VLSI Physical Design

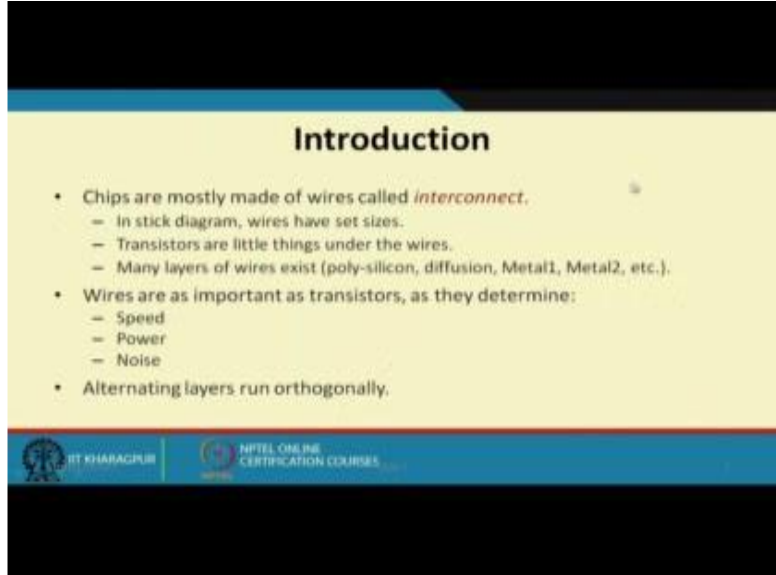
Prof. Indranil Sengupta
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 43
Interconnecting Modeling (Part 1)

So, welcome in this week we start with the lecture on interconnecting modeling. Let me try to understand what interconnecting model is all about. Now we have seen in physical design so when we create the layout of the circuit what does it contain? It contains the basic active components like the transistors the gates which are built using transistors and their interconnections.

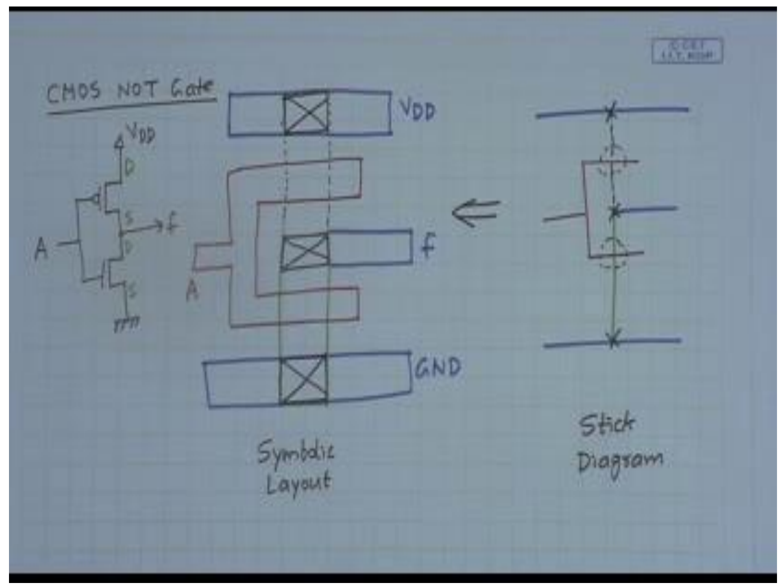
Now interconnect modeling what it means, that when some wires or interconnections are run between 2 points on the silicon floor. So, what are the properties of those interconnections? So mainly we shall be looking at those properties from 2 angles, one would be the resistive properties and the second would be the capacitive properties. Now this resistive and capacitive properties of these wires could be important to understand or analyze what kind of delay characteristics and noise coupling this will be seen later. This kind of characteristics can be involved with that interconnection line. So there are lot of issues regarding the sizing of the lines should the lines be narrow should it be wider or should it be of some other type. So, this interconnection modeling we shall be looking basically into these issues.

(Refer Slide Time: 01:56)



So this is what I was trying to say. So in a modern VLSI chip of course, there are transistors, but a very good percentage of the chip area is dominated by the so called interconnects. Nowadays the transistors have become very small and the interconnections the way we want to connect the transistors to build our functional circuits they tend to dominate in terms of the chip area and also in terms of the delay this we have mentioned. Now there are several ways to represent a layout. So here I have used the term stick diagram. So I would like to explain to you of course, we shall be again coming back to this later.

