Integrated Circuits and Applications
Prof. Shaik Rafi Ahamed
Department of Electronics and Electrical Engineering Indian Institute of Technology, Guwahati

Digital CMOS Circuits
Lecture – 43
Transient Response of CMOS NAND and NOR Gates

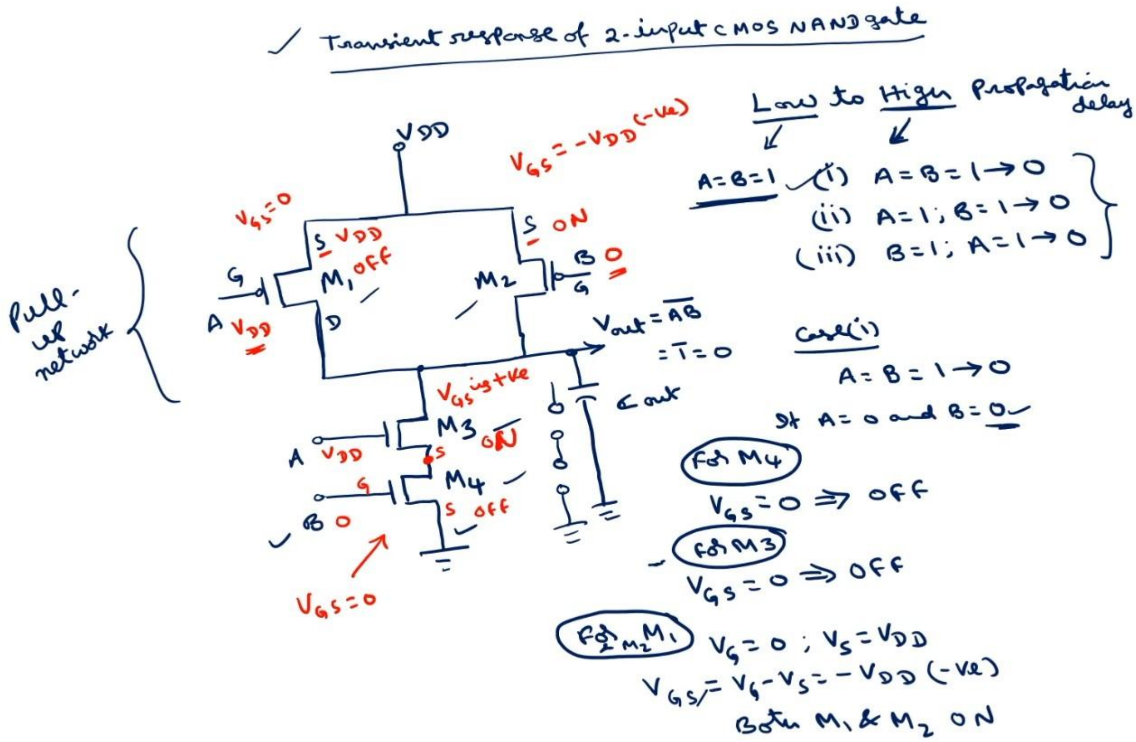
So, in the last lecture, we have discussed the operation of CMOS 2 input NAND gate. Today, we will discuss the transient response of this 2 input CMOS NAND gate. This is the circuit diagram of 2 input NAND gate. In the pull up network, we will be having 2 N-mass devices and pull down 2 P MOS devices in series, and here we are going to take the output either Y or you can call also called as v_{out}, which is V_{DD}. So, in the transient response, we know that there will be 2 transitions: high to low and low to high. Low to high, we will call as rise time rather or if I take the time taken to change from 0 to 50% of the final V_{DD}, then it is called as a propagation delay.

From here onwards we will consider only the propagation delays. So, low to high propagation delay. So, initially, it was low, now it has to charge it to high. So, under which conditions the output of 2 input NAND gate will be low? For the output of the 2 input NAND gate to be low both the input should be 1.

This will be low only if $A = B = 1$, output $V_0 = \overline{AB}$. If both are 1, this will be $\overline{1} = 0$. If any one of the input is 0, output is 1. So, low to high transition means $A = B = 1$. Now to convert to this high state so, what are the possibilities? We have 3 possibilities.

Initially in the low state means $A = B = 1$. One possibility is, if both A and B changes from 1 to 0, then there will be a low to high transition. Second case, is $A = 1$, whereas, B changes from 1 to 0. In this case also output will be high. The third case is $B = 1$, A changes from 1 to 0.

So, in these 3 cases, the output will changes from low to high. So, what are the propagation delays in each of these cases? Initially if I take the case 1, $A = B = 1$ to 0 transition. If both changes from 1 to 0 so, what happens to this PMOS and NMOS device? Both $A = 0$, $B = 0$. What happens to this transistors? We call this as M_1, M_2, M_3, M_4. For M_4, what is a V_{GS}? This gate and the source both are at the same potential so, 0 implies OFF.



