

## MODULE 5 - PART A

### DELAY MODELS OF CMOS GATES AND INTERCONNECTS

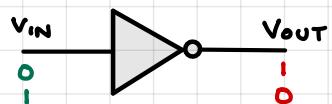
In this MODULE, we are going to look at the characteristics of DELAY in CMOS LOGIC GATES and INTERCONNECTS.

In particular, we will be looking at the DELAY of an INVERTER in details.

So what do we mean by DELAY and why is it important?

In ECE 2020, we had studied LOGIC GATES at a FUNCTIONAL level only.

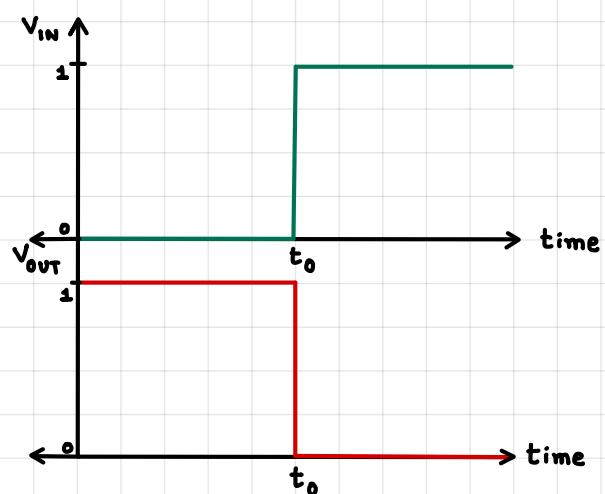
For example, if we consider an INVERTER, when the INPUT is



0, the OUTPUT is a 1 and when the INPUT is 1, the OUTPUT is a 0. This essentially means that if

we have a TIMING DIAGRAM, where we plot  $V_{IN}/V_{OUT}$  on the y-axis vs time on the x-axis, and if at a given time ' $t_0$ ',  $V_{IN}$  goes from 0 to 1, then we assume that, at that same time ' $t_0$ ',  $V_{OUT}$  goes from 1 to 0,

as shown below:-



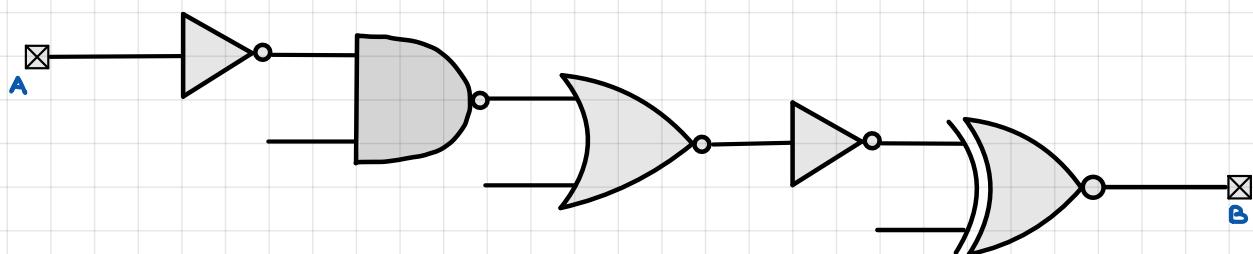
Here we did not think about what happens at the transition (time = ' $t_0$ ') when  $V_{IN}$  goes from 0 to 1. Does  $V_{OUT}$  immediately go from 1 to 0? Or is there a lag?

In reality, there is definitely a LAG or a DELAY and the OUTPUT will not follow the

INPUT immediately. This DELAY between the INPUT stimulus and the OUTPUT response is what we are going to look at and try to model.

So how does this DELAY affect circuits ?

Let us consider a chain of gates between points A and B as shown below :-



When signals propagate from A to B , the individual DELAYS of the GATES will add up and the TOTAL DELAY between points A and B can be significant .

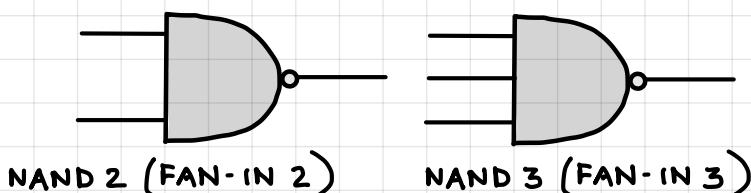
We are going to look at the DELAYS of the individual gates in details and then we can determine the total DELAY in such cases .

But before we start modeling the DELAY , we will look at a couple of definitions which you already know from ECE 2020 :-

① FAN-IN → FAN-IN is defined as the maximum number of input signals that are connected to the input of a LOGIC GATE .

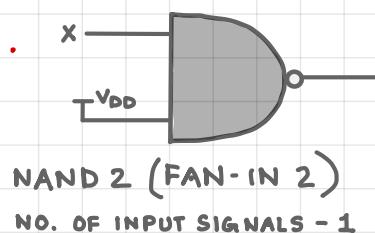
Say , for example , we have a NAND 2 gate and a NAND 3 gate as shown .

The FAN-INS for these gates are 2 and 3 respectively .



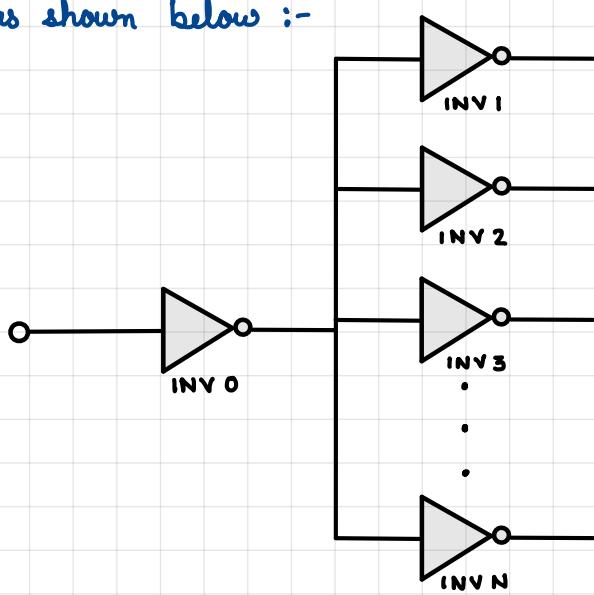
But , why do we define the FAN-IN as the MAXIMUM number of input signals

connected to the gate ? This is because some gates may have inputs connected to V<sub>DD</sub> or GND and in those cases the number of inputs connected to signals will differ from the FAN-IN .



② FAN-OUT → FAN-OUT is defined as the maximum number of output signals that are driven by a LOGIC GATE.

Say, for example, we have an inverter (INV0), the output of which is connected to 'N' other inverters (which may or may not be identical) as shown below :-



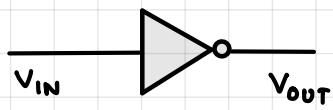
In this case the FAN-OUT of inverter INV0 is 'N' since 'N' other inverters are connected to its output.

We will use these concepts of FAN-IN and FAN-OUT extensively in our course and also in VLSI DESIGN in general.

Now before we proceed with calculating the DELAY, let us go over some definitions.

