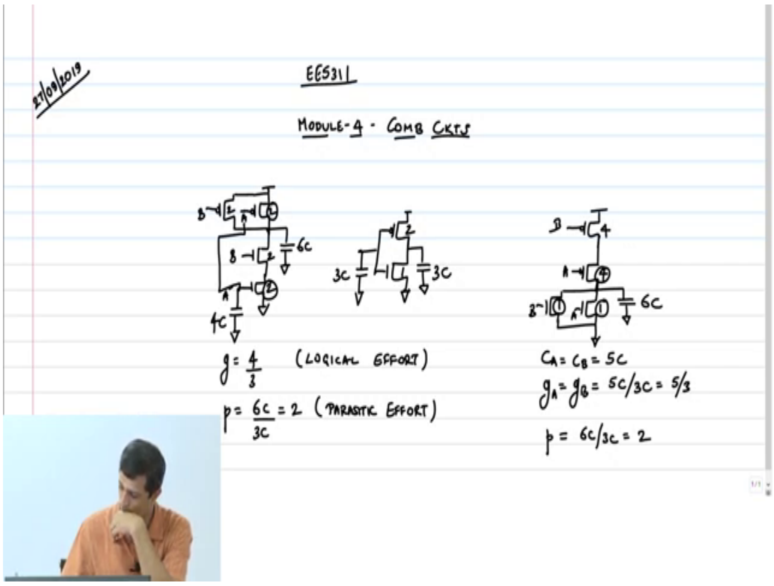
Digital IC Design
Prof. Janakiraman Viraraghavan
Department of Electrical Engineering
Indian Institute of Technology, Madras

Lecture - 41
Gate Delay

Good morning. Let us proceed with our discussion on the logical effort ok.

(Refer Slide Time: 00:22)



27/09/2019

EE5311

MODULE-4 · COMB CKTS

$g = \frac{4}{3}$ (LOGICAL EFFORT)

$p = \frac{6C}{3C} = 2$ (PARASITIC EFFORT)

$C_A = C_B = 5C$

$g_A = g_B = 5C/3C = 5/3$

$p = 6C/3C = 2$

NAND 2: A B, A and B right. Then size is 2, 2, 2 and 2. So, what is the net capacitance on input A? It is sum of both NMOS and PMOS capacitance right. So therefore, you will get 4 C: 2 C from here and 2 C from here right. And logical effort is defined as the ratio of input gate capacitance to the gate capacitance of a static CMOS inverter with identical drive strength ok.

So, because we have sized the NAND gate already to have 2, 2, 2, 2, wherein we have made sure that the resistances are the same already our reference inverter here is just going to be the standard reference CMOS inverter which is going to be 1, 2. So, this capacitance will be 3 C. So, the logical effort will be how much?

Student: 4.

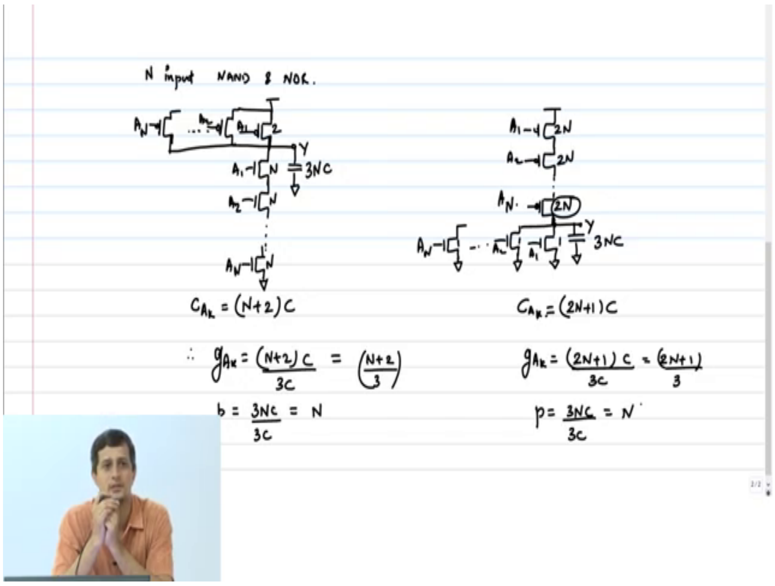
4 by 3. What about the parasitic effort or the parasitic delay? Right. If you look at the output capacitance, your diffusion capacitance it is 6 C and here it is 3 C, right. Therefore, p is going to be 6 C by 3 C that is equal to 2. Logical effort we denote by g, parasitic effort we are going to denote by p right. The parasitic delay we also normalize parasitic delay we also call parasitic effort p ok. This is logical effort and this is ok.