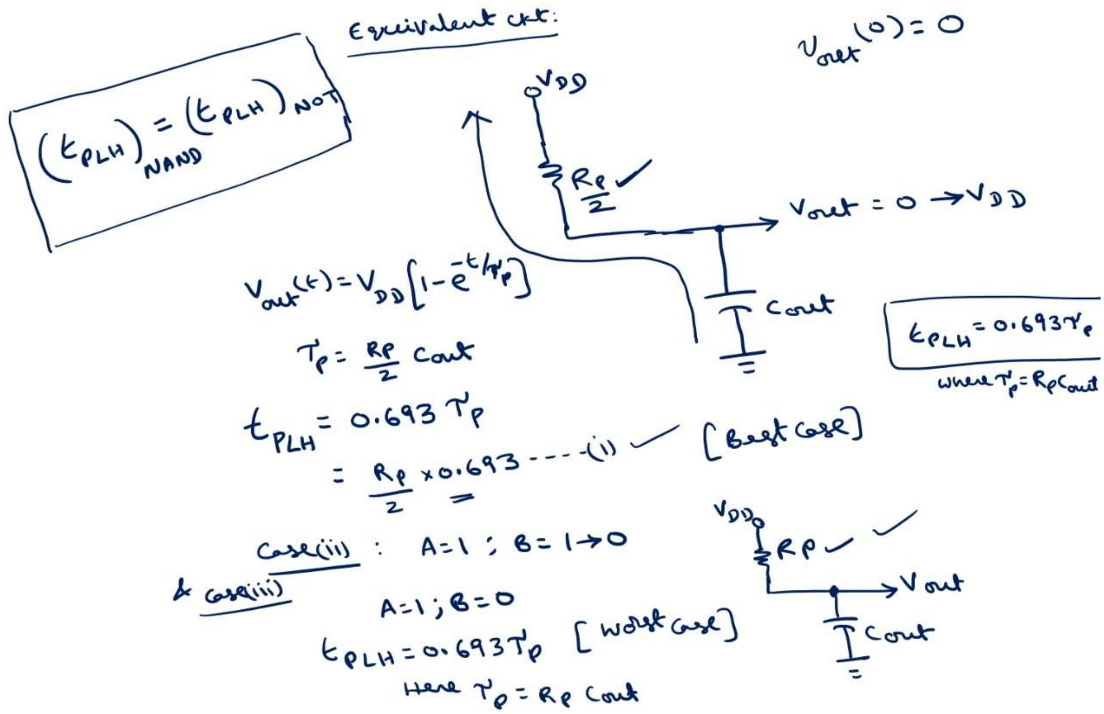
(Refer to the slide at 13:28)

For M_3, V_{GS} is again 0, implies OFF. For M_1, what is V_{GS}? This is source, this is drain, this is gate. V_{GS} is because this is 0, V_{GS} = 0, V_S = V_{DD} for both the M_1 as well as M_2. V_{GS} = 0 because $A = B = 0$, V_S = V_{DD}. So, for both the cases, what is V_{GS}? V_G - V_S, which = -V_{DD}.

So, PMOS device V_{GS} is negative means both will be ON. So, these two will OFF; these two will ON. Now, in order to obtain the transient response, you have to take the equivalent circuit. So, we know that from the RC model of this transistor, this MOSFET, if a device is OFF, it will be acts as open circuit. So, these two will be acts as open circuit, means you can remove this part, whereas, this pull up network there will be ON resistances of R_p, and the switch will be closed.

So, what will be equivalent circuit now? This is V_{DD}; here, this will be ON resistance, this is PMOS transistor. So, R_p switch will be closed, that I am not showing here, this is the output v_{out} then this part will be rounded. And in order to find out this transient response actually you have to take a capacitance here C_{out}, which is V_{DD} and here we have C_{out}. As we are discussing about the low to high transition, initially, this was low that means, $v_{out}(0) = 0$. Then, it will charges towards V_{DD} through this path.

If I take this equivalent circuit this R_p, R_p are in parallel. So, resultant will be $\frac{R_p}{2}$; you can make this equivalent circuit as $\frac{R_p}{2}$. Now, this capacitor will charges in this direction. So, initially, this was 0; now it will charges to V_{DD}, that is logic 0 to logic 1, which is low to high transition. So, what is the charging expression? We know that $v_{out}(t) = V_{DD}(1 - e^{\frac{-t}{\tau_p}})$, where $\tau_p = \frac{R_p}{2} C_{out}$ because the equivalent resistance is $\frac{R_p}{2}$.



(Refer to the slide at 17:45)

So, the propagation delay t_{PLH} low to high is equal to this, we have derived as this is approximately equal to 0.7 times, the exact value is $0.693\tau_p$. So, this is equal to $\frac{R_p}{2} \times 0.693$. So, this is the propagation delay from low to high in the first case, where both A and B are changing from 1 to 0. So, in the other case, if A = 1, B changes from 1 to 0, or B = 1, A changes from 1 to 0, what will be the t_{PLH}? If I consider the second case, A = 1, B changes from 1 to 0 means A = 1, B = 0. So, what happens to the transistors M_1, M_2, M_3, M_4? So, A is 1, this is V_{DD}, B is 0, this is V_{DD}, this is 0. Now, you can see the V_{GS} of each of the transistors. The V_{GS} of this transistor is because this is gate and source. Both will be at ground potential. So, V_{GS} is 0 implies this is OFF. What about V_{GS} of M_3? This is almost at ground potential source whereas $V_G = V_{DD}$. So, V_{GS} of M_3 is positive. This is n channel MOSFET; V_{GS} is positive, means this will be ON. Now, what is V_{GS} of M_1? This is V_{DD}, source is also at V_{DD}.

