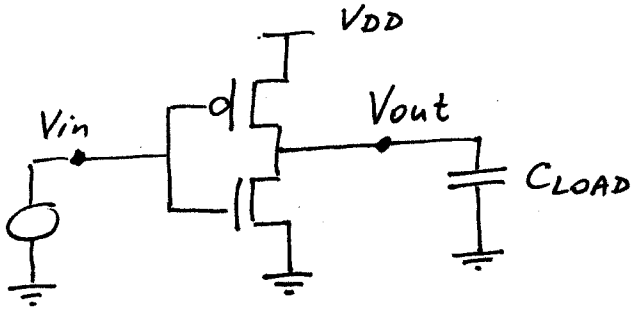
CMOS Inverter Switching Characteristic

The switching characteristics of a gate determine the speed at which the gate can operate. The switching speed of a CMOS gate can be measured in terms of the time required to charge and discharge a capacitive load C_{LOAD}.



The switching characteristics of a gate tell you how fast a gate is able to react to a change (switch from one steady state to another) in the input signal.

There are usually 4 parameters used to characterize the switching (or transient or dynamic) behavior of a CMOS gate:

rise time
fall time
propagation delay from high to low
propagation delay from low to high
(average propagation delay)

Rise time → t_{RISE}

is defined as the time taken by the output voltage to rise from 10% to 90% of its steady state value.

NOTE

fall time → t_{FALL}

is defined as the time taken by the output voltage to fall from 90% to 10% of its steady state value.

In order to make things a little bit easier we assume that the steady state values are always 0 and V_{DD}

NOTE: not everybody use the same percentages to define rise and fall time (another common pair is 20%-80%)

propagation delay from high to low → t_{PHL}

It determines the input-to-output delay during the high-to-low transition of the output.

It is defined as the time difference from the 50% of the input voltage and the time required for the output voltage to fall to the 50% of its steady state value.

(another common percentage pair used in practice is 50% - 35%)

propagation delay from low to high → t_{PLH}

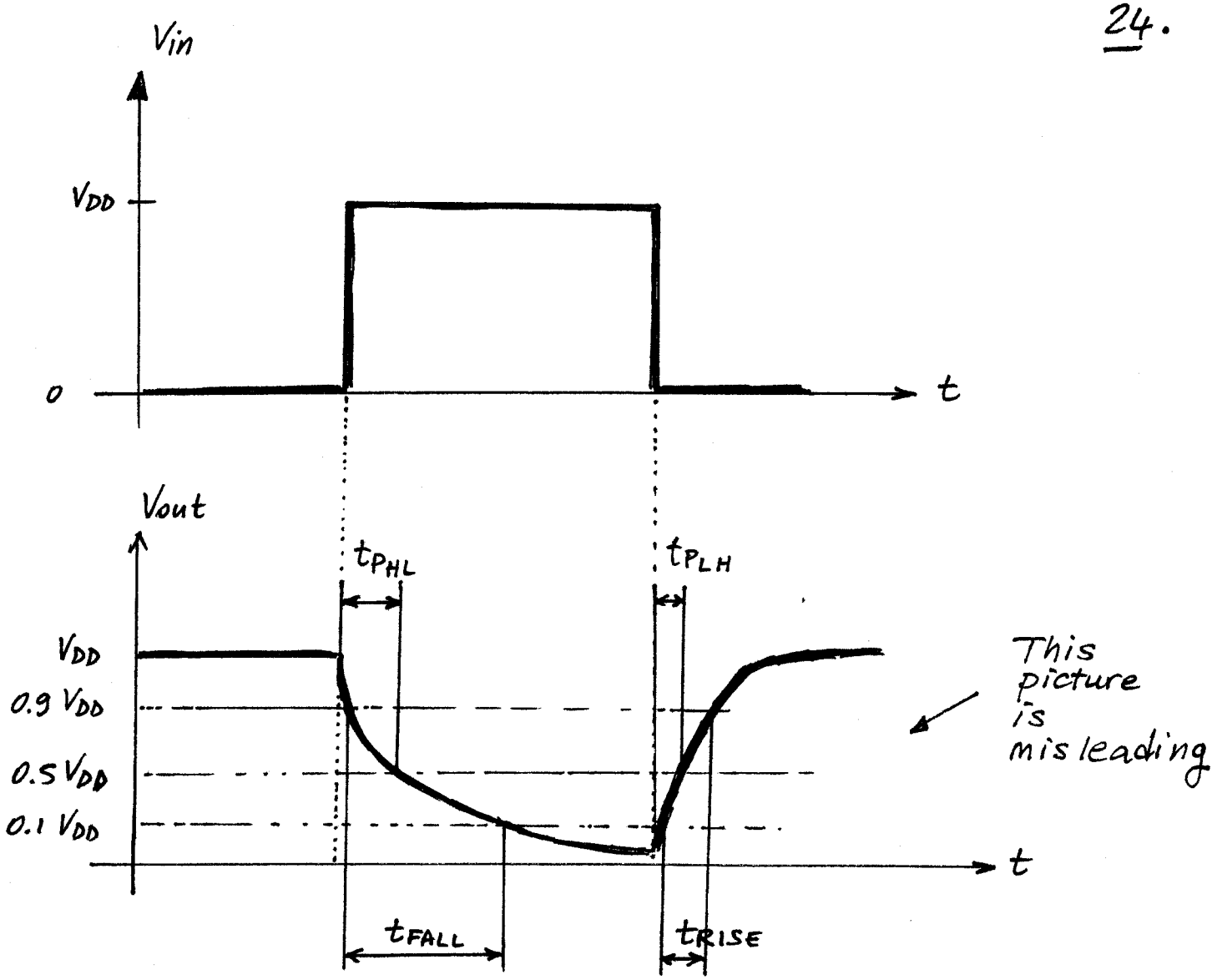
It is defined as the time difference from the 50% of the input voltage and the time required for the output voltage to rise to the 50% of its steady state value
(another common percentage pair is 50%-65%)