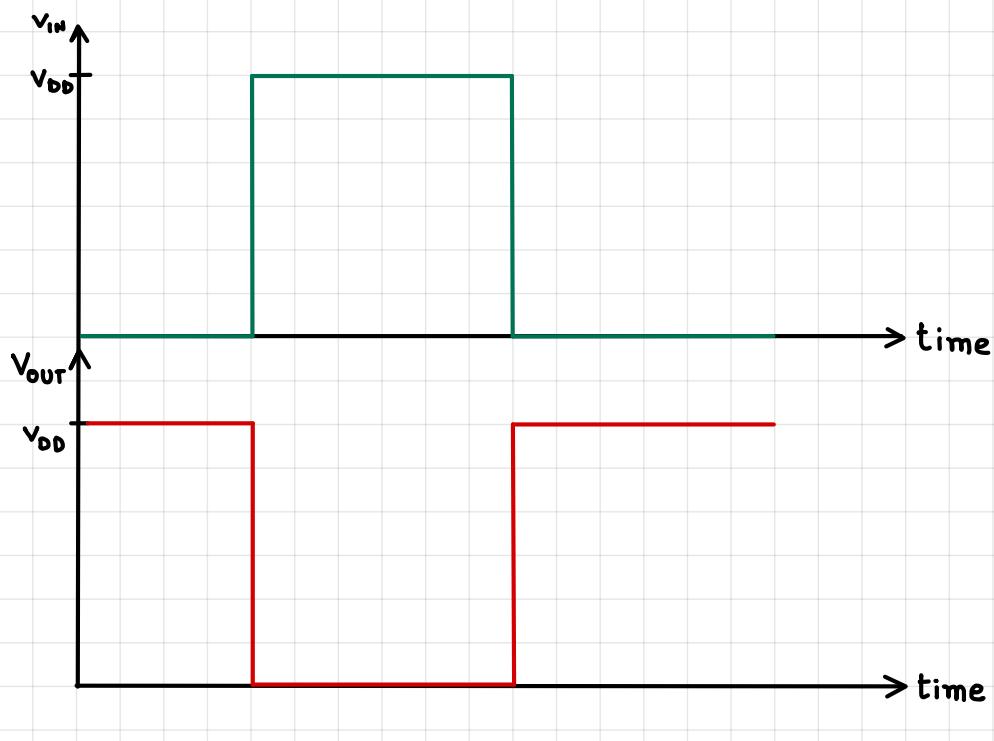
#### DELAY DEFINITIONS :-

Let us first consider an inverter and now we know that for this inverter, the output will not follow the input signal immediately, but there will be a certain time lag or DELAY. In reality, we also find that the input and output signals, do not rise to '1' or fall to '0' immediately, either. These signals ( $V_{IN}$  and  $V_{OUT}$ ) will gradually rise to '1' and also, gradually fall to '0'.

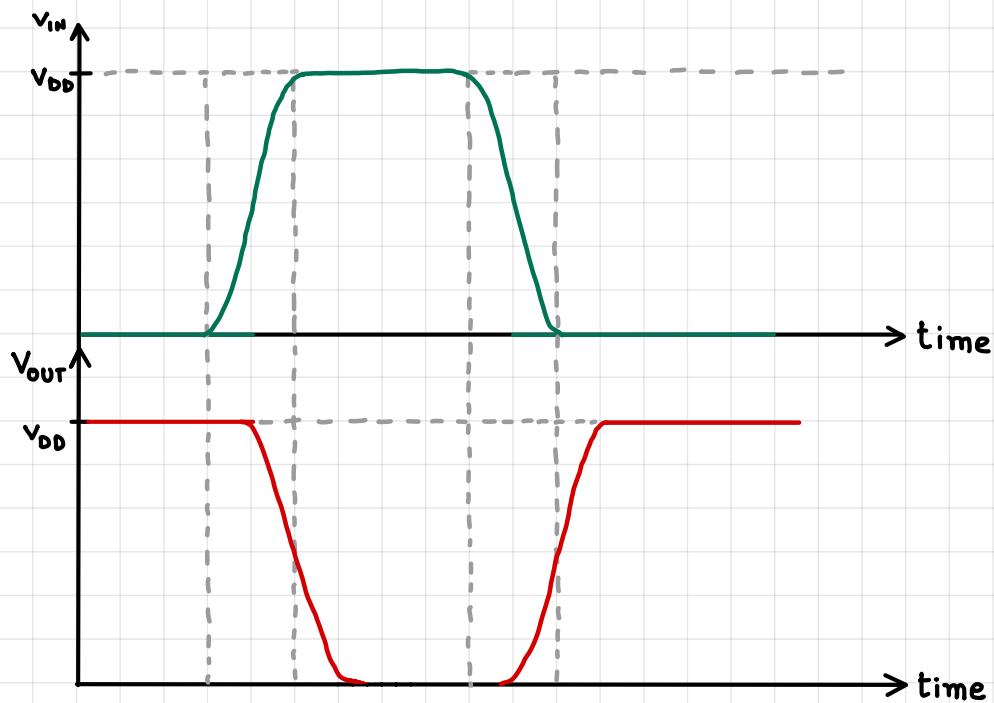


So if we have an input signal  $V_{IN}$  of an inverter that rises to '1' and then falls to '0' after a while, we can draw the

Timing diagrams as below :-



IDEAL INVERTER



INVERTER WITH DELAY

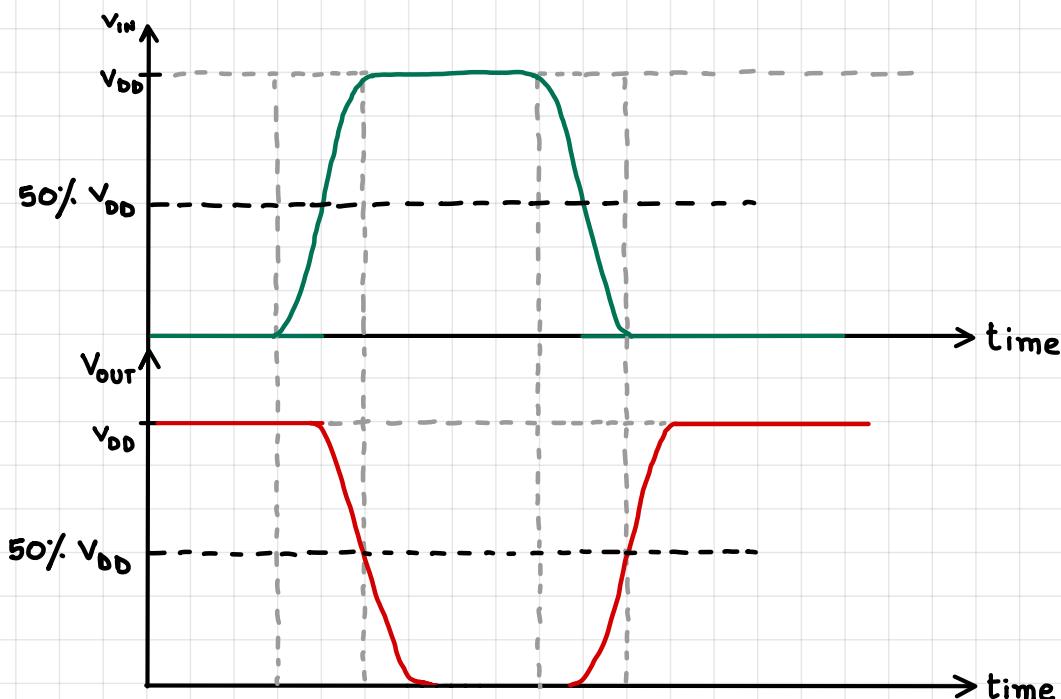
In reality, no input can make such a sudden transition from 0 to 1. So we have to draw a more realistic waveform that gradually rises to 1 and gradually falls to 0 as shown above. The exact equation of this curve we will go over next but for now let us assume that this is how our  $V_{IN}$  looks like.

The  $V_{OUT}$  or output voltage corresponding to this  $V_{IN}$  would also of course, fall gradually to 0 and then rise slowly to 1 again as shown in the figure above. But what we also notice from the figure is that, the  $V_{OUT}$  does not respond immediately to  $V_{IN}$ , i.e. it does not start falling as soon as  $V_{IN}$  starts rising. There is some DELAY associated with it. Similarly,  $V_{OUT}$  will not start rising immediately after  $V_{IN}$  starts falling.

Before defining the DELAYS let us look at some of the critical parameters.

First is  $\frac{V_{DD}}{2}$  or 50% of  $V_{DD}$  and we have seen before that this is often the TRIP POINT or  $V_T$  of a well-designed inverter. This is critical because this is where we will distinguish between a 1-level or a 0-level.

