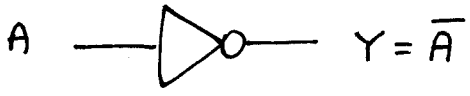
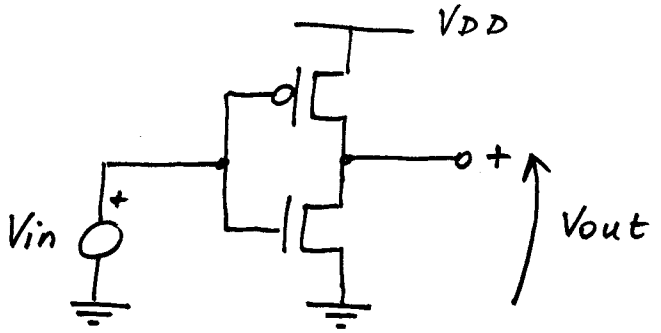
CMOS Inverter [W-E] [K-L] [U]

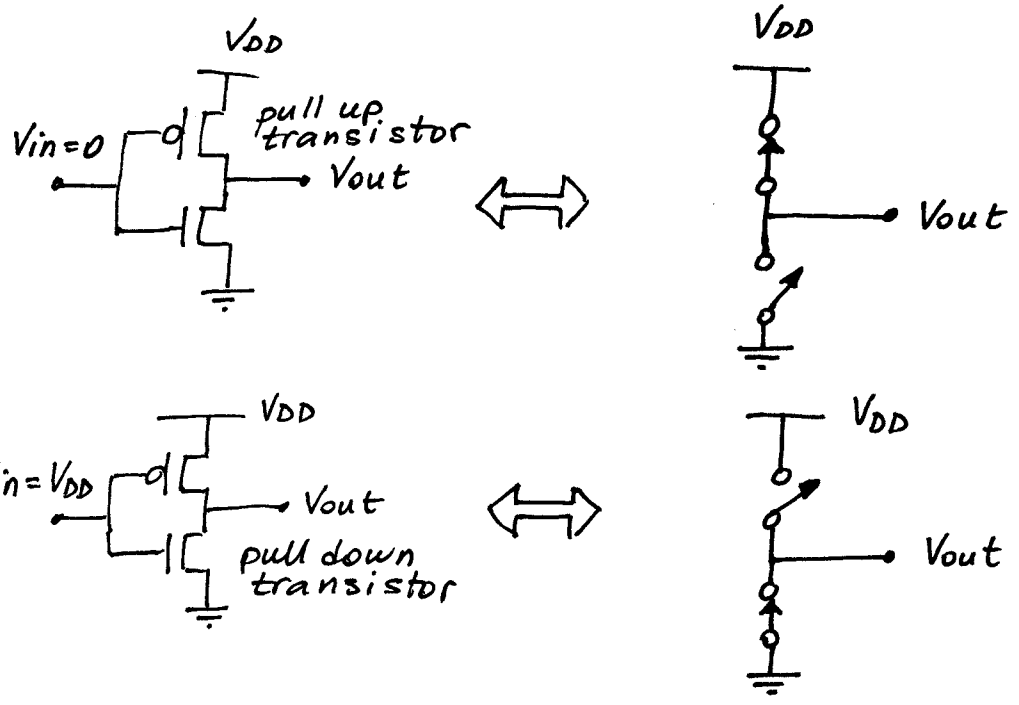
The inverter is the easiest logic gate we can build.