average propagation delay → t_p

"It is the average time needed for the output to respond to a change in the input" [U]

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

In order to make things a little bit easier we can assume that the input signal is an "ideal" pulse waveform (rise time and fall time of the pulse are zero).



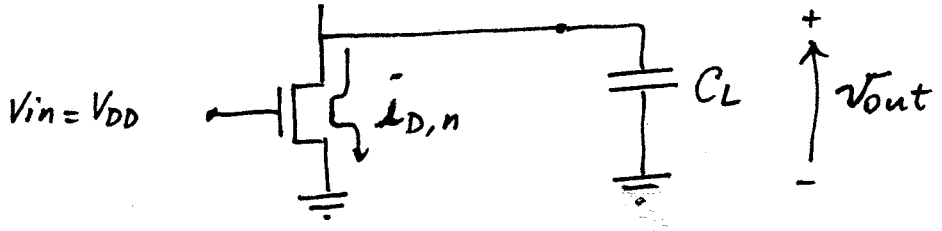
in reality

notice that ↓ $t_{FALL} < t_{RISE}$ and $t_{PHL} < t_{PLH}$

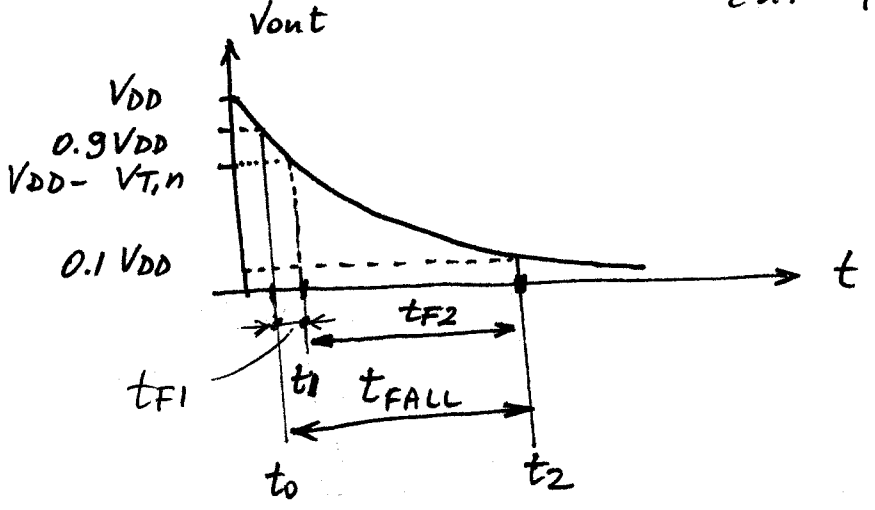
(we'll see the reason soon!)

Now we can try to find out some expression for the 4 switching parameters.