(Refer Slide Time: 02:53)



So, let me try to explain this with an example. We take the example of a CMOS NOT gate. So this schematic diagram of a CMOS NOT gate will look like this. There will be 2 transistors this will be p type this will be n type these are the input A and this will be output f and this will be your power supply V DD and this is ground. Now we are looking at the layout of this CMOS NOT gate well of course, as I said we will be coming into the layout aspects later, but with respect to how the layout looks like. I am just showing you a schematic side by side.

You see so for these 2 transistors p type and n type, let us say if this is your drain this will be your source you can just reverse with the convention, if this is the drain this is the source. So this source of one transistor it is connected to the drain of the other transistor. And this channel is formed on the diffusion layer. Now usually there is a color coding we shall be looking to the color coding later, but let me tell you so it looks something like this this is the drain to source diffusion layer for the pull down transistor and for pull up transistor we usually show it by a different color.

So, let me show it doted. Typically, it is shown in brown. And this source and drain has to be connected. So there has to be a contrast connection this is typically represented by a big x kind of a notation indicating that here we have a contact. So these are the 2 diffusion regions this is for the p type transistor. This is for the n type transistor and on top V DD and the bottom the ground line should be running. So V DD and ground lines they are represented by again another color. Let us say this is your V DD metal line this is your ground line. And again these diffusion regions have to be connected to this V DD guard so there is another connection here. And there is another connection here fine. And this this output has to be taken from here. So this is the point where the 2 source and the drains are connecting.

So, from here you have to take an output and this output can be taken on different layers. Let us take suppose that we are taking it on metal. So blue is the convention for the metal so here I get the output f. Now means about the inputs. The inputs have to be connected to the gates of this 2 transistors. This is represented like this. Using again another different color convention, red, so this will represent your A. So this is how the layout of this CMOS invertor will look like. Now this width of these different segments I have shown these are interconnects I am talking these are the lines or wires you can say some kind of wires because they will be carrying some signals also. Some of the wires are

even metal layer the blue are the metal the red are the poly silicon which forms the gates and the greens and the dotted regions are the so called diffusion regions.

Now, this is this kind of a notation or this kind of a diagram is sometimes called as schematic or we also sometimes refer to as a symbolic layout, but usually symbolic means each of our regions are represented by either a color or different shade. Now the next question is if we already know how much should be the width and the separation of these lines, then we need not required to draw such elaborate diagram. We can simply represent each of the regions by a single line. So now, the diagram will look something like this. So I am using the same color convention as is used here. And with red there will be a connection like this, and regarding this contacts I represent them by small this x notations. This is an equivalent layout and this is called a stick diagram.

This is called a stick diagram because each line is represented by a line of 0 width you can say. So it is like a stick, but you know blue stands for metal. So when I draw a line like this it is implicitly known that what should be the minimum width of this line. We shall come to these issues later this minimum width and separation rules. So if there is a poly silicon line and metal line running parallel to each other what should be the minimum separation between them these rules are already known.

So, once here I have drawn the stick diagram, it is rather straight forward to convert the stick diagram into this layout. So usually when a designer creates a layout, so usually it is the stick diagram which is created, so let us come back so in stick diagram as I have said so the lines that we draw they are set sizes. Set sizes means the width I mean the width and also the separations. There are many layers of wires because wires are not necessarily on metal like in this diagram as you can see some of the wires are on metal some are on diffusion and some are on poly silicon. So we denote them by different color conventions. And transistors exist under some of these wires like the convention is that wherever poly silicon and a diffusion line cross there is one cross here and there is another cross here.

So, a transistor is formed so I am showing it as a circle. So there will be a transistor here there will be a transistor here. So this is another convention. So in your stick diagram layout wherever you find that a poly silicon and a diffusion line is crossing there are 2 orthogonal mutually perpendicular lines, so there will be one transistor here one

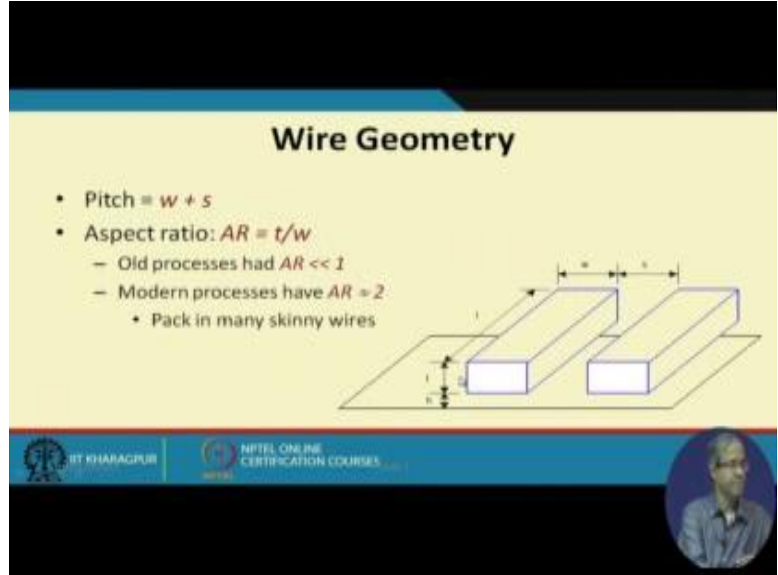
transistor here. So you can easily see that this stick diagram of this layout this actually corresponds to the transistor level diagram or netlist which I have shown on the left side right fine.