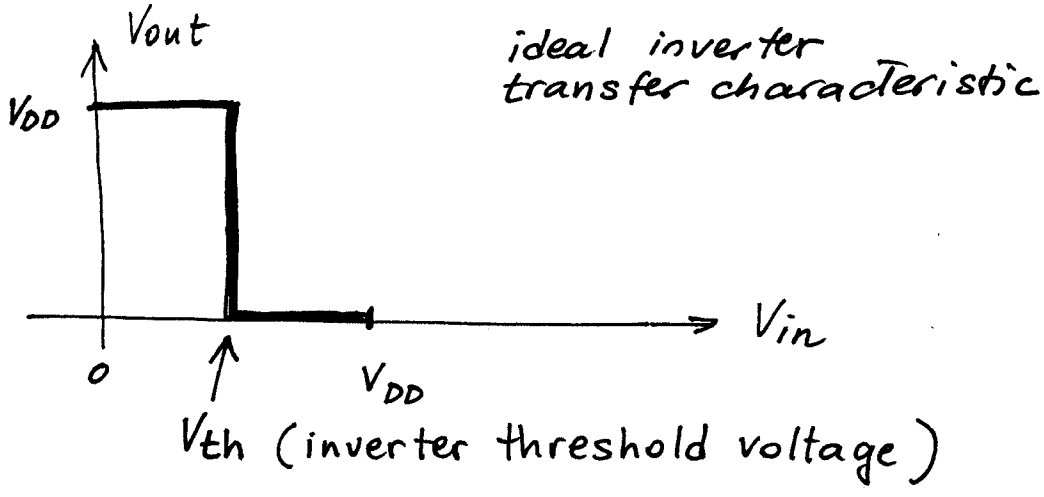
A	Y
0	1
1	0

The ideal behavior of the inverter can be found modeling the MOS transistors as switches

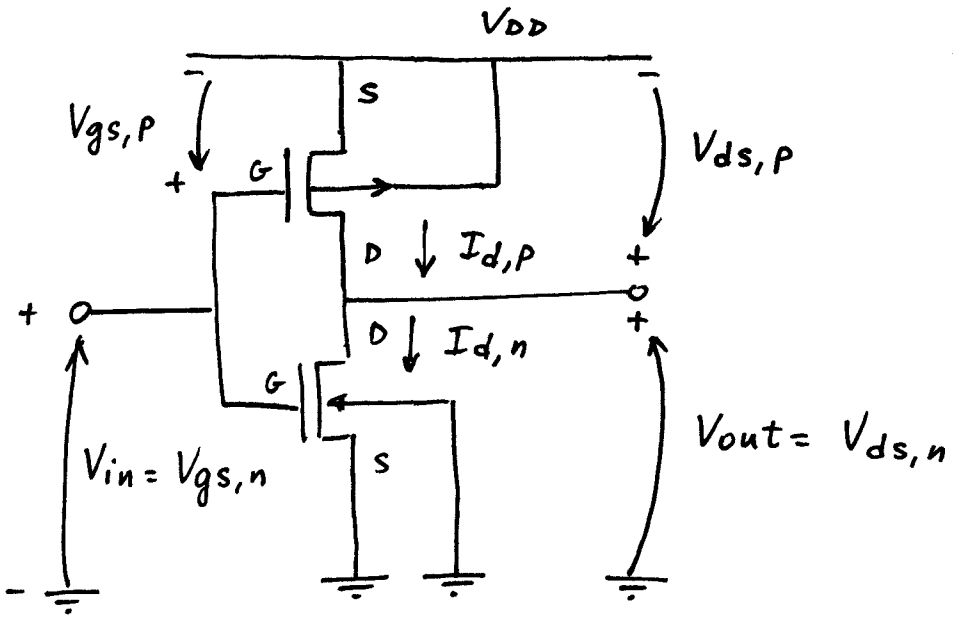




In this "idealized" analysis we assume that an input voltage $0 < V_{in} < V_{th}$ is "translated" by the ideal inverter as a logic 0 while an input voltage $V_{th} < V_{in} < V_{DD}$ is seen as a logic 1.



The nice thing about the inverter is that it is easy enough to allow a rigorous theoretical analysis.
The goal we want to achieve through the rigorous analysis of the inverter is to identify a general methodology and a set of definitions we can apply to characterize any logic gate.



$$V_{gs,n} = V_{in}$$
$$V_{ds,n} = V_{out}$$
$$V_{gs,p} = -(V_{DD} - V_{in}) = V_{in} - V_{DD}$$
$$V_{ds,p} = -(V_{DD} - V_{out}) = V_{out} - V_{DD}$$