1. Let's start with the fall time.

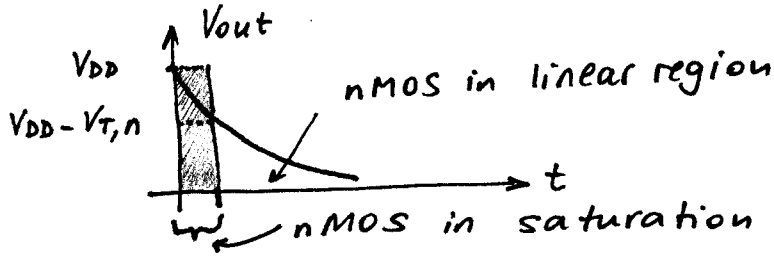


the pMOS is in cut-off



$$C_L \cdot \frac{dv_{out}}{dt} = - i_{D,n}$$

When nMOS starts conducting we are in saturation region, but after a while the output voltage goes below $V_{DD} - V_{T,n}$ so the nMOS goes in linear region.



$$C_L \cdot \frac{dv_{out}}{dt} = - i_{D,n}$$

$$i_{D,n} = \frac{\beta_n}{2} (V_{DD} - V_{T,n})^2 \quad \text{for} \quad V_{DD} - V_{T,n} < V_{out} \leq V_{DD}$$

↑ saturation

$$i_{D,n} = \beta_n \left[(V_{DD} - V_{T,n}) V_{out} - \frac{V_{out}^2}{2} \right] \quad \text{for} \quad V_{out} \leq V_{DD} - V_{T,n}$$

↑ linear

in saturation:

$$\int_{t_0}^{t_1} dt = - C_L \int_{V_{DD}}^{V_{DD} - V_{T,n}} \frac{1}{i_{D,n}} \, dv_{out} =$$

$$= - \frac{2 C_L}{\beta_n (V_{DD} - V_{T,n})^2} \int_{0.9 V_{DD}}^{V_{DD} - V_{T,n}} dv_{out}$$

↕

$$t_{F_1} = \frac{2 C_L}{\beta_n (V_{DD} - V_{T,n})^2} \cdot (V_{T,n} - 0.1 V_{DD})$$

in linear region:

$$\int_{t_2}^{t_2} dt = -C_L \int_{0.1\,V_{DD}}^{V_{DD}-V_{T,n}} \frac{d v_{out}}{\beta_n (V_{DD}-V_{T,n}) v_{out} - \beta_n \frac{v_{out}^2}{2}}$$

$$-t_{F2} = \frac{C_L}{\beta_n} \int_{0.1\,V_{DD}}^{V_{DD}-V_{T,n}} \frac{d v_{out}}{-(V_{DD}-V_{T,n}) v_{out} + \frac{v_{out}^2}{2}}$$

$$t_{F2} = \frac{-C_L}{\beta_n (V_{DD}-V_{T,n})} \int_{0.1\,V_{DD}}^{V_{DD}-V_{T,n}} \frac{d v_{out}}{-v_{out} + \frac{v_{out}^2}{2(V_{DD}-V_{T,n})}} =$$

$$= \frac{-C_L}{\beta_n (V_{DD}-V_{T,n})} \int_{0.1\,V_{DD}}^{V_{DD}-V_{T,n}} \frac{d v_{out}}{\frac{v_{out}^2}{2(V_{DD}-V_{T,n})} - v_{out}}$$

Here we need some math handbook:

$$\int \frac{dx}{ax^2+bx} = \frac{1}{b} \ln\left(\frac{x}{ax+b}\right) + \text{Const}$$

I call:

$$a = \frac{1}{2(V_{DD}-V_{T,n})}$$

$$b = -1$$

8.

Then:

$$t_{F2} = \frac{-C_L}{\beta_n (V_{DD} - V_{T,n})} \left[\frac{1}{b} \ln \left(\frac{v_{out}}{a v_{out} + b} \right) \right]_{0.1 V_{DD}}^{V_{DD} - V_{T,n}} =$$

$$= \frac{+ C_L}{\beta_n (V_{DD} - V_{T,n})} \left[\ln \left(\frac{V_{DD} - V_{T,n}}{\frac{V_{DD} - V_{T,n}}{2(V_{DD} - V_{T,n})} - 1} \right) + \right.$$

$$\left. - \ln \left(\frac{0.1 V_{DD}}{\frac{0.1 V_{DD}}{2 (V_{DD} - V_{T,n}) \cdot} - 1} \right) \right] =$$

$$= \frac{+ C_L}{\beta_n (V_{DD} - V_{T,n})} \left[\ln \left(\frac{V_{DD} - V_{T,n}}{-\frac{1}{2}} \right) + \right.$$

$$\left. - \ln \left(\frac{0.1 V_{DD}}{\frac{0.1 V_{DD} - 2 (V_{DD} - V_{T,n})}{2 (V_{DD} - V_{T,n})}} \right) \right] =$$

