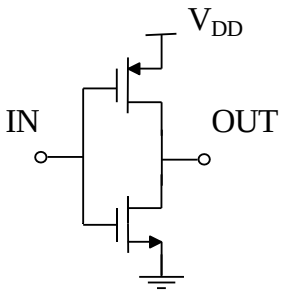
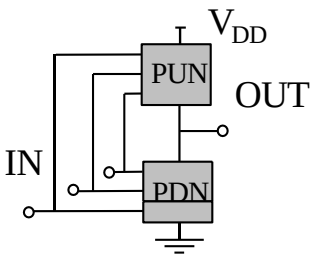
CMOS Logic

Inverter

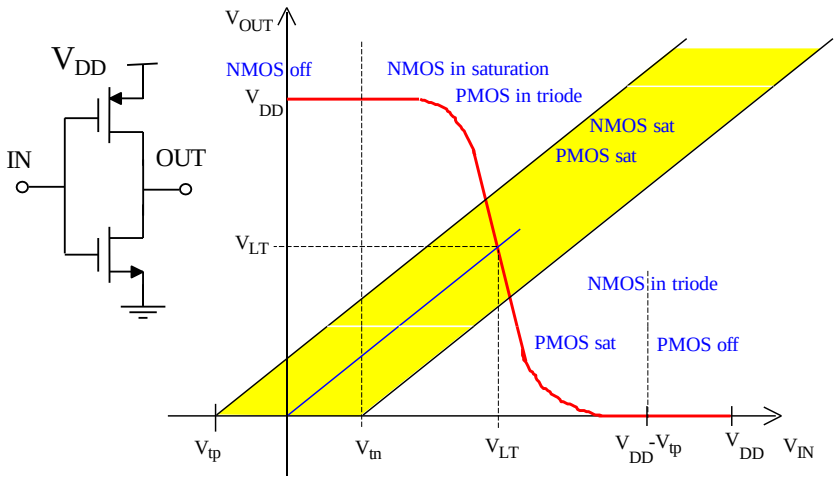


Complex function



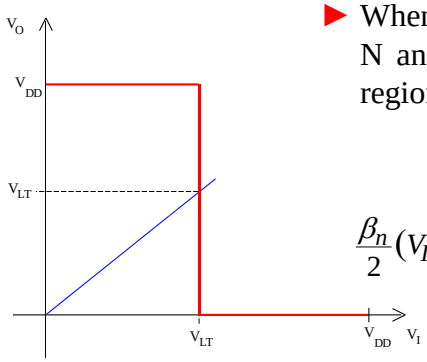
- The static power dissipation is ideally zero
- The high and low output value are V_{DD} and ground, respectively; ⇨ the voltage swing is equal to the power supply.
- Since the logic levels do not depend upon the relative device size, PDN and PUN transistors can be designed separately ⇨ all of them could be minimum size.
- A well designed CMOS gate has a low output impedence ⇨ it is less sensitive to noise disturbance.

CMOS inverter transfer characteristic



Logic threshold

Neglecting the channel length modulation the transfer characteristic simplify into



- When the output is equal to the input, N and P transistors are in saturation region (their drain current is equal)

$$\frac{\beta_n}{2}(V_{IN}-V_{tn})^2=\frac{p}{2}\left(V_{DD}-V_{IN}-|V_{tp}|\right)^2$$

- Solving for V_{IN}

$$V_{LT} = \frac{V_{DD} + \sqrt{\frac{\mu_n (W/L)_n}{p (W/L)_p}} V_{tn} - |V_{tp}|}{1 + \sqrt{\frac{\mu_n (W/L)_n}{\mu_p (W/L)_p}}}$$

- A symmetrical characteristic ($V_{LT} = 0.5 V_{DD}$) maximizes the noise margin. Assuming $V_{tn} = - V_{tp}$

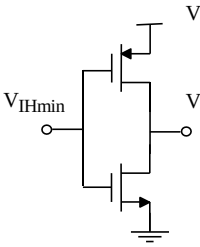
$$\frac{(W/L)_p}{(W/L)_n} = \frac{\mu_n}{\mu_p}$$