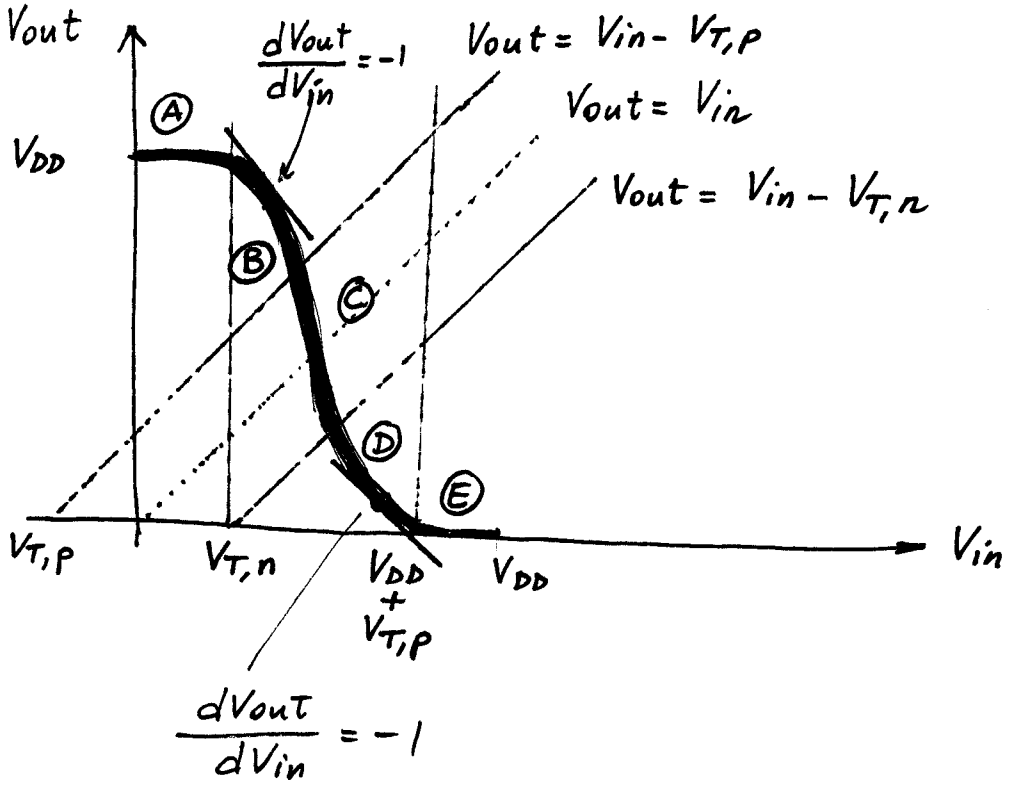
Equations (*)

where V_{in} is in the range: $0 \le V_{in} \le V_{DD}$

The static (DC) circuit characteristic can be derived by the voltage transfer curve (VTC → Vout as a function of V_{in}).

V_{OH} → maximum logic 1 output voltage
V_{OL} → minimum logic 0 output voltage
V_{IH} → ~~maximum~~ minimum logic 1 input voltage
V_{IL} → ~~minimum~~ maximum logic 0 input voltage
V_{th} → switching voltage

The DC analysis consists in plotting Vout as a function of Vin. In order to plot the VTC we need to solve the equation of the drain current I_d for various input voltage values.
Varying Vin the nMOS and pMOS will operate in different regions.



Region A

$0 \leq Vin \leq V_{T,n}$ $(I_{ds,n} = 0)$

The nMOSFET is in cut-off and the pMOSFET is in linear region.

Since $I_{ds,n} = -I_{ds,p}$ the current for the pMOSFET is also zero. As consequence $V_{ds,p} = 0$ then from (*)

$$-V_{ds,p} = V_{out} - V_{DD}$$

$$V_{out} = V_{DD} \triangleq V_{OH}$$ high output voltage

Region B

$$V_{T,n} \leq V_{in} < \frac{V_{DD}}{2}$$

The nMOS is in saturation and the pMOS in linear region.

$$I_{ds,n} = \beta_n \frac{(V_{in} - V_{T,n})^2}{2}$$

where $\beta_n = \frac{\mu_n \varepsilon_{ox}}{t_{ox}} \cdot \frac{W_n}{L_n}$

$$I_{ds,p} = -\beta_p \left[(V_{gs,p} - V_{T,p}) V_{ds,p} - \frac{V_{ds,p}^2}{2} \right]$$

$$= -\beta_p \cdot \left[(V_{in} - V_{DD} - V_{T,p}) \cdot (V_{out} - V_{DD}) - \frac{(V_{out} - V_{DD})^2}{2} \right]$$

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

Recalling that $I_{ds,n} = - I_{ds,p}$:

$$\beta_n \frac{(V_{in} - V_{T,n})^2}{2} = \beta_p \left[(V_{in} - V_{DD} - V_{T,p})(V_{out} - V_{DD}) + - \frac{(V_{out} - V_{DD})^2}{2} \right]$$

and carrying on manipulating the equality we can find out an espression for V_{out}.