So, V_{GS} is 0, implies OFF that here V_{GS}, $V_G = 0$, V_{GS} is V_{DD}. So, $V_{GS} = 0 - V_{DD}$, negative means ON. So, M_2 and M_3 are ON and M_1 and M_4 are OFF. What will be the equivalent circuit? Instead of two paths, only one path is ON, this is R_p, and here you are taking v_{out} and here C_{out} capacitance. Here anyhow one is open circuit and another is having R_n.

Because of this open circuit, this part can be removed from the equivalent circuit. Now, this is the equivalent circuit. So, what will be expression for this $t_{PLH} = 0.693\tau_p$, but here $\tau_p = R_p C_{out}$. Because here the resistance is R_p, here resistance is $\frac{R_p}{2}$.

This is in case 2, case 3 is also similar. In case 3, what happens is so, this will be OFF, this will be ON. Anyhow, we have here one is short circuit, another is open circuit, another is R_n. So, this part can be removed. Whereas, here, between these two, this was ON; now, this will becomes OFF, the third case, this will be ON.

So, there will be R_p resistance through this path. So, the equivalent circuit is exactly same as this. This will be exactly same in both the case 2 as well as case 3. Now, if you see these two different values of the t_{PLH} in one case, this is $R_p \times \frac{R_p}{2} \times 0.693$, in our case, $0.693 \times R_p \times C_{out}$. So, one is called worst case, another is called best case. So, which one is worst and which one is best? Best case is the case where the propagation delay is less; this is called best case. So, here, in this case, propagation delay is less. So, this is best case. Whereas here, the propagation delay is more when compare with the previous case this is called worst case. But in the design, you have to consider the worst case. If we design a circuit which works well with worst case conditions also means it will work with best case also. So, normally, it is a practice to design the circuits which can operates in the worst case also. So, we will take this t_{PLH} as the worst case value, which is $0.693\tau_p$, where $\tau_p = R_p C_{out}$. Now, we have derived this t_{PLH} for the inverter also. So, inverter you have got as same $0.693\tau_p$. So, one conclusion here is $(t_{PLH})_{NAND}$ gate is same as $(t_{PLH})_{INV/NOT}$ gate this is one of the important conclusion.

Now, we will consider t_{PLH} high to low propagation delay. Now, what is the condition for this high to low? v_{out} was initially high, means for what values of A B output will be high for NAND gate means one of the input should be 0. There are three conditions where the output of the NAND gate is high. So, what are the three conditions? $A = B = 0$; if both are 0s 0 into 0 0 0 bar is 1. If A = 0, B = 1. In this case, also 0 into 1 is 0, 0 bar is 1, B = 0, A = 1.

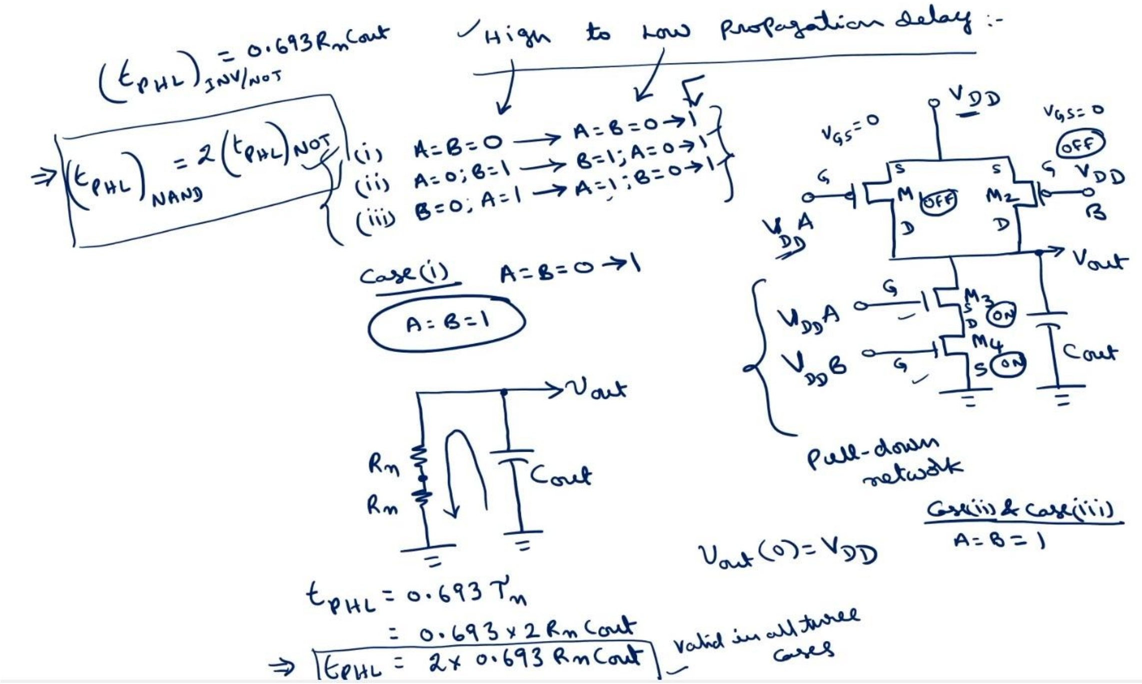
So, there are three cases in which the output is high. If you want to change to low so, the first condition is both you can change from 0 to 1. A and B are changing from 0 to 1. So, that now both A and B will be 1 1 into 1 is 1 1 bar will be 0 for the low this is the one condition corresponding to this. Corresponding to this, how to change from high to low? So, in these three cases, it was in high now, you have to change to corresponding low. So, in this case, what you have to do? If you want to change to 1 B, you fix it 1 already 1 A, you change from 0 to 1.

