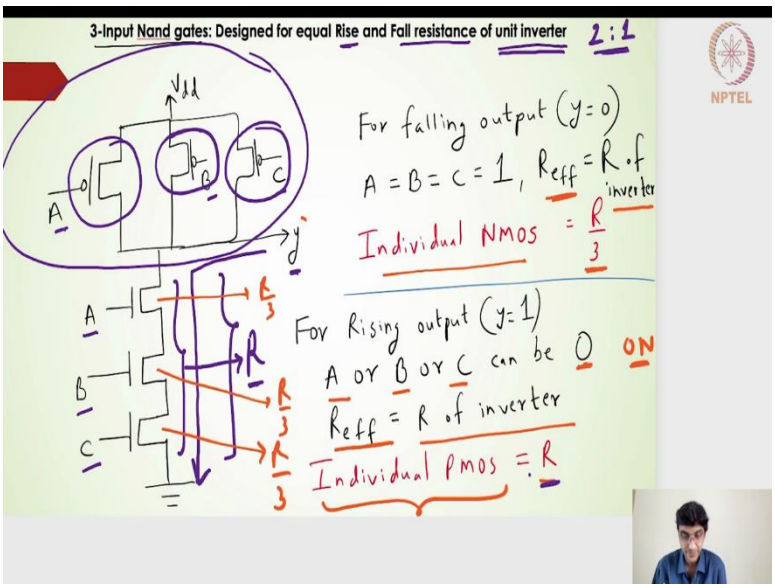
Design and Analysis of VLSI Subsystems
Dr. Madhav Rao
Department of Electronics and Communication Engineering
The International Institute of Information Technology, Bangalore

Lecture - 27
Extracting capacitances of 3-Nand gate for delay estimation

(Refer Slide Time: 00:16)



Hello, students. Let us continue with this lecture on the 3-input NAND gates. What we are trying to look into is a 3-input NAND gate schematic and then appropriately find its propagation delay falling as well as the rising delay.

In this particular aspect because there are multiple transistors, which are connected in parallel and then series especially in series we have to use the Elmore delay method or the Elmore delay model to evaluate or to arrive at the propagation delay falling or the rising parameter.

This is particular schematic of a 3-input NAND gate. By definition what we are going to do is design the 3-input NAND gates such that it is rising and then falling resistance is equal to that of the unit inverter. Unit inverters rising and falling resistance whatever is the unit inverters rising and falling resistance is we are going to match with that of the 3-input NAND gates rising and falling resistance.

For a 3-input NAND gate, I have given the inputs A, B and C. A, B, C on the pull down side and then on the pull up side. On the pull up side, we have three of the PMOS transistors in parallel and on the pull down side we have this three of the NMOS transistors in series and this is my output node.

For falling output what we need is an output node $y = 0$, that is the falling output. It will start from 1 and then it will do the transition from 1 to 0 and then it will reach a steady state of 0, that is what the falling output is. During this particular falling output what we want is an equivalent resistance of R. The R equivalent resistance is coming from the unit inverters falling resistance.

If I consider the unit inverter when the output is actually falling the NMOS transistor gives us the switching resistance which is nothing but the falling resistance. If I use an unit inverter which is 2:1 ratio, I know that the switching resistance for the falling output is nothing but unit R. Similarly, if it is a 2:1 inverter the switching resistance for a rising output is also R.

In this case for a 3-input NAND gate all the three transistors has to be operating on and that is only we will get the output to be discharge to 0. In that case if all the three series transistor has to be in the equivalent of all these three resistance should be equal to R, that means I can have individual transistors giving a resistance of R/3, R/3 and R by/3.

The individual resistance I can say that it should give me R/3, it should give me R/3 and then if it gives me R/3 here switching resistance, then the equivalent is the total of these three transistors switching resistance will be R.

$$R_{eff} = R_{inverter}f$$

The individual NMOS transistors for the 3-input NAND gates in the pull down side, the 3-input NAND gates will give me three transistors in series. That is why the individual NMOS transistors is R/3 is the switching resistance.

On the rising side, it is slightly different. On the rising side for the 3-input NAND gate one of this PMOS transistor has to be on and if the other two are off, because the output will get charged through one of the on transistors. The three transistors are in parallel if the two

transistors are off the other transistor is on and then it will get charged from the output through the on transistors which is connected to the V_{dd}.