Similarly, we will have a 50%  $V_{DD}$  in the case of  $V_{OUT}$ , as marked below :-



INVERTER WITH DELAY

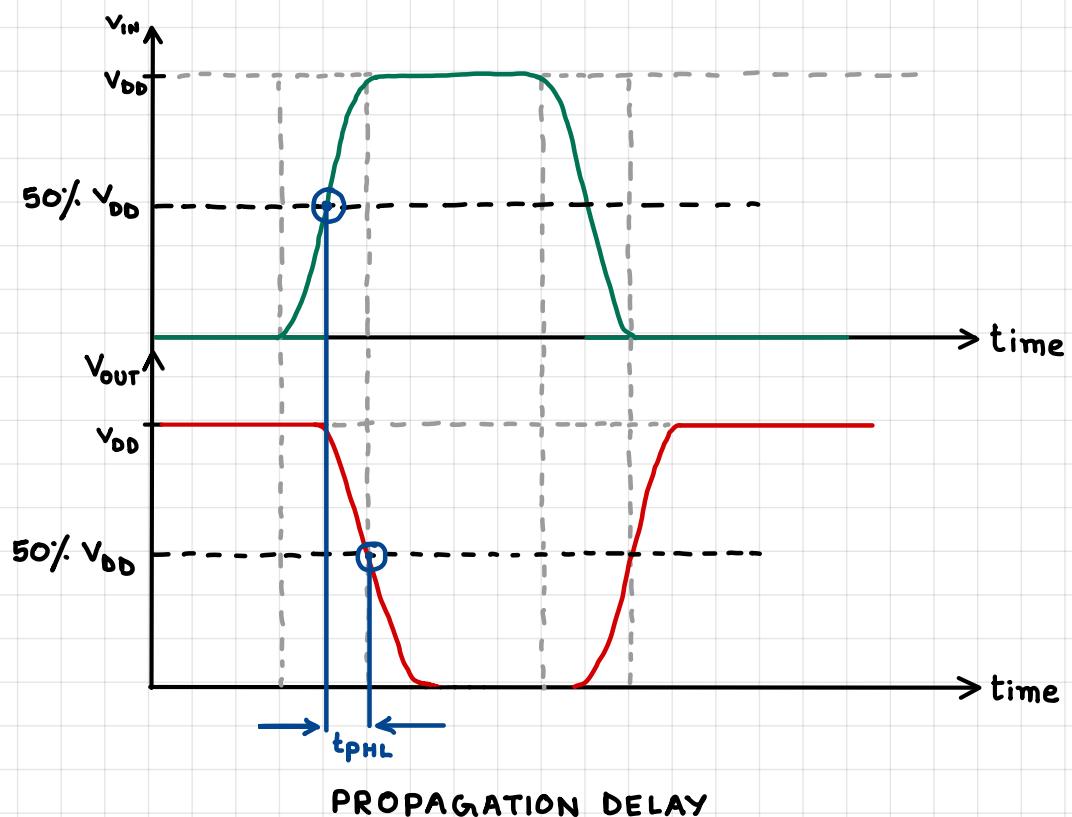
## ① PROPAGATION DELAY :-

This is the first delay that we will define. Propagation Delay can be of two different types:-

### a) HIGH TO LOW PROPAGATION DELAY ( $t_{PHL}$ ) →

The HIGH TO LOW PROPAGATION DELAY is defined as the delay between 50%  $V_{DD}$  at the input and the 50%  $V_{DD}$  at the output, when the output transitions from high to low.

This means that, we are looking at the first transition of our figure.

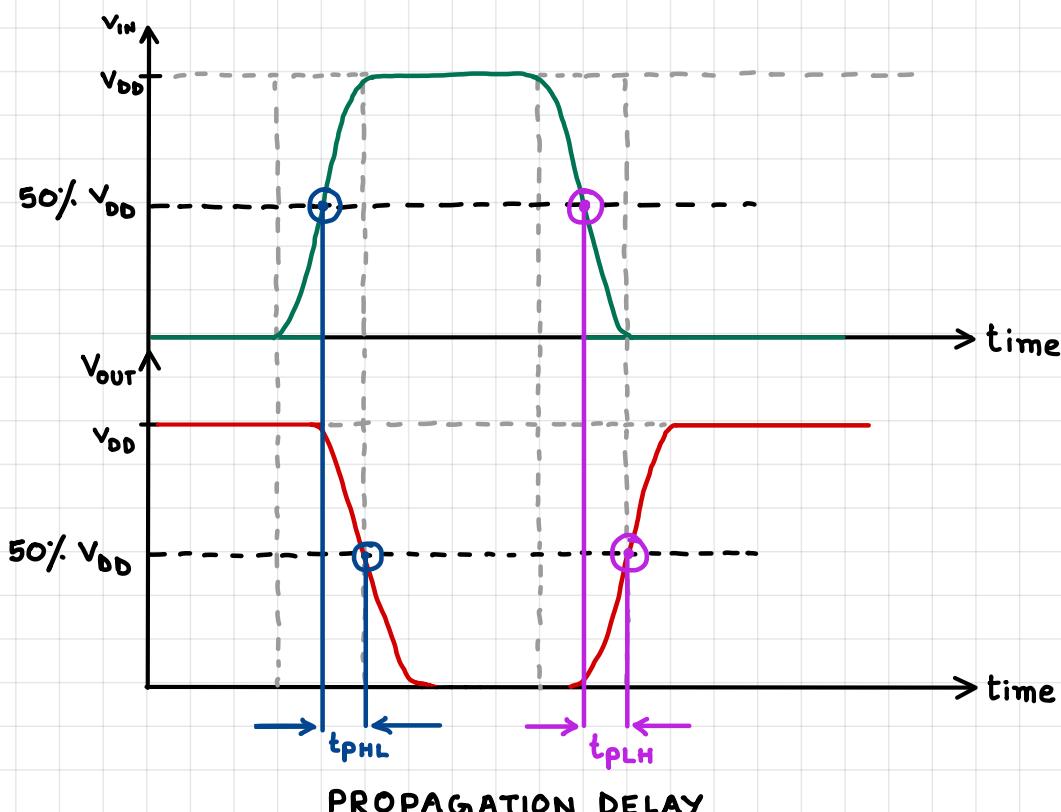


(b) LOW TO HIGH PROPAGATION DELAY ( $t_{PLH}$ ) →

The LOW TO HIGH PROPAGATION DELAY is defined as the delay between 50%  $V_{DD}$  at the input and the 50%  $V_{DD}$  at the output, when the output transitions from low to high.

This means that, we are looking at the second transition of our figure.

\* NOTE → The PROPAGATION DELAY is defined with respect to the output (i.e. as the output transitions from high to low ( $t_{PHL}$ ) or low to high ( $t_{PLH}$ ))



AVERAGE PROPAGATION DELAY ( $t_p$ ) →

Sometimes we also use the average value of  $\frac{t_{PHL} + t_{PLH}}{2}$

This is called the average propagation delay and is denoted by  $t_p$ .