$$= \frac{+ C_L}{\beta_n (V_{DD} - V_{T,n})} \left[\ln \left(\frac{V_{DD} - V_{T,n}}{-\frac{1}{2}} \right) + \right.$$

$$\left. - \ln \left(\frac{0.1 V_{DD} \cdot 2 \cdot (V_{DD} - V_{T,n})}{0.1 V_{DD} - 2 V_{DD} + 2 V_{T,n}} \right) \right] =$$

$$= + \frac{C_L}{\beta_n (V_{DD} - V_{T,n})} \left[\ln \left(-2 (V_{DD} - V_{T,n}) \right) + \right.$$

$$\left. - \ln \left(\frac{0.2 V_{DD} (V_{DD} - V_{T,n})}{-1.9 V_{DD} + 2 V_{T,n}} \right) \right] =$$

$$= + \frac{C_L}{\beta_n (V_{DD} - V_{T,n})} \left[\ln \left(\frac{-2 (V_{DD} - V_{T,n})}{\frac{0.2 V_{DD} (V_{DD} - V_{T,n})}{-1.9 V_{DD} + 2 V_{T,n}}} \right) \right]$$

$$= + \frac{C_L}{\beta_n (V_{DD} - V_{T,n})} \left[\ln \left(\frac{-10 \cdot (-1.9 V_{DD} + 2 V_{T,n})}{V_{DD}} \right) \right]$$

$$= + \frac{C_L}{\beta_n (V_{DD} - V_{T,n})} \cdot \ln \left(\frac{19 V_{DD} - 20 V_{T,n}}{V_{DD}} \right)$$

So finally:

$$t_{FALL} = t_{F1} + t_{F2} =$$

$$= \frac{C_L}{\beta_n (V_{DD} - V_{T,n})} \cdot \frac{2 (V_{T,n} - 0.1 V_{DD})}{(V_{DD} - V_{T,n})} +$$

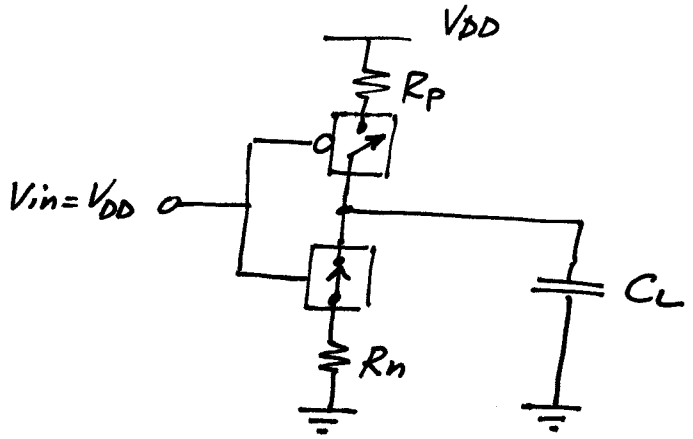
$$+ \frac{C_L}{\beta_n (V_{DD} - V_{T,n})} \cdot \ln \left(\frac{19 V_{DD} - 20 V_{T,n}}{V_{DD}} \right)$$

$$t_{FALL} = \underbrace{\frac{C_L}{\beta_n (V_{DD} - V_{T,n})}}_{\tau_n} \cdot \left[\frac{2(V_{T,n} - 0.1 V_{DD})}{(V_{DD} - V_{T,n})} + \right.$$

$$\left. + \ln\left(\frac{19 V_{DD} - 20 V_{T,n}}{V_{DD}} \right) \right] =$$

$$= \tau_n \cdot \left[\frac{2(V_{T,n} - 0.1 V_{DD})}{(V_{DD} - V_{T,n})} + \ln\left(\frac{19 V_{DD} - 20 V_{T,n}}{V_{DD}} \right) \right]$$