Now, I said that wires with technology scaling are becoming narrower I mean they are becoming thinner. And as they are becoming thinner the impact of the resistive and other effects are increasing. So the interconnect delays we have already talked about the RC delays and other things. So those delays will determine the speed of operation and of course, there will be an impact on power consumption because if there are 2 wires running parallel to each other if there is some signal transition on one wire, there can be an capacitive effect on the other and there can be also a resulting signal transitioning to the other wire. And whenever there is a signal transition we shall again be seeing this later when we talk about power issues that whenever there is a signal transition some dynamic power dissipation work occurs and also noise.

so noise might get captured. So means just the example I had said whenever there is some change in this signal in one line some noise might get capacitively coupled to some other line. And just to reduce the impact of the noise or this capacitive of coupling so the convention that is followed is that alternative layers run orthogonally. See what this means is that when I create the layout of a chip there are several layers which are constructed. First so on top of the silicon we create the diffusion then poly silicon then several layers of metal.

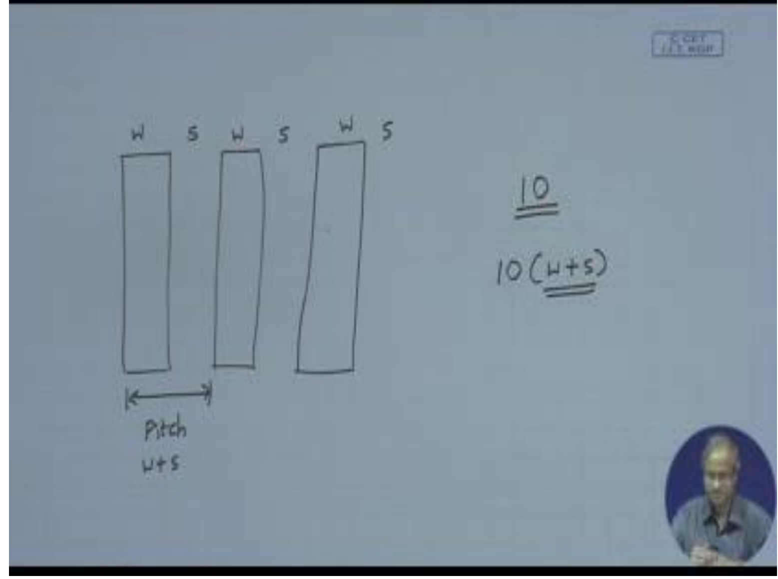
Now, across 2 consecutive layers the capacitive affect will be the maximum because if there are further if away the value of the capacitance will be less. So across consecutive layers the impact of the capacitance or the capacitive affects will be more. So as a matter of convention what we say is that if there are 2 wires running on 2 different layers let they be perpendicular to each other like this. So that the area of overlap between these 2 wires is minimum and as a result the capacitive affect will also be minimized right. So you can see here in this layout we have followed a convention like that like this diffusion and the poly silicon layers are running perpendicular to each other. This diffusion and this metal they are also running perpendicular to each other.

(Refer Slide Time: 13:51)



So, roughly this is a convention, which the designer tries to follow so that the capacitive effect is minimized. Now talking about the wire geometries, we refer to the diagram here, where we show 2 wires running parallel to each other. Now this wires are laid out on the silicon floor. Now actually a wire cannot be of 0 thickness. There is a finite thickness so it will be like a rectangular you can say 3 dimensional shape which is created as the wire. So each of this shape will be having a length l, a width w and a thickness t, and 2 such wires running parallel to each other will be having a separation s.