So similarly, let us do one more for the NOR 2: A B, A and B this is VDD; sizes are 1, 1, 4, 4. What is the net output gap? Again 6 C: 4 coming from here, 1 coming from each of these; 6 C is the output cap. Capacitance C A equal to C B equals how much? 5 C: 4 coming from PMOS, 1 coming from the NMOS; so, net capacitance is.

So, logical effort g; actually remember logical effort can be different for different inputs, in these cases just they just happen to be the same; g B equals how much? 5 C divided by 3 C: that is 5 by 3. And parasitic effort, parasitic effort is obviously not depend on the input its only the output diffusion capacitance that we are counting, right; p is going to be 6 C divided by 3 C you get a parasitic delay effort of 2 ok.

So, what is the logical effort for an N input NAND gate and an N input NOR gate? Can you calculate that now?

(Refer Slide Time: 04:56)



N input NAND & NOR.

$C_{A_k} = (N+2)C$

$\therefore g_{A_k} = \frac{(N+2)C}{3C} = \left(\frac{N+2}{3}\right)$

$b = \frac{3NC}{3C} = N$

$C_{A_k} = (2N+1)C$

$g_{A_k} = \frac{(2N+1)C}{3C} = \frac{(2N+1)}{3}$

$p = \frac{3NC}{3C} = N$

Yeah.

Student: Same as (Refer Time: 05:20) same.

Why? Same as what?

Student: (Refer Time: 05:28) NAND gate should be (Refer Time: 05:31).

Is that so? So, if you take a NAND you have N NMOS transistors on the stack size is N, N, N; and PMOS is all 2, like that you go right Y. So, the net capacitance here is how much? 3 N 3 N into C right, 2 C coming from each of the N PMOS's and N C coming from the top

NMOS; so, you get 3 N C. And what is the input capacitance? C A k. So, this is A 1 I mean or A 1, A 2, A N; A 1 A 2 A N is N plus N plus 2 into C.

Therefore, logical effort on each input is N plus 2 into C divided by 3 C because, I have to now compare this to a reference inverter with size 1 2; that is how we have size the. Later you will see that I can actually do different sizing for the NMOS and PMOS in which case the comparison with the reference inverter will be different, right. That will be a different inverter and therefore you will get a different logical effort, ok.

Anyway, so you get now the answer as N plus 2 divided by 3, ok. What about the NOR?. So, this is size 1, 1, 1. By the way what is the parasitic effort here on the NAND?

Student: (Refer Time: 08:28).

It is?

Student: (Refer Time: 08:30) N.

3 N C divided by 3 C which is equal to N, ok. Now, what is the input capacitance on each input A 1 A 2 A N; A 1 A 2 and A N. So, what is the PMOS size by the way?

Student: (Refer Time: 09:03).

Huh.

Student: 2 N.

2 N; 2 N. So, what is the net capacitance?

Student: 2 N.

2 N.

Student: plus 1.

Plus 1 into C: 2 N coming from PMOS one coming from NMOS. Therefore, the logical effort g A k is how much? 2 N plus 1 into C by 3 C; that is equal to 2 N plus 1 divided by 3. What about the parasitic effort? You get 2 N coming from the PMOS and N C coming from the NMOSs, so you get again 3 N C. So, this is 3 N C divided by 3 C equals back to N.

