Digital integrated circuit design
Notes

Professor:
Andrea Bonfanti

Student:
Matteo Baldo

June 11, 2017

...and so on and so forth ...

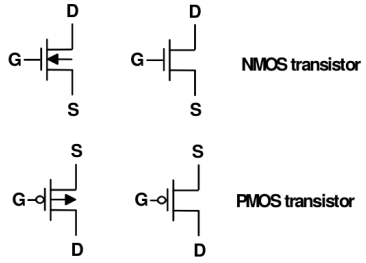
Contents

Chapter 1

The MOS transistor

1.1 Symbols

The symbols used to draw n and p MOS transistors in digital electronics are the one show in figure below



1.2 Working regimes

We are intrested in the use of MOS transistors as swithces and not as amplifiers.
In our analysis we will consider the following operating regions with the corrisponding voltages

1) **CUT OFF**
$V_{GS} - V_T \leq 0$

2) **LINEAR or TRIODE**
$V_{GS} - V_T > 0$
$\left.\begin{array}{l} V_{DS} < V_{GS} - V_T \\ V_{DS} < V_{DSAT} \end{array}\right\} V_{DS}$ is the lowest

3) **PINCH-OFF SATURATION**
$V_{GS} - V_T > 0$
$\left.\begin{array}{l} V_{GS} - V_T < V_{DS} \\ V_{GS} - V_T < V_{DSAT} \end{array}\right\} V_{OV}$ is the lowest

4) **VELOCITY SATURATION**
$V_{GS} - V_T > 0$
$\left.\begin{array}{l} V_{DSAT} < V_{DS} \\ V_{DSAT} < V_{GS} - V_T \end{array}\right\} V_{DSAT}$ is the lowest

and to assest the right expression of the current we will use the following unified model

$$V_{GS} < V_T \Rightarrow I_{DS} \cong 0$$

otherwise:

$$I_{DS} = \mu C'_{OX} \left(\frac{W}{L}\right) V_{\min} \left(V_{GS} - V_T - \frac{V_{\min}}{2}\right)(1 + \lambda V_{DS})$$

with:

$$V_{\min} = \min(V_{GS} - V_T, V_{DS}, V_{DSAT})$$

In the reference technology of this course (bulk CMOS $0.25\mu m$) this are the characteristics parameters

	V_{T0} (V)	γ ($V^{0.5}$)	V_{DSAT} (V)	k' (A/V^2)	λ (V^{-1})
NMOS	0.43	0.4	0.63	115×10^{-6}	0.06
PMOS	+0.4	+0.4	+1	$+30 \times 10^{-6}$	+0.1

1.2.1 Body effect

Since the V_t depends on the source-body potential if this value is different form 0 we get

$$V_T \cong V_{T0} + \gamma\left(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|}\right)$$

where γ is the body effect coefficient.

1.2.2 Subthreshold regime

When the overdrive voltage is equal to 0 the current in the device is not exactly 0 but the device enters in the so called subthreshold regime where the current behaves like in an BJT junction

$$I_{DS} \simeq I_S e^{\left(\frac{V_{GS} - V_T}{nU_T}\right)} \left[1 - e^{\left(\frac{-V_{DS}}{U_T}\right)}\right]$$

this current is crucial in the power consumption in off mode.

1.3 Equivalent resistance

We can define an equivalent resistance of the MOS transistor as

$$R_{EQ} = \frac{1}{2}\left[\frac{V_{DD}}{I_{DSAT}(1 + \lambda V_{DD})} + \frac{V_{DD}/2}{I_{DSAT}\left(1 + \lambda \frac{V_{DD}}{2}\right)}\right] = \frac{3}{4}\frac{V_{DD}}{I_{DSAT}}\left(1 - \frac{5}{6}\lambda V_{DD}\right)$$

where

$$I_{DSAT} = \mu C'_{ox}\left(\frac{W}{L}\right) V_{DSAT}\left[V_{DD} - V_T - \frac{V_{DSAT}}{2}\right].$$

the last term in parentesis $\left(1 - \frac{5}{6}\lambda V_{DD}\right)$ is almost negligible since is $\simeq 1$.
The expression for the current is the one of velocity saturation beacuse the nmos stays in velocity saturation during the transient since we always consider $V_{DD} = 2.5 > 0.63 * 2$. If we change the supply voltage the current expression has to be changed consistently.
We can make three considerations:

1)the resistance is inversely proportional to the (W/L) ratio of the device

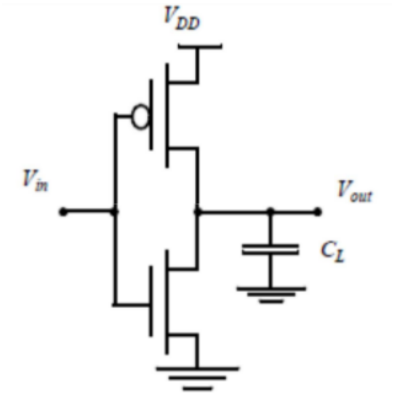
2)for $V_{DD} >> V_T + V_{DSAT}/2$, the resistance becomes virtually independent of the supply voltage

3)once the supply voltage approaches $V_{TE} = 0.745V = V_T + V_{DSAT}/2$, a dramatic increase in resistance can be observed,even if the model adopted (considering the velocity saturation) is no longer valid

In digital electronics once we have the equivalent resistance of the MOS, we can evaluate the propagation time on the basis of the RC model, as $\ln(2)RC$.

Chapter 2

The CMOS inverter



We will refer to CMOS inverter implemented in a static FC-CMOS logic. FC-COMS means fully-complementary logic, which is a particular class of static gates. Static refers to the fact that these gates have an output node always connected to GND or VDD through a low impedance path. Fully-complementary means that the gate is composed by a pull-down network and a complementary pull-up network.

2.1 Static behavior

Independently of the transistor size the high and low output levels are equal to V_{DD} and GND that in out reference technology are 2.5V and 0V

$$V_{OH} = V_{DD} = 2.5V \tag{2.1}$$

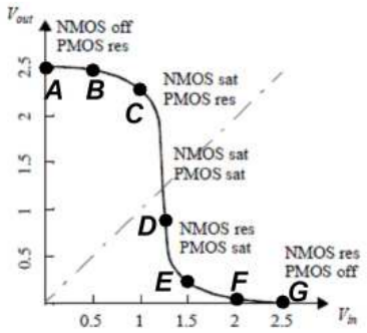
$$V_{OL} = GND = 0V \tag{2.2}$$

This two values are indipentent form the relative device size; gates with this propriety are called ratioless (gates without this propriety are ratioead).