In that case A or B or C transistors can be 0, that means the transistors will be on in the PMOS side giving an input of 0 represents that that PMOS transistor is on.

$$R_{eff} = R_{inverter}f$$

Individual PMOS=R

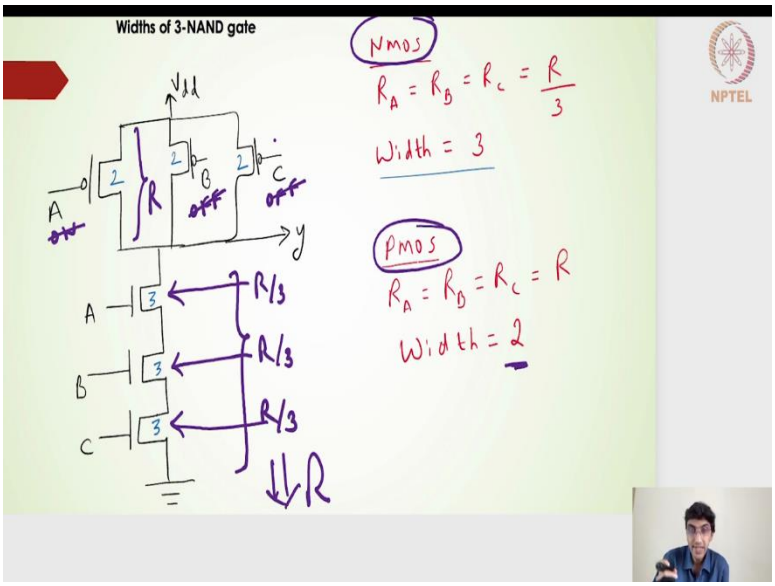
What we are doing is in the pull up side, we are taking the worst condition to design the equivalent resistance. The worst condition is when one of them is on when all of them is on actually in the pull up side when the PMOS transistor A in parallel with the PMOS transistor B in parallel with the PMOS transistor C is on.

Effectively my resistance all of them will be in parallel and then my equivalent resistance in the pull up side will be very very less parallel resistance. A very less switching resistance will give me a very less delay, RC will give me a very less delay and which is very very good for our digital circuit. What we really want is, we want to estimate for the worst case condition. Worst case timing condition is when A will be the signal for the A arrives, but B and C does not arrive at all.

In that case what is the propagation delay? In that sense, we wanted to find out for the worst case condition, still our resistance for the pull up side should be equal to that of the to the resistance of the pull up side of the unit inverter. Hence we have taken one of the conditions no one of the input conditions or the one of the transistors on the pull up side to be on, either A will be on or B will be on or C will be on.

Considering that we need to have it equated or expressed or you make it equal to that of the unit inverters pull up slide the switching resistance. In that sense the individual PMOS transistor we are saying that it will be having a resistance of R. Hope you have understood that, moving ahead.

(Refer Slide Time: 07:03)



What should be the width of the 3-input NAND gate in that sense, on the NMOS side, on the pull down side we know that all the three transistors which are in series has to be on and then only the y output will get discharged to 0. All the three transistors on the NMOS side has to be on and we had said that each of this transistors should have a individual switching resistance of R/3.

Finally, we will get a switching resistance of R which is equal to that of the unit inverters for falling resistance. If it is R/3, then my width has to be 3, that I will get a resistance of R/3. Similarly, the second transistor the B transistor should also have a width of 3, the C transistor should also have a width of 3, so that I will get the switching resistance of R/3.

On the pull up side, on the PMOS side one of the transistors will be on, the other two will be off and in that case my equivalent resistance should be that of matching with that of the unit inverters rising resistance. The unit inverters rising resistance is R on the PMOS side and if one of them is on the width if I consider A to be on and the B to be off and C to be off, in that sense the resistance here should still be R on the PMOS side.

