidation rate in water, in 1 O2 is a lot faster. You see, the whole dashed line is higher. So if I really, really wanted to enhance the diffusion of boron, what would I do? I could put it in the substrate, and I would subject the substrate to wet oxidation at a given temperature. And that would really boost up. I could make the boron diffuse a lot faster. And generally, you want to slow things down.

So this-- as you can see, this depends primarily upon temperature. There's a little influence of the wet versus dry. But the big influence here is on temperature. And the interstitial supersaturation ratio is much larger at low oxidation temperatures. That's because ci star is going down rapidly while you continue to inject a lot of interstitial ci. So we expect the enhancement in diffusion-- or diffusivity to be small at high temperatures, like, at 1,200, where the supersaturation ratio is only a factor of 2.

But to be large at low temperature, say, 800, very large, where you can get ratios of 10, or 100, 100 times faster diffusion than you would get under equilibrium non-oxidizing conditions. So this tells us where OED, where ORD is going to be most prevalent at low temperatures. So how do people model the interstitial and vacancy components of diffusion? Well, here's-- again, I just want to show you some experimental data, some SIMS plots with both arsenic and antimony in the sample at the same time.

So the red here is shown under inert conditions. So no oxidation. You can see the arsenic is abrupt. If you oxidize it, the arsenic diffuses a little faster. So it's being influenced by the interstitials. Antimony, at the same time under inert conditions, is a little broader. But if you oxidize it, it maintains-- it diffuses less. So here's OED, oxidation enhanced diffusion of arsenic taking place at the same time as ORD, oxidation retarded diffusion, of antimony.

So both types of point defects, interstitials and vacancies, are important in diffusing in silicon. So what people do is-- it's somewhat empirical but it works-- is you say that the dopants diffuse with a certain fraction, f sub i, of interstitial type diffusion, and a certain fraction f sub c-- f sub b, which is just 1 minus fi, of vacancy type. So we're just going to apportion-- for any given dopant we're going to say, well, X percent, or x fraction, f sub i fraction, is associated with it moving with interstitials.

So we write in a very generalized form the diffusivity of any dope, da, is da star, where da star is the normal equilibrium diffusivity measured under inert conditions-- no oxidation, no thermal nitridation. We're not perturbing the surface in any way. So that's da star times this quantity, f sub i, which is a number between zero and 1, times ci over ci star, plus f sub v, times cv over ci star-- cv star. Again, the star means equilibrium, no oxidation or nitridation.

So you can see I can enhance the diffusivity just by enhancing ci over ci star, assuming f sub i is greater than zero. So if I have a dopant like boron, people believe that f sub i is 1. If I pump up ci over ci star, then I get a great enhancement proportional to ci or ci star in the diffusivity. So again, oxidation injects interstitials, so it's going to raise ci star, and it reduces vacancy. So this goes down, cv over cv star by a recombination mechanism. And nitridation does exactly the opposite. So this is the mathematical formulation we can use to express these observations.

And in fact, you go on to slide 27, people have tried to measure-- they have measured the enhancement of the diffusion, or the retarded diffusion, under different conditions. And these are the f sub i and f sub v values that people-- that are roughly in SUPREM. So what do we see? Well, for boron, f sub i is 1. So they're saying it diffuses entirely by interstitials. So that's roughly what people believe.

Phosphorous is close to 1. Arsenic, which is our most popular n-type dopant, unfortunately, is mixed. It diffuses both by interstitial mechanism and by vacancy mechanism. Antimony is just the opposite, entirely by vacancies. So these [INAUDIBLE] numbers. Of course, you can modify them at will in the simulator, but these are the ones that are programmed in into SUPREM IV.

So let's go on to slide 28. Again, this is a general formulation for how we write the diffusivity in terms of si, fi and fv. How does this actually relate to our previous description? We keep making-- now, that I've made the model more atomistic, how does it actually relate back to the more macroscopic description? Well, this is the macroscopic way we wrote it, right? We said da effective is just da zero, some number times e to the minus e over kt.

Well, I can rewrite this expression on top under inert conditions. Inert meaning, ci over ci star is 1. cv over cv star is 1. So there's no oxidation or nitrogen. Then da is just sum of two terms. The diffusivity of the paired species ai, plus the diffusivity of the paired species av. And in fact, I can break this down even further where I write this diffusivity of the paired ai as a little d, diffusivity of ai times the concentration of c of these ai pairs, divided by ca, plus a comparable expression analogous for vacancies.