(Refer Slide Time: 14:59)

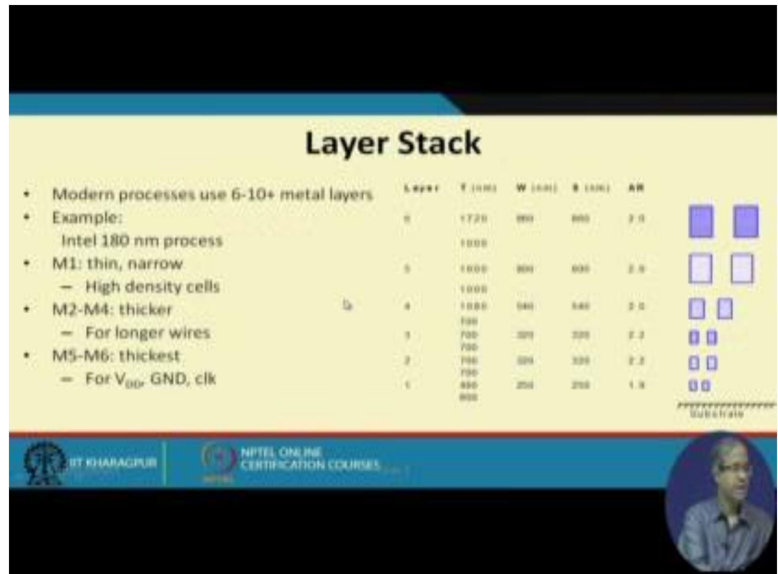


So, on the same layer we are saying that a wire has to be of width s and has to be of separation s from another layer. So what I mean to say is that if I view from the top there will be one wire running here there can be another parallel wire running like this there can be another parallel wire running like this so we are saying so the width is w and the separation is s. So similarly here there will be s so this w plus s taken together is called the pitch. So what is the physical you can say implication of pitch. Pitch means in a particular area or within a particular length means how many lines you can draw. So we must have this w plus s amount of space left for the wire itself and also the separation before the next wire can be layout laid out.

So, if there are 10 such parallel wires we have to layout, the total amount of space you have to give will be 10 into w plus s. So this refers to the pitch. So pitch is one thing and another aspect of this wires is something called the aspect ratio. So aspect ratio represents the relative geometry in terms of the thickness and the width, t by w is defined as the aspect ratio. Now in the older process technologies so earlier what you had we had that the wires were pretty narrow, but they are much wider, they are much wider in geometry, but very narrow, so the aspect ratio t by w was very small. This was what we had in earlier.

But with technology scaling so means we now have more accurate mechanism for fabricating these basic wires and this segments of these layouts. So now, it is possible to create wires which are much thinner in terms of their width w. So w has gone down and in modern processes it is quite practical to have aspect ratio of the order of 2; that means, if thickness is t then width can be t by 2, so even less than the thickness.

(Refer Slide Time: 17:33)



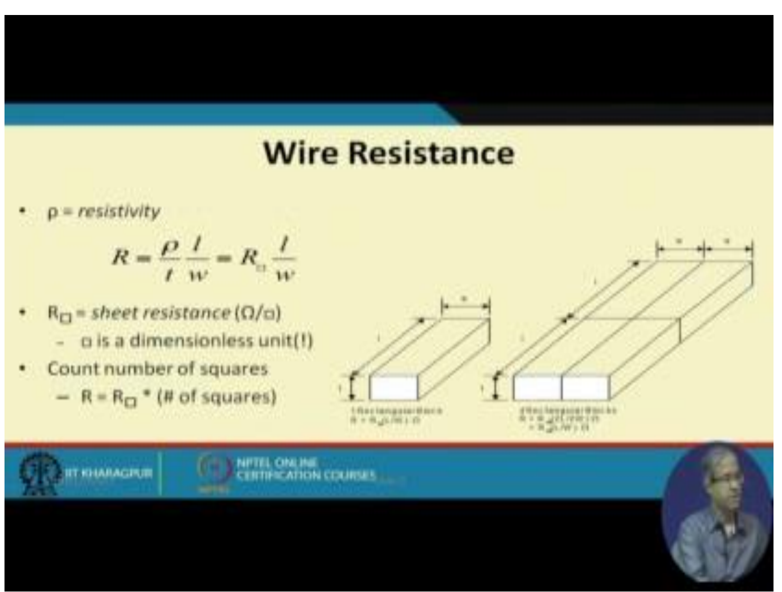
This is what we have today. And just to look at the layering, like I am showing several layers like say you think of the substrate. So I am taking a means a particular example of a 180 nanometer process which is again not very new, but this will help us in understanding the implications what I am trying to say. So there can be several metal layers I am showing the metal layers, in particular. There can be at least 6 metal layers even more. So typically so as the metal layers go up the lowest metal layer we call it M1 then M2 then M3 then M4 like that. So the rule is that when you are lower towards the substrate we can lay the wires much more accurately and thinner.