where τ_n is the discharge time constant for the circuit

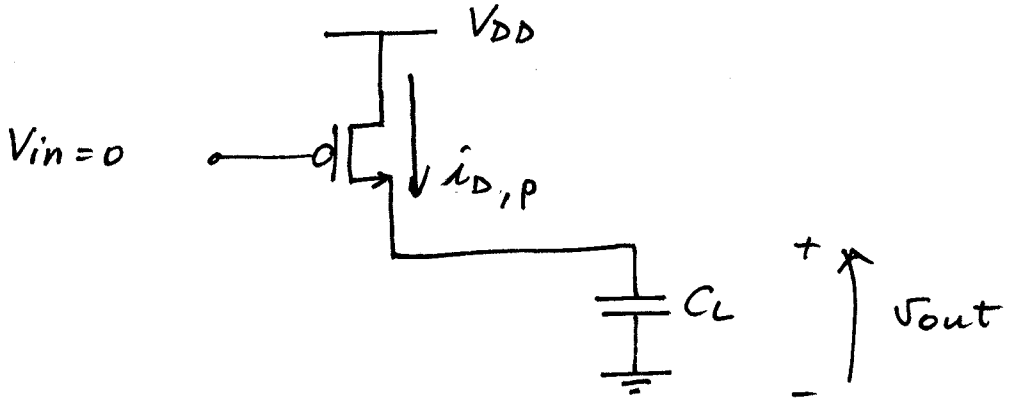


$$\tau_n = R_n \cdot C_L$$

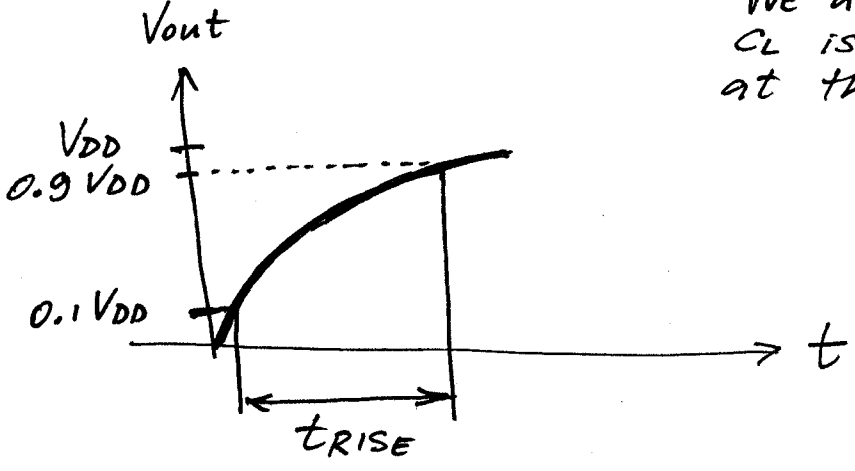
where the n-MOS drain-to-source resistance is approximate by

$$R_n \simeq \frac{1}{\beta_n (V_{DD} - V_{T,n})} \quad [\Omega]$$

2. Rise time



We assume that C_L is uncharged at the beginning



$$i_{D,p} = C_L \frac{dV_{out}}{dt}$$

Following exactly the same approach we find out that:

$$t_{RISE} = \tau_p \left[\frac{2\left(|V_{T,p}| - 0.1\,V_{DD}\right)}{\left(V_{DD} - |V_{T,p}|\right)} + \ln\left(\frac{19\,V_{DD} - 20\,|V_{T,p}|}{V_{DD}} \right) \right]$$

with $\tau_p = R_p \cdot C_L$

where $R_p \simeq \dfrac{1}{\beta_p \left(V_{DD} - |V_{T,p}|\right)}$

At this point we can notice that if we build the n-MOS and the p-MOS in the same way (same aspect ratio):

$$\left(\frac{W}{L}\right)_n = \left(\frac{W}{L}\right)_p$$

since typically $\mu_n \simeq 2\,\mu_p$ we have that:

$$\beta_n = \mu_n \frac{\varepsilon_{ox}}{t_{ox}} \frac{W_n}{L_n} \simeq 2 \cdot \beta_p$$

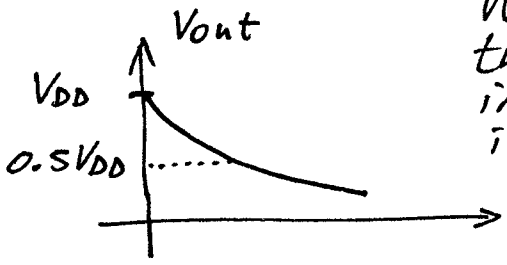
$$\beta_p = \mu_p \frac{\varepsilon_{ox}}{t_{ox}} \cdot \frac{W_p}{L_p}$$

therefore assumming that $V_{T,n} \simeq |V_{T,p}|$ we get that

$$t_{FALL} \simeq \frac{t_{RISE}}{2}$$

Since in the reality we want the gate to behave in the same way for both transitions ($0 \to 1$, $1 \to 0$) usually the p-MOS is build with an aspect ratio 2 times bigger than the nMOS aspect ratio.

