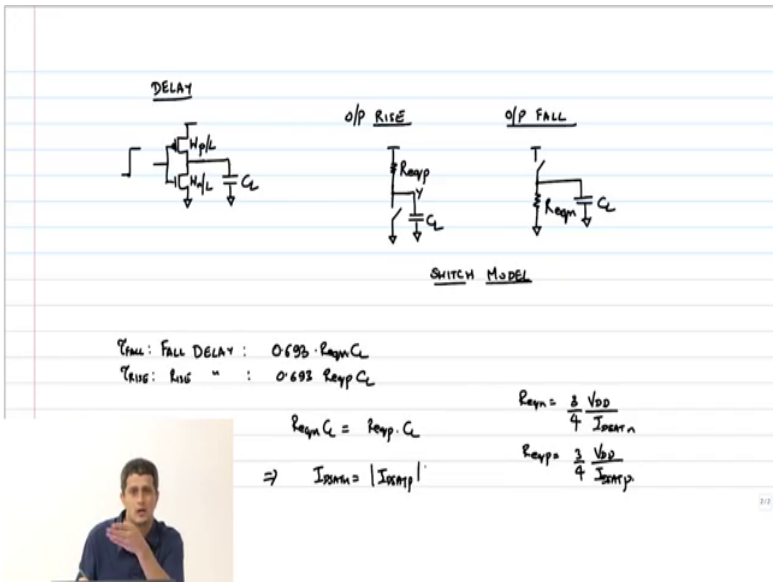
Digital IC Design
Prof. Janakiraman Viraraghavan
Department of Electrical Engineering
Indian Institute of Technology, Madras

Lecture – 27
Inverter: Transient Responsea

(Refer Slide Time: 00:17)



So, today let us move on to the heart of this subject delay ok. So, I have a CMOS inverter; this is W n this is W p and of course, some channel length is there right, but it is the same length. So, going forward I will not keep mentioning this L. I will just drop it lets assume to be the minimum channel length at 2 lambda we will just assume and there is a load capacitor here and my input is assumed to be rise instantaneously from 0 to V DD ok.

So, what is the good model in order to find out how long it takes to go to the other state and how do I model these two transistors in both cases? So, it turns out where for output rise,

output rising. The rising will happen through the PMOS transistor right and you remember that we calculated a resistance for the NMOS transistor when you discharge the capacitor right. So, similarly the PMOS will have a resistance R equivalent of p ok. This is my output Y. What happens to the NMOS transistor when you are raising? It will be off ok.

Similarly, when the output has to fall the PMOS transistor will be off and its going to get pull down to ground through a NMOS transistor and of course, I have this load capacitor C L here right, this is my switch model ok. So, earlier we said that it does not matter what the size of the NMOS and PMOS transistors widths of the NMOS and PMOS transistors are, the output logic level will swing from 0 to VDD right.

So, does it mean that we do not have to worry about the sizes, well that is not entirely true because when you come to delay you the delay indeed depends on the width of these two transistors right. So, how do I evaluate the fall delay? Output is falling NMOS, PMOS has turned off, NMOS has turned on. And what did we derive this expression to be? Yeah.

Student: (Refer Time: 03:42).

Sorry.

Student: (Refer Time: 03:45).

Yeah. So, we basically said this was 0.693 into R equivalent of p into C L right. This is right now, similarly what is the rise delay? Is 0.693 R equivalent n into C L.

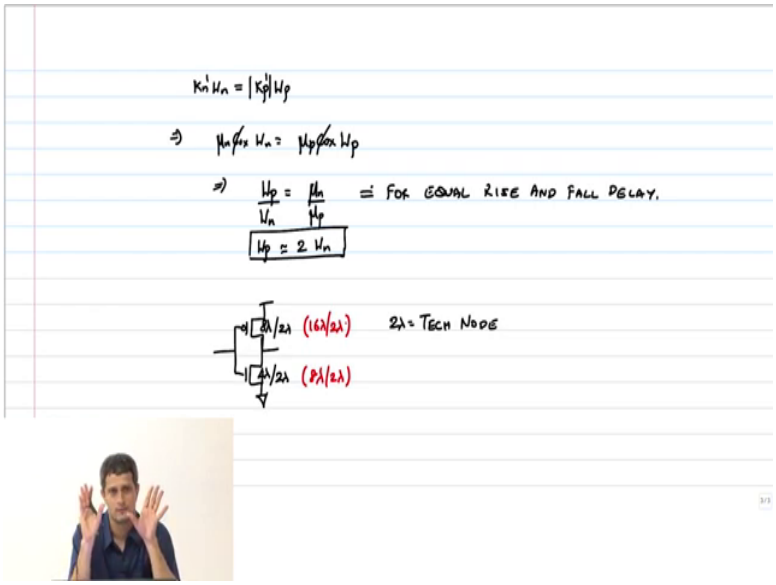
Student: (Refer Time: 04:22).