So, as we go up move up the wires tend to become thicker because there will be more irregularity in the surfaces and laying out those thin wires on the higher layers will be not that reliable. So as we move up the width of the wires also will be getting wider and wider. So the first metal layer M1 will typically be thin and narrow so this wire will be used to connect the transistors and the high density cells etcetera and the higher level wires will be for interconnections like M2 to M4 is 3 layers they are thicker than M1 and they can be run for longer wires for interconnections and say M5 and M6 will be even thicker and they are used to run the critical nets like your power supply and clock.

So, here in this table you see the typical values for this 180 nanometer process. What will be the thickness t of the wire width w and the aspect ratio also? So you come back to this, s is the separation t is the thickness and w is the width. So here we show these

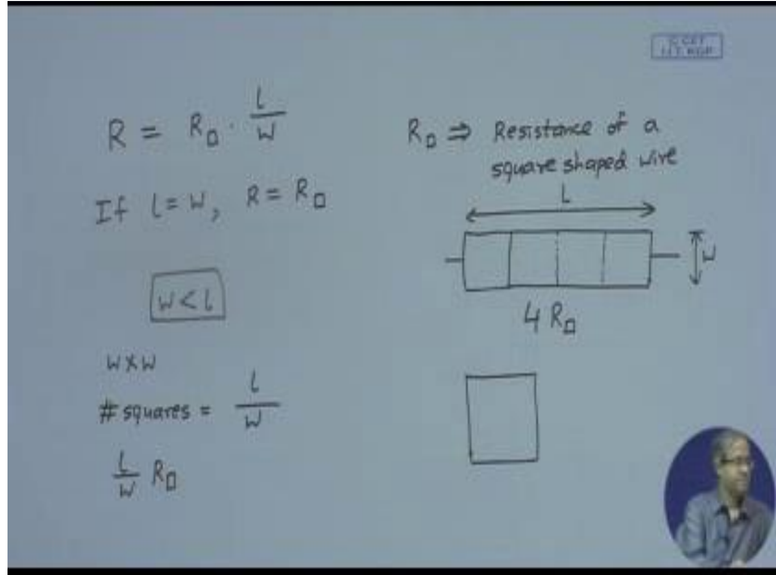
figures separation width thickness and also the aspect ratios. So as you see that as you are moving from layer one up to layer 6 sorry, so the wires are becoming thicker and thicker, the width is becoming larger. The separation is also increasing. So the aspect ratio if you do the simple calculation it is 1.92 0.2 and 2 so approximately 2. So aspect ratio remains 2.

(Refer Slide Time: 20:26)



Next comes the wire resistance now this wires are nothing but this kind of solid rectangular blocks with a thickness with a length with a width. Now for such a block you know that the resistance from one end to the other is given by an equation like this, rho is the constant called the resistivity. So resistance will be proportional to l the length it will be inversely proportional to the thickness t, and also inversely proportional to the width w. Now resistance per unit thickness, because t is usually constant for a particular layer so rho by t, we define as a particular value called R box. So what we are saying is that this R is equal to R box l by w.

(Refer Slide Time: 21:22)



So, what does R box indicates? You see if l is equal to w, then R is the same as R box. So what you can say that what is R box? R box is the resistance of a wire which is in the shape of a square, square shaped wire because for a square shaped wire l will be equal to w so the wire will look like this, l and w are equal say, these are my 2 terminals. Say length is this much width is also the same. So well means like it really does not matter whether square is like this or this square is bigger. So as long as l and w are equal it is resistance value will be the same so as per this equation, right.

This is what we mean by R box. Now suppose my line is like this. Longer which consists of several such smaller boxes so between the terminals what will be the resistance each box will be having resistance R box so this will be having a resistance of 4 times R box. So there is a simple mechanism of estimating the resisting of a wire you just take the smallest square shaped feature and count how many such squares are appearing, like say if in a particular case your width is w, w can be anything this is w and your length is l naturally w is less than l, so here you assume that you have a square of size w by w, imagine that each square is w by w, and how many squares number squares number of squares will be given by l divided by w. So your resistance will be l divided by w times R box which is given by this equation, right.