## ② RISE TIME ( $t_r$ ) →

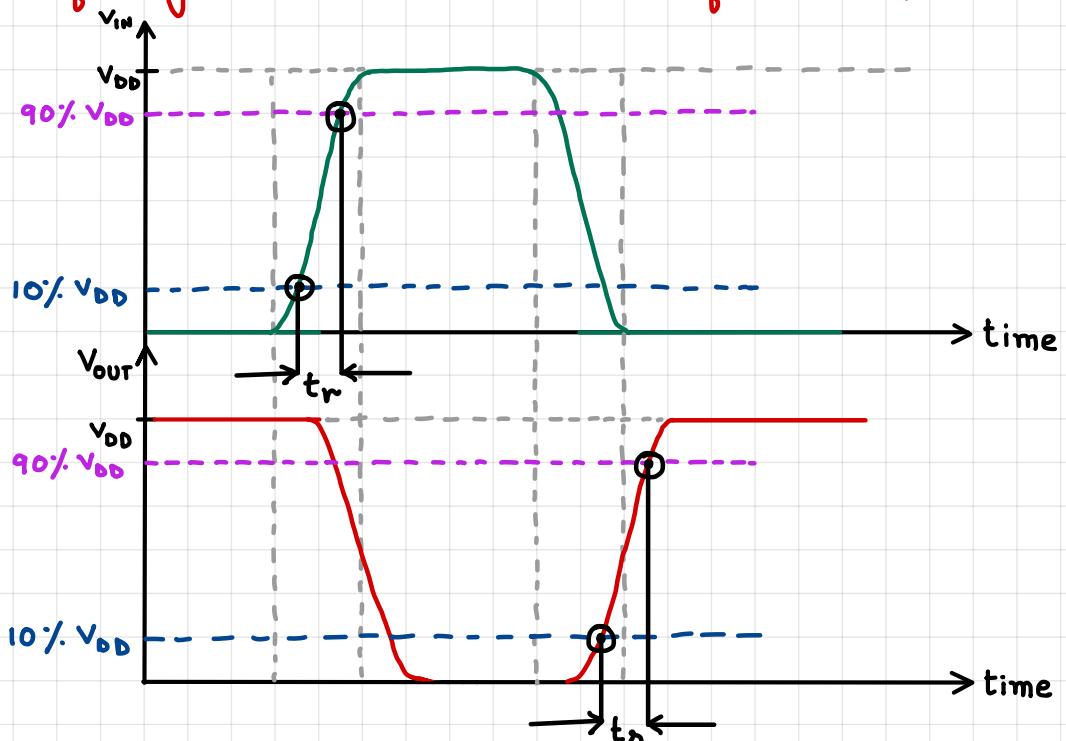
The next delay that we are interested in, is called the RISE TIME or  $t_r$ .

Rise time refers to the time when a signal switches from 10% to 90% of the maximum value (which is typically  $V_{DD}$  for CMOS logic)

Why is 10% and 90% significant here?

This is because, if we try to go back to our discussions on  $V_{IL}/V_{OL}$  and  $V_{IH}/V_{OH}$ , these are the lowest input/output and highest input/output values where the gain  $A_v = -1$ . So if the input value is less than 10%, we can safely say that it is a 0 and if it is more than 90%, we can safely say that it is a 1. Similarly, we can assume this for the output too.

\* NOTE → The RISE TIME is not defined for the input or the output voltages specifically. It can be calculated for the input or the output



The RISE TIME ( $t_r$ ) of the input signal will be the time it takes  $V_{IN}$  to go from  $0.1 V_{DD}$  to  $0.9 V_{DD}$  as shown in the figure above and ,

The RISE TIME ( $t_r$ ) of the output signal will be the time it takes  $V_{OUT}$  to go from  $0.1 V_{DD}$  to  $0.9 V_{DD}$  as shown in the figure above.

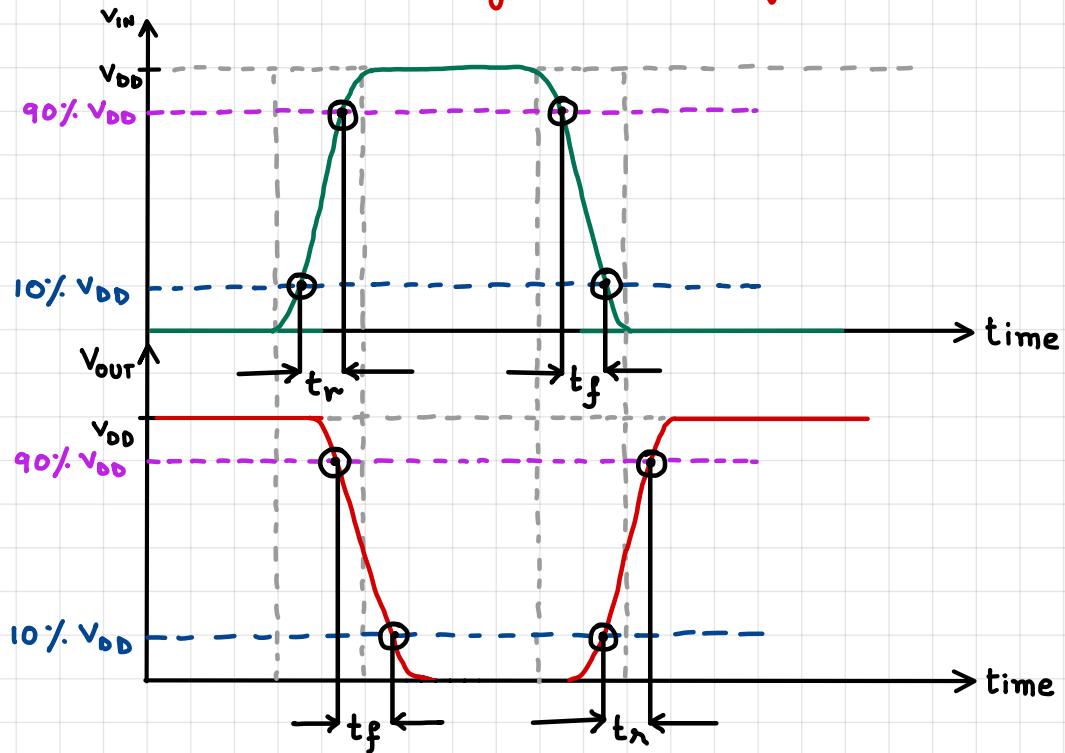
**NOTE :-** The  $t_r$  of the input and output signals need not be equal. Every signal can have its own rise time .

### ③ FALL TIME ( $t_f$ ) →

The last delay that we are interested in , is called the FALL TIME or  $t_f$  .

Fall time refers to the time when a signal switches from 90% to 10% of the maximum value ( which is typically  $V_{DD}$  for CMOS logic )

Again , this is applicable to any node or signal ( input or output )



This value again can be different for the input and the output signals, just like the RISE TIME.

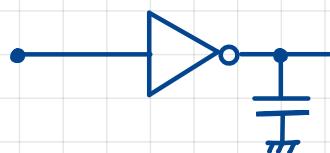
NOTE :- In some designs we use 30% - 70% limits (as opposed to 10% - 90% limits) to define RISE TIME and FALL TIME. But unless mentioned specifically, we will assume the 10% - 90% definitions.

Next, let us try to mathematically model these delays.

MATHEMATICAL MODEL OF DELAYS ( $t_p$ ,  $t_r$ ,  $t_f$ ) :-

REMEMBER → Every node of a circuit has a capacitance associated with it, even if it is not explicitly drawn.

Eg:-



SINGLE INVERTER

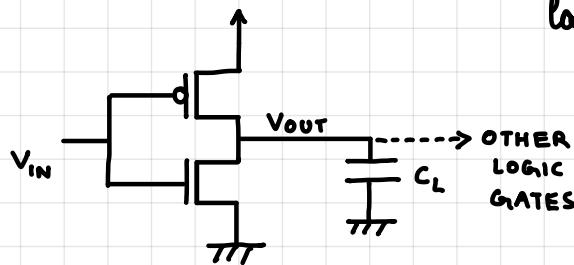


CHAIN OF GATES

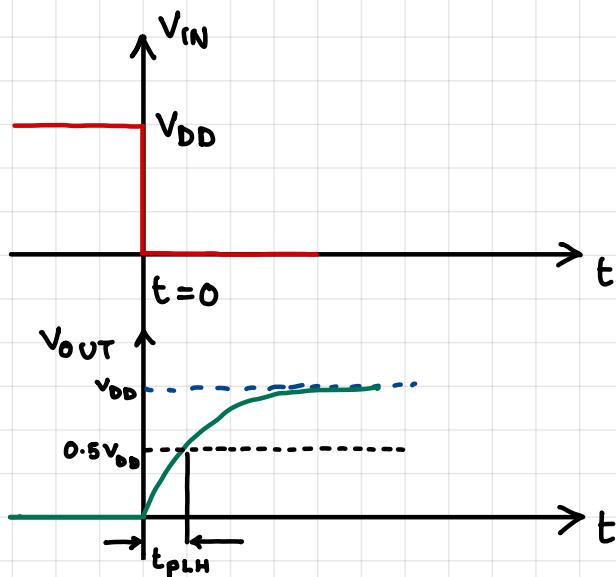
Whether it is a single inverter or a chain of gates, there are capacitors associated with every node. So whenever we are looking into an INV, NAND, XOR etc., we are looking into the gate of a transistor and there is always going to be a capacitance associated with that. We will learn how to model this capacitance. But realize that whenever we talk about a logic gate 'switching', we talk about switching an output capacitance.