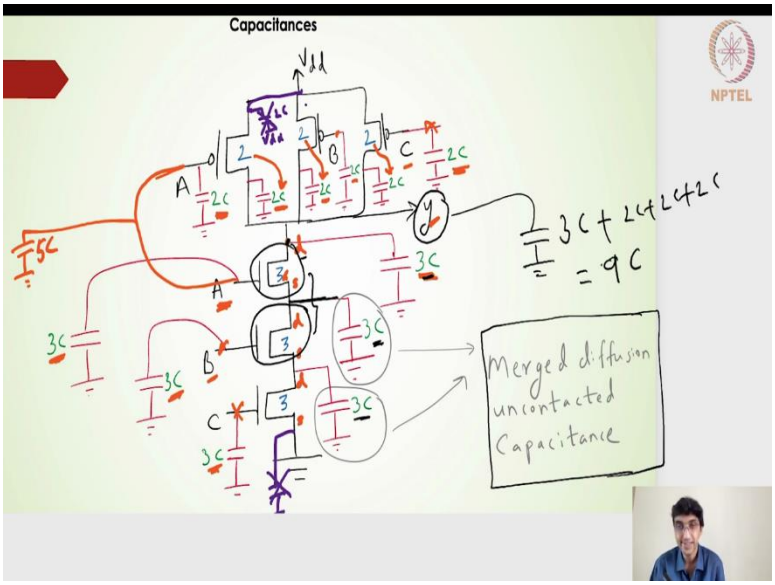
The width here for the 2:1 inverter was 2, which will give me an switching resistance or the rising resistance is R, therefore the width here for the 3-input NAND gate on the pull up side should also be 2, a width of 2 for a PMOS transistor will give me a resistance of R.

That is why in a transistor A, I have written the width of 2. For transistor B, when this one will become on and then A will become off and then C will become off, in that sense I will have to have a width of 2, that it gives me an equivalent resistance of R. Finally, for C, if B and A are off and then C is on I should have a width of 2, that it will give me an equivalent resistance or rising resistance of R.

That implies that all the three transistors on the pull up side should have a width of 2. That individually they should be able to give a rising resistance of R for the worst case condition, what we are assuming is one of the inputs will be on the other two will be off. Hope this is clear.

The dimensions or the designing of this particular 3-input NAND gate is now clear because we have done this width or the dimensions scaling in such a way that it will give me an equivalent resistance on the pull up and the pull down side equivalent to that of the two is to one inverter.

(Refer Slide Time: 10:05)



Moving ahead, if I have this particular scaling on each of these transistors we need to understand what is the capacitances associated with each of these transistors. There are three capacitances the gate capacitances, the depletion capacitances or the diffusion capacitances.

Put together I have made all the capacitance arranged except this non switching capacitances, non-switching capacitances is this capacitance. I have to put 2C and then V_{dd} here, this terminal is also connected to V_{dd} and on the other side it is also connected to the V_{dd}, I have not taken this into account.

Similarly, there will be one capacitances here there will be one capacitances here, which will not switch based on the changes in the input or the changes in the output, as the input changes the output should change, but this non switching capacitances on each of this transistors on the PMOS side and on the NMOS side we will have a C transistors which will have a one non switching transistor here on both the sides it is ground. This we will not accommodate or incorporate in our design. Because that will not have any effect on our calculation of the delay because on both the terminals we have the DC source, that is something I have not accommodated in this particular diagram.

We have the gate capacitances, the gate capacitance is much easier to understand. Now, if the width is 2, the gate is the gate capacitance sees a capacitance of 2C, the gate capacitance from the B of the PMOS is 2C and then the gate capacitance on the C transistors on the pull up side is 2C. The gate capacitance is on the pull down side for a transistor of width 3 is nothing but 3C, 3C and 3C, this is a gate capacitances.

Remember that this input A and then this particular input are actually tied together, input B this particular input and then this particular input on the pull down side is tied together and on the C input and the C input and then the C input on the pull up, pull up and pull down side are tied together.

The input as such if I am talking about the input A it sees an overall capacitance of 5C, this actually sees a capacitance of 3C + 2C which will be 5C. Similarly, the B input also sees that overall capacitance of 5C and C input also sees an overall capacitance of 5C. On the input side it is clear now.

Now, let us accommodate or let us try to understand the diffusion capacitances in this particular circuit or the parasitic capacitances. The PMOS side there will be 2 diffusion capacitances on for the transistor A here this being a non-switching capacitance. We are not going to write or we are not going to worry about it.

There is one which is connected to the output node here. This becomes the drain of the PMOS transistor and that will be the 2C with respect to the body is connected to the V_{dd} and for our transient analysis any kind of a V_{dd} source or dc source we will make it ground, that is why we have this 2C capacitance. Coming from the depletion capacitance of this particular transistor there is one more 2C coming from the B transistors, 2 is the width. I will have the 2C capacitance here and then the C transistor will give me one more because its scaling is 2, it will give me a 2C capacitances here.