So, this is how resistance value is estimated. So this diagram actually tells you that. So you simply count the number of squares so whether your wire is of shape like this or you

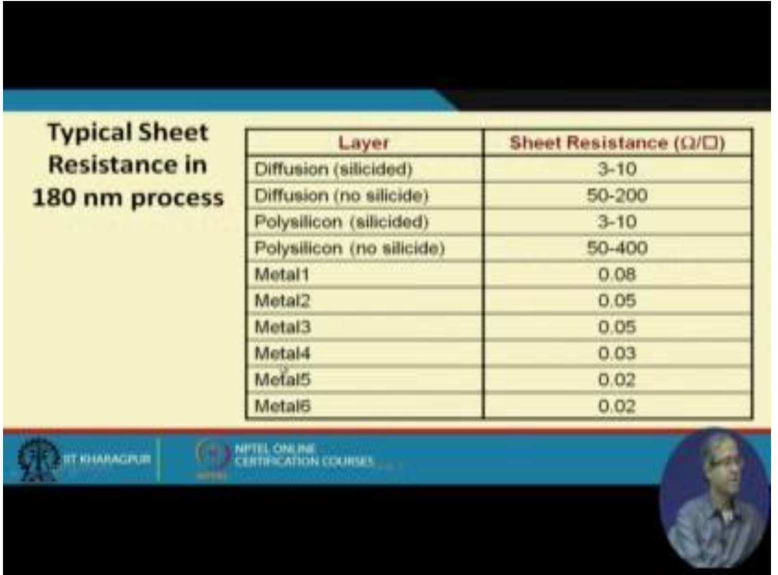
increase the wire in terms of the length and the width by the same multiplicative factor make the width double and make the length also double your resistance value will remain the same. This is the idea. So when I say that I am scaling up or scaling down a gate I am making a gate bigger or smaller so the concept that comes in to the picture is this. The channel area the diffusion and poly silicon overlap area that increases or decreases and just by noting down how many squares are there we can estimate the exact value of resistance in terms of the basic resistance R square R box, fine.

(Refer Slide Time: 25:02)

Now, in terms of the interconnections, interconnections are typically laid out on metal. So here in earlier we used aluminum for the interconnections. These are the resistivity. So as you can see that aluminum has a resistivity of 2.8 micro ohms' centimeter, but materials like gold copper and silver has better resistivity. Gold is expensive, but copper is cheaper, but copper has a better you can say 1.7 silver is comparable. So modern processes they typically use copper, but one problem with copper is that the copper atoms they tend to diffuse into silicon thereby damaging the transistor.

So, proper care has to be taken so that a diffusion barrier is created and this copper layers are wires must be surrounded by the diffusion barrier.

(Refer Slide Time: 26:07)



Typical Sheet Resistance in 180 nm process

Layer	Sheet Resistance (Ω/□)
Diffusion (silicided)	3-10
Diffusion (no silicide)	50-200
Polysilicon (silicided)	3-10
Polysilicon (no silicide)	50-400
Metal1	0.08
Metal2	0.05
Metal3	0.05
Metal4	0.03
Metal5	0.02
Metal6	0.02

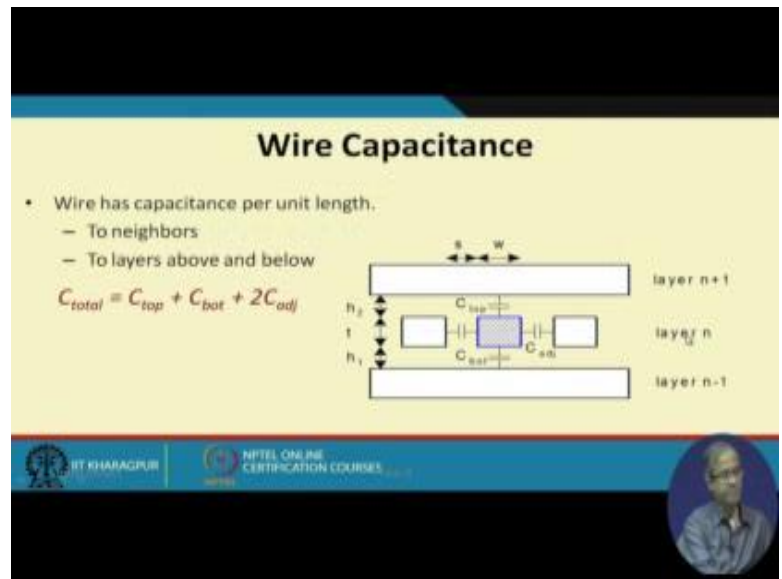
So, these techniques are used and practiced. These are the typical resistance values for the different layers, like I talked about the different metal layers' sheet resistance. Sheet resistance is ohm per box. This is your basic R box. R box value you see metal 1 because the lines are narrower resistance is 0.08, but as you move up the lines become thicker so the resistance value is decreasing 0.02. Similarly, poly silicon has much higher than metal 3 to 10 diffusion has 3 to if it is sillicided. Sillicided is a special way of creating, but if there is no sillicide which is much simpler, then the resistance value increases.

So, the resistance value changes like this. So as you move up your values are changing move down values are increasing. So another issue is the contacts resistance. Whenever you want a contact between 2 layers like here.