If we didn't have this capacitance, we wouldn't have had the notion of DELAY.

Let us look at the inverter with a load capacitance  $C_L$  at the output. We will learn about the origin of this capacitance  $C_L$  later.  $V_{OUT}$  is connected to other logic gates.



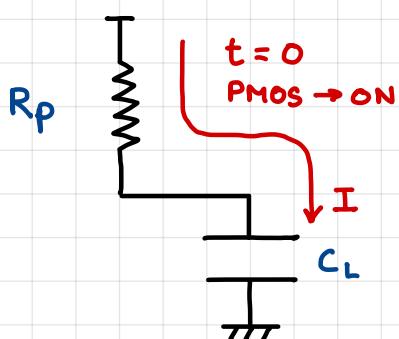
abruptly to 0 as this will keep our math simple. This happens



Let us assume for now that the input  $V_{IN}$  changes at  $t = 0$ . Here, as  $V_{IN} = 0$ , the pMOS turns ON and the nMOS is OFF. So,  $V_{OUT}$  is connected to  $V_{DD}$  through the pMOS and let us assume that the capacitor  $C_L$  slowly charges up to  $V_{DD}$  as shown in the figure on the left.

Let us first try to calculate the Propagation Delay  $t_{PLH}$  as shown above.

In reality, as the PMOS charges up  $C_L$ , it will go through all the 3 regions of operation and the current and the resistances associated with this will also change and will not be constant.



But to make our calculations simple,

we will assume that the PMOS offers a CONSTANT resistance  $R_P$  when it is 'ON'.

This is a big assumption but for now, it will make our calculations

simple.

We can write down the current flowing through the capacitor  $C_L$  at  $t=0$  when the pMOS is ON, using the principles we have learnt in ECE 2040 :-  $I = I_{C_L} = I_{RP}$

$$I = C_L \frac{d V_{OUT}}{dt} = \frac{V_{DD} - V_{OUT}}{R_p}$$

(since the same current  $I$  is flowing through the resistor and the capacitor)

Note that we are trying to find an equation for  $V_{OUT}$ .

$\therefore$  We have,

$$C_L \frac{d V_{OUT}}{dt} = \frac{V_{DD} - V_{OUT}}{R_p}$$

$$\text{or, } \frac{d V_{OUT}}{V_{DD} - V_{OUT}} = \frac{1}{R_p C_L} dt$$

Integrating both sides, we have,

$$\text{or, } \int_{V_{OUT}=0}^{V_{OUT}(t)} \frac{d V_{OUT}}{V_{DD} - V_{OUT}} = \frac{1}{R_p C_L} \int_{t=0}^t dt$$

When  $t=0$ ,  $V_{OUT}=0$  and at time 't', let us assume that  $V_{OUT}$  is  $V_{OUT}(t)$

Also, we know,  $\int \frac{dx}{x} = \ln x + c$

$$\therefore \int \frac{dx}{a-x} = -\ln|x| + c$$

∴ We have,

$$-\ln(V_{DD} - V_{OUT}) \Big|_0^{V_{OUT}(t)} = \frac{1}{R_p C_L} t$$

$$\text{or, } -\ln(V_{DD} - V_{OUT}(t)) + \ln(V_{DD} - 0) = \frac{t}{R_p C_L}$$

$$\text{or, } -\ln(V_{DD} - V_{OUT}(t)) + \ln V_{DD} = \frac{t}{R_p C_L}$$

$$\text{or, } \ln \frac{V_{DD}}{V_{DD} - V_{OUT}(t)} = \frac{t}{R_p C_L}$$

$$\left( \because \ln A - \ln B = \ln \frac{A}{B} \right)$$

$$\frac{V_{DD}}{V_{DD} - V_{OUT}(t)} = e^{\frac{t}{R_p C_L}}$$

$$\left( \because \log_a b = c \Rightarrow b = a^c \right)$$

$$\text{or, } \frac{V_{DD} - V_{OUT}(t)}{V_{DD}} = e^{-\frac{t}{R_p C_L}}$$

(inverting both sides)

$$\text{or, } V_{DD} - V_{OUT}(t) = V_{DD} e^{-\frac{t}{R_p C_L}}$$

$$\text{or, } V_{OUT}(t) = V_{DD} (1 - e^{-\frac{t}{R_p C_L}})$$

This is a very important equation that connects  $V_{OUT}$  to time. If we plot this, it will look like our sketch that we had drawn before.

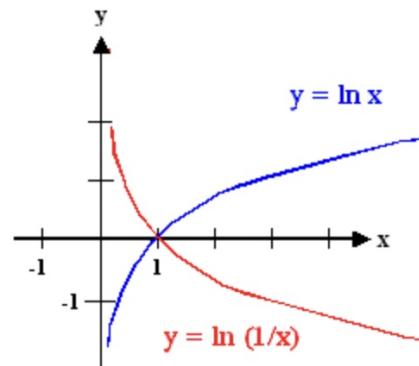
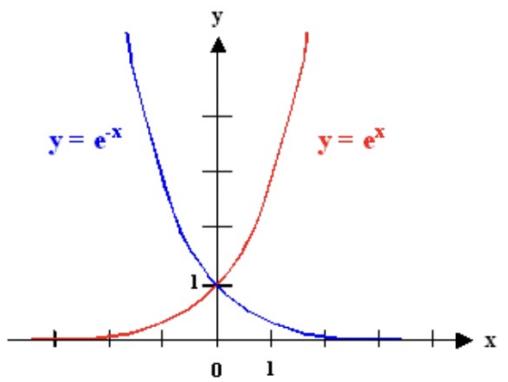
When  $t = 0$ ,

$$V_{\text{OUT}}(0) = V_{\text{DD}}(1 - e^0) = 0$$

and, when  $t \rightarrow \infty$ ,

$$V_{\text{OUT}}(\infty) = V_{\text{DD}}(1 - e^{-\infty}) = V_{\text{DD}}$$

\* RECAP :- In case you have forgotten the basic  $e^x$  and  $\ln(x)$  functions, see the graphs below :-



When  $t = R_p C_L$ ,

$$V_{\text{OUT}}(R_p C_L) = V_{\text{DD}}\left(1 - e^{-\frac{R_p C_L}{R_p C_L}}\right)$$