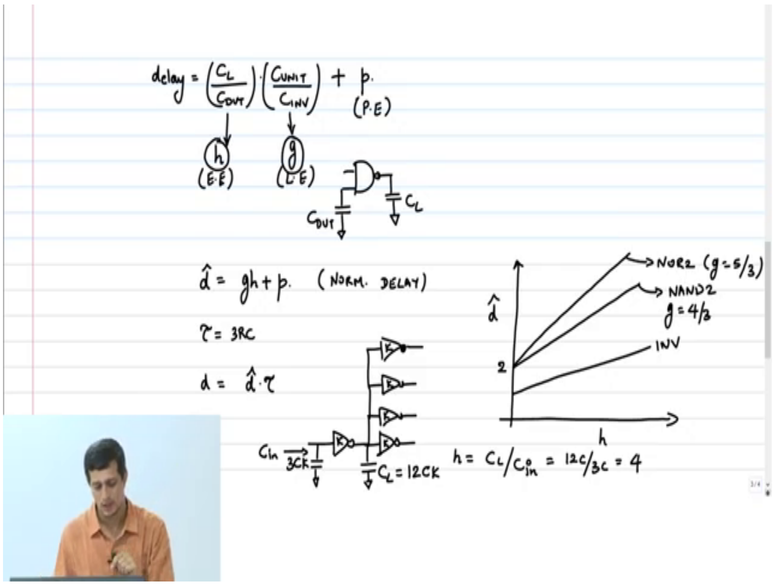
So, again if you look at this, the logical effort as I increase the stack be it NMOS or PMOS is going to increase, right. So, apart from the fact that your parasitic delay went as square of N; earlier you remember we did this derivation the parasitic delay goes as N square. So, putting these two reasons together its obvious that you do not want to have a very large stack.

The moment you have a large stack your logical effort you go up, your parasitic delay will start going up significantly ok. This of course, this approximation does not hold their counting just the output diffusion capacitance, you have to calculate the actual Elmore delay, right. This approximation is also assuming that the stack is not very large ok.

So therefore, do not make transistors with more than 4; do not make logic gates with more than 4 transistors on either stack. Clearly, it is much worse to stack transistors on the PMOS rather than on the NMOS, because the logical effort is now going up as two N plus 1 by 3. Again what does this logical effort mean? It means that in order to achieve the same resistance I need to pay a higher price on the input capacitance.

This 2 N plus 1 by 3 compared to N plus 2 by 3 of the NANDs tells me that I need to pay a higher price on the input capacitance for NORs compared to NANDs in order to achieve the same resistance, right. And there by the similar delay of charging load capacitance.

(Refer Slide Time: 12:15)



$$\text{delay} = \left(\frac{C_L}{C_{DUT}}\right) \cdot \left(\frac{C_{UNIT}}{C_{INV}}\right) + p.$$

$\boxed{h}$ (E.E) $\quad \boxed{g}$ (L.E) $\quad$ (P.E)

$\hat{d} = gh + p.$ (NORM. DELAY)

$\tau = 3RC$

$d = \hat{d} \cdot \tau$

$h = C_L / C_{in} = 12c/3c = 4$

So, let us just get back to our delay that we talked about yesterday ok. We said that we had let us say some device under test and it was driving a load capacitance C L and its a capacitance was C DUT, right. That is the gate capacitance that it will represent after up sizing and all that, not the unit gates capacitance; what the gate actually presents in the circuit. So, this reset we could write as C L by C DUT into C unit by C inverter right plus the parasitic delay p. We define this to be the logical effort g.

And this as I told you yesterday is nothing but the electrical effort h right. So, this is parasitic effort, this is logical effort, and this is called electrical effort. The electrical effort depends on the drive strength of the gate. If I have A 1 x gate then it presents less capacitance, if I have a 10 x gate say gate has been up size by a factor of 10 then it presents a larger capacitance and therefore the electrical effort will change, ok.

So, this is the guy that actually captures the sizing part of the gate. The g only captures the logical part, whatever logic function it is implementing it captures in some abstract form that complexity, right a tradeoff between resistance and capacitance it captures. Now the presence of the actual capacitance in the circuit, because of upsizing that gate is captured by the electrical effort E. And therefore, the normalized delay I will call it d hat can be written as g h plus p, ok.

And how do you get the absolute delay from this you just have to multiply it by the delay of the reference inverter that we divided by. So, let me call the reference inverter delay as tau, this was 3 RC, right under no load condition inverter driving no load then the delay was 3 RC. So, if you want to get the actual delay you multiply the normalized delay by tau, normalized here.

So, what is this saying: if I plot my normalized delay as a function of the electrical effort h I will get a straight line with some offset on the y axis, right. So, if I take a NAND gate it would be something like this. For an inverter right, obviously it is simply going g is 1 by definition, p is also 1 right.

So, the slope will be the best for an inverter right. And this is for a NAND where the delay starts at 2, because the parasitic delay itself is 2 right; even if you put no load you will get that. Now for an inverter it would be something like this. Lower slope this is an inverter, this is a NAND 2; a slope is g equals 4 by 3 logical effort. A NOR 2 would have the same parasitic delay. So, it would start at the same point, but the slope will be what.