(Refer Slide Time: 27:07)

I am showing a connection between red and blue. Red represents poly silicon blue represents metal. Normally I need a cut or contact. Now each such contact typically will be having a resistance value which can vary from 2 to 20 ohm. So higher the resistance greater will be the means RC delay for some other issues. So if I want to reduce the resistance of the contacts what is done is that instead of a single contact you can use many small contacts, like here where the 2 wires are running side by side you can have some small this 8 contacts in this window. And similarly where there is a corner so here you see that there are 16 – 4 by 4 contacts are there. The result is that the net resistance value of this contact will be less. This is what is normally done.

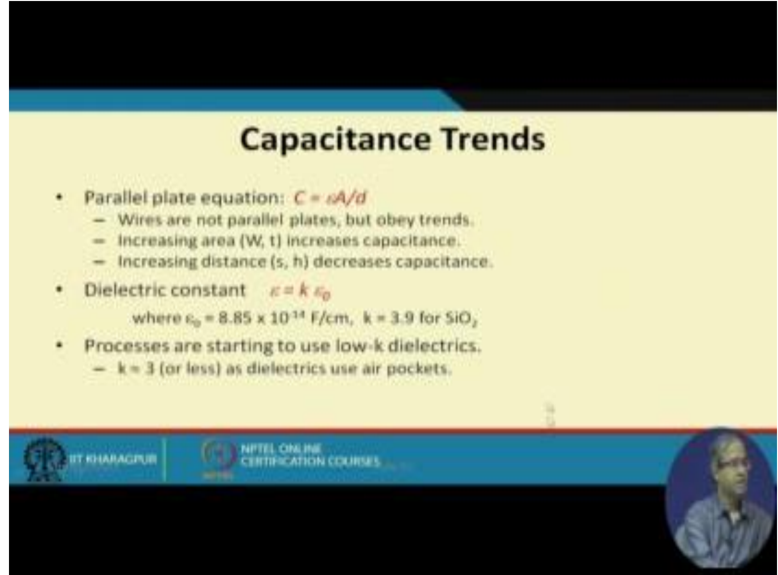
(Refer Slide Time: 28:12)



Now, talking of the wire capacitance, so again suppose I consider a wire which is running on a particular layer, I am looking at a side view. So there can be other wires running in parallel on the same layer.

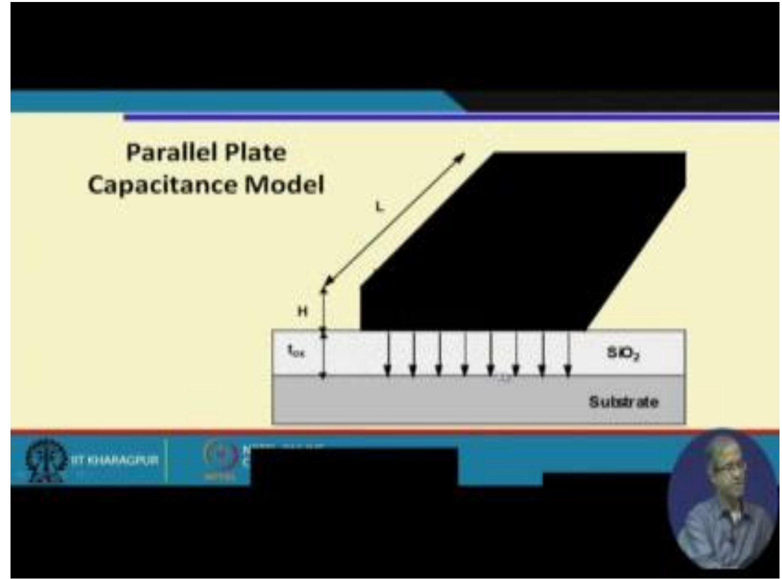
There can be some wires running on the layer above. There can be some wires running on the layer below. So when I talk about the total capacitive effect there can be capacitance between this adjacent wires I call it a C adjacent there is one on the right one on the left so the contribution is twice. The capacitance with the line on the above layer C top and C bottom, so the total resistance or the total capacitance that this wire experiences will be C total. So when you want to make a good estimate we will have to look into this. Now the reason why adjacent layer wires are orthogonal to each other is to reduce the values of C top and C bottom.