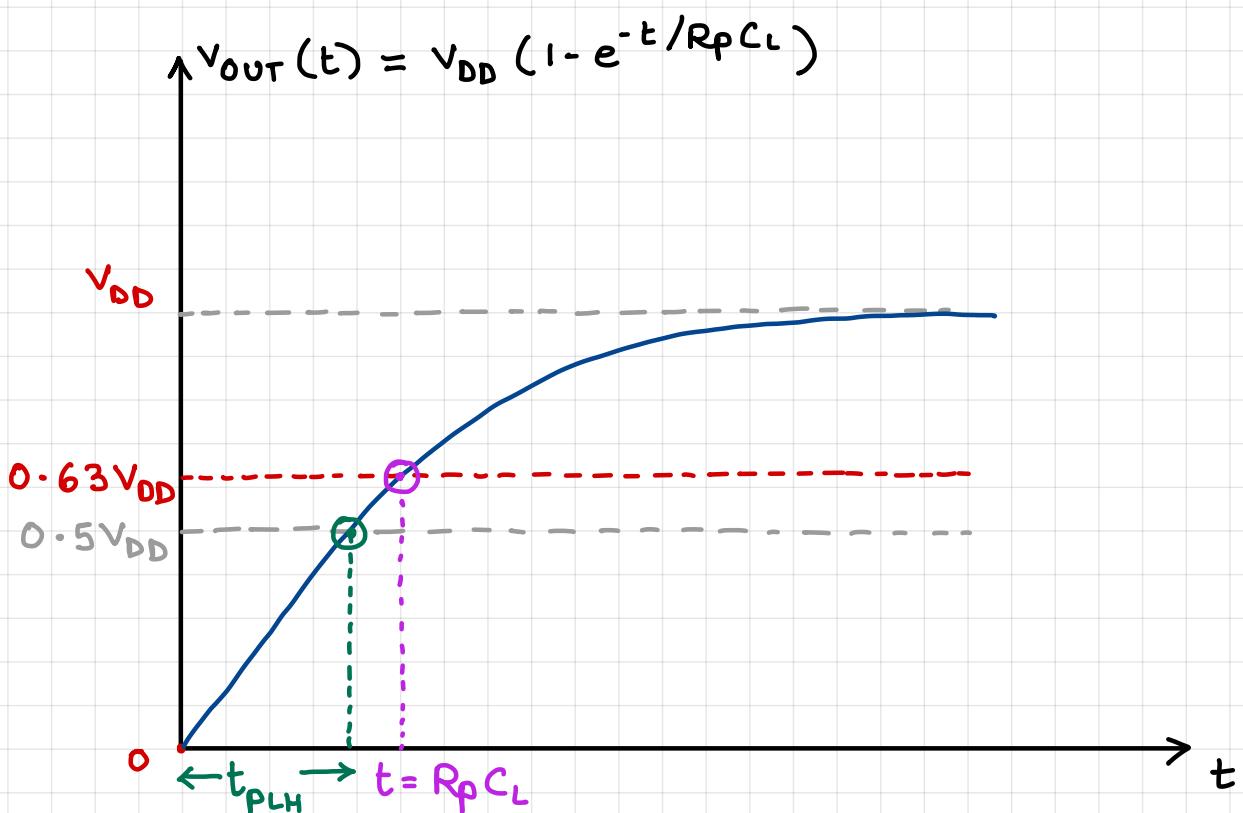
$$\text{or, } V_{\text{OUT}}(R_p C_L) = V_{\text{DD}}(1 - e^{-1})$$

$$e = 2.718 \quad \therefore \frac{1}{e} = e^{-1} = 0.3679$$

$$\therefore 1 - e^{-1} = 1 - 0.3679 = 0.63$$

$$\therefore V_{\text{OUT}}(R_p C_L) = 0.63 V_{\text{DD}}$$

Putting these values together, we can sketch this equation as below :-



NOTE :-  $R_p$  is the LUMPED AVERAGE PMOS RESISTANCE. We are not considering the various operating regions of the PMOS and the corresponding resistances separately, instead we just use an average resistance,  $R_p$ .

Now, we want to find the Propagation Delay  $t_{PLH}$  as marked in the figure above. So we substitute  $t = t_{PLH}$  in the equation for this curve. Note that  $V_{OUT}(t_{PLH}) = V_{DD}/2$

$$\therefore \frac{V_{DD}}{2} = V_{DD} \left(1 - e^{-t_{PLH}/R_p C_L}\right)$$

$$\text{or, } 1 - e^{-t_{PLH}/R_p C_L} = 0.5$$

$$\text{or, } e^{-t_{PLH}/R_p C_L} = 0.5$$

$$\text{or, } e^{t_{PLH}/R_p C_L} = 2$$

$$\text{or, } \ln_e 2 = \frac{t_{PLH}}{R_p C_L}$$

$$\text{i.e. } \ln 2 = \frac{t_{PLH}}{R_p C_L}$$

$$\text{or, } t_{PLH} = \ln 2 R_p C_L$$

$$\text{or, } t_{PLH} = 0.69 R_p C_L$$

This is a very important result that connects the CHARGING TIME CONSTANT ( $R_p C_L$ ) to the LOW TO HIGH PROPAGATION DELAY ( $t_{PLH}$ )

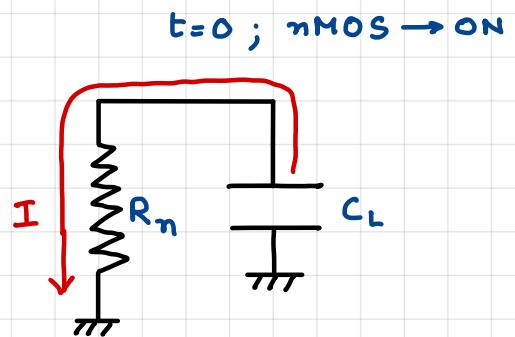
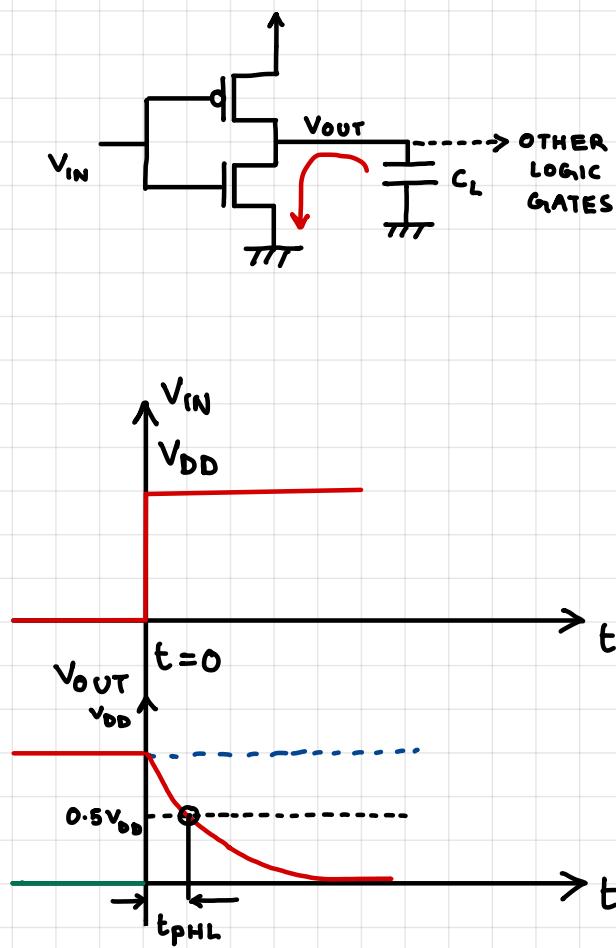
So the assumptions we have made here are :-

- ① The input makes an abrupt transition from 1 to 0 at  $t = 0$
- ② The output charges up gradually to  $V_{DD}$  through the LUMPED AVERAGE PMOS RESISTANCE  $R_p$  and has a Charging time constant equal to  $R_p C_L$

Now that we have an expression for  $t_{PLH}$ , we will derive an expression for  $t_{PHL}$  too, using a very similar process.

If we go back to our INVERTER CIRCUIT, we realize that now the nMOS will be responsible for the discharging process as the output goes from high to low.

Therefore, we will assume a similar LUMPED AVERAGE NMOS RESISTANCE,  $R_n$ , through which the LOAD CAPACITANCE  $C_L$  will now discharge.



The input is now assumed to make an abrupt transition from 0 to 1 (again, to keep the math simple).

The output will discharge gradually through the nMOS to 0.