3. t_{PHL}



We assume that the input pulse is ideal

if we want to compute t_{PHL} we follow the same approach used for t_{FALL} but this time we need to integrate from V_{DD} to $0.5V_{DD}$

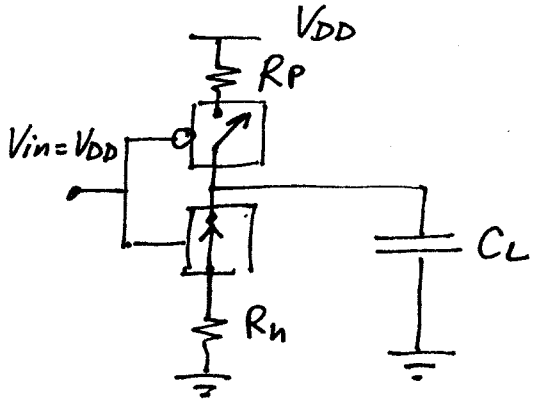
$$-i_{D,n} = C_L \frac{dv_{out}}{dt}$$

$$\int_0^{t_{PHL}} dt = -C_L \int_{V_{DD}}^{0.5V_{DD}} \frac{dv_{out}}{i_{D,n}}$$

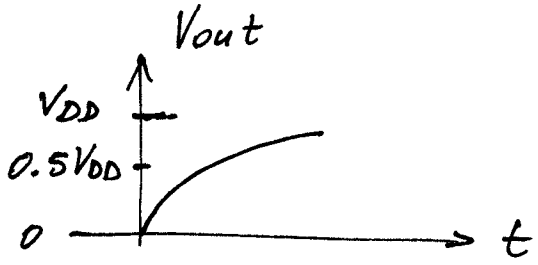
and we need to keep in mind that from V_{DD} to $V_{DD}-V_{T,n}$ the n-mos is in saturation and after in linear region.

$$t_{PHL} = \tau_n \left[\frac{2V_{T,n}}{V_{DD}-V_{T,n}} + \ln\left(\frac{3V_{DD}-4V_{T,n}}{V_{DD}} \right) \right]$$

$$\tau_n = \frac{1}{\beta_n (V_{DD}-V_{T,n})} \cdot C_L \approx R_n$$

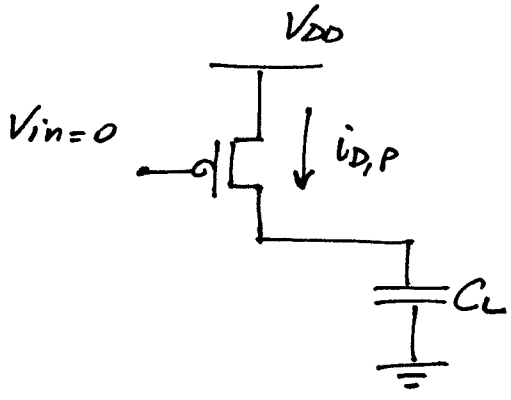


4. t_{PLH}



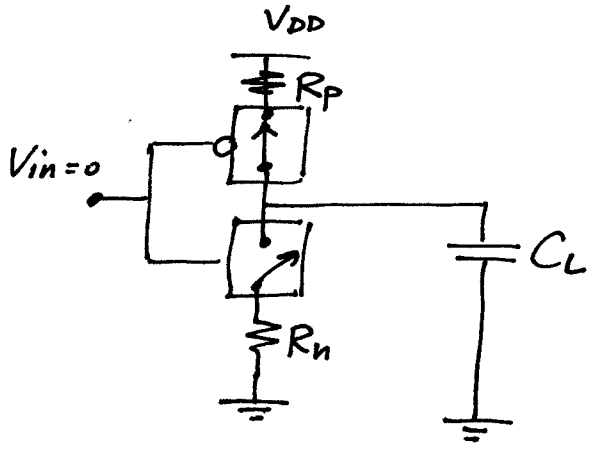
ideal input pulse

Usual approach:



$$i_{D,P} = C_L \cdot \frac{dV_{out}}{dt}$$

$$\int_0^{t_{PLH}} dt = C_L \int_0^{V_{DD}/2} \frac{dV_{out}}{i_{D,P}}$$

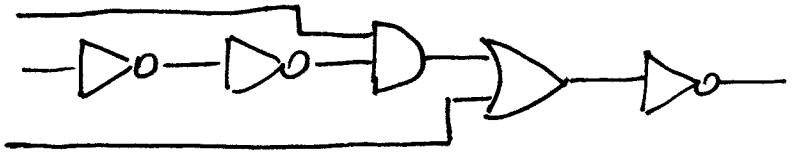


$$t_{PLH} = \tau_P \left[\frac{2|V_{T,P}|}{V_{DD} - |V_{T,P}|} + \ln\left(\frac{3V_{DD} - 4|V_{T,P}|}{V_{DD}} \right) \right]$$