(Refer Slide Time: 29:27)



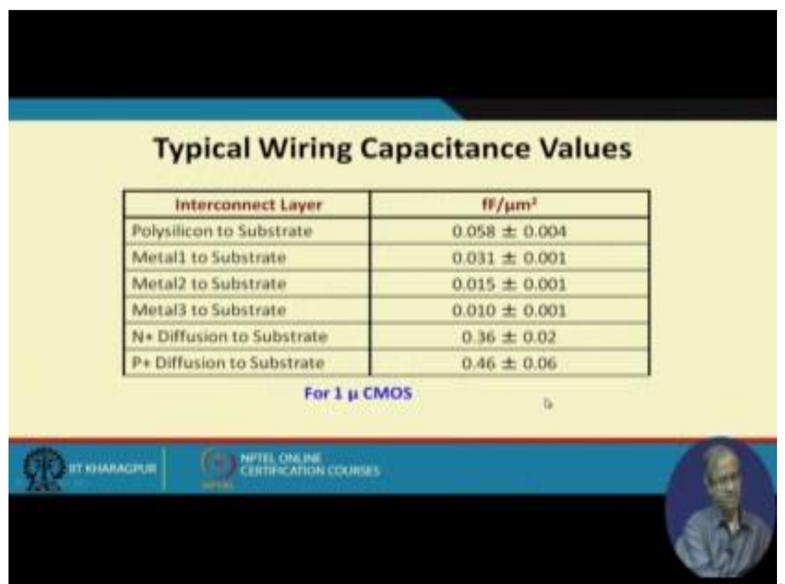
If they are running absolutely parallel, then the capacitance value will be higher all right. So that is why they are made orthogonal. So the capacitance value of a parallel plate capacitor if you recall is given by this expression where epsilon is a dielectric constant of the medium, A is the area of the plates and d is the separation. So greater the overlap greater is the area higher will be the capacitance, greater the distance between the wires or the plates lower will be the capacities. So this is what we typically do. And epsilon in turns is given by some parameter k multiplied by the dielectric constants of free space this is epsilon 0. And for silicon dioxide the value of k is typically 3.9.

(Refer Slide Time: 30:31)



So, lower the value of this k this epsilon will be smaller, so capacitive effect will be smaller. So nowadays we try to use some material which has a lower value of k. So this is what is used means parallel plate capacitance is already I have mentioned across layers the capacitance this block, the solid block shown as black. This represents a plate of the capacitor. So there can be layers above and below which forms a capacitance.

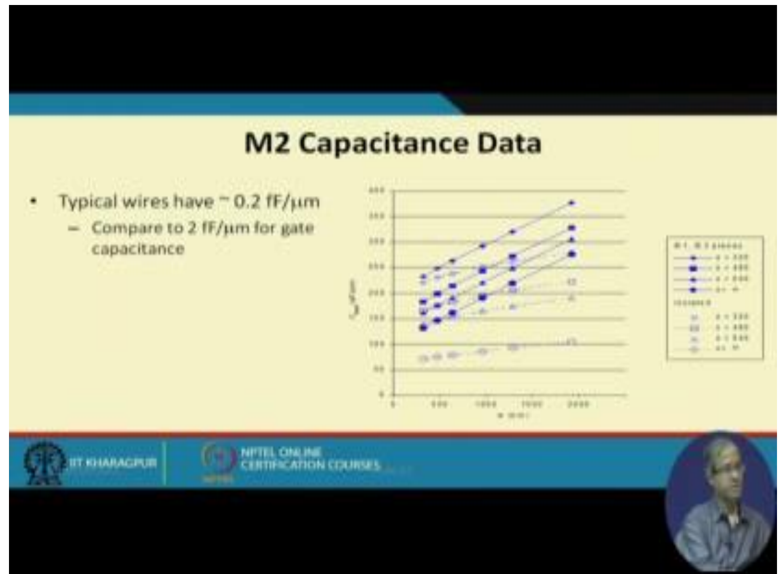
(Refer Slide Time: 30:47)



Now, typically wiring capacitance values are also shown here. Poly silicon to substrate is this is femto farad pi micrometer square micron square. 0.058 metal 1 to substrate is less metal 2 is even less metal 3 is less, but between the diffusion layers to the substrate it is significantly high as you can see - 0.36 and 0.46. These figures are for such a older process one micron CMOS process, but as the technology has scaled down the values, the relative values are still similar. That diffusion capacitance dominates.

And this graph actually shows you for this second metal layer M2, so how the capacitance value changes with w the width of the wires.

(Refer Slide Time: 31:32)



So on this side I show the capacitance and this is the width. And this different plots that are shown they represent the different pitch that means, the separation so if the wires are separated widely, say here the biggest dot shows s equals to infinity. Here the capacitance is least, but as the wires are becoming closer and closer the capacitance values are increasing. So this general trend holds for all the layers. So I have just shown this diagram just to give you this, this idea.

So, with this we come to the end of the lecture. In the next lecture we shall be continuing with some more issues on this interconnect modeling.

Thank you.