In this case too, we are trying to derive an expression for  $V_{OUT}(t)$

Again, we will equate the current flowing through the resistor and the capacitor,  $I = I_{R_m} = -I_{C_L}$

$$- C_L \frac{dV_{OUT}}{dt} = \left( \frac{V_{OUT} - 0}{R_n} \right)$$

$$\text{or, } \frac{dV_{out}}{V_{out}} = -\frac{1}{R_n C_L} dt$$

Integrating both sides, we have,

$$\frac{dV_{OUT}}{V_{OUT}} = -\frac{1}{R_n C_L} dt$$

$V_{OUT}(t)$        $t$   
 $V_{OUT} = V_{DD}$        $t=0$

Here, since at  $t = 0$ ;  $V_{out}$  is high, we will adjust the limits of integration accordingly.

$$\text{or, } \ln \frac{V_{\text{out}}(t)}{V_{\text{DD}}} = - \frac{1}{R_n C_L} t \Big|_0^t$$

$$\text{or, } \ln(V_{\text{OUT}}(t)) - \ln V_{\text{DD}} = - \frac{1}{R_n C_L} (t - 0)$$

$$\text{or, } \ln \left( \frac{V_{\text{OUT}}(t)}{V_{\text{DD}}} \right) = - \frac{t}{R_n C_L}$$

$$\text{or, } \frac{V_{\text{OUT}}(t)}{V_{DD}} = e^{-t/R_n C_L}$$

$$\text{or, } V_{\text{OUT}}(t) = V_{\text{DD}} e^{-t/R_m C_L}$$

When  $t = 0$ ,

$$V_{\text{OUT}}(0) = V_{\text{DD}}(e^0) = V_{\text{DD}}$$

and, when  $t \rightarrow \infty$ ,

$$V_{\text{OUT}}(\infty) = V_{\text{DD}}(e^{-\infty}) = 0$$

when  $t = R_n C_L$ ,

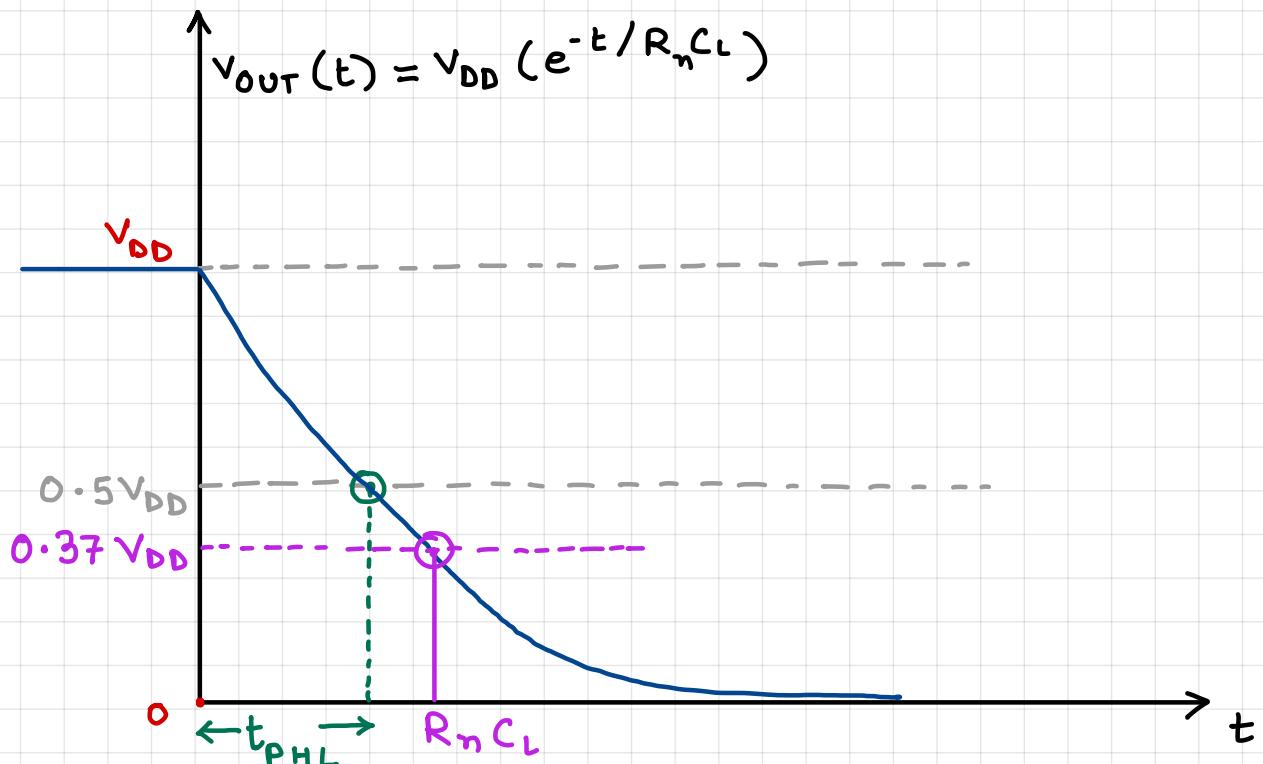
$$V_{\text{OUT}}(R_n C_L) = V_{\text{DD}}\left(e^{-\frac{R_n C_L}{R_n C_L}}\right)$$

$$\text{or, } V_{\text{OUT}}(R_n C_L) = V_{\text{DD}}(e^{-1})$$

$$e = 2.718 \quad \therefore \frac{1}{e} = e^{-1} = 0.3679$$

$$\therefore V_{\text{OUT}}(R_n C_L) = 0.36 V_{\text{DD}}$$

Putting these values together, we can sketch this equation as below :-



Now, let us calculate the value of  $t_{PHL}$

$$\frac{V_{DD}}{2} = V_{DD} \left( e^{-t_{PHL}/R_n C_L} \right)$$

$$\text{or, } e^{t_{PHL}/R_n C_L} = 2$$

$$\text{or, } \ln 2 = \frac{t_{PHL}}{R_n C_L}$$

$$\text{or, } t_{PHL} = \ln 2 R_n C_L$$

$$\text{or, } t_{PHL} = 0.69 R_n C_L$$

We can see that the expressions for  $t_{PLH}$  and  $t_{PHL}$  are very similar but while  $t_{PLH}$  is a function of  $R_p$   $t_{PHL}$  is a function of  $R_n$ .

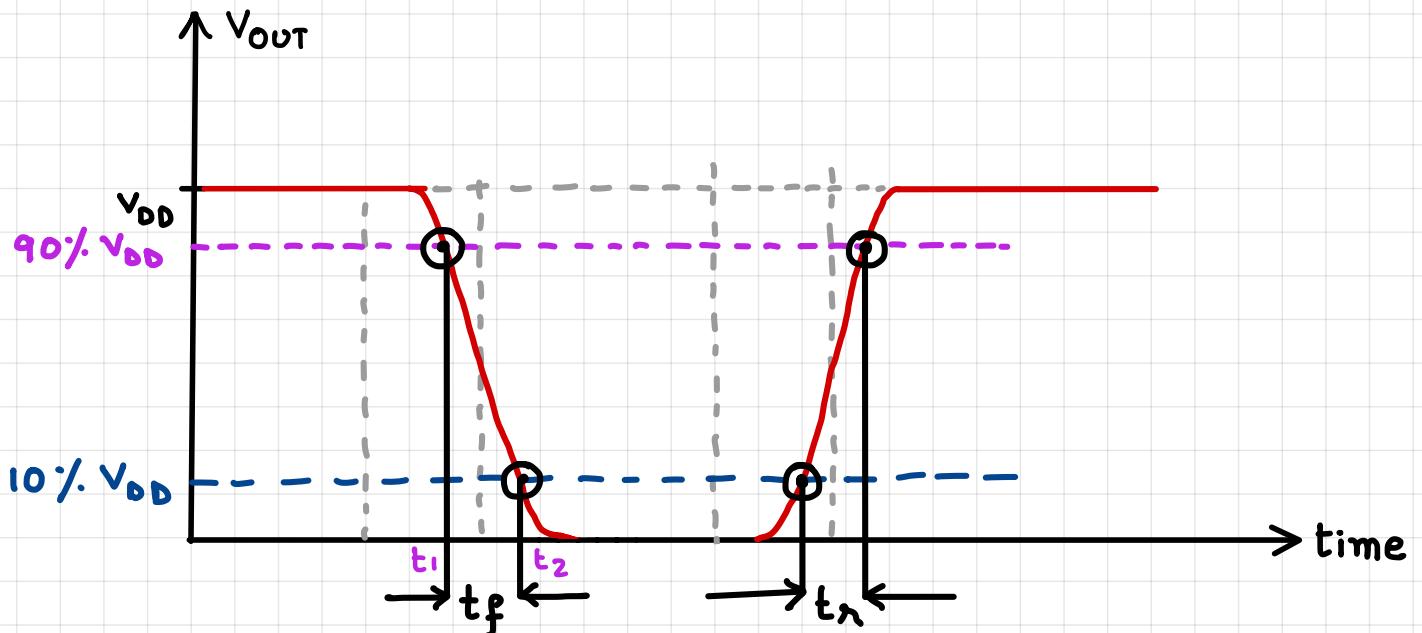
Now, the AVERAGE PROPAGATION DELAY ( $t_p$ ) is :-

$$t_p = \frac{t_{PLH} + t_{PHL}}{2}$$

$$\text{or, } t_p = 0.69 \left( \frac{R_p + R_n}{2} \right) C_L$$

which gives us the value of the propagation delay through an average of the nMOS and the pMOS resistances  $\left( \frac{R_n + R_p}{2} \right)$

Let us now try to measure the  $t_f$  and  $t_r$  of the output signal.