Overall, on the pull up side is contributing $2C + 2C + 2C = 6C$ capacitance connected to the output node because all these drain terminals of A, B, C are connected to the output node terminal y. Now, coming to the pull down side I have transistor A, I have the transistor B and I have the transistor C which are connected in series all these transistors also have the depletion of the parasitic or the diffusion capacitances associated with its terminals drain and source with respect to the body.

If I consider this transistor here on this particular side which is the drain side it incorporates or it shows a capacitance of 3C here, because the scaling is 3C. It will have a capacitance of 3C connected to the body which is nothing but grounded on this particular side.

If I consider this to be the drain and then this is the source and this is the drain and then this is the source this is the drain and this is the source. The drain to body capacitance 3C we have seen and source to body I should have another 3C capacitance for this transistor A and for transistor B the drain to body the drain to body should also have one more capacitance of 3C.

In this particular node C and D, which is a basically a merged node merge diffusion node as such coming from the transistors A and B, I should get here instead of 3C I was expecting around 6C capacitance because one 3C is accommodated from the source to the body capacitance coming from the transistor A and another one is the drain to body capacitance, drain to the body capacitance coming from the transistor B, but I have stated here 3C capacitance.

Similarly, on another node transistor B source to body and transistor C drain to body I was expecting around a 6C capacitance, but I have written it as 3C. That is something we will have to look into it and I am saying that this is nothing but a merged diffusion uncontacted capacitance.

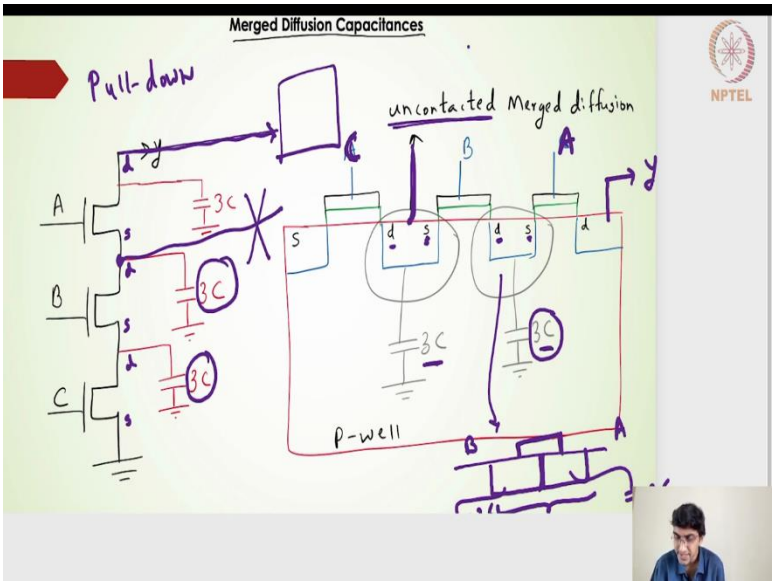
What I am saying here is because of the diffusions being merged the transistor A, B and C are in series. The source diffusion of the transistor A gets merged with that of the drain diffusions of the transistor B. I will have only one diffusion now of the width of the 3 size, that is why I am writing it as 3C instead of two separate 3C making it as a 6C capacitance.

Similarly, for the transistor B the source to body and then the transistor C drain to body I should accommodate 3C and 3C capacitance, but it should have been 6C capacitance, but I am writing it as 3C, because I am saying that this is a merged diffusion un-contacted capacitance, that means the drain of one transistor and then a source of another transistor are actually merged together and giving me a single pocket diffusion.

At this particular slide the overall capacitance is been drawn, the gate capacitance is kind of clear to understand and this particular drain to body capacitance is clear to understand and there is two more merge diffusion capacitance which we will see in detail in the next slide. To come back to this y output node.

The drain to body of this transistor A on the pull up the pull down side is going to contribute to 3C and the pull down side is going to contribute $2C + 2C + 2C = 6C$. Overall the output y node of this particular individual stage sees a capacitance of 3C coming from the pull down side plus $2C + 2C + 2C$ coming from the pull up side. Overall, it sees a total capacitance of 9C. Hope this is clear.