Student: 5 by 3.

5 by 3. So, you will have NOR 2 g equals 5 by 3 ok. So, let us look at this thing. So, if I have an inverter driving 4 copies of itself, you remember this we called as fan out of 4 delay right because there are 4 that the output is fanning out into 4 different gates and it is driving identical copies of itself; that is what do you mean by fan out of 4. So, what is the net kept load capacitance on this load capacitance C L? Sorry.

Student: (Refer Time: 18:25).

No, only load capacitance not parasitic capacitance.

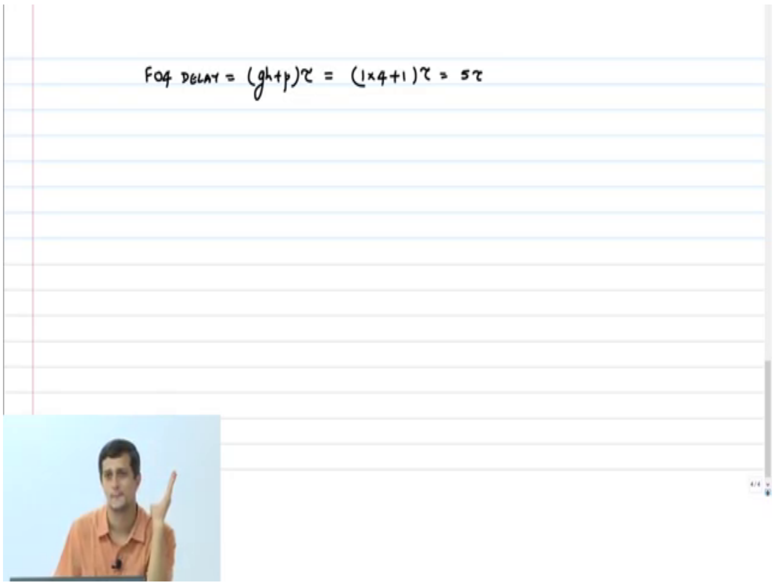
Student: (Refer Time: 18:32).

Yeah 4 times. So, let us assume that this guy is presenting a capacitance of 3 C ok. So, what is C L equal to then?

Student: 12.

12 C right, because all are identical inverters. Therefore, my h is going to be what output capacitance C load divided by C input; whatever input capacitance you are going to see here this is C in and that is 3 C. So, you can write this as 12 C by 3 C equals 4.

(Refer Slide Time: 19:22)



FO4 DELAY = $(gh+p)\tau$ = $(1\times4+1)\tau$ = 5τ

So, therefore FO4 delay is g h plus p into tau the actual delay. So, this I can write as 1 into 4 plus 1 into tau equals 5 tau; just 5 tau ok. Now you think about this. Suppose this inverter were not the unit inverter and instead I upsize everything by a factor of k; everything the whole thing went up by a factor of k, then what would happen all capacitances would go up by a factor of k, correct. That is the only thing that the only thing that is going to affect the delay is h because of the gate sizing right, other things are just constant irrespective of the gate size.

So, if you write this you will find that this also goes up right the h will be 12 k C by 3 k C that is again equal to 4. So, it does not matter when your if I tell you fan out of 4 delay in a technology it does not matter what the size of the inverter was. You understand the fan out of 4 delay is the same whether you did it for a unit inverter or for an up sized inverter or for an inverter with higher drive strength. Because, the delay comes out to be the same the h is effectively the same, because I am scaling up all capacitances by the same factor.

So, if I tell you in a technology the fan out of 4 delay is some 100 picoseconds, then you do not have to ask any further questions, you know that. Therefore, the inverter delay is just one-fifth of that, clear. So, this FO4 delay is an important characterization that people will do in VLSI. When you go to a new technology first thing they will characterize is what is the FO4 delay of the inverter ok.

Just a thumb rule this FO4 delay if you are in if you are in a technology 180 nanometer then the FO4 delay is typically between half and one-third of the technology mode in picoseconds. If you are in 180 nanometer half of that is 90, one-third is 60. So, the FO4 delay will typically be between 60 to 90 picoseconds ok. And then you can calculate what the inverter delay is; which is just one-fifth of the FO4 delay ok.