- A symmetrical characteristic leads to have equal the rise and fall time (symmetric time behavior). Indeed, the upper and lower transistors has the same drive capability
- $_n/\ _p$ typical values are 2.5 - 2.9 ⇒ $\frac{(W/L)_{..}}{(W/L)_n} = 3$
- V_{LT} is not much sensitive to β_n/β_p ratio (small variation on the ratio does not disturb the transfer characteristic)

  to reduce area and capacitances ⇒ $\frac{(W/L)_p}{(W/L)_{..}} = 2$

  this choice is further justified by secondary effects such as channel-length modulation and velocity saturation, but we lose time response symmetry

V_{IHmin} evaluation



Saturation region

Hp. $V_t = V_{tn} = -V_{tp}$; $\beta_n = \beta_p$

Triode region

$$(V_{IN} - V_t)V_{OUT} - \frac{V_{OUT}^2}{2} = \frac{1}{2}(V_{DD} - V_{IN} - V_t)^2 \quad (*)$$

Deriving with respect V_{IN}

$$V_{OUT} + (V_{IN} - V_t)\frac{dV_{OUT}}{dV_{IN}} - V_{OUT}\frac{dV_{OUT}}{dV_{IN}} = -(V_{DD} - V_{IN} - V_t)$$

By setting $\frac{dV_{OUT}}{dV_{IN}} = -1$ ⇒ $V_{IH\,\min} = V_{OLMAX} + \frac{V_{DD}}{2}$

Substituting in (*)

$$\left(\frac{V_{DD}}{2} - V_t\right)V_{OLMAX} + \frac{V_{OLMAX}}{2} = \frac{1}{2}\left(-V_{OLMAX} + \frac{V_{DD}}{2} - V_t\right)^2$$

$$V_{OLMAX} = \frac{1}{4}\left(\frac{V_{DD}}{2} - V_t\right) \Rightarrow V_{IH\,\min} = \frac{1}{4}\left(\frac{5V_{DD}}{2} - V_t\right)$$

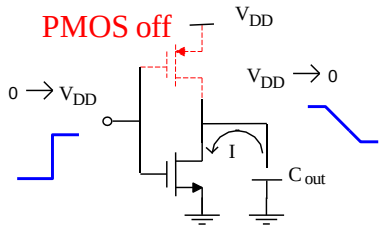
- Due to the symmetry around V_{TL}

$$V_{OH\min} = V_{DD} - V_{OLMAX} = \frac{1}{4}\left(\frac{7V_{DD}}{2} + V_t\right)$$

$$V_{ILMAX} = V_{DD} - V_{IH\min} = \frac{1}{4}\left(\frac{3V_{DD}}{2} + V_t\right)$$

- Noise margin $\quad NM = NM_H = NM_L = \frac{V_{DD}}{4} + \frac{V_t}{2}$

Fall time



$$C_{out}\frac{OUT}{dt} = -I$$

- $V_{OUT} > V_{DD} - V_t$, NMOS in saturation region $\Rightarrow$ $I = \frac{\beta_n}{2}(V_{DD} - V_t)^2$

$$\tau_{f1} = -C_{out}\int_{0.9V_{DD}}^{V_{DD}-V_t}\frac{dV}{I} = \frac{C_{out}}{I}[V]_{V_{DD}-V_t}^{0.9V_{DD}} \approx \frac{C_{out}V_t}{\frac{\beta_n}{2}(V_{DD} - V_t)^2}$$