(Refer Slide Time: 18:26)



Now, what do we mean by this merged diffusion capacitances, which we had seen earlier, which we had seen in the last slide which is nothing but talking about this 3C and then this 3C. If I draw the cross-sectional diagram of the pull down circuit of the 3-input NAND gate this is what we will get.

On the A side, I think this is slightly wrong in the sense, this should be A, this is drain, this is source, this is drain, this is source, this is drain, this is source. On the pull down side, this is my pull down circuit of the 3-input NAND gate and this is a cross sectional diagram, this is A, B and then this should be C.

On the A side, there is a drain diffusion which is nothing but output node y. On the other side, it is the source diffusion and then the B, we have the transistor B NMOS transistor having the drain diffusion. Both the diffusion pockets are nothing but the same it is merge diffusion. I will not have two separate diffusion pockets. I will have only one diffusion pocket here, which will give me a capacitance of 3C.

Similarly, on the other side of the B it is a source here and then the C transistors the drain diffusion. Both of them will get merged and then you will give me the 3C capacitance, that is what we had said, if I were to draw two diffusions. If I were to draw two diffusions for the B transistors, this is for the B transistor and then this is coming for the A transistors what is likely to happen is.

The width will be 3, I will have this overall footprint is going to increase because now I have two diffusions instead of merged diffusion, if I have unmerged diffusion showing not two different diffusions for B and A, but also it is connected because I need this particular node is anyway is connected.

I need to have this and then I have this connected by a metal line or something like this. It is an unnecessary metal line coming or it is an unnecessary area that is being consumed. In addition to having instead of 3C, I am going to have this particular diffusion is going to give me 3C and then this particular diffusion is going to give me 3C. It is going to add to my capacitance, which will be 6C capacitance.

Then 6C capacitance is likely to take more time to charge till 50 percent in a halfway point. The overall delay is also going to be increased my performance is going to drop, if I have two separate diffusions my area is going to be more because of two separate diffusions of

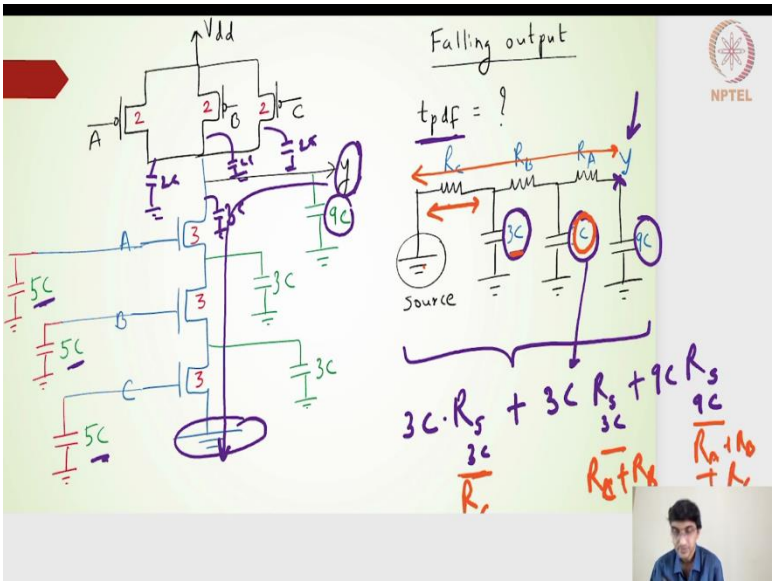
the same width and then lastly I will have to have one more metal line, which is going to connect both of them together.

All these three is making our design inappropriate, in the sense I wanted to optimize it over performance or the footprint or the area reduce the processing steps as well. Any additional processing steps is likely to take more cost. In that sense putting all the three apart, if we can have a single merge diffusion that is what we have drawn here. It will take away all those things and make our design little bit more optimized towards the performance towards the area and then towards the steps also.

In that sense we have this merge diffusion capacitance, whenever you see the series transistors we will have the merge diffusion capacitances also. Even in an 3-input NOR gate or a 2-input NOR gate we will have two of the PMOS transistors connected in series, whenever we see a series transistors we will have a merge diffusion un-contacted capacitances..

Un-contacted represents that so, this is called un-contacted because there is no metal line here which is connecting. What it signifies is we are only interested in the output of this particular 3-input NAND gate connected to some other circuits, we are not interested in this connected to any other circuit. That is why there is no metal line that is flowing along these particular diffusions. This is not there we have only this output connected to the another part of the circuit and that is why it is called as an un-contacted merge diffusions.