So, that the output becomes low. In the third case, already A = 1. Now you change from you change B from 0 to 1. So, these are the three different cases. So, similar to this low to high transition, we have three case out of these three cases. In one case you will get worst delay. In other cases, you will get the worst delay. Now, here, which is the best case and which are the worst cases? So, if I consider the first case, case 1, both A and B are changing from 0 to 1, that means, $A = B = 1$. So, for $A = B = 1$ what is the status? This is equal to 1, this is also 1. This is also 1 means V_{DD}, this is also V_{DD}.

What is V_{GS} of this transistor? If I call this transistors as M_1, M_2, M_3, M_4, this is gate, this is source, this is gate, this is source, this is drain of M_4, this is drain of M_1, drain of M_2, source of M_1, source of M_2, gate of M_1, gate of M_2. So, what is V_{GS} of M_4 is positive, n channel positive means on this is also V_{GS} is positive on whereas, V_{GS} here will be because G also at V_{DD} S also at V_{DD}. So, this will be V_{GS} is 0 implies off this is also V_{GS} is 0 this will be off. So, it will be equivalent circuit; off means will be having open circuit. So, only now pull down network will be active.

So, we have this output two resistors, and this is capacitance C_{out}; this is v_{out} high to low, means initially, v_{out} of 0 was V_{DD} high. Now, this will discharge the capacitor through the resistors to ground this is R_{in}. Of course, here also between this point and ground also there will be some capacitance I am not considering that case. Now, here what will be the propagation delay $t_{PHL} = 0.693\tau_n$ here what is $\tau_p = 2R_nC_{out} = 2 \times 0.693R_nC_{out}$, this is in case 1. We will see what happens in case 2 and case 3. In case 2, what happens, B = 1 A also will becomes 0 to 1 this is same as case 1 case 3 also same as case 1. Because in all these three cases, finally, ultimately, we are going to get A = 1, B = 1, A = 1, B = 1 B = 1. A = 1. Ultimately A = B = 1 this is same case as case 1. So, in all the three cases, this expression is valid.

There is no worst case and best case in case of high to low transition. But we know that for the inverter what is t_{PHL} of inverter which you have derived in the previous lecture is simply $0.693R_nC_{out}$. You know this is twice that of this. So, one important conclusion is $(t_{PHL})_{NAND} = (t_{PHL})_{NOT/INV}$.



(Refer to the slide at 26:56)

That means, if I connect two NMOS devices in series like here. So, the propagation delay is going to be increased means the circuit becomes slower. So, we have to normally avoid to connect the two NMOS devices in series. Now, in case of the sizing, how this transistors has to be sized so, that the propagation delays of the NAND gate and inverter are same. This is the NAND gate these two input NAND gate this is inverter or NOT gate.

Now, we know that the propagation delay of the inverter is. In one case, it is R_n, high to low, $t_{PHL} = 0.693 R_n C_{out}$. Because high to low is which I mean network is responsible for the high to low is pull down network. So, pull down network means n type MOSFET so, R_n is there.

Here as t_{PHL} means low to high means pull up network has to be active. So, pull up network consists of the p type. So, this will be $0.693 R_n C_{out}$. In any case, the propagation delay is proportional to the on resistance.

For modeling purposes, we will assume that $R_n = \frac{1}{\mu_n C_{ox} (\frac{W}{L}) n [V_{DD} - V_{T,n}]}$. This is $\frac{W}{L}$ of n type this one and $R_p = \frac{1}{\mu_p C_{ox} (\frac{W}{L}) p [V_{DD} - V_{T,p}]}$. Because $V_{T,p}$ is negative so, we will normally take modulus. We can see that the propagation delay is directly proportional to on resistance, but this is inversely proportional to $\frac{W}{L}$ ratio. So, propagation delay is directly proportional to $\frac{L}{W}$ ratio.