No you are right this is n and this is p yeah ok. Now, there is no reason to make the rise delay lesser than the fall delay because in any general circuit both edges are equally important for me. When the input rises I want the output to fall, when the input falls I want the output to rise, but both should happen quickly enough.

There are certain special cases like in memories for example, where you might want to turn on the memory cell very quickly, but it is we turn it off slowly. In those cases yes the rise delay may be more important than fall delay, but in general without loss of generality it is the safest assumption to make is both edges are equally important and both delays should be similar. No reason to pick one over the other therefore, I would say R equivalent n equals p into C L right.

So, what was R equivalent n? 3 by 4 V DD by I DSAT n right, R equivalent p you can check it out is it is very similar I DSAT yeah I DSAT and I DSAT p right. So, I want these two to be the same which implies that the charging current effectively if you cancel out all the common term C L V DD everything you will find that I DSAT n should be equal to mod I DSAT p ok, leaving out all the sign there because I am just talking about magnitude of current here.

(Refer Slide Time: 06:47)



Now, for this to happen you will find that if the technology is symmetric V D n equal to V D p and all that all these VDD minus V T V DSAT will all cancel out right. All of those terms will simply go away and you will find that K n prime W n should be equal to K p prime W p ok. I will say mod here. So, what is K n prime by the way? Mu and Cox W n equal to mu p Cox W p of course, you can cancel this out which implies that W p by W n should be equal to mu n by mu p; this for equal price fall delay.

So, typically this ratio is approximately equal to 2 W p is twice W n is the standard gate sizing that will be done in digital logic. You will see it everywhere because, I want rise delay and fall delay to be the same approximately equal. Of course, if your technology is very different we decide V T n and they are all very different then you have to take ratio of the I DSAT in order to find this appropriate sizing. Any questions here? So, this two is a sort of a magic number that will be used.

So, just to clarify the sign convention not the sign convention the sizing convention, we say that the length is 2 lambda right. So, 2 lambda equal to technology node. So, if it is 180 nanometer its 2 lambda is 180 that means lambda is 90. If it is a 14 nanometer technology then 2 lambda is 14 nanometer right that is the convention. Then typically this will be 4 lambda and this will length will be 2 lambda, this will be 8 lambda ok.

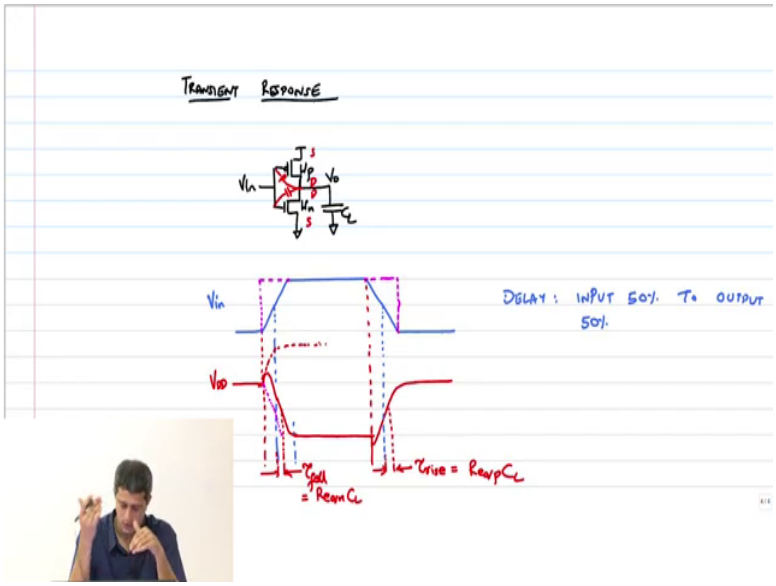
So, you will construct what is known as a reference inverter first ok. This reference inverter is what is going to give you the reference to size all other gates in your in your circuit later on. So, for example, I might need an inverter with larger drive strike. So, with twice the drive strength then I will just double both NMOS and PMOS bits right. This reference inverter it need not be that the width should always be 4 lambda and the PMOS width should always be 8 lambda.

In some technologies it could be for example, this could be 8 lambda by 2 lambda and this could be 16 lambda by 2 lambda. Then minimum width that is allowed for the NMOS right, which is 8 lambda or 4 lambda is determined by other factors in terms of manufacturability,

in terms of variability and all that. In the end of this module I will tell you I will try to give you a short primer on variability.