(Refer Slide Time: 23:22)



Proceeding further now I have the sizes, now I have the capacitances, now I have the capacitances seen at the output node y which is 9C, I have the gate capacitances which is nothing but overall the pull up and pull down side for the individual inputs turns out to be 5C and then I have the merge capacitance which is nothing but 3C and 3C.

I should be able to find out, what is the falling output what should be the propagation delay for the following 3-input NAND gate. I have drawn an equivalent circuit here all A, B and C should be 1, then only the output will discharge to 0. If A, B and C are 1 all the transistors on the PMOS side will be will not be operating at all.

Although its capacitance of 2C here is anyways there which will get incorporated into the y and that is why I have returned the y as a capacitance of 9C. All this 2C, 2C and 2C and then this particular 3C capacitances here also which is accommodated into the 9C capacitance, this 9C capacitance has to discharge completely to 0 and that is when we will find out the propagation delay for you.

Using the Elmore delay method, what I have drawn is the source is now ground here because it is discharging and then we have drawn this is the ground here. This is the resistance closest to the source is nothing but the switching resistance of the transistor C. We have the switching resistance of RC, the next node capacitance is nothing but 3C, that is what I have written 3C here and then we have the switching resistance of the transistor B which is RB and then the node capacitance is 3C.

The next resistance is the R_A switching resistance of the transistor A and then the output node y, which sees a capacitance of 9C. This is my overall RC circuit or I mean whatever the 3RC circuits which are cascaded together and here we should be able to apply the Elmore delay method to find out what is the overall propagation delay falling for the y output.

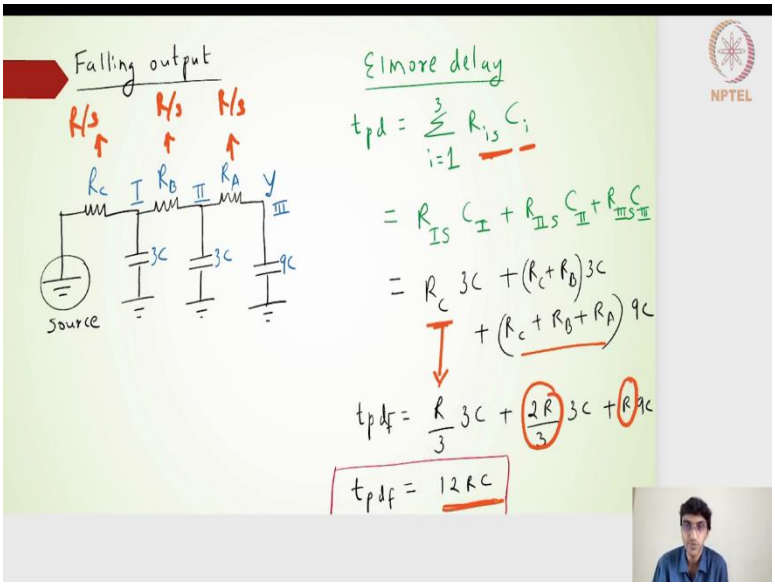
My node of interest is y and each of this capacitance has to be accommodated. I am going to write 3C capacitances, which is nothing but I am considering this capacitances I have to find an equivalent capacitance seen for this 3C. I need to find out the shared resistance.

$$3CR_{s_{3C}} + 3CR_{s_{3C}} + 9CR_{s_{9C}}$$

This is what I need to find out which will give us the propagation delay for the falling output. Hope this is clear.

The shared resistance for the 3C capacitance is here will be because it is a shared one we need to identify two paths, one path is from the source to the output, it will be $R_C + R_B + R_A$ that is one path. The second part is with respect to the individual nodes where the capacitances are seen, I will have this 3C, from source to 3C this capacitance I will have only R_C as a component. The common between the two paths of resistance is nothing but R_C, for the next node I have this $R_C + R_B + R_A$ has one path resistance and then from source to this particular 3C capacitance I have now $R_C + R_B$. The shared resistance is now $R_A + R_C + R_B$ and finally for 9C from the source to the 9C node and the node y I see you have a resistance of R_C, R_B, R_A. In terms of the output node to source path also gives me RC, R_B and R_A. The overall resistance here is $R_A + R_B + R_C$.