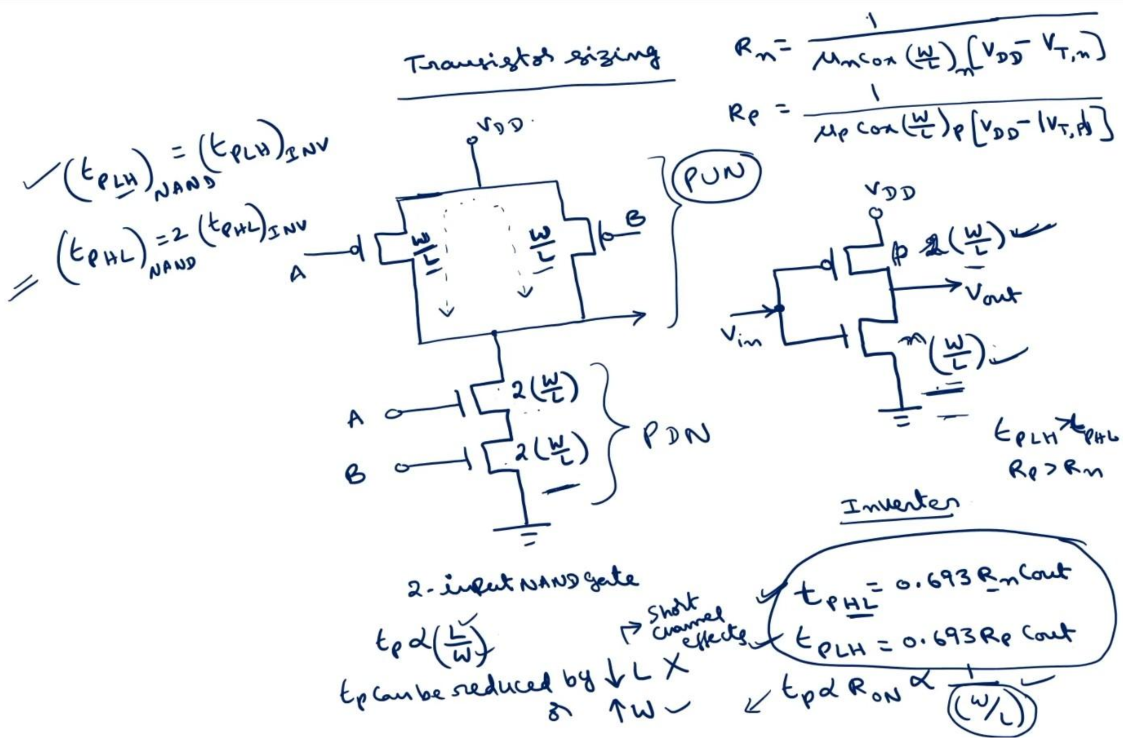
So, in order to reduce this propagation delay, either we have to reduce L, or you have to increase the W by reducing the L or increasing the W. Reducing the L is ruled out because if I reduce the channel length, there will be some short channel effects which you may have studied in your electronic circuits. So, this is avoided due to short channel effects. So, the only option is you can increase the width of the transistors. So, if I increase the width, then propagation delay will also will increase.

So, if I assume that these two will be having $\frac{W}{L}$ ratio same $\frac{W}{L}$ ratio, then $(t_{PLH})_{NAND} = (t_{PLH})_{NOT/INV}$. Whereas, $(t_{PLH})_{NAND} = 2(t_{PLH})_{INV}$. So, here in order to have this $\frac{W}{L}$ ratio. So, in this pull up network, we call this as pull up network, and this as pull down network. So, we can fix this also as $\frac{W}{L}, \frac{W}{L}$ because the current either flow in this direction or in this direction.

So, because of that, this propagation delay of NAND gate and inverter from low to high transition will be maintained. So, for low to high transition so, of this output. So, this pull up network is active. So, we can fix this $\frac{W}{L}$ ratio of the P channel MOSFETs, which is similar to that of the inverter because this propagation delays are equal. Now, the problem here is in case of the high to low transition of NAND gate is twice that of inverter.

So, to make this equal the propagation delays of this also should be equal to that of the inverter. Now, to choose this $2\frac{W}{L}$ of that of this N channel MOSFET. So, if I choose twice this $\frac{W}{L}$ ratio T P is inversely proportional to $\frac{W}{L}$; if I increase this $\frac{W}{L}$, then what happens this will decrease. This propagation delay was twice that of this inverter, by choosing the twice, the $\frac{W}{L}$ width of these two N type transistors then you can make the propagation delays equal. This is one important conclusion, if you want to have the same propagation delay as that of inverter.

In pull up network, you have to choose the same $\frac{W}{L}$ whereas, in pull down network you have to choose twice that of $\frac{W}{L}$ in case of two input NAND gate. Here also normally to maintain these two same here this low to high and high to low propagation delays of inverter are different. And we also shown that t_{PLH} is greater than t_{PHL} because R_p is greater than R_n because the mobility of electrons is greater than mobility of holes. R_p will be greater than R_n. In order to have the same t_{PLH} and t_{PHL} for this inverter also, normally, we will choose this is equal to twice that of the $\frac{W}{L}$ ratio of the N channel MOSFET. Because the mobility of electrons is approximately equal twice that of the holes to maintain the same t_{PHL} and t_{PLH} normally, we will choose this has a $2\frac{W}{L}$.