There is always in steady state a finite resistance between the output node and GDN or VDD (that isn't the equivalent resitance of the previous chapter that is useful to assest the propagation delay).

For VDD=2.5V we get

$$r_{on} = \frac{4.2k}{(W/L)_n} \qquad r_{on} = \frac{15.9k}{(W/L)_p} \tag{2.3}$$



2.1.1 Switching threshold

Let's suppose that during the transistion both transistor are in velocity saturation region (this is not correct but leads to a negligible error) to assest the threshold voltage V_M of the inverter we have to put the 2 currents of the mos equal doing this we obtain

$$V_M = \frac{\left(V_{Tn} + \frac{V_{DSATn}}{2}\right) + r\left(V_{DD} - V_{Tp} - \frac{V_{DSATp}}{2}\right)}{1+r}$$

with

$$r = \frac{k_p'\left(\frac{W}{L}\right)_p V_{DSATp}}{k_n'\left(\frac{W}{L}\right)_n V_{DSATn}},$$

We can oslo have the following inverse relationship

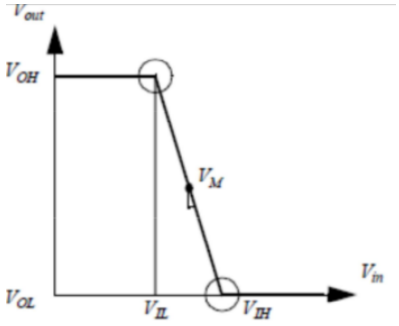
$$r = \frac{V_M - V_{TEn}}{V_{DD} - V_M - V_{TEP}}$$

For $V_M = 1.25V$ that is the best case we get

$$\frac{(W/L)_p}{(W/L)_n} = 3.5 \simeq 3 \tag{2.4}$$

2.1.2 Noise marigns

To compute V_{IL}, V_{IH} we consider the piecewise linear approximation for the VTC, where the transition region is approximated by a straight line having the same negative slope of original VTC in the threshold point.



In this way the straight line that approximates the VTC near the threshold is

$$V_{OUT} = V_{IN}g + (1-g)V_M \qquad (g<0) \tag{2.5}$$

we can compute g considering that $g = -(g_{mn} + g_{mp})r_{0p}//r_{0p}$ we obtain

$$g \cong \frac{k_n'\left(\frac{W}{L}\right)_n V_{DSATn} + k_p'\left(\frac{W}{L}\right)_p V_{DSATp}}{\lambda_n I_{DSATn}(V_M) + \lambda_p I_{DSATp}(V_M)}$$

The gain is apporoximately constant if the velocity saturation is taken into account.
From this we get

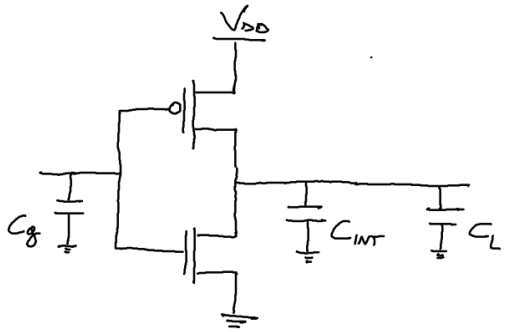
$$V_{IH} = V_M + \frac{V_{OL} - V_M}{g}$$

$$V_{IL} = V_M + \frac{V_{OH} - V_M}{g}$$

$$NMH = V_{OH} - V_{IH}$$

$$NML = V_{IL} - V_{OL} =$$

2.2 Dinamic behavior



We define the following parameters

W_n, W_p the width of the drain junctions (that is the same of W/L)
L_D Drain length (that isn't the same as W/L)
$A_d = W * L_D$ the area of the drain
$P_d = 2L_D + W$ the perimeter of the drain

2.2.1 Intinsic capacitance C_{int}

$$C_{int} = C_{overlap} + C_{dbn} + C_{dbSWn} + C_{dbp}C_{dbSWp} \tag{2.6}$$

Overlap capacitance

$$C_{ovelap} = 2(C'_{ovp}W_p + C'_{ovn}W_n) \tag{2.7}$$

Drain-bulk n capacitances

$$C_{dbn} = \frac{C'_{jn}A_{dn}}{2}\left(\frac{1}{(1+\frac{V_{DD}}{\phi})^{0.5}} + \frac{1}{(1+\frac{V_{DD}}{2\phi})^{0.5}}\right) \tag{2.8}$$

$$C_{dbSWn} = \frac{C'_{jSWn}P_{dn}}{2}\left(\frac{1}{(1+\frac{V_{DD}}{\phi})^{0.44}} + \frac{1}{(1+\frac{V_{DD}}{2\phi})^{0.44}}\right) \tag{2.9}$$

Drain-bulk p capacitances

$$C_{dbp} = \frac{C'_{jp}A_{dp}}{2}\left(1 + \frac{1}{(1+\frac{V_{DD}}{2\phi})^{0.48}}\right) \tag{2.10}$$

$$C_{dbp} = \frac{C'_{jSWp}P_{dp}}{2}\left(1 + \frac{1}{(1+\frac{V_{DD}}{2\phi})^{0.32}}\right) \tag{2.11}$$