If we differentiate both sides with respect to V_{in} :

$$\frac{d}{dV_{in}} \left[\frac{\beta_n}{2} (V_{in} - V_{T,n})^2 \right] = \beta_n (V_{in} - V_{T,n})$$

$$\beta_p \cdot \frac{d}{dV_{in}} \left[(V_{in} - V_{DD} - V_{T,p})(V_{out} - V_{DD}) + - \frac{(V_{out} - V_{DD})^2}{2} \right] =$$

$$\beta_p \frac{d}{dV_{in}} \left[(V_{in} - V_{DD} - V_{T,p}) V_{out} - V_{DD}(V_{in} - V_{DD} - V_{T,p}) + \right.$$
$$\left. - \frac{(V_{out} - V_{DD})^2}{2} \right] =$$

$$\uparrow$$

$$\frac{d}{dx} f(x) g(x) = f'(x) \cdot g(x) + g'(x) \cdot f(x)$$

$$= \beta_p \left[(V_{in} - V_{DD} - V_{T,p}) \cdot \frac{dV_{out}}{dV_{in}} + V_{out} - V_{DD} + \right.$$
$$\left. - (V_{out} - V_{DD}) \cdot \frac{dV_{out}}{dV_{in}} \right]$$

Then:

$$\beta_n (V_{in} - V_{T,n}) = \beta_p \left[(V_{in} - V_{DD} - V_{T,p}) \frac{dV_{out}}{dV_{in}} + \right.$$
$$\left. + (V_{out} - V_{DD}) - (V_{out} - V_{DD}) \frac{dV_{out}}{dV_{in}} \right]$$

Substituting $V_{in} = V_{IL}$ and $\left.\frac{dV_{out}}{dV_{in}}\right|_{V_{IL}} = -1$

$$\beta_n (V_{IL} - V_{T,n}) = \beta_p \left[-(V_{IL} - V_{DD} - V_{T,p}) + V_{out} - V_{DD} + \right.$$
$$\left. + V_{out} - V_{DD} \right]$$

$$\beta_n (V_{IL} - V_{T,n}) = \beta_p \left[2V_{out} - V_{IL} + V_{T,p} - V_{DD} \right]$$

$$V_{IL}(\beta_n+\beta_p) = \beta_p\left[2V_{out} + V_{T,p} - V_{DD}\right] + \beta_n V_{T,n}$$

$$V_{IL} = \frac{\beta_p\left[2V_{out} + V_{T,p} - V_{DD}\right] + \beta_n V_{T,n}}{\beta_p + \beta_n}$$

Region C

In this region both nMOS and pMOS operate in saturation.

$$I_{ds,p} = -\frac{\beta_p}{2}\left(V_{gs,p} - V_{T,p}\right)^2$$

$$\downarrow$$

$$I_{ds,p} = -\frac{\beta_p}{2}\left(V_{in} - V_{DD} - V_{T,p}\right)^2$$

$$I_{ds,n} = \frac{\beta_n}{2}\left(V_{gs,n} - V_{T,n}\right)^2$$

$$\downarrow$$

$$I_{ds,n} = \frac{\beta_n}{2}\left(V_{in} - V_{T,n}\right)^2$$

Since $I_{ds,n} = -I_{ds,p}$

$$\frac{\beta_n}{2}\left(V_{in} - V_{T,n}\right)^2 = \frac{\beta_p}{2}\left(V_{in} - V_{DD} - V_{T,p}\right)^2$$

$$\beta_n\left(V_{in} - V_{T,n}\right) = -\beta_p\left(V_{in} - V_{DD} - V_{T,p}\right)$$