So, that you understand why these arguments come in right, but in the technology they will find out what that minimum width is, that is that you can successfully manufacture across devices across wafers and consistently give you the same current and same manufacture width. Then based on that you will size the other guy to be twice that; so, nothing sacrosanct about 4 lambda clear yeah.

(Refer Slide Time: 11:03)



So, let us now look at the transient response, let me just tell you exactly what this delays are; this tau rise and tau fall that we just calculated now ok. So, maybe I should just draw the innovator again CL V in V out and this is my W n W p.

So, my V in right is initially assumed to be in reality by the way let us just look at the real case like this my V in V out. If V in is 0 what is V out? V DD. So, it will start at V DD ok. So, let me just get these, it will start at 0 at V DD and then it has to fall right. Now question is will it just start falling like this? What will happen?

Student: (Refer Time: 12:50).

Student: (Refer Time: 12:58).

No, it will fall with the delay yes. So, I am saying yeah it will fall with the delay. So, maybe yeah I can write till here right and then is or you want me to draw it more to scale I can do that. So, that yeah is this how it will be. Why?

Student: (Refer Time: 13:33).

Ok.

Student: (Refer Time: 13:44).

Fine.

Student: (Refer Time: 13:48) has gone to V DD So, it (Refer Time: 13:50).

So, if you remember this is the VDD by two point right. Now this is so, there is a delay clearly right. So, I am asking you is that all that you happen. What are the capacitances now? Between the input and the output is there any capacitance? Which capacitance is there between the input and the output?

Student: (Refer Time: 14:19).

Yeah.

Student: Gate to.

Gate to.

Student: Drain.

Drain of both NMOS and PMOS right because remember for both of them this is drain, source, source. So, there is a CGD which is sitting between input and output like this right. Therefore, this cannot happen like that as soon as this input rises the voltage will not change instantaneously across that capacitor the CGD capacitor. Therefore, this will go up slightly and then drop all the way down; similarly when this input falls instantaneously this would go below ground and then come up, is that right? You agree with me or not?

You do not, why not? It is a instantaneous thing that is going, very instantaneously it will go to a small value, especially you see later you will see that when the nodes are actually floating and it is not being driven to a value either V DD or ground. This voltage can jump significantly above V DD. So, you should do this simulation and see for yourself the reason why this bump is there above V DD and below ground is because of this coupling capacitance between gate and the drain of both NMOS and PMOS transistors right.

So, therefore, now where is this was my 50 percentage point for the input. This is the 50 percentage point for the output. This delay is my tau fall, similarly this is my input here output delay, tau rise. The delay from input going to 50 percentages to output going to 50 percentage is defined as the delay ok. Input 50 percentage to output 50 percentage ok. However, remember that the expressions that we derived earlier 0.6963 R equivalent into C was assuming that my input went instantaneously from 0 to V DD, this was the assumption that we made right.

So, if you look at that delay I mean that case the delay would happen earlier like this. So, it turns out that by dropping that 0.693 we can capture this actual delay case where the input itself is rising a little slowly, input when it has a finite rise time within reasonable limits then the output will you know take slightly longer. How much longer? It turns out that if you drop that 0.693 R equivalent into C that is a fair estimate of the delay.

And therefore, we will simply say that this is equal to R equivalent n into CL and this is equivalent to R equivalent p into C. Hence forth, we will not put 0.693 in our delay expression for this reason. Its only an approximation, it is not true really, but anyway we are only looking at back of the envelope calculation estimates right. Ultimately whatever so, what does the you know you know we have we have derive quite a few expressions till now. So, it is important to understand where we intend to use them.

I am not for example, trying to use these expressions of delay and other things to figure out if inverter A is 5 percent faster than inverter B. That I can never do because the actual current model is so much more complex. I can never capture all that in my equations here, in these back of the envelope calculations. But, with these expressions I can always tell you if inverter A is 50 percent faster than inverter B or maybe 30 percent faster.

That level of estimate I can give you with these back of viewer calculations where we have simplified the model got an understanding we get some delay estimates, I can tell you yes this configuration is going to be 50 at least 30 percent faster whether it is 27 percent faster or 33 percent faster you have to do a simulation ok; that is the philosophy of these derivations that we are doing.

So, now, if you ask me is that 0.693 necessary well honestly in both inverter A and inverter B if I drop the 0.693 it does not matter, I am only looking at telling you which one is faster and even the 30 percentage faster will still hold even if I drop the 0.693 right. Keep that in mind for all the derivations throughout this ultimately you have to resort to simulation for accurate estimates, no choice there ok. So, any questions on delay? Yeah.

Student: (Refer Time: 20:33) has a capacitor (Refer Time: 20:35) they can additional capacitor they will be are in the (Refer Time: 20:40) right the output changes, why the output changes?

No the.