So what this is saying is, I can sum up the diffusivity. And if I just look at this one term, it's looking like the diffusivity of the pair, say, the dopant paired with the interstitials, times the ratio of its concentration, to the total concentration of the arsenic, or whatever the dopant is. So if I make this go up, then this will go up. So at an atomistic level, we can decompose this effect of diffusivity into these two different mechanisms.

So let's go on to slide 29. And, say, there's another way people can look at this atomistic scale reactions and diffusions. And people do it through a chemical reaction. If you're familiar with chemistry, this makes sense to you. If you're not, you think, why am I going to all this effort when I can express it mathematically a little bit differently? But what we say is, the reaction where a substitutional dopant atom a interacts with an interstitial silicon atom to form a mobile species.

So the simple reaction says that a, substitutional, plus i, goes to ai. Now, the important thing to realize about this equation is that on the left-hand side, a and i are both immobile. So a is immobile by itself. We're assuming that anytime the atom, the arsenic or the boron, is substitutional on the lattice at high temperature, it can't move. That the only way it can become mobilized is when it's in a pair form with an a right next to an i, or next to a vacancy if you want to do it in terms of vacancies.

So the substitutional species themselves are immobile, only as a pair are they immobile. And this is what the model is saying. So we can actually be used to explain a lot of different phenomena at a distance that people have observed. For example, if you have an interstitial supersaturation. So you pump up i a lot. This is going to drive more dopant atoms into the mobile state by shifting the equation to the right, and enhance the dopant diffusivity. And that's OED.

So example, if you're a chemist and I tell you I flood the reactor-- I flood the silicon with a lot of i, well, this reaction tends to get-- if we add more i, this reaction gets pushed to the right. So I form more pairs. If I have more pairs, arsenic can diffuse more readily-- or the dopant, and you get OED. Now, the interesting thing is that there's can-- this equation also predicts some effects even under inert conditions.

So this can indicate that the interior of the silicon will be injected by this mobile ai species as it diffuses in. And that's going to drive the equation to the left, to this way, releasing interstitials in the interior when the dopant regains its substitutional position. So interestingly, this chemical reaction tells us that silicon interstitials can be pumped into the interior of the sample by dopant diffusion.

So for example, let's say I have arsenic, it diffuses in by pairs, it finds its way and settles into a certain position where it's now substitutional. And then what happens, these interstitials are released. So all of a sudden, the arsenic is carried in with it, all these excess interstitials, there's a bunch of interstitials released, they could then impact the dopant-- the diffusion of a dopant nearby. And in fact, that's exactly-- that's exactly what happens in this profile on Page 30 of phosphorus.

For years, it was observed that high concentration phosphorus had a kink and a tail. It was kinked region here. And then at lower concentrations, it had a long tail. It was into the substrate. And people thought of all kinds of mechanisms to explain this. Well, one mechanism in terms of this chemical equilibrium formulation that we're talking about today is to say, all right, the phosphorus diffuses with interstitials so they diffuse as a pair. And then, eventually, phosphorus finds a substitutional site. It stays there. And then it's going to release them into the bulk. So all of a sudden, I have a flux here, I have a flux of a pair. And then I get a flux of extra interstitials.

This, in turn, enhances the tail diffusivity of the phosphorus profile. So the reason the tail is enhanced, people believe, is because the phosphorus itself is pumping in a whole bunch of interstitials and then releasing them somewhere down in this depth. So that's a way to use this to explain qualitatively the tail region of phosphorus diffusion.

On slide 31, there's a famous effect that people call emitter push. Again, they thought of lots of reasons to explain this. What is emitter push? Well, when you're making a bipolar transistor, you have a high concentration emitter-- it could be phosphorus or arsenic. And then you have a more lightly dope base. And then you have a more lightly dope collector. What people observed is, wherever the emitter, the phosphorus was being diffused, right underneath it, the boron base was pushed out, almost like the emitter was pushing the boron faster, to diffuse faster.

Far away from the emitter over here, it didn't diffuse quite so much. But underneath it, it diffused quite a bit. And again, there were a lot of models people came up with to explain this. Well, again, people could say that this high concentration of phosphorus is pumping interstitials because p and i are diffusing together. So the interstitials themselves are being carried in towards the base. We get a high supersaturation of these interstitials. They get released when the phosphorus stops diffusing, and they're released into the boron base.