↑ I take the positive solution of the sqrt. (on both side)

Therefore:

$$V_{in} - V_{T,n} = \sqrt{\frac{\beta_p}{\beta_n}} \left(-V_{in} + V_{DD} + V_{T,p}\right)$$

$$V_{in}\left(1 + \sqrt{\frac{\beta_p}{\beta_n}}\right) = \sqrt{\frac{\beta_p}{\beta_n}}\left(+V_{DD} + V_{T,p} + \sqrt{\frac{\beta_n}{\beta_p}}\, V_{T,n}\right)$$

$$V_{in} = \frac{\sqrt{\beta_p/\beta_n}\left(V_{DD} + V_{T,p} + \sqrt{\beta_n/\beta_p}\; V_{T,n}\right)}{1 + \sqrt{\beta_p/\beta_n}}$$

From that solution we see that the region C exists only for one value of V_{in}. (n-MOS transistor and pMOS transistor are both in saturation for only one value of V_{in}).
The VTC corresponding to region C is completely vertical (in reality the slope is very steep, but not infinite) → we are neglecting the channel length modulation effect).

The possible values of V_{out} in region C can be derived from the saturation condition

$$nMOS: \quad V_{ds,n} > V_{gs,n} - V_{T,n}$$

$$\downarrow$$

$$V_{out} > V_{in} - V_{T,n}$$

pMOS: $V_{ds,p} < V_{gs,p} - V_{T,p}$

$$\downarrow$$

$$V_{out} - V_{DD} < V_{in} - V_{DD} - V_{T,p}$$

$$V_{out} < V_{in} - V_{T,p}$$

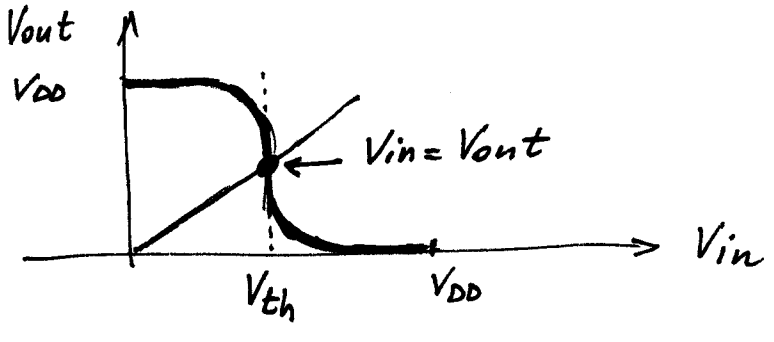
combining the inequalities:

$$V_{in} - V_{T,n} < V_{out} < V_{in} - V_{T,p} \quad (+)$$

The equation for V_{in} is used to define the switching threshold V_{th} of the inverter

$$V_{th} = V_{in} = \frac{\sqrt{\beta_p/\beta_n}\left(V_{DD} + V_{T,p} + \sqrt{\beta_n/\beta_p}\, V_{T,n}\right)}{1 + \sqrt{\beta_p/\beta_n}}$$

which correspond to the case where $V_{in} = V_{out} = V_{th}$ (without violating the disequalities (+))



For $\beta_p = \beta_n$ and $V_{Tn} = |V_{Tp}| \Rightarrow V_{th} = \frac{V_{DD}}{2}$

Region D

$$\frac{V_{DD}}{2} < V_{in} \leq V_{DD} - V_{T,p}$$

The pMOS is in saturation while the nMOS is in linear region

$$I_{ds,p} = -\frac{\beta_p}{2}\left(V_{gs,p} - V_{T,p}\right)^2$$

$$\downarrow$$

$$I_{ds,p} = -\frac{\beta_p}{2}\left(V_{in} - V_{DD} - V_{T,p}\right)^2$$

$$I_{ds,n} = \beta_n\left[\left(V_{gs,n} - V_{T,n}\right)V_{ds,n} - \frac{V_{ds,n}^2}{2}\right]$$

$$\downarrow$$

$$I_{ds,n} = \beta_n\left[\left(V_{in} - V_{T,n}\right)V_{out} - \frac{V_{out}^2}{2}\right]$$

$$I_{ds,p} = -I_{ds,n}$$

$$+\frac{\beta_p}{2}\left(V_{in} - V_{DD} - V_{T,p}\right)^2 = \beta_n\left[\left(V_{in} - V_{T,n}\right)V_{out} - \frac{V_{out}^2}{2}\right]$$