2.2.2 Gate capacitance C_g

$$C_g = C'_{ovn}W_n + C'_{ovp}W_p + C'_{ox}(W_nL_n + W_pL_p) \tag{2.12}$$

2.2.3 Self-loading factor

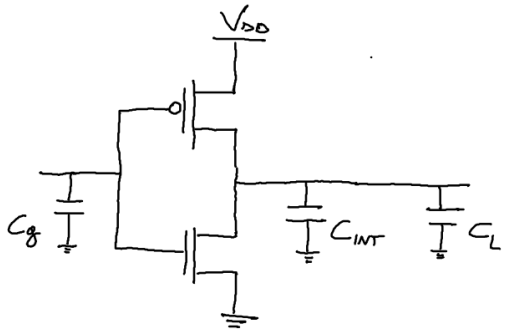
We can define the self-loading factor γ as

$$C_{int} = \gamma C_g \tag{2.13}$$

In our reference technology we have $\gamma = 1$ and $C_g^{(1)} = 2fF$ where the suffix (1) is the size of the inverter (that is the W/L of the n-mos).
If we have an inverter of size α it's gate capacitance will be

$$C_g^{(\alpha)} = \alpha C_g^{(1)} = \alpha \cdot 2fF \tag{2.14}$$

2.3 Propagation delay



Propagation delay form high to low

$$\tau_{pHL} = \ln(2) \cdot R_N \cdot (C_{int} + C_{ext}) = 0.69 \cdot \frac{13k\Omega}{\left(\frac{W}{L}\right)_n} \cdot (C_{int} + C_{ext})$$

Propagation delay from low to high

$$\tau_{pLH} = \ln(2) \cdot R_P \cdot (C_{int} + C_{ext}) = 0.69 \cdot \frac{31k\Omega}{\left(\frac{W}{L}\right)_p} \cdot (C_{int} + C_{ext})$$

We define intrinsic propagation delay as the average between this 2 results when $C_L = C_{ext} = 0$

$$\tau_{p0} = \frac{\tau_{HL} + \tau_{LH}}{2} \tag{2.15}$$

2.3.1 Fan-out

Considering an external capacitance we get the folowing expression

$$\tau_p = \ln(2) R_{eq}(C_{int} + C_{ext}) = \tau_{p0}\left(1 + \frac{C_{ext}}{C_{int}}\right) = \tau_{p0}\left(1 + \frac{C_{ext}}{\gamma C_g}\right) = \tau_{p0}\left(1 + \frac{f}{\gamma}\right) \tag{2.16}$$

Where we used the fan-out f and the self-loading factor defined as

$$f = \frac{C_{ext}}{C_g} \qquad C_{int} = \gamma C_g \tag{2.17}$$

2.4 Chain of inverters

To optimize the propagation delay of an inverter chain with N elements we have to set all propagation delay of the inverters equal $f_i = f_i + 1 \quad \forall i = 0, ..., N$.
To do this we have to calculate the path fan-out that is the load over the first gate capcitance

$$F = \frac{C_L}{C_{g,1}} \tag{2.18}$$

And from this we calculate the optimum fan-out

$$f_{opt} = \sqrt[N]{F} \tag{2.19}$$

given the first inverter size s_1 all the other inverters' dimensions are fixed

$$s_n = s_1 \cdot (f_{opt})^n \quad \forall n = 1, ..., N \tag{2.20}$$

and the propagation delay is

$$\tau_p = N\tau_{p0}\left(1 + \frac{f_{opt}}{\gamma}\right) \tag{2.21}$$

If the number of inverter isn't fixed we can compute for our reference technology the optimum number of stages as

$$N_{opt} = \frac{\ln(F)}{\ln(3.6)} \tag{2.22}$$

beacuse the best fan-out that we can have is

$$f_{opt}^{ideal} = 3.6 \tag{2.23}$$

Of course the number of stages N has to ben an integer number.

Chapter 3

Inverter power consumption

3.1 Dynamic power consumtion

Only in case of charge of a capacitance so in a LOW to HIGH transition at the output of an inverter.
The most general formula for the power dissipated in this process is

$$P_{dyn} = C_L V_{DD}^2 f_{1\to 0} \tag{3.1}$$

where $f_{1\to 0}$ is the frequency of the 0-1 transition at the output node. This term can be deconposed in 2 factor the clock of the circuit and the switching activity α_{SW} that is the probability to have a transition from 0 to 1 at the output.

In case of a square wave we have

$$f_{1\to 0} = f_{clk} \cdot \alpha_{SW} = f_{clk} \cdot \frac{1}{2} \tag{3.2}$$

In case of a randoom signal

$$f_{1\to 0} = f_{clk} \cdot \alpha_{SW} = f_{clk} \cdot \frac{1}{4} \tag{3.3}$$

In the end we can write this 2 final equation for dynamic power dissipation

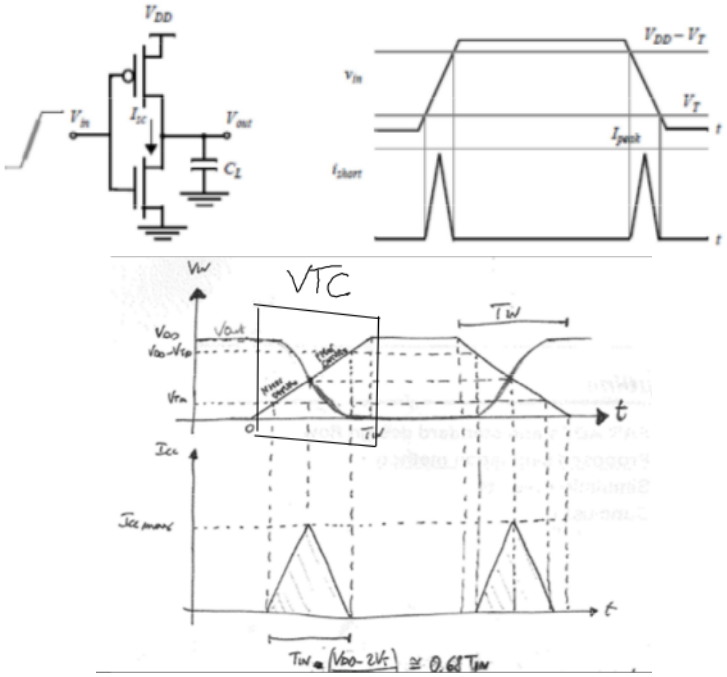
$$P_{dyn} = C_L V_{DD}^2 f_{clk} \cdot \alpha_{SW} \tag{3.4}$$

$$E_{dyn} = C_L V_{DD}^2 \alpha_{SW} \tag{3.5}$$

In case of a chain of inverters C_L is the sum of all the capacitance at the output nodes of the single inverters.

3.2 Cross-conduction power consumption

Due to the finite slope of the input signal there is a finite time when both p-mos and -mos are on creating a low impedance path between V_{DD} and GND.