This excess interstitials then enhances the boron diffusivity and causes it to push in, because we know that we're-- boron has an f sub i of 1. So this is an interesting effect. We said we could inject interstitials by oxidation and enhance boron. We can also inject interstitials by other processes just by the presence nearby of a high concentration diffusion of another species like phosphorus. So this is called full coupling.

Full coupling means that the diffusion of the dopants is affected by the interstitials. And likewise, the diffusion of the interstitials is affected by the presence of dopants. And where those interstitials end up in your wafer, they could impact some other process. So that's what they mean by full coupling in the SUPREM model.

So if we go on to slide 32, again, this is that same equation I showed before. And if you're a chemist and you assume chemical equilibrium between these dopants a and the defects i, you can write a law of mass action that says the concentration of the products, c of ai, is just a constant at any given temperature of the multiply, or the product of the reactants ca times ci. So you can actually write this in a chemical equation.

And then the neat thing is, given this relationship that it's just the product of ca times ci, I can apply Fick's first law to this mobile species. So if I want to differentiate cai by dx, so this is Fick's first law, it says the flux of the mobile pair is just some constant times the concentration gradient of the mobile pair. Well, I can now apply the chain rule to this. And applying the chain rule, I can see that the flux of ai depends on some diffusivity times the term that goes like the gradient of the arsenic, or the dopant, plus a term that goes like the gradient in the interstitials.

So what this is saying is that grains in the defects, as well as gradients in the dopants cause dopant diffusion. When we talked about Fick's law earlier, we said, well, we have a gradient of arsenic, and that's gradient of arsenic is what drives the diffusion. Well, not only will the gradient of arsenic, but somehow if you create by some other mechanism a gradient of interstitials, that will also drive diffusion. So there's a hidden term. And that's because we're doing a pair model. We're saying that arsenic, or boron, or whatever, has to diffuse by means of pairing and so it gives you this extra term.

And there are a lot of interesting ways you can accidentally create gradients of interstitials, not even realizing it, and then end up driving dopant diffusion at a faster rate. So on Page 33, this is actually-- I'm not going to derive this, but this is the actual overall flux equation that SUPREM uses, and it's discussed in your text. It's a fairly complicated-- you say that the total flux of boron interstitial pairs is the product of all these terms.

Well, there are a-- we can look at the terms and make sense out of them. This dvi star, that's just the inner-- the star indicates its inert low concentration diffusion driven by the dopant gradient, the usual good old Fick's diffusion. And then all the rest are correct in factors. In these large parentheses, we see the high concentration effects due to the Fermi level. So that's what this beta time is due to.

The interstitial supersaturation, ci over ci star, again, if I inject interstitials, we know that will cause an enhancement in diffusivity. And this term over here at the very end, partial partial x of the ln of p over ni, that's the electric field effect. So all of these are lumped together in order to calculate the total flux.

OK. So let me just summarize. We talked about Fermi level effects. They apply, they're important when the carrier concentration is greater than ni. The diffusivity is dependent upon the local carrier concentration. It's either determined by the diffusion species itself, or by the background doping, whichever is higher. And we tend to get diffusivities of this form, this formulation. This leads to very box-like profiles.

We talked about segregation at the oxide interface. It determines the boundary conditions. Boron segregates into the oxide, it's depleted from the silicon. Arsenic, and phosphorus pile up. They go out of the oxide and they go into the silicon. There's also interfacial dopant pileup, which is different from segregation at the oxide silicon interface. And this results in dramatic dose loss, particularly, for shallow source drains. We know from OE-- we know OED, ORD, and growth and shrinkage of stacking faults can be explained by this atomic scale diffusion picture.

We said the boron and [? phosph ?] diffuse primarily with interstitials and ammonia, primarily, by vacancy. This has been determined by a lot of experiments. And if we use this chemical equation formula for dopant defect interaction, it can explain a lot of action at a distance effects, like, OED, phosphorus tail, emitter push, things that people had a hard time explaining for many years.

So that's about all I have to say today. I know it's a pretty dense lecture. But we'll finish up Chapter 7 on Thursday. And Thursday, remember, your homework is due.