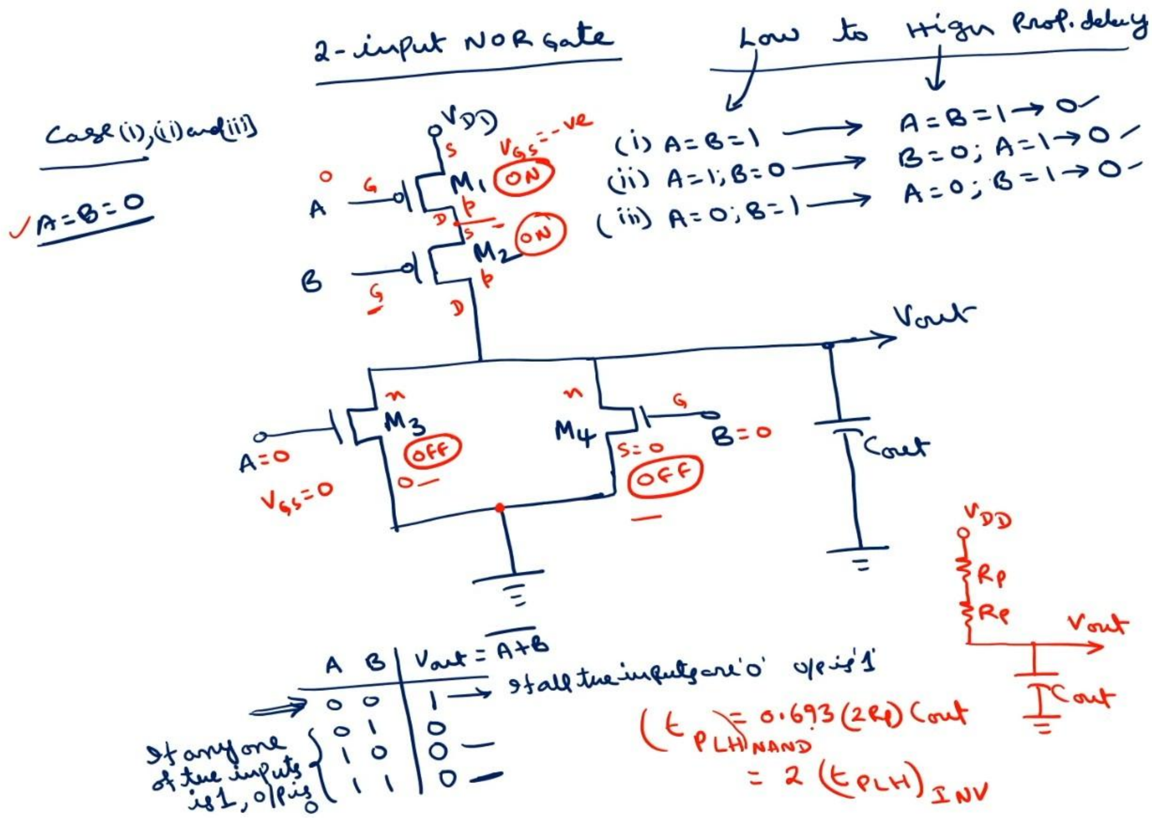


(Refer to the slide at 35:48)

So, this is about the transistor sizing. Now coming for the two input NOR gate. The circuit diagram consists of the two P channel devices connected in series, whereas N channel will be in parallel. Directly we will consider the transient response in that we will discuss the operation also. If I first consider low to high transition propagation delay. For the NOR gate, you know that if A B are the inputs, v_{out} is the output 0 0 0 plus 0 is 0 0 bar is 1 because $v_{out} = \overline{A+B}$. 0 1 also 1 plus 0 is 1 1 bar becomes 0 1 0 also 1 plus 0 is 1 1 bar is 0 1 1 is 1 plus 1 is 1 1 bar is 0. That is, if all the inputs are 0, then the output is 1. If any one of the input is 1, output is 0. Whereas, here, if all the inputs are 0, output is 1. So, initially this was in low is what will be the possibilities of A and B.

So, that output is low. Again, there are three cases here both A and B can be 1, in that case, also output is low. Second case is A is 1 and B is 0. In this case third is A is 0 B = 1. In all the three cases, it will be low to get the output as high. So, what are the three different transitions that we have to consider to get high all the input should be 0. So, we can change both A and B from 1 to 0, whereas, in the second case, B is already 0.

So, you change A from 1 to 0 here, A is already 0 B you change from 1 to 0. So, these are the three cases. So, we will see which case is best or worst, or all the three cases are same, we have to verify. If I consider the case 1. So, both A and B are 0, of course, you can see here from that case 1, case 2, case 3 all are same.