Differentiating both side with respect to V_{in} :

$$\beta_p (V_{in} - V_{DD} - V_{T,p}) = \beta_n \left[(V_{in} - V_{T,n}) \frac{dV_{out}}{dV_{in}} + \right.$$

$$\left. + V_{out} - V_{out} \cdot \frac{dV_{out}}{dV_{in}} \right]$$

For $V_{in} = V_{IH}$ and $\left.\frac{dV_{out}}{dV_{in}}\right|_{V_{IH}} = -1$

$$\beta_p (V_{IH} - V_{DD} - V_{T,p}) = \beta_n \left[-V_{IH} + V_{T,n} + 2 V_{out} \right]$$

$$V_{IH} (\beta_p + \beta_n) = \beta_n (V_{T,n} + 2 V_{out}) + \beta_p (V_{DD} + V_{T,p})$$

$$V_{IH} = \frac{\beta_n (V_{T,n} + 2 V_{out}) + \beta_p (V_{DD} + V_{T,p})}{\beta_p + \beta_n}$$

Region E

$V_{in} \geq V_{DD} - V_{T,p}$

the pMOS is in cut-off ($I_{ds,p} = 0$) while the nMOS is in linear region.

Since $I_{ds,n} = -I_{ds,p} = 0$ must be $V_{ds,n} = 0$

Therefore since $V_{ds,n} = V_{out}$

$V_{out} = 0 \triangleq V_{OL}$ — low output voltage

[K-L]

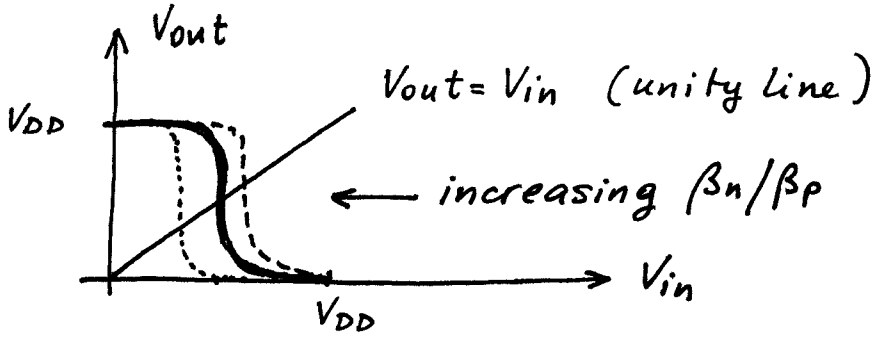
Physical interpretation of the DC parameters

We have seen how To define and derive the static parameters of an inverter, but we did not really understand why we have chosen those parameters.
The goal is To understand what is their physical meaning so that we can appreciate why we can use them to characterize a logic gate.
We have defined 5 parameters:

- V_{th} → switching threshold (gate threshold V)
- V_{OH} → maximum high output voltage
- V_{OL} → minimum low output voltage
- V_{IL} → maximum logic 0 input
- V_{IH} → minimum logic 1 input

gate threshold voltage V_{th}

It represents the midpoint of the switching characteristic. V_{th} is given by the intersection of the VTC and the unity line.