$$\tau_P = \frac{1}{\beta_P \left(V_{DD} - |V_{T,P}| \right)} \cdot C_L$$

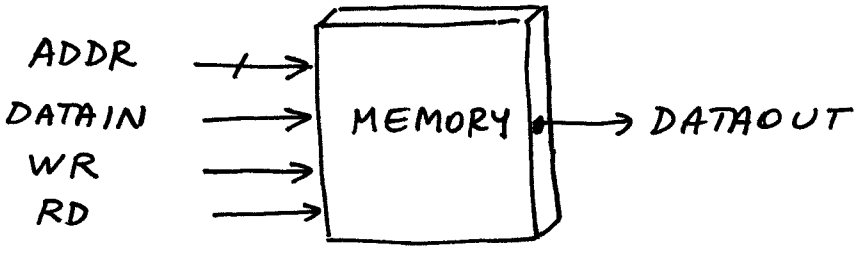
Some physical considerations about switching char.

The propagation delay time t_p gives the time taken by the gate to propagate a change in the input to the output. From a very practical point of view this means that any change in the input will take effect only after a certain delay. So if we need several level of logic gates to perform the function we want to achieve, it may take a long time before the output is ready (has responded to the changes in the inputs)



5 level of logic

If the propagation delay through our network is too long, the output won't be ready on time causing an unexpected behavior



(e.g. I may write a wrong data in memory)

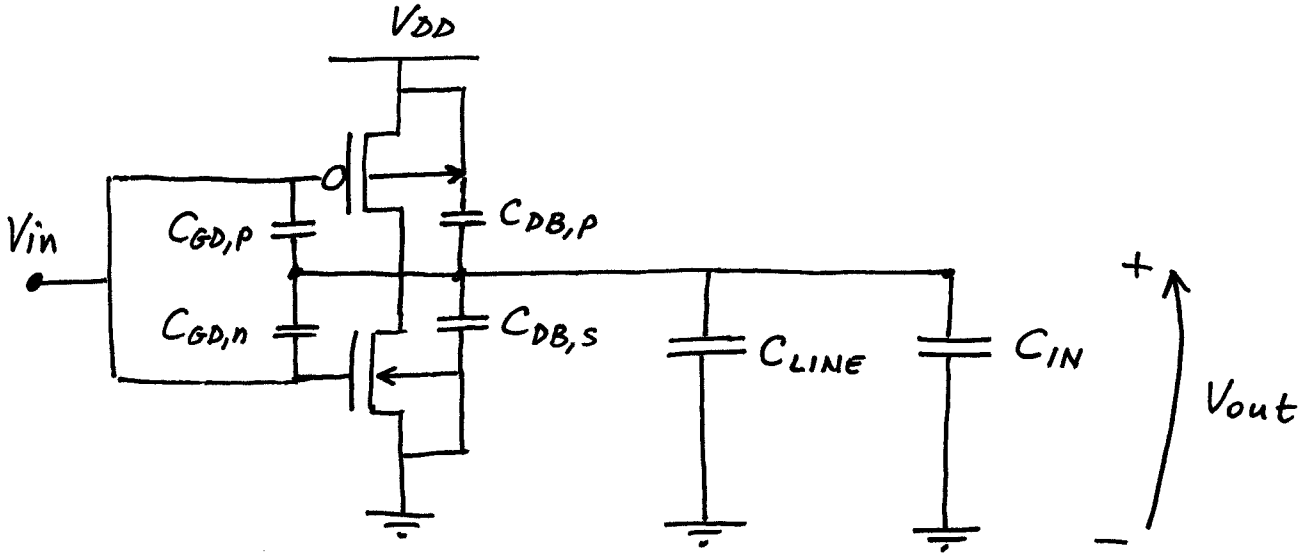
The propagation delay affects the performances (the speed) of the system we are able to build.

It's important to understand what are the factors that Limit the propagation delay.

For that purpose we look at the issue from a slightly different perspective.

If we look at the calculations done so far we see that all transient times are proportional to the capacitance C_{LOAD}.

In the reality the capacitance C_{LOAD} is composed by several contributes.



$$C_{LOAD} = \underbrace{C_{GD,n} + C_{GD,p} + C_{DB,n} + C_{DB,p}}_{C_{intrinsic}} + \underbrace{C_{LINE} + C_{IN}}_{C_{extrinsic}}$$

There are two main contributes to the C_{LOAD} one given by some of the parasitic capacitance of the mos transistors (some have no effect on the dynamic behavior) and one given by the sum of interconnect capacitance C_{LINE} and input capacitance of the mosFETs in the next stage

$$C_{IN} = \sum_{i=1}^{FO} \left(C_{g,n} + C_{g,p} \right)_i \qquad FO = \text{fan out}$$

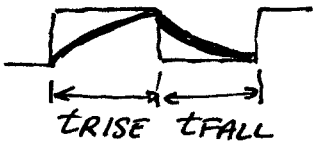
The C_{LOAD} however is not the only factor that affects the propagation delay. There is another factor that we have completely neglected so far: the slope of the input ramp (also called slew rate of the input).