(Refer to the slide at 43:56)

In all the three cases finally, we have to make both A B as 0 0 then only output is 1. So, if both A B are 0 0. So, what happens to this four transistors if we call as M_1, M_2, M_3, M_4 these are n channel these are p channel. What is V_{GS} of M_3? This is 0, this is also 0. So, V_{GS} is 0 implies off this also this is gate this is 0 this is source this is also 0.

So, V_{GS} is 0 off whereas, this is gate source drain for this is gate source drain. So, what is V_{GS} of M_1? VG is VA, which is equal to 0. This is 0, S is V_{DD}. So, V_{GS} becomes 0 - V_{DD} negative for p channel negative means ON. So, the entire V_{DD} will come here because this is almost a short circuit with some resistance. There will be some negligible drop.

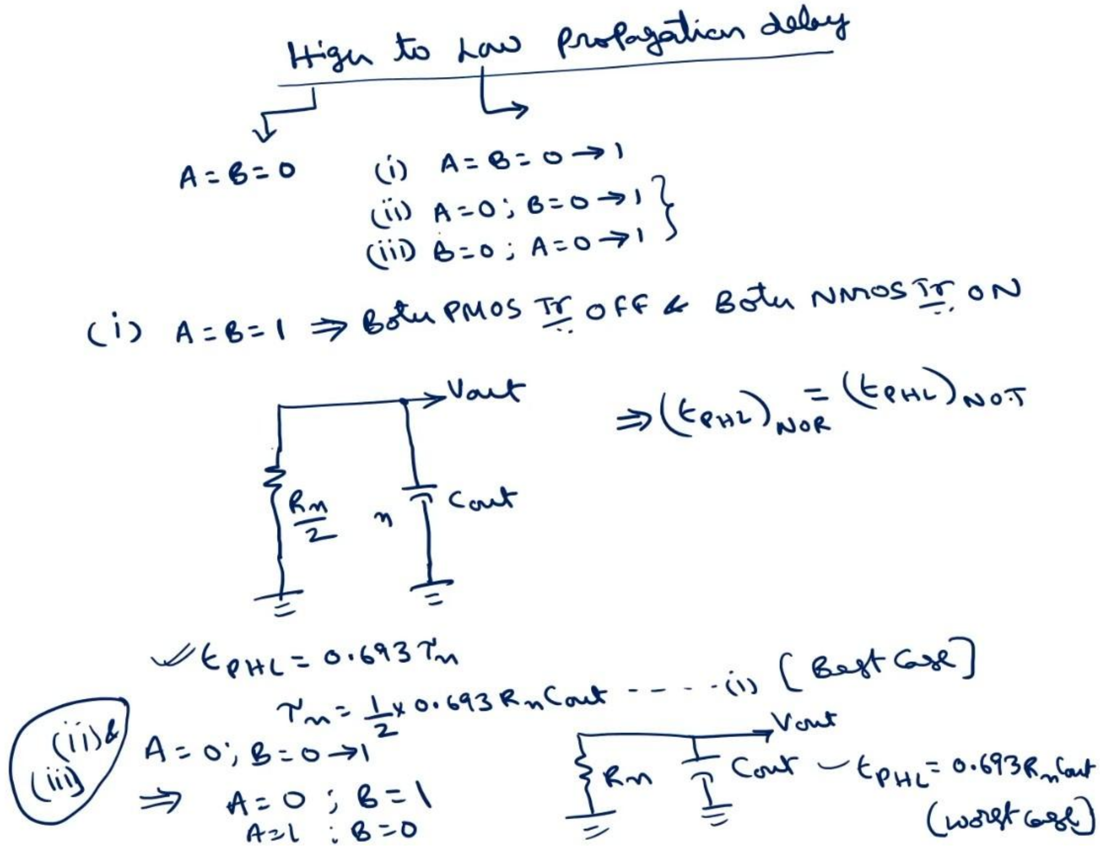
Now, what about V_{GS} of this one? Again, this is 0. This is almost V_{DD}. So, negative. So, this is also on. So, what will be the equivalent circuit? So, these two will be off open circuited this part will not be there only pull up network will be there, this is V_{DD} R_p R_p and v_{out} C_{out}. What is the expression for the propagation delay? $t_{PLH} = 0.693\tau_p, \tau_p = 2R_nC_{out}$, this is equal to $2(t_{PLH})_{NOT/INV} = (t_{PLH})_{NAND}$, this is reverse.

In case of NAND gate, t_{PLH} of both the inverter and NAND gate are same whereas, t_{PLH} is twice whereas, here this is $(t_{PLH})_{NAND} = 2(t_{PLH})_{INV}$. Now, we will see what happens to high to low propagation delay. High means for the NOR gate if all the inputs are 0s then it will be high means A = B = 0. Low means if any one of the input is 1. So, three possibilities one is both A and B changes from 0 to 1, second case is A is 0, but B changes from 0 to 1, third case is B is 0, A changes from 0 to 1.