Students: (Refer Time: 20:45) capacitors are delayed with (Refer Time: 20:47).

True.

Student: Yeah. So, the (Refer Time: 20:48) capacitor are delayed will increase your (Refer Time: 20:49).

True.

Student: Yeah. So, the (Refer Time: 20:52) capacitors just greater than increase the delay.

Correct.

Student: But then why does the (Refer Time: 20:57).

No, see the voltage cannot change instantaneously across the capacitor right. So, the C DV by DT you cannot have an instantaneous change which means that there is an infinite current always correct. So, to maintain that the other end will also follow the input for a short while right, this is a sort of a very high frequency switch that is happening when the input changes instantaneously. Therefore, it will the other plate of the capacitor will follow this nod a little bit and that is why that initial bump is happening there.

Student: Why frequency fills the other (Refer Time: 21:38).

No that is why it is not following exactly its not for example, this node is just not going up like this right. The input is actually going all the way to VDD. That is not happening with just that momentary thing it will go and then of course, the other NMOS now the NMOS transistor turns on and so, therefore it will fall. So, there are multiple effects happening there right. One is the capacitor which is causing this coupling other is the NMOS transistor now turning on. So, now, that will take on and then pull that node voltage downright. However, if it was not if the NMOS did not turn on let us say in this process then this node voltage would jump to a much higher value.

And, it would actually remain there after that. it can remain at a value much higher than V DD after that because, nothing is there to pull it down. This is what I meant where when it is called a floating node, I will come to this later you will see this happen in your simulations also yeah.

Student: Does the (Refer Time: 22:48).

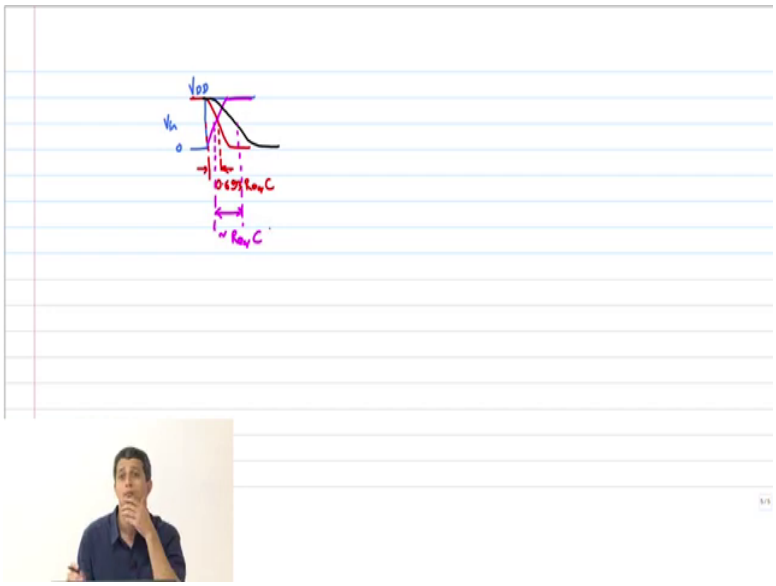
No. So, we have lumped all the capacitances as CL for now actually, later I will show you I will show you how we are going to account for all these capacitances; it does come into the picture.

Student: Here it (Refer Time: 23:09).

Yeah, actually it is just a representative capacitor for now, I will tell you what these elements are when I move on into the course.

Student: (Refer Time: 23:18) why do we drop the 0.693 (Refer Time: 23:19).

(Refer Slide Time: 23:26)



Yeah. So, see what I said was when we did that 0.693 derivation the assumption was V in went from 0 to V DD instantaneously ok. Now, you leave all this output you know rising and all that forget all that, output will now fall like this right. So, this delay we said was 0.693 R equivalent into C. Now what am I saying? I am saying the input itself is rising slowly; that means, the output is actually going to take that much longer.

I have run out of colors because someone told me that green is not visible from there; so, I will use black. So, this output when with the magenta input where the input is rising slowly will take longer right. So, in that case the delay is actually from here till here and it is longer delay. How much longer? 1 by 0.693 times longer is all we are saying. So, this we are saying is approximately R equivalent into C.

Student: Rising and falling across the input crosses are V t input.

Output.

Student: After the V T in the when input crosses are not allowed that (Refer Time: 24:59).

Correct in the sense you are right. So, here the you are saying that somewhere here is V Tn and only then it will fall correct. So, what will all right you are right. So, what will happen is until V Tn that capacitor will be floating in the sense the output will start output will just follow the input then the NMOS turns on, then it will start dropping you are right. So, again it is too cluttered for me to show all of that here, but you are right you have to look at the point at which at V Tn, up to V Tn it will follow. Output will follow the input and then start falling. Any other questions on delay? Ok.