- $V_{OUT} < V_{DD} - V_t$, NMOS in linear region $\Rightarrow$ $I = \beta_n\left(V_{DD} - V_t - \frac{V_{OUT}}{2}\right)V_{OUT}$

$$\tau_{f2} = -C_{out}\int_{V_{DD}-V_t}^{0.1V_{DD}} \frac{dV}{I} = \frac{C_{out}}{\beta_n(V_{DD}-V_t)}\left[\int_{V_{DD}-V_t}^{0.1V_{DD}} \frac{-\frac{1}{2}dV}{V_{DD}-V_t-V/2} + \int_{V_{DD}-V_t}^{0.1V_{DD}} \frac{-dV}{V}\right] =$$

$$= \frac{C_{out}V_t}{\beta_n(V_{DD}-V_t)}\left[\ln(V_{DD}-V_t-V/2) - \ln V\right]_{V_{DD}-V_t}^{0.1V_{DD}} =$$

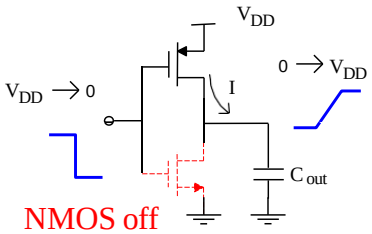
$$= \frac{C_{out}}{\beta_n(V_{DD}-V_t)}\ln\frac{2(V_{DD}-V_t)-0.1V_{DD}}{0.1V_{DD}}$$

$$\tau_f = \tau_{f1} + \tau_{f2} = \frac{C_{out}}{\frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)_n}\frac{f(V_{DD},V_t)}{V_{DD}-V_t}$$

- We can set the fall time with the NMOS W/L
- reducing the power supply NMOS is almost in saturation region

$$\tau_f \approx -C_{out}\int_{0.9V_{DD}}^{0.1V_{DD}} \frac{dV}{I} = \frac{0.8C_{out}V_{DD}}{\frac{\beta_n}{2}(V_{DD}-V_t)^2}$$

Rise time



$$\tau_r = \tau_{r1} + \tau_{r2} = \frac{C_{out}}{\frac{pC_{ox}}{2}\left(\frac{W}{L}\right)_p} \frac{r(V_{DD},V_t)}{V_{DD}-V_t}$$

$$r(V_{DD},V_t) \approx \frac{V_t}{V_{DD}-V_t} + \frac{1}{2}\ln\frac{2(V_{DD}-V_t)-0.1V_{DD}}{0.1V_{DD}}$$

- We can set the rise time with the PMOS W/L
- reducing the power supply PMOS is almost in saturation region

$$\tau_r \approx 1.6\frac{C_{out}V_{DD}}{\beta_p(V_{DD}-V_t)^2}$$

Propagation delay

- If the current is constant $t = \frac{C_{out}(V_2 - V_1)}{I_{av}}$

$$\tau_{PD} = \frac{C_{out}V_{DD}}{2I_m}$$

- I_{av} is the average current $I_{av} = \frac{I|_{V_o=V_{DD}} + I|_{V_o=V_{DD}/2}}{2}$

when $V_{OUT} = V_{DD}$ the NMOS in saturation
when $V_{OUT} = V_{DD}/2$ the NMOS in triode (it depends on the power supply value)

$$I_{av} = \frac{\frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)_n \left[(V_{DD} - V_t)^2 + 2\left(V_{DD} - V_t - \frac{V_{DD}}{4}\right)\frac{V_{DD}}{2}\right]}{2} =$$

$$= \frac{{}_nC_{ox}}{2}\left(\frac{W}{L}\right)_n \left(\frac{7}{4}V_{DD}^2 - 3V_{DD}V_t + V_t^2\right)$$

- I_{av} can have a simpler expression if we assume NMOS transistor almost in saturation

$$PD = \frac{C_{out}V_{DD}}{\mu_n C_{ox}\left(\frac{W}{L}\right)_n (V_{DD} - V_t)^2}$$

The error is generally lower than 10%

The propagation delay and the other time constraint can be reduced

- reducing the load capacitance which means to reduce:
  - the internal diffusion capacitamce
  - the interconnect capacitance
  - the fan-out
- increasing W/L transistors. This need cautions since leads to increase both the diffusion capacitances and the fan-out of the previous gate
- increasing V_{DD}, but this determines a square increase in the power consumption, and it is generally fixed