The slope of the input waveform can modify the delay of a gate considerably (this is due to the fact that "as the voltage change on the gate of a transistor, so does the capacitance of the gate terminal" [W-E]).

Summarizing the propagation delay depends on:

- slope of the input waveform
- intrinsic capacitance
- extrinsic capacitance

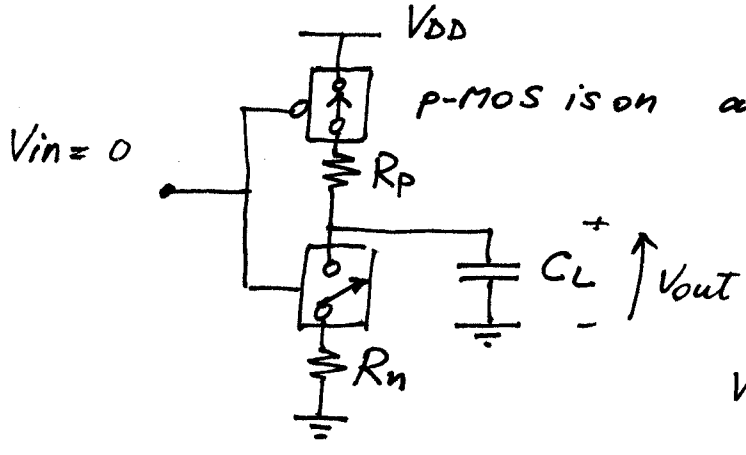
The sum of fall time and rise time ($t_{FALL} + t_{RISE}$) can be interpreted as the minimum time needed for a gate to go through a complete switching cycle (from 0 → 1 and 1 → 0).



$$\text{maximum switching frequency} = \frac{1}{t_{RISE} + t_{FALL}}$$

"If we have a system, the max switching frequency is set by the slowest gate in the network." [U]

The analysis we have done to find out the switching characteristics of a gate is way too complicated. Usually simple RC models are more than enough to get reasonable estimates of the switching time.



Charging

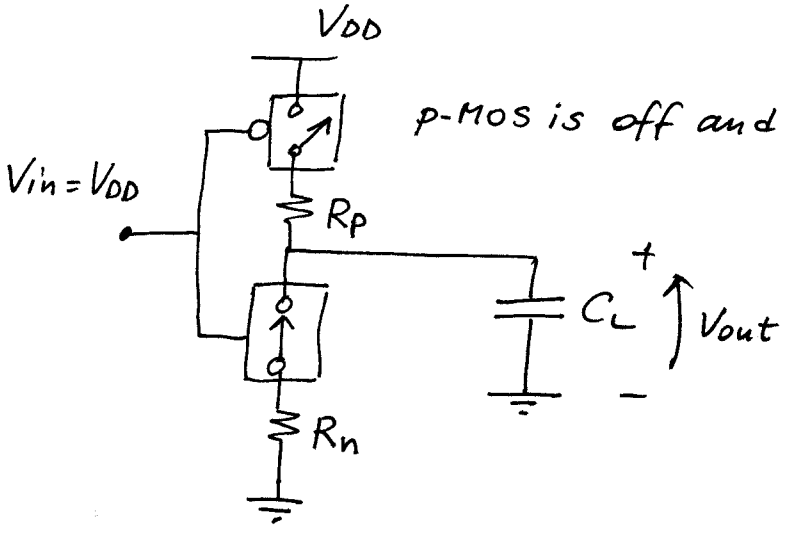
$$V_{out}(t) = V_{DD}\left[1 - e^{-t/\tau_p}\right]$$

R_P, R_n → drain-to-source equivalent resistance

where R_p and R_n are approximated with:

$$R_n \simeq \frac{1}{\beta_n (V_{DD} - V_{Tn})}$$

$$R_p \simeq \frac{1}{\beta_p (V_{DD} - |V_{T,p}|)}$$



p-MOS is off and n-MOS is on

discharging ↓

$$V_{out}(t) = V_{DD}\, e^{-t/\tau_n}$$

$$\tau_p = R_p \cdot C_L \quad ; \quad \tau_n = R_n \cdot C_L$$