(Refer Slide Time: 27:45)



The Elmore delay method for the shared resistance and then the individual capacitance is

$$t_{pd} = \sum_{i=1}^{3} R_{is} C_i$$

$$= R_{IS} C_I + R_{IIS} C_{II} + R_{IIIS} C_{III}$$

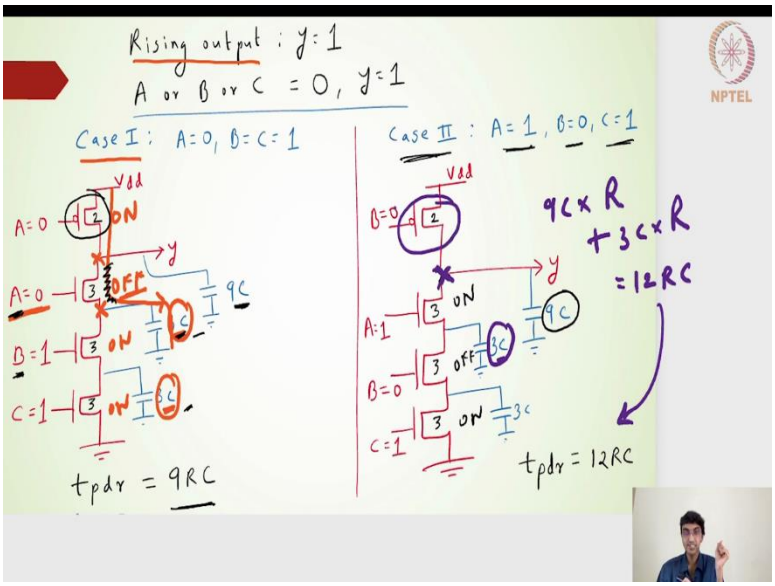
$$= R_C 3C + (R_c + R_B)3C + (R_c + R_B + R_A)9C$$

Individual resistance is R_C, R_B and R_A are nothing but R/3 that is coming from the switching resistance of the NMOS transistor which has a width of 3. This each one of them is R/3, each one of them is R/3 and R_A is also R/3.

$$t_{pdf} = \frac{R}{3}3C + \frac{2R}{3}3C + R9C = 12RC$$

Finally, I will get 12RC as my propagation delay falling assuming the inputs are not a step input. It is a very conservative approach, that is why we have not used any log2. Hope you know this is clear that the falling output is clear.

(Refer Slide Time: 28:46)



Next, we have to find out the rising output, for the rising outputs because there are three transistors on the PMOS side which are in parallel. What we can have is a three different combinations. Either one of it we can have in fact five different 5, 6, 7 different combinations where the first three combination could be the one of the transistors is on the other two transistors are off.

Then the next two combination could be two of the transistors being on then the other one being off and then the last one, the overall six combinations. The sixth combination is when all the transistors will be on the PMOS side. Out of the six combinations the first three combinations is likely to give me a larger delay because only one transistor is on if any two combinations or three combinations are on the overall equivalent resistance will be smaller because all of them are in parallel.

In that sense, if I have only one transistor being on, the first three combinations where one of the transistor is on, that we will consider and then see what is the output what is the propagation delay rising.

I have written case 1, where $A = 0$ and then $B = C = 1$. $A = 0$ represents that PMOS side it is on. This particular transistor on the PMOS side is on, the other two PMOS transistors on the pull up side is off.

On the pull down side this is 0 represents this is off and the other two is on. Other two is on, but this is off what it means is the connection between the output node to the next diffusion is off. I do not have to worry about this 3C capacitance and 3C capacitance being charged, because this charging path for this particular 3C capacitance and then this particular 3C capacitance through the V_{dd} path if I want to charge this. I have to have this particular path, the 3C should get charged, but this path because this transistor is off, this path is not there. That there is no channel, hence this 3C capacitance and then this particular 3C capacitance and this particular 3C capacitance is not likely to charge at all what is likely you need to charge is only this 9C capacitance. The 9C capacitance is going to charge with the propagation delay rising will be nothing but 9C into the switching resistance of this transistor which is R. I will get 9 RC.

$$t_{pdr} = 9RC$$

Considering this as the first case where $A = 0$ and then $B = C = 1$, what if I have B=0 and A=1; that means, there is a channel existing. There is a channel existing for A, but B is off, that means this 3C will be able to see a channel for charging, this three will also has to be charged now.