Defining T_{in} as the time needed by the input signal to transition from GND to V_{DD} and vice-versa and I_{peak} the velocity saturation current of the n-mos (when the input and the output are at V_m) we can compute the cross-conduction energy as (assuming $V_{tp} = V_{tn}$)

$$E_{cc} = \frac{1}{2} V_{DD} I_{peak} \cdot 0.68 \cdot T_{in} \tag{3.6}$$

If we consider an input signal with an update frequency of f with continuos switch HL LH the power is

$$P_{cc} = \frac{V_{DD} I_{peak} \cdot 0.68 \cdot T_{in}}{2} f \tag{3.7}$$

With capacitve loads the cross-conduction power becomes smaller than without.
The cross conduction power can be neglected when

$$T_{in} < 10\tau_p \tag{3.8}$$

with τ_p propagation delay of the stage.

3.3 Static power consumption

It's due to the subthreshold current of the MOS devices.

Chapter 4

The wires

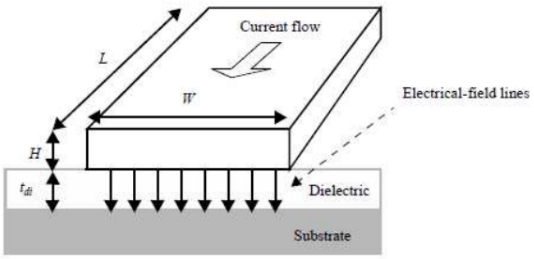
To simplify the analysis of the wires parasitics effects we introduce 3 simple assumption;

→ Inductance can be neglected if the wire resistance is large or if the rise/fall time of the input signal is large.

→ When the wire is short and when the equivalent resistance of the driver is large, the wire resistance can be serenely neglected.

→ When the separation between nearby wires is large or when the wires run in parallel for a short distance, the inter-wires capacitance can be neglected.

4.1 Capacitance



We divide the overall capacitance of a wire in 2 main contribution; parallel plate capacitance and fringing capacitance.
In the ideal case where the parameter W/t_{di} is very large the fringing capacitance contribution can be neglected but since in our reference technology we can have $W/H > 1$ the fringing or border capacitance is the dominant contribution.

For our reference technology process the following table is given reporting the parallel-plate and the fringing capacitance contributions for a wire in a certain layer with respect to another wire in another layer.
The parallel-plate capacitance is reported in the white rows expressed in $aF/\mu m^2$ of overlapping area, while the shaded rows report the fringing capacitance contribution in $aF/\mu m$ of perimeter (that is quite always $\simeq 2L$).

	Field	Active	Poly	Al1	Al2	Al3	Al4
Poly	88						
	54						
Al1	30	41	57				
	40	47	54				
Al2	13	15	17	36			
	25	27	29	45			
Al3	8.9	9.4	10	15	41		
	18	19	20	27	49		
Al4	6.5	6.8	7	8.9	15	35	
	14	15	15	18	27	45	
Al5	5.2	5.4	5.4	6.6	9.1	14	38
	12	12	12	14	19	27	52

When using this table we have always to look what tipe of metal is our wire (rows) and in what material is made the ground plane or the other wire (columns) in order to get the correct values.

To evaluate che capacitance of 2 nearby wires implemented in the same layer at minimum distance we get the following table with the vales of the capacitances for unit lenght expressed in $aF/\mu m$

Layer	Poly	Al1	Al2	Al3	Al4	Al5
Capacitance	40	95	85	85	85	115

4.2 Resistance

Resistance of a wire can be assested as

$$R = \rho \frac{L}{WH} = R_{sh} \frac{L}{W} \tag{4.1}$$

where the last part represent the resistance per square multiplied by the number of squares.

At very high frequency, the resistance tends to increase due to the skin effect. In practice the current tends to flow in the peripheral part of the wire. Considering a wire with a width W and a height H, the current flows almost entirely in a peripheral section characterized by a depth

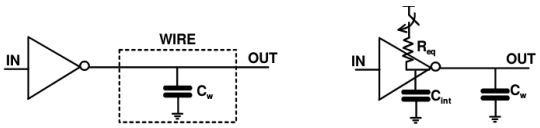
$$\sigma = \sqrt{\frac{\rho}{\pi f \mu}} \tag{4.2}$$

if the wire is smaller than the skin effect at a given frequency there are no difference in the resistance. It's an effect that affects only wide wires.

4.3 Models for wires

4.3.1 Lumped C model

Since the resistance of the wire is much smaller wrt the driving resitance we can model the wire with only it's capacitance.

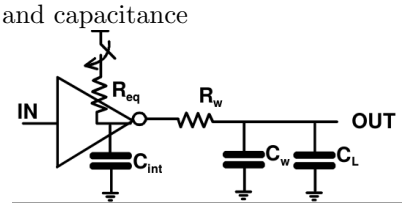


The wire capacitance has to be added to the intrinsic capacitance of the inverter in order to correctly estimate the propagation delay that in this case is

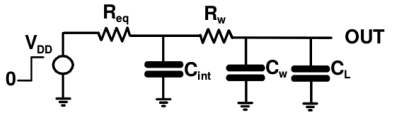
$$\tau = \ln(2) R_{eq}(C_{int} + C_w) \tag{4.3}$$

4.3.2 Lumped RC model

When the resistance is no more negligible a first odrer approximation is to model the wire as it's single equivalent resistace and capacitance

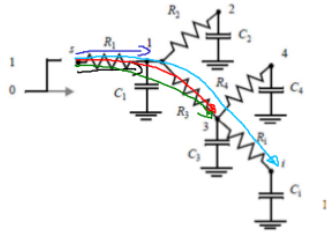


To estimate the overall propagation we can model the circuit as follows and use the Elmore theorem



Elmore theorem

With a network that has a single input node ,all capacitance between a node and ground and no resistive loops we can assest the propagation delay of the line as calculating for every capacitance C of the network the so called shared-path resistance. This resistance represents the resistance shared between the path from the source of the signal and the output and the path form the source to the capacitance.