CMOS power consumption

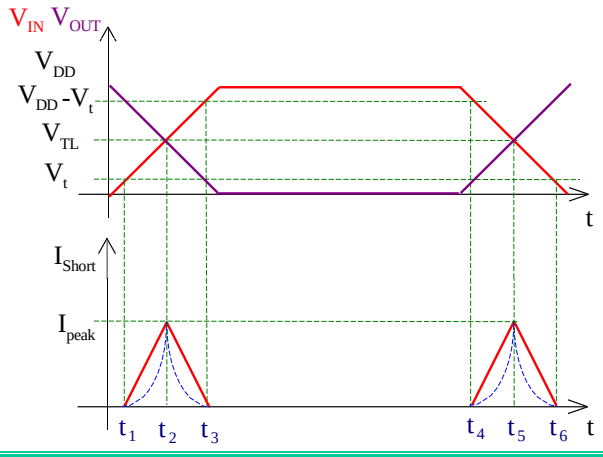
- Main power consumption: Dynamic power

$$P_d = \frac{E_d}{T} = C_{OUT} V_{DD}^2 f$$

- Second order dissipation: Consumption due to Direct-**Path** Currents (generally < 20%)

A direct current path exists (for a short period) during switching while NMOS and PMOS are conducting simultaneously

Hp. 1) The input has a finite slope
2) The output follow the input
output capacititance was considered in the dynamic power



$$P_{dp} = 2\frac{V_{DD}}{T}\left[\int_{t_1}^{t_2} i_D(t)dt + \int_{t_2}^{t_3} i_D(t)dt\right] = 4\frac{V_{DD}}{T}\int_{t_{on}}^{\tau_{PD}} \frac{\beta_n}{2}\left[v_{GS}(t) - V_t\right]^2 dt =$$

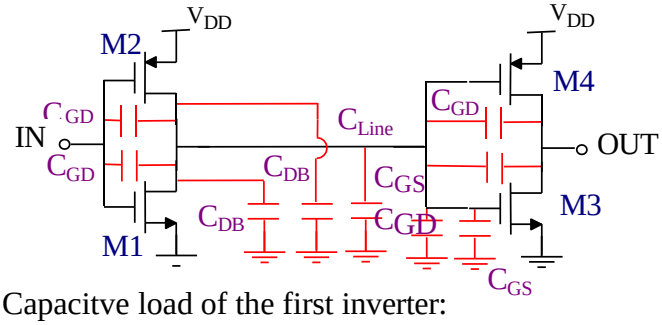
$$= 2\frac{V_{DD}}{T}\beta_n \int_{t_{on}}^{\tau_{PD}} \left(\frac{V_{DD}}{2\tau_{PD}}t - V_t\right)^2 dt = \frac{4}{3}\frac{\tau_{PD}}{T}\beta_n\left[\left(\frac{V_{DD}}{2\tau_{PD}}t - V_t\right)^3\right]_{t_{on}}^{\tau_{PD}}$$

* Since $\frac{V_{DD}}{2\tau_{PD}}t_{on} - V_t = 0$

$$P_{dp} = \frac{4}{3}\frac{\tau_{PD}}{T}\beta_n\left(\frac{V_{DD}}{2} - V_t\right)^3 = \frac{8}{3}\frac{\tau_{PD}}{T}I_{peak}\left(\frac{V_{DD}}{2} - V_t\right)$$

- Third order dissipation: Static consumption:
  - * due reverse-biased source or drain and substrate junction;
  - * due to sub-threshold current the interconnect capacitance (it is negligible when V_t>0.4 V)

Load capacitance evaluation

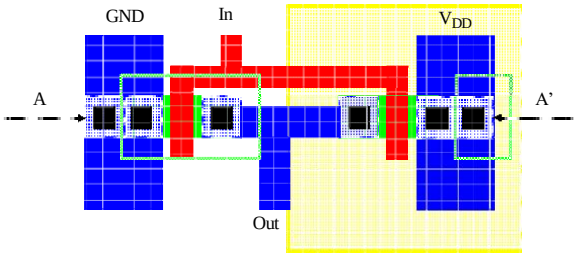


- Capacitve load of the first inverter:

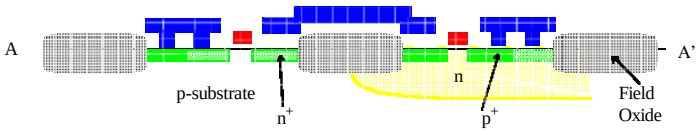
$$C = C_{DB1} + C_{DB2} + (1-1/A_V)(C_{GD1} + C_{GD2}) + C_{line} + C_{GS3} + C_{GS4} + (1-A_V)(C_{GD3} + C_{GD4})$$

$A_V = -1$ (Miller Effect)

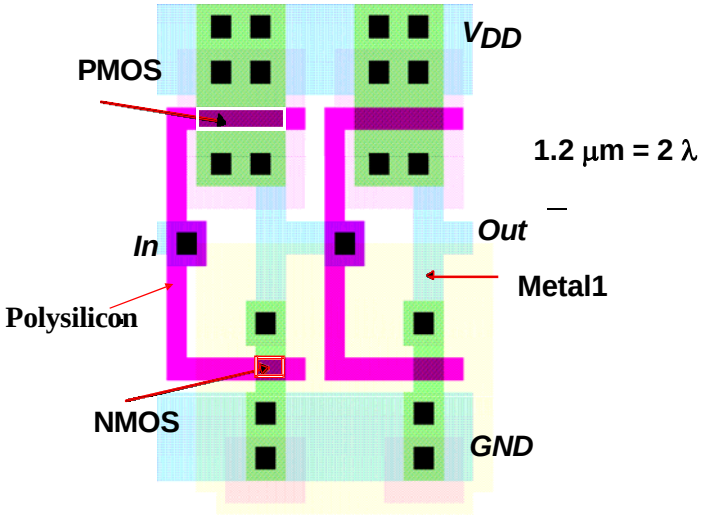
Inverter Layout



(a) Layout

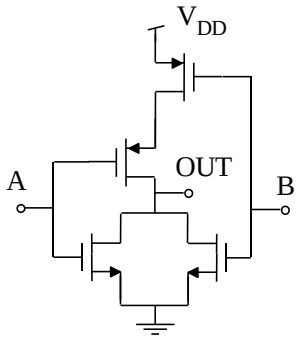


(b) Cross-Section along A-A'