In the 2nd case I have A =1, B=0 and C=1, that means A is equal to on, C is equal to on, but B is equal to off.

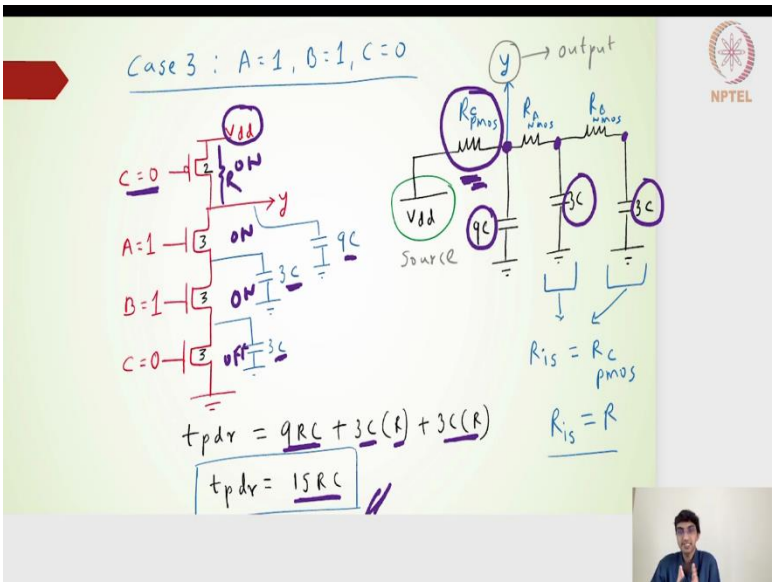
Now I have this 3C getting a path to charge, but my output node of intersect is only this. According to the Elmore delay what we will have is this 9C, this 9C multiplied by the shared resistance which will be nothing but the switching resistance of the transistor B on the PMOS side plus I will have this particular 3C, which now sees a path through the transistor A on the NMOS side the switching resistance of the NMOS side is nothing but R/3.

Then the switching resistance of this pull up side B transistor is R, but because my node of interest is only y the common shared resistance path is nothing but the B transistor of the PMOS switching resistance which is nothing but R, put together I will get 12 RC, that is what I have written here. Hope you have understood this.

$$t_{pdr} = 12RC$$

According to the Elmore delay, even though it sees a channel here the output node of intersect is y. The shared resistance is nothing but the switching resistance of this B transistor.

(Refer Slide Time: 33:12)



Case 3, when A=B=1, that means C=0, that means this is on in the pull down side and then this is on the pull down side this is off. Now I will have this 3C now, C is a charging path for the V_{dd} for connecting to the V_{dd} source. The overall Elmore delay I have written for the charging.

The source is V_{dd} but this y node is my intersect output node, I will have RC of PMOS as the switching resistance and R_A and R_B connecting to the 3C and 3C capacitance on the NMOS transistor. This is the NMOS transistor which I have stated here and then this is R_B is the NMOS transistor, which connects to the 3C capacitance here.

The shared resistance for this particular capacitance is nothing but still the R_C PMOS and the shared resistance for this 3C is also the R_C PMOS because my output is the y is my output node of intersect. I will have the propagation delay rising as,

$$t_{pdr} = 9RC + 3C(R) + 3C(R) = 15RC$$

This is my worst case the maximum delay I will get because of this particular combination. I am going to consider this in my characteristics for a 3-input NAND gate, because we will always design our gates at a transistor level such that I will get the worst case condition to satisfy the timing constraints or the timing requirements. In that sense I will have a propagation delay falling and propagation delay rising. The propagation delay rising for a 3-input NAND gate turns out to be 15RC and the propagation delay falling it turns out to be 12RC as per our previous slides.

Hope you have understood this lecture very important to understand the 3-input NAND gates, how do we design the width and how do we apply the algorithm method to find out the worst-case delay in this particular combination the worst case it turns out to be on the rising side and 3-input NAND gates on the rising side we get the worst case delay as 15RC the following delay all the three transistors for a 3-input NAND gate has to be on.

There is no worst-case delay on the falling side because all of them has to be on otherwise the output will not discharge to 0 and thereby all the three transistors has to be on and then we will get the following delay has to be 12RC.