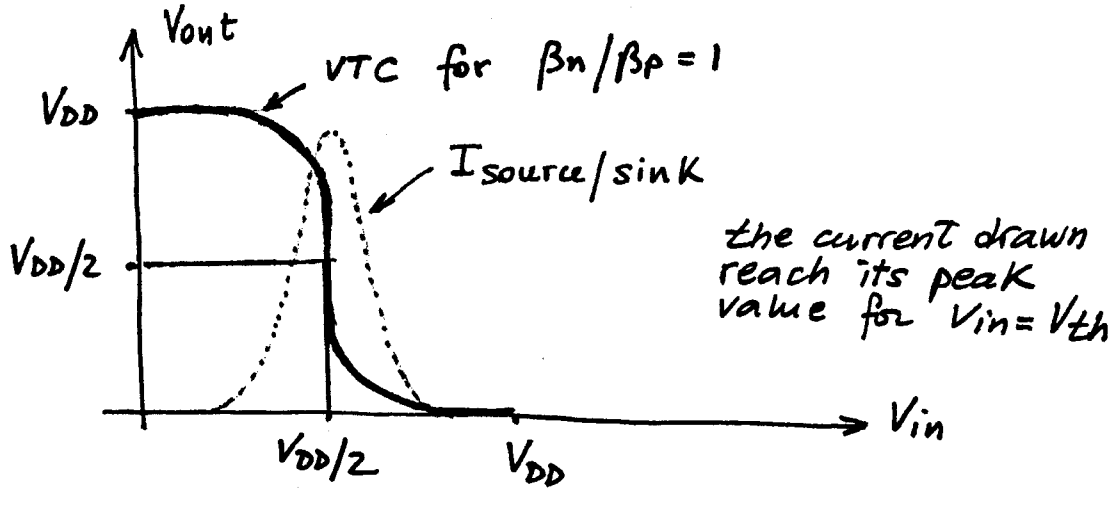
The designer can change the logic gate threshold voltage through the ratio β_n/β_p

$$V_{th} = \frac{(\beta_p/\beta_n)^{+1/2}\left[V_{DD} - |V_{T,p}| + (\beta_n/\beta_p)^{1/2} V_{T,n}\right]}{1 + (\beta_p/\beta_n)^{1/2}}$$

the ratio β_n/β_p can be changed working on the aspect ratios W_n/L_n and W_p/L_p of the transistors.

As the ratio β_n/β_p is decreased the transfer curve is shifted from left to right.

Typically we try to build the inverter as symmetric as possible since it allows a capacitive load to be charged and discharged in equal times (For $\beta_n/\beta_p = 1$ we have equal source and sink currents)



It's also important to notice that the temperature has an effect on the VTC.
We have to remember that as the temperature is increased the carrier mobility μ decreases (therefore there is a decrease in β as well)

$$\beta \propto \mu \propto T^{-3/2}$$

Summarizing the gate threshold voltage can be intrepreted as a measure of the "symmetry" of a gate behavior.

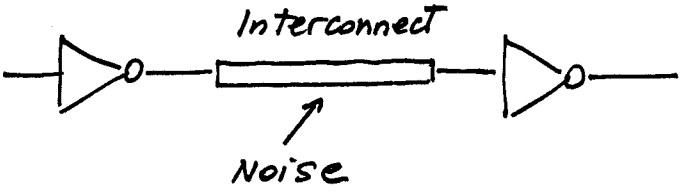
Maximum high Vout & Minimum low Vout

Here the physical interpretation of the two parameters is straightforward.
The interesting thing to notice is that the maximum high VouT is given when the nMOS is in cut-off region while the minimum low Vout is given when the pMOS is in cut-off region.
In both cases there is no current drawn from the power supply (therefore in those steady states the power dissipation is negligible).

No current flow means no voltage drops so we can rely on a full output voltage swing between logic 0 and logic 1 (full output voltage swing between 0 V and V_{DD}).

V_{IL} and V_{IH}

The maximum logic 0 input V_{IL} and the minimum logic 1 input express the ability of the gate to interpret an input voltage within a certain range as a logic 0 or as a logic 1.
This ability allows a certain tolerance to external noise



V_{IL} is the maximum allowable voltage at the input of a gate that can still be interpreted as a logic 0.

V_{IH} is the minimum allowable voltage at the input of a logic gate that can be interpreted as a logic 1.

The two parameters V_{IL} and V_{IH} definition (using the slope condition $\frac{dVout}{dVin} = -1$) is based on noise considerations.