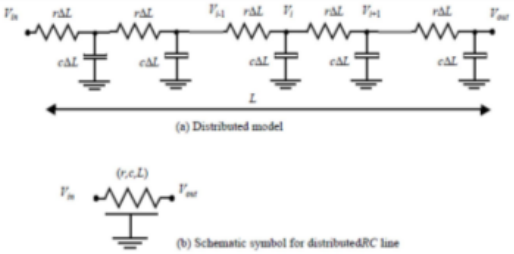


$$\tau = C_1 R_1 + C_2 R_1 + C_3(R_1 + R_3) + C_4(R_1 + R_3) + C_j(R_1 + R_2 + R_j) \tag{4.4}$$

This is an approximation that brings us to an overestimation of the propagation delay (factor 2) beacuse we are supposing all parasitic terms concentrated in one point and not distributed over a line.

4.3.3 Distributed model

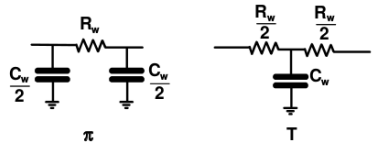
With the distributed model we divide the wire into lot of wires of length ΔL with $\Delta L \to 0$.



In this way through the resoluton of a differential equation for the voltage over the line we get that the delay of the wire is

$$\tau_w \simeq \ln(2)\frac{R_w C_w}{2} \tag{4.5}$$

That is the correct value wrt the lumped RC model.
To correctly take into account the effects of the distribution we will use the following 2 models that give us both the same result of the distributed model



4.4 Buffering of a line

If the time constant of a line it's the dominant contribution for delay in our digital circuit it's worth in order to get a faster respnce to cut the wire in N pieces adding buffer in the middle. This is convenient if

$$L \geq \sqrt{\frac{16 C_{int} R_{eq}}{cr}} \tag{4.6}$$

where C_{int}, R_{eq} are parameters of the driving circuit and c,r are the specific resistance (Ω/m) and capacitance (F/m) of the wire. The optimum number of division N (and so the number of inverter to be added N-1) is

$$N = \sqrt{\frac{R_w C_w}{4 C_{int} R_{eq}}} = L\sqrt{\frac{rc}{4 C_{int} R_{eq}}} \tag{4.7}$$

And the contribution to τ_p of the wire $(R_w C_w/2)$ decreases by a factor N.
The propagation delay of an optimized line that is broken in N peices is

$$\tau_p = \ln(2) N \left(R_{eq} C_{int} + (R_{eq} + \frac{R_w}{2N})\frac{C_w}{N} + (R_{eq} + \frac{R_w}{N}) C_{int} \right) \tag{4.8}$$

In this expression we supposed inverters with the same size and $\gamma = 1$.
This optimization of the line it's indipendent form the size of the stage present in the circuit.

4.5 Inductance

Parasitic inductances can be neglected if the time of flight of the signal is much less than the minimum propagation delay of the circuit. With a line of length L and the maximum propagation delay of τ_{max} this translates as

$$t_{flight} = \frac{L}{\frac{c}{\sqrt{\mu\varepsilon}}} \quad << \quad \tau_{max} \tag{4.9}$$

where tipically $\varepsilon = 3.9$.

Another condition that makes the inductance negligible is to have

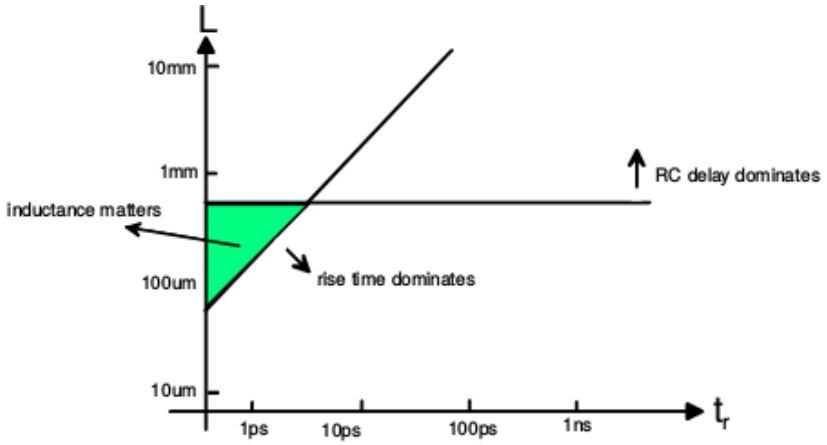
$$L \quad << \quad \frac{1}{r}\sqrt{\frac{l}{c}} \tag{4.10}$$

where r, l and c are the specific resistance inductance and capacitance.

This translates in the followin inequality

$$\frac{t_r}{\sqrt{lc}} << L << \frac{1}{r}\sqrt{\frac{l}{c}} \tag{4.11}$$

where t_r is the rise time of the input signal. If L satisfy this conditions we can neglect the parasitic inductance as rapresented in the graph below.

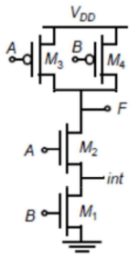


Chapter 5

FC-CMOS gates

5.1 Sizing of single transistor inside a logic gate

Once the transistor are placed in order to compute the requested logic function the sizing of the single transistor has to follow this simple rule: the equivalent resistance in the worst case state has to be equal to the one of an inverter of the same size of the gate.
Transistor in parallel show a resistance half of a single transistor so considering the following NAND gate with a size equal to 1 the p-mos will have an aspect ratio of 3 and the n-mos of 2 (two transistor of size 2 in series are the equivalent of a transistor of size 1).



If the searched size was s=2 we will have p-mos 6 and n-mos 4 of aspect ratio.

5.2 Equivalent capacitance

To estimate the equivalent internal capacitance C_{int} of a generic FC-CMOS gate we have to look to the dimesnions of the transistors connected to the output node. The sum of the aspect ratios of this transistors divided by 4 it's how many $C_g^{(1)}$ the internal capacitance is.

To better assest the propagation delay we should also take into account the inter-nodes capacitance inside our gate; this leads us to a more complex analysis since we have to consider whitch parasitic capacitance are charged an whitch are discharged and we shoulde use the Elmore theorem for every transistion.