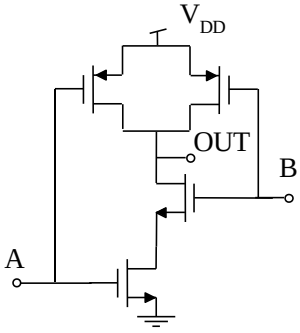
CMOS Gates

2-input NOR



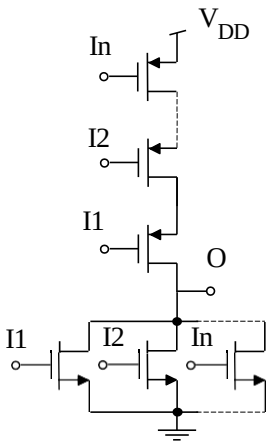
$OUT = \overline{A + B}$

2-input NAND



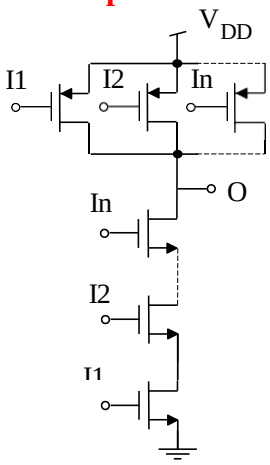
$OUT = \overline{A \cdot B}$

n-input NOR



$OUT = \overline{I1 + I2 + ... + In}$

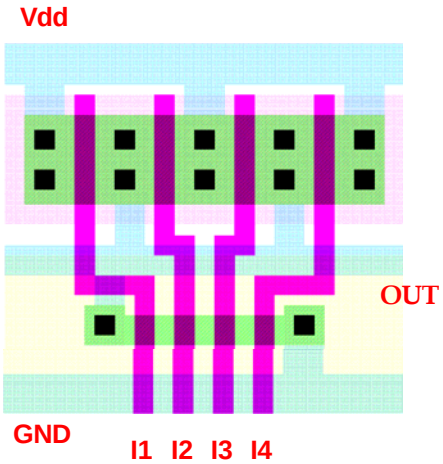
n-input NAND

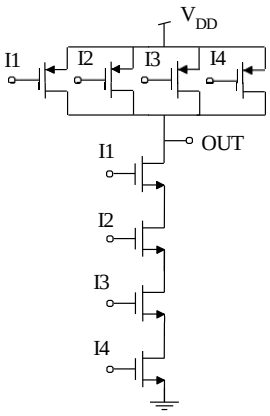


$OUT = \overline{I1 \cdot I2 \cdot ... \cdot In}$

Prof. Orazio Aiello 25

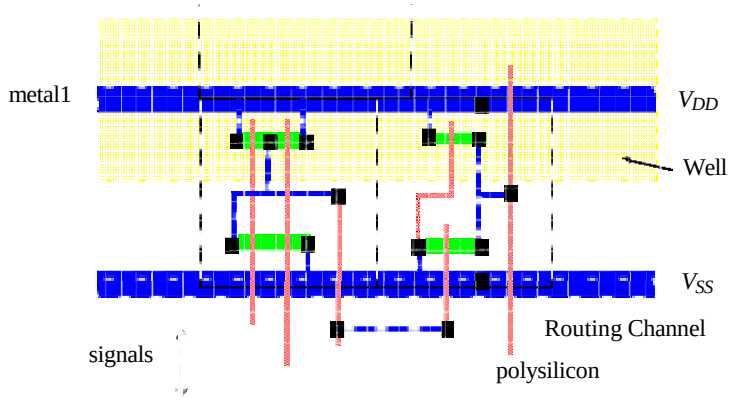
4-input NAND Layout



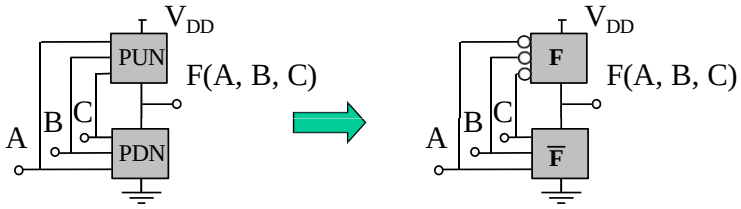


Prof. Orazio Aiello 26

Standard Cell Layout

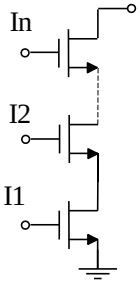


Logic formation

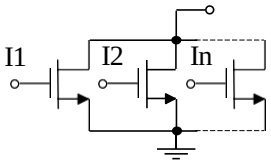


- Logic function are synthesized by adding series- and parallel connected transistor combination
- Each input is connected to the gate of both an NMOS and PMOS transistor
- The NMOS and PMOS circuits are complements of each other

- NMOS in parallel implement the OR operation
- NMOS in series implement the AND operation

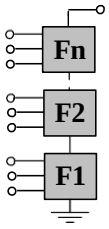


$$F = I1 \cdot I2 \cdot \ldots \cdot In$$

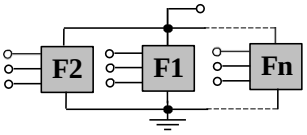


$$F = I1 + I2 + \ldots + In$$

- Parallel NMOS branches OR the individual branch function
- Logic function in series are ANDed together



$$F = F1 \cdot F2 \cdot \ldots \cdot Fn$$

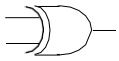


$$F = F1 + F2 + \ldots + Fn$$

- The output is the complement of the NMOS logic
- At the end add the p-channel upper part which is dual of the NMOS circuit.