The output voltage of an inverter, in steady state conditions is a function of the input voltage

$$Vout = f(Vin)$$

If the input voltage is altered by a noise the output voltage will be also altered

$$V_{out}^{(altered)} = f(V_{in} + V_{noise})$$

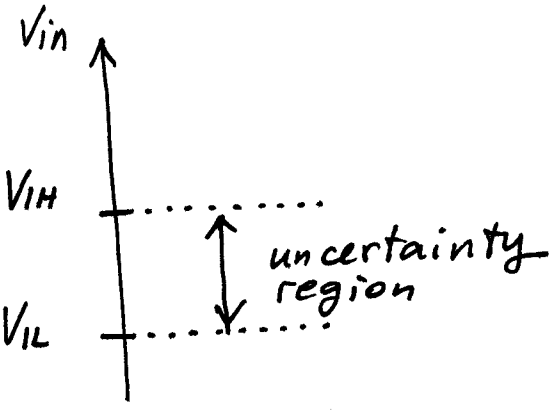
If we approximate the altered output through Taylor series expansion:

$$V_{out}^{(altered)} = f(V_{in}) + \frac{dVout}{dVin} \cdot \Delta V + \ldots$$

↑ higher order terms

Since $\frac{dVout}{dVin}$ represents the voltage gain of the inverter (for a small signal centered around the operating point Vin), we can

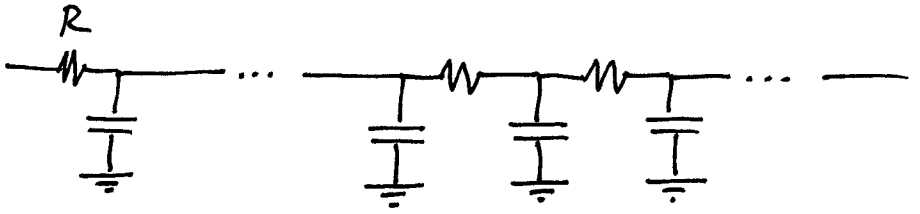
easily see that if the magnitude of the voltage gain is smaller than unity the input perturbation is not amplified and consequently the output perturbation remains quite small.
(this explain why we use as boundary condition $\frac{dVout}{dVin} = -1$)



The voltage range $V_{IH} < V_{in} < V_{IL}$ represents a region that we do not know how to interpret (logic 0 or logic 1 ?)
for that reason, in practice, one of the design objectives is to try to reduce the width of the uncertain region as much as possible.
[K-L]

Example [W-E]

The propagation of a signal along a wire depends mainly on the distributed resistance and capacitance of the wire.
If the interconnect is very long it is no longer possible to neglect the distributed nature of R and C and we need to model the interconnection through a transmission line (rather than a simple lumped RC model)



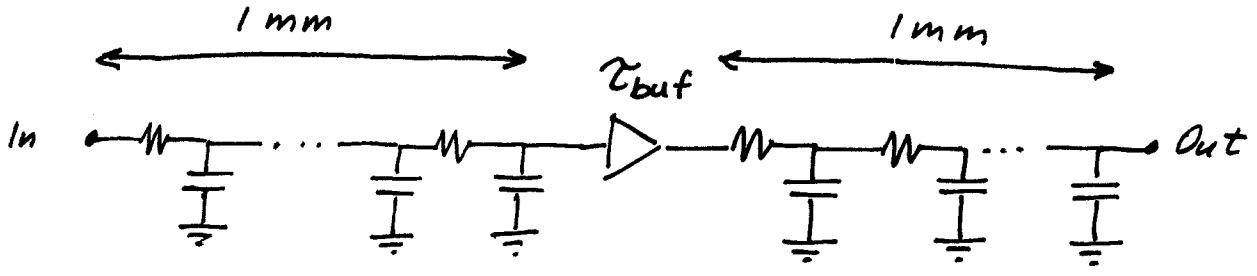
$$rc\,\frac{dV}{dt} = \frac{d^2V}{dx} \qquad (*)$$

x = distance of the interconnect
r = resistance per unit length
c = capacitance per unit length

The solution of (*) for a voltage step shows that the propagation delay t_p for a wire of length x is:

$$t_p \approx \frac{rc}{2}x^2$$

↑ quadratic with the distance



$r = 20\ \Omega/\mu m$

$c = 4 \times 10^{-4}\ pF/\mu m$

t_p = (without buffer) $16\ ns$

t_p = (with buffer) $4\ ns + \tau_{buf} + 4\ ns = 8\ ns + \tau_{buf}$

Interconnect delay can be a major issue in trying to achieve your circuit timing goals.
This is especially true for sub-micron technologies where the interconnect delay dominates the logic gates delay.