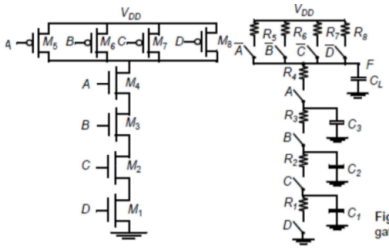
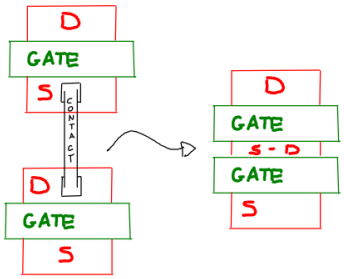


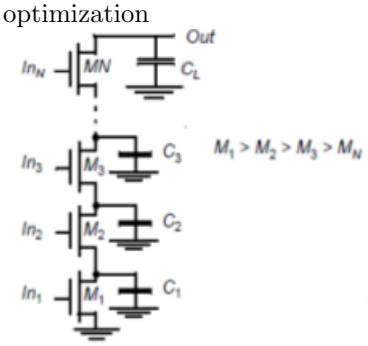
Figure 6.11 Four input NAND gate and its RC model.

This inter-nodes capacitance can't be estimated as before with a value of 2fF but , since 2 transistor in series are integrated with common drain-source, with 1fF capacitance.



We can neglect the effect of the inter-nodes capacitance since our gates have small fan-in (3-4 max).

In case of large fan-in and a lot of transistors in series it's possible to adopt clever sizing "inspired" by the inverter chain optimization



Another good rule of thumb is that the last signal coming to the gate (the slowest signal at the input of the gate) has to be connected to the transistor closest to the output.

5.3 Propagation delay

5.3.1 Equivalent resistance

We define the equivalent resistance of a generic gate of size s (built with the right internal proportion described before) as

$$R_{eq} = \frac{R_{eq}^{(1)}}{s} = \frac{11.6k\Omega}{s} \tag{5.1}$$

5.3.2 "p" factor

The factor "p" or intrinsic delay factor (OUTPUT) it's a constant that connect the intrinsic capacitance of our gate with the intrinsic capacitance of an inverter of the same dimensions and can be calculated as

$$p = \frac{C_{int}(s=1)}{C_{int}^{(1)}} \tag{5.2}$$

$$p = \frac{\sum \frac{W}{L}|_{mos\ connected\ to\ y}}{4 * s} \tag{5.3}$$

and it's releted with the intrinsic capacitance of a generic size gate as

$$C_{int} = spC_{int}^{(1)} \tag{5.4}$$

This parameter it's indipendent on the size of the gate

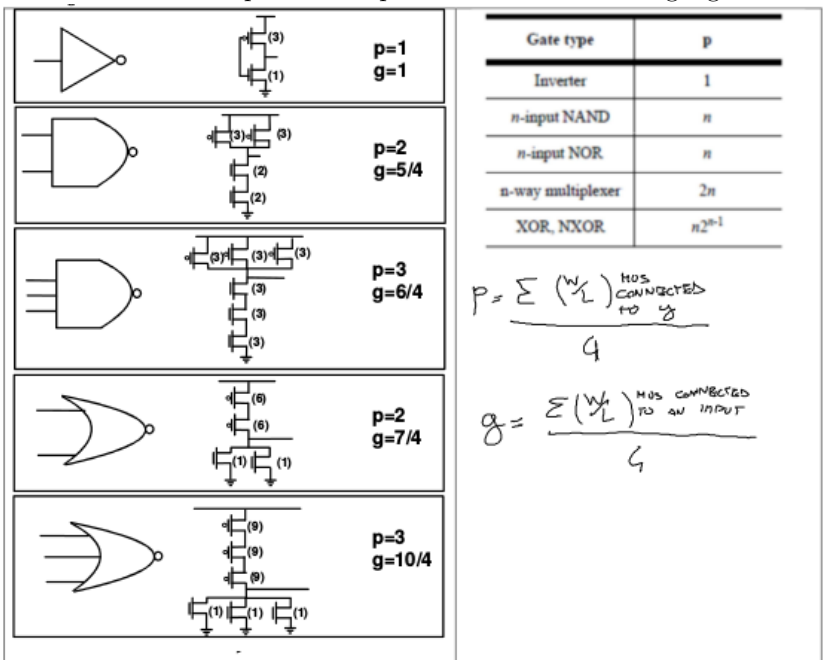
5.3.3 "g" factor

The factor "g" is the logical effort (INPUT) that quantifies the complexity of the gate with respect to the inverter in the "input" direction.
Also this parameter is indipendent on size and has to be calculated considering single couple of inputs

$$g = p = \frac{C_g(s=1)}{C_g^{(1)}} \tag{5.5}$$

$$g = \frac{\sum \frac{W}{L}|_{mos\ connected\ to\ A}}{4 * s} \tag{5.6}$$

Here some examples of this parameters on famous logic gates



Gate type	p
Inverter	1
n-input NAND	n
n-input NOR	n
n-way multiplexer	$2n$
XOR, NXOR	$n2^{n-1}$

$p = \frac{\sum \left(\frac{W}{L}\right)_{\text{MOS CONNECTED TO } y}}{q}$

$g = \frac{\sum \left(\frac{W}{L}\right)_{\text{MOS CONNECTED TO AN INPUT}}}{6}$

5.3.4 Propagation delay

The expression of the propagation delay of a generic gate is

$$\tau_p = \tau_{p0}\left(p + \frac{fg}{\gamma}\right) \tag{5.7}$$

where we can distinguish two contributions; the term τ_{p0} is the intrinsic delay and $fg\tau_{p0}$ is the effort delay strictly related to the load capacitance. The product fg it's called stage effort or h.

5.4 Size optimization of τ_p

The delay that counts and that we are going to optimize it the one related with the critical path, the delay of the longest path.
To optimize the propagation delay we need that all the gates have the same stage effort h.

We define the path fan-out as

$$F = \frac{C_L}{C_{g,1}} \tag{5.8}$$

and the path logical effort as

$$G = \prod g_i \tag{5.9}$$

To optimize the propagation delay we all the stages has to have the following h_{opt}

$$h_{opt} = \sqrt[N]{H} \tag{5.10}$$

This corresponds to a propagation delay of the overall chain of

$$\tau_{p,opt} = \tau_{p0}\left((\sum p) + \frac{Nh_{opt}}{\gamma}\right) \tag{5.11}$$

in this way the dimension of the j-th stage will be

$$s_j = \frac{s_1 g_1}{g_j}\prod_{i=1}^{j-1} f_i \tag{5.12}$$

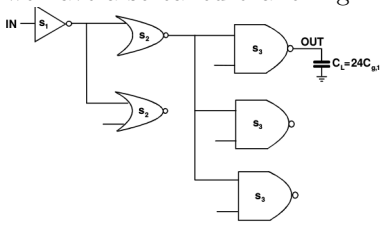
or better we can assest every dimension form the first stage to the last one using the formula of h_{opt}

$$h_{opt} = f_i g_i \quad \forall i \tag{5.13}$$

using this formula on the i-th stage we will end up with the size of the i+1 stage.