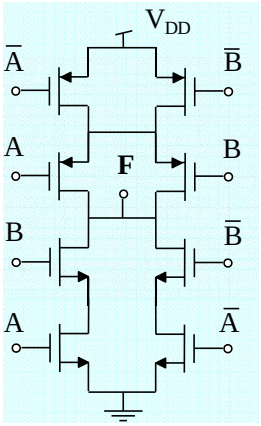
  inputs to NMOS which are in series (parallel) become parallel-(series-)connected in the PMOS section

Example 2-input XOR $F = A \oplus B = A\overline{B} + \overline{A}B$



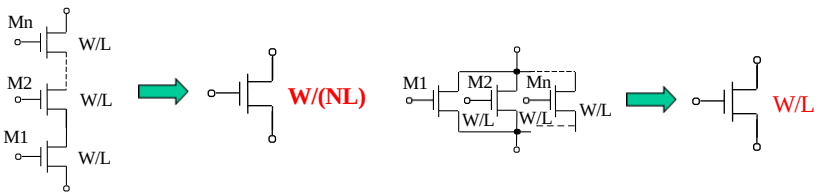
A	B	A⊕B
0	0	0
1	0	1
0	1	1
1	1	0

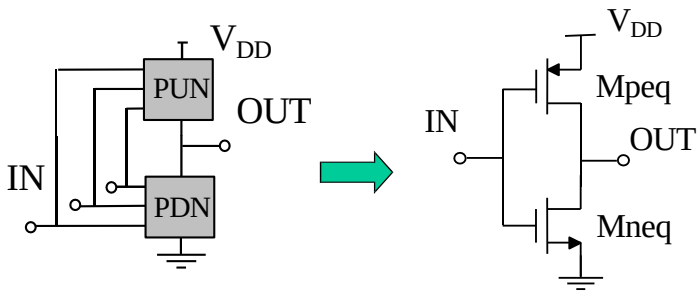
$\overline{F} = AB + \overline{A}\overline{B}$



Design of a CMOS gate

- The design of a complex gate can be simply reduced to that of an equivalent inverter (like for ratioed gate)
- Series of N transistors with W/L is equivalent to a transistor W/(NL)
- Equal transistors in parallel correspond to a single transistor

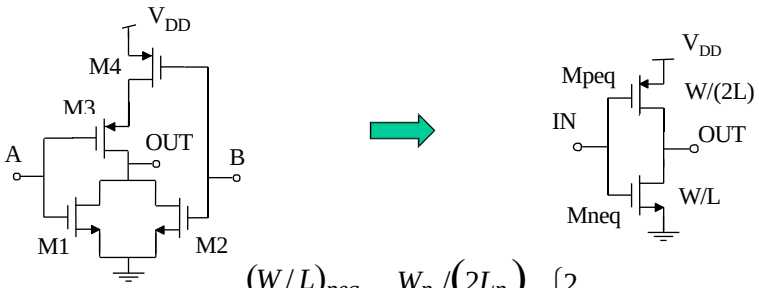




- To have an almost symmetrical characteristic set equal β_{neq}/β_{peq}

$$\frac{(W/L)_{peq}}{(W/L)_{neq}} = \begin{cases} 3 \\ 2 \end{cases}$$

Example 2-input NOR

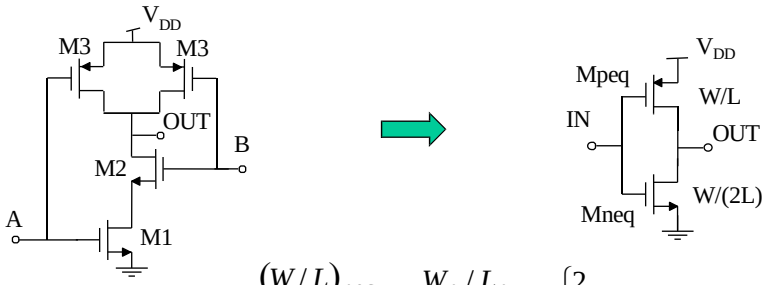


$$\frac{(W/L)_{peq}}{(W/L)_{neq}} = \frac{W_p/(2L_p)}{W_n/L_n} = \begin{cases} 2 \\ 3 \end{cases}$$