We have seen that when the output signal is falling, we can write the equation as :-

$$V_{\text{OUT}}(t) = V_{\text{DD}} e^{-t/R_n C_L}$$

Let,  $t_1 \rightarrow$  Time when the output signal has fallen to 90% of  $V_{\text{DD}}$

and,  $t_2 \rightarrow$  Time when the output signal has fallen to 10% of  $V_{\text{DD}}$

$$\therefore t_f = t_2 - t_1$$

$$\therefore V_{\text{OUT}}(t_1) = 0.9 V_{\text{DD}} = V_{\text{DD}} e^{-t_1/R_n C_L}$$

$$\text{or, } e^{-t_1/R_n C_L} = 0.9$$

$$\text{or, } \ln 0.9 = -\frac{t_1}{R_n C_L}$$

$$\text{or, } t_1 = -\ln 0.9 R_n C_L \dots (\text{i})$$

$$\text{and, } V_{\text{OUT}}(t_2) = 0.1 V_{\text{DD}} = V_{\text{DD}} e^{-t_2/R_n C_L}$$

$$\text{or, } e^{-t_2/R_n C_L} = 0.1$$

$$\text{or, } \ln 0.1 = -\frac{t_2}{R_n C_L}$$

$$\text{or, } t_2 = -\ln 0.1 R_n C_L \dots (\text{ii})$$

Subtracting (i) from (ii), we have,

$$t_2 - t_1 = -R_n C_L (\ln 0.1 - \ln 0.9)$$

$$\text{or, } t_f = R_n C_L \ln \left( \frac{0.9}{0.1} \right)$$

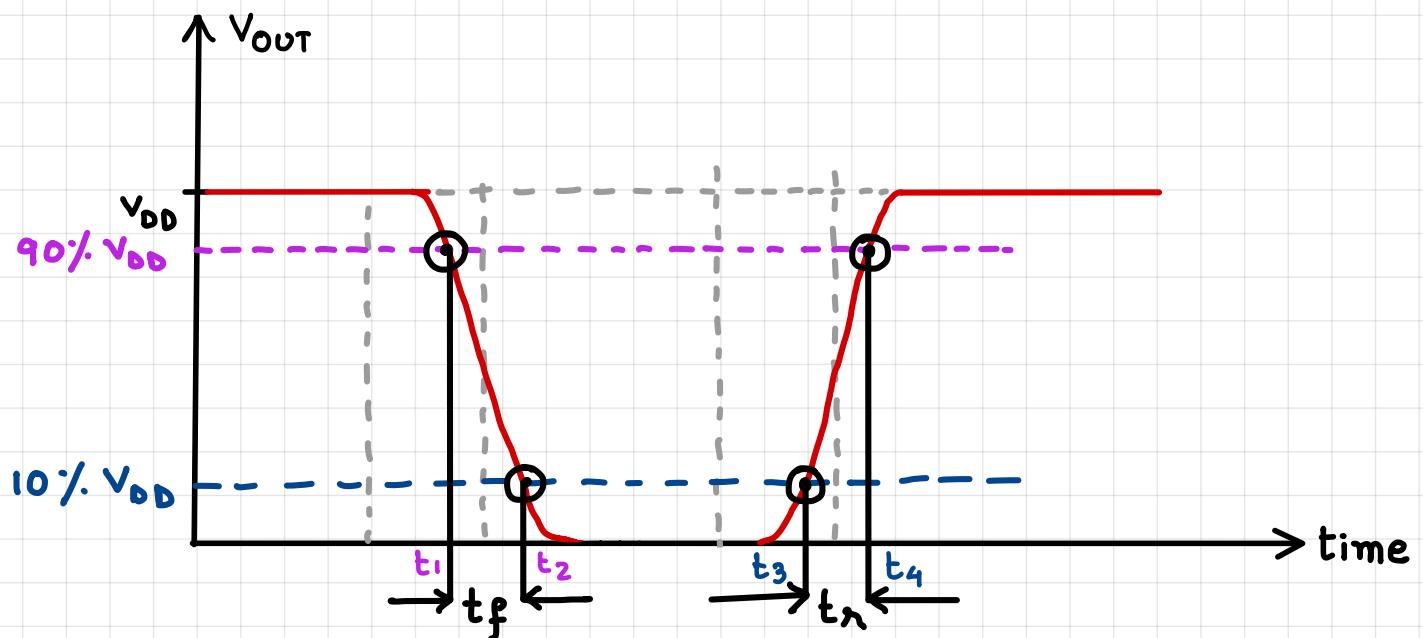
$$\text{or, } t_f = R_n C_L \ln 9$$

$$\text{or, } t_f = \ln 9 R_n C_L$$

$$\text{or, } t_f = 2.19 R_n C_L$$

- \* We will consider  $t_r$  and  $t_f$  only for  $V_{\text{OUT}}$  (output signal) in our calculations as we have assumed our input signal to be an ideal step function.

We will now proceed to calculate  $t_r$  similarly :-



We have seen that when the output signal is rising , we can write the equation as :-

$$V_{out}(t) = V_{DD} \left( 1 - e^{-t/R_p C_L} \right)$$

Let,  $t_3 \rightarrow$  Time when the output signal has risen to  $10\% \text{ of } V_{DD}$

and,  $t_4 \rightarrow$  Time when the output signal has risen to  $90\% \text{ of } V_{DD}$

$$\therefore t_r = t_4 - t_3$$

$$V_{out}(t_3) = 0.1 V_{DD} = V_{DD} \left( 1 - e^{-t_3/R_p C_L} \right)$$

$$\text{or, } 0.1 = 1 - e^{-t_3/R_p C_L}$$

$$\text{or, } e^{-t_3/R_p C_L} = 1 - 0.1 = 0.9$$

$$\text{or, } \ln 0.9 = -\frac{t_3}{R_p C_L}$$

$$\text{or, } t_3 = -\ln 0.9 R_p C_L \dots (\text{iii})$$

Similarly, we can write,

$$V_{out}(t_4) = 0.9 V_{DD} = V_{DD} (1 - e^{-t_4/R_p C_L})$$

$$\text{or, } 0.9 = 1 - e^{-t_4/R_p C_L}$$

$$\text{or, } e^{-t_4/R_p C_L} = 1 - 0.9 = 0.1$$

$$\text{or, } \ln 0.1 = -\frac{t_4}{R_p C_L}$$

$$\text{or, } t_4 = -\ln 0.1 R_p C_L \dots (\text{iv})$$

Subtracting (iii) from (iv), we have,

$$t_4 - t_3 = -\ln 0.1 R_p C_L + \ln 0.9 R_p C_L$$

$$\text{or, } t_x = R_p C_L (\ln 0.9 - \ln 0.1)$$

$$\text{or, } t_x = \ln 9 R_p C_L$$

$$\text{or, } t_x = 2.19 R_p C_L$$

which is again similar to the expression of  $t_f$ , but while  $t_f$  is a function of  $R_n$ ,  $t_x$  is a function of  $R_p$  since we know that the pMOS is responsible for charging the output to  $V_{DD}$ , while the nMOS is responsible for discharging the output to GND.