5.4.1 Branching

In the case illustrated in figure we have a so called branching in the path.



There are more stages in parallel (NAND in this case). Under the assumption that all the parallel elements have the same dimensions (all NAND with equal size) we can define the branching factor refering to the extrinsic capacitance as b

$$b = \frac{C_{path} + C_{off-path}}{C_{path}} \tag{5.14}$$

that in practice is the number of stages in parallel (3 in our case).

In this cases we can optimize the chain introducing the path branching factor B

$$B = \prod_{i=1}^{N} b \tag{5.15}$$

than the way to optimize the chain is the same as before but the H is now defined as

$$H = FGB \tag{5.16}$$

the number of stages n does not takes into account the braching stages and so to find the size of the single stages we can use the relation

$$h_{opt} = \sqrt[n]{H} = f_i g_i b_i \tag{5.17}$$

5.5 Power dissipation

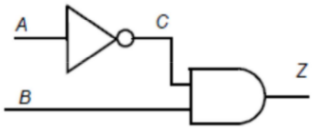
The consideration on power dissipation are the same for the simple NOT circuits we've done in the previous chapter. The difficult topic here is to assest the parameter α_{sw} and to reason if some states are correleted each other. Assuming that the inputs are independent and uniformly

distributed, any N-input static gate has a transition probability that corresponds to

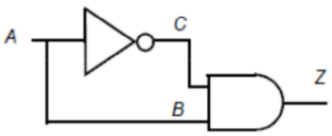
$$\alpha_{sw} = \frac{N_0 N_1}{2^{2N}} = \frac{N_0(2^N - N_0)}{2^{2N}} \tag{5.18}$$

where N_0 and N_1 are the number of zero and one, respectively, in the truth table of the logic function. Here follows an example where the above formula doesn't work since the 2 inputs are

correlated



(a) Logic circuit without reconvergent fanout



(b) Logic circuit with reconvergent fanout

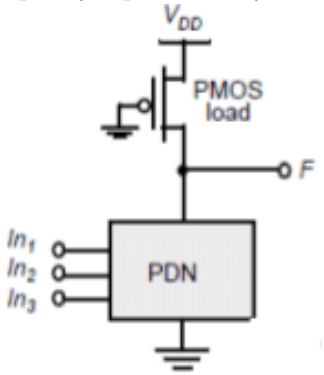
Chapter 6

Ratioed logic

Ratiohead logic means that the characteristics of the circuit (statics and dynamics) depends on the ratio between the pull-up and the pull-down network.
This type of circuits the advantage of being less area consuming and with small output capacitance.

6.1 Pseudo-NMOS

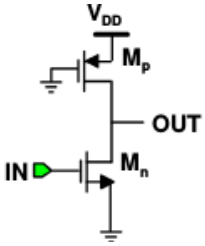
The basic idea behind this logic is to use the same pull down network of an FC-CMOS implementation but using as pull up only a p-mos always on (with the gate grounded)



Obviously to pull down the output node the pull down network have to be stronger than the p-mos.
We drastically reduced the number of transistors form 2N+1 of the FC-CMOS logic to N+1 and doing so also the input and output capacitance.

6.1.1 Static characteristics

Let's take as example the pseudo-NMOS inverter



V_{OH} remains at VDD but V_{OL} isn't at 0V since both transistors are on when the pull-down network is enabled.
Therefore we can estimate that V_{OL} it's close to ground so we can found it's value supposing the pmos in velocity saturation and the nmos in ohmic region and comparing the currents when the input is at VDD and out at V_{OL}

$$k'_n \left(\frac{W}{L}\right)_n V_{OL}(V_{DD} - V_{tn} - \frac{V_{OL}}{2}) = k'_n \frac{W}{L}|_p V_{satp}(V_{DD} - V_{tp} - \frac{V_{satp}}{2}) \tag{6.1}$$

V_{OL} stricktly depends on the relative ratio between nmos and pmos. Weaker pmos smaller the output voltage low.

Regarding the switching threshold we can suppose it's near the middle and write the ballance of currents with both transistors in velocity saturation where we find the term V_M only in the nmos current

$$k'_n \left(\frac{W}{L}\right)_n V_{satn}(V_M - V_{tn} - \frac{V_{satn}}{2}) = k'_n \left(\frac{W}{L}\right)_p V_{satp}(V_{DD} - V_{tp} - \frac{V_{satp}}{2}) \tag{6.2}$$

then after getting the first value we can verify if our supposion was right and iterate.

Main drawback of this technology is the static power consumption that we have in correspondance of a logic 0.
This power can be evaluated as

$$P_{static} = P(0)V_{DD} \cdot k'_n \left(\frac{W}{L}\right)_p V_{satp}(V_{DD} - V_{tp} - \frac{V_{satp}}{2})(1 + \gamma_p(V_{DD} - V_{OL})) \tag{6.3}$$

6.1.2 Dynamic characteristics

The intrinsic propagation delay can be evaluated in the same way as the FC-CMOS logic restoring the idea of equivalent resistance.
In the pull-up time we have to consider that it's an overestimation since we don't start from 0V but from V_{OL}.
In the pull-down time we're doing an underestimation since the pmos is on and try to charge the capacitance. In order to take into account this effect we can consider the pull-down resistance as

$$R_{pull-down} = \frac{R_{eq,n}}{1 - \frac{R_{eq,n}}{R_{eq,p}}} \tag{6.4}$$

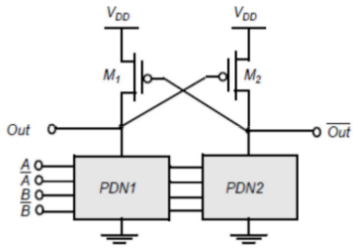
Hidden in this relation there is the fact that if the pmos is too strong we aren't able to pull down the y node (for $R_n < R_p$, R_{pd} becomes negative).