$L_1 = L_2 = L_3 = L_4 = L_{min}$;

$W_1 = W_2 = W_{min}$; $W_3 = W_4 = \mathbf{4}W_{min}$

Example 2-input NAND

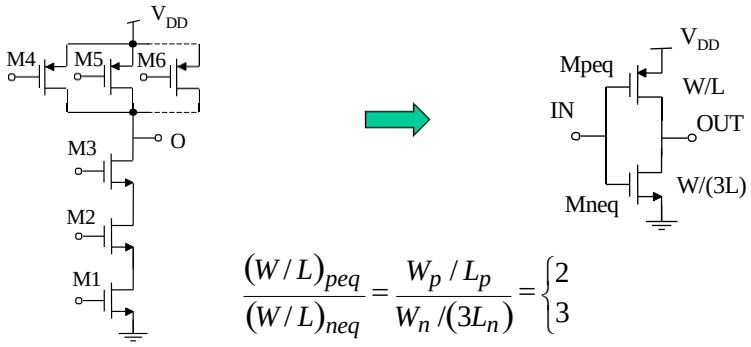


$$\frac{(W/L)_{peq}}{(W/L)_{neq}} = \frac{W_p / L_p}{W_n /(2L_n)} = \begin{cases} 2 \\ 3 \end{cases}$$

$$L_1 = L_2 = L_3 = L_4 = L_{min};$$

$$W_1 = W_2 = W_3 = W_4 = W_{min}$$

Example 3-input NAND



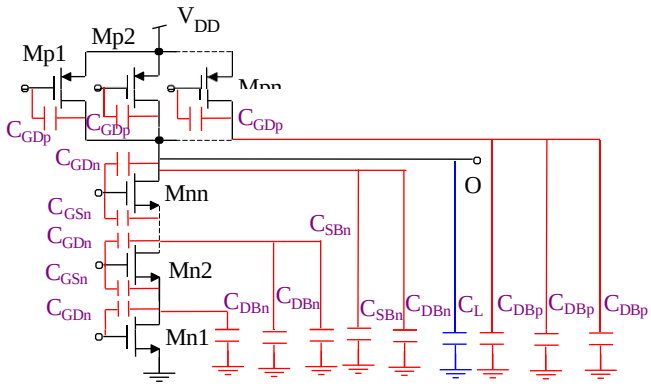
$$\frac{(W/L)_{peq}}{(W/L)_{neq}} = \frac{W_p / L_p}{W_n /(3L_n)} = \begin{cases} 2 \\ 3 \end{cases}$$

$$L_1 = L_2 = L_3 = L_4 = L_5 = L_6 = L_{min};$$

$$W_1 = W_2 = W_3 = W_4 = W_5 = W_6 = W_{min}$$

- 2-inputs and 3-input NAND with all minimum transistor are almost symmetrical
- Symmetrical NAND gates are more advantageous than NOR gates
  - require less area
  - provide better speed performance since has less parasitic capacitance
- Speed performance (worst case) of complex gates can be evaluated on the equivalent inverter and setting a proper equivalent capacitance

Capacitance evaluation



$$C = (n\text{-}1/A_V)(C_{GDn} + C_{GDp}) + (n\text{-}1)C_{GSn} + n(C_{DBn} + C_{DBp}) + (n\text{-}1)C_{SBn} + C_L$$

- C_{GD} of transistor ON can be assumed to be the value in triode region
- C_{GD} of transistor OFF is only due to ovelap contribute
- Capacitances of Switching transistors can be approximated with their final values
- Speed performance of gates strongly depend on the fan-in (*FI*) and fan-out (*FO*).

$$\tau = a_1 FI + a_2 FI^2 + a_3 FO$$

Proportional reduction of drive capability to the transistors in series