So, which case is worst and which is best? We can consider now case 1, which is A = B = 0 to 1, means will be 1. So, we can see that when A = 1, B = 1, for 0, A is 0, this is OFF. Now, for A = 1, this will be ON, this will be ON. For A = 0 B = 0, both will be ON whereas, now it becomes OFF means M_1 M_2 are OFF whereas, M_3 M_4 are ON. So, in the pull up network, both will be open circuited, pull down circuit both will be short circuited, or both will be having some on resistance.

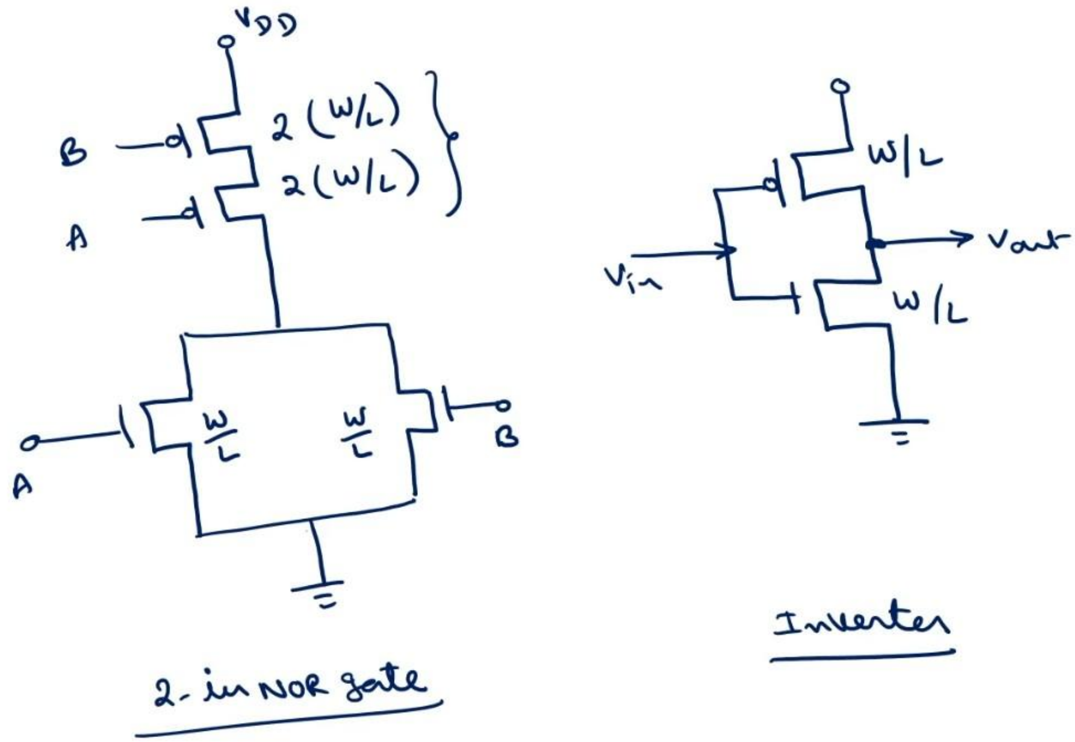
This is R_n R_n v_{out} C_{out} this is the equivalent circuit in this case because both PMOS transistors will be OFF and both NMOS ON. So, what will be this $t_{PHL} = 0.693\tau_n$ where $\tau_n = R||2R$. So, if I write a single R_n then this will be $\frac{R_n}{2}$, this is equal to $\frac{1}{2} \times 0.693R_nC_{out}$, whereas, the other two cases they are same A = 0, B = 0 to 1 implies A = 0 B = 1.

You can see that here, between these two, one of the transistor is ON, another is OFF. So, we will have something like this situation: either this or this. In one case, this is an open circuit. This is having R_p. In other case, this is R_p; this is an open circuit. In any case, this you can remove. Whereas, in the pull down network, this is R_n, and this is open circuit, or this is R_n, this is open circuited, in any case only one R_n is presented. So, you can see that now here, the equivalent circuit will be in these two cases, case 2 and case 3, either A = 1, and B = 0. So, upper network pull up network will be removed. Pull down will be having only single resistance; therefore, what is $t_{PHL} = 0.693R_nC_{out}$. So, between this and this, this is having lower value. So, this is best case, this is worst case, but in the design, normally, we will consider the worst case. So, worst case means t_{PHL} of inverter and NOR gate are same is equal to $(t_{PHL})_{NOR/INV}$. This is NOR. So, the final conclusion is $(t_{PLH})_{NOR} = 2(t_{PLH})_{INV}$ t_{PLH}, whereas, t_{PHL} is same as that of inverter. So, while sizing, what you have to do? This is two input NOR gate; this is inverter if I choose the inverter same $\frac{W}{L}$ ratio for both NMOS and PMOS. This you have to choose twice $\frac{W}{L}$ ratio because this low to high transition of NOR gate is twice that of inverter, and this you have to keep as it is.



(Refer to the slide at 49:09)

Here, also you have to avoid to connect either two NMOS devices or more than two NMOS devices in series or two or more PMOS devices in series.



(Refer to the slide at 51:04)

This is about the sizing of this one. So, in the next lecture we will solve some problems and the relation of various Boolean functions using CMOS circuits. Thank you.