There is a trade off between static and dinamic proprieties as the following scheme shows

$(W/L)_p$	V_{OL} (V)	V_M (V)	$P_{low}(\mu W)$	τ_{pLH}	τ_{pHL}
0.5	0.128	1.1	74	39.6ps	8.5ps
1	0.26	1.41	147	17.6ps	11.9ps
1.5	0.395	1.63	218	12.75ps	18.7ps
2	0.537	1.77	287	9.6ps	32.7ps

6.2 DCVSL

This technology is an improvement of the pseudo-NMOS logic that eliminates the problem of static power consumption and makes the $V_{OL} = 0$ using a positive feedback and a differential structure.
The basic structure is shown in figure



The two pull-down network are complementary and we have a differential output.
At steady state there is no conductive path between the voltage supply and ground and the low output voltage is restored due to the feedback loop.
Although the logic is still ratioed since if the pull down network is not stronger than the PMOS transistor, the corresponding output node cannot be driven low.

This logics has still some problems since we have cross-conduction current the switching activity is increased ($\alpha_{sw} \simeq 1$) and the cross-connection between the two pmos as the differential naure can be troublesome.

Chapter 7

Other static logics

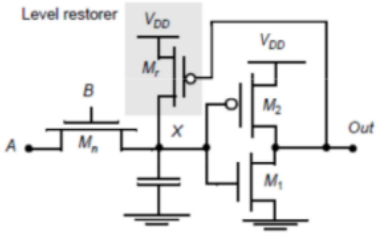
7.1 Pass-transistor

This technology aim is to reduce the number of transistor with the idea to drive the mos by the gate and olso by the drain/source trerminal. It's a technology suited to implement adders and multipilers.
The main idea is to wathc the truth table of our function and think how to implement it as a multiplexer like in figure

The signal B is used as a selector.
This type of implementation has a lot of drawback since the nmos transistors are not good at charging capacitance they are affected by body effect so on the output node we have $V_{DD} - V_{tn}*$ that creates crossconduction current if it's connected to an inverter and it's the reason why it's forbidden to cascade pass-transistor.

7.1.1 Pass-transistor with level restorer

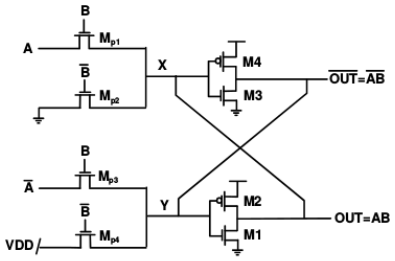
First try to fix this issues is to use a level restorer in order to restore the voltage only to ground or power supply without consuming static power consumption using a classical inverter and a pmos



The logic is still ratioed since the ratio of dimensions between the pass-transistor and the pmos it's foundamental since the pass-transistor have to be able to drive the X node low.

7.1.2 Swing restored pass transistor logic

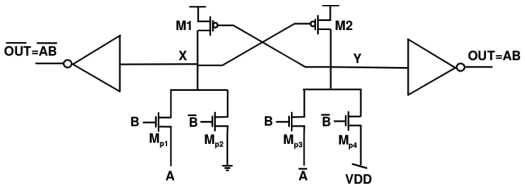
If we want a differential implementation of the pass-transistor logic we can employ two inverters in a crosscoupled fashon instead of the voltage restores and two complementary pass-transistor network.



The problem is again that this is still a ratioed implementation and a right sizing is mandatory to let this circuit work fine.

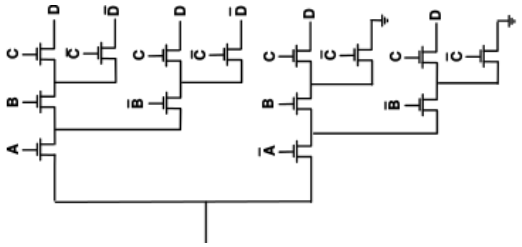
7.2 Complementary pass-transistor logic CPL

The final solution to the problems of the pass-transistor logic is the complementary pass transistor logic where we have two level restorers with gate connected to the opposite pass-transistor block and and two inverters that operates as buffer to permit the cascade of other logic ports.

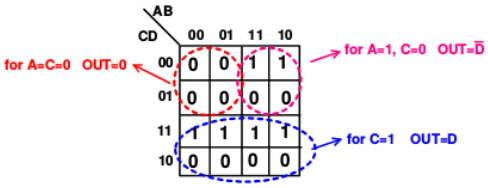


This logic is ratioless since the feedback loop heps the transition and the on-off state of the two pmos transistors.

To build the pull down networks we have 2 different method: mux way or intuituve way.
In the multiplexer way we decide witch signals activate the mos and the signal that will be used as output than we can do some semplification limiting ourself to a tree structure like in figure



The second way is to find on the k-map some groups of bits that corresponds to some condition like in figure



7.2.1 Propagation delay

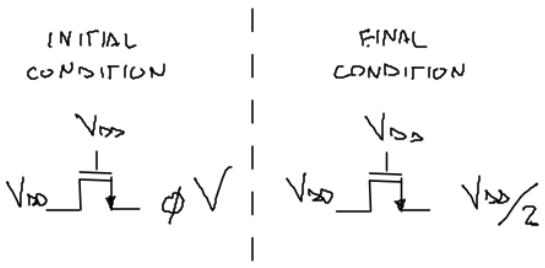
Low to high τ_{lh}
A low to high transition on the output Y corresponds to a high to low transition before the inverter so the nmos are discharging the intrinsic capacitance C_{int}.
The time taken by the system to discharge this capacitance is

$$\tau' = \ln(2) R_n C_{int} \tag{7.1}$$

where $R_n = 13k$ as usual.
So the oveall time for a low to high transition is τ' plus the low to high propagation delay of the inverter that is 12ps

$$\tau_{lh} = \ln(2) R_n C_{int} + \tau_{plh}^{inv} = \ln(2) 13k C_{int} + 12ps \tag{7.2}$$

High to low τ_{hl}
A high to low transition on the output Y corresponds to a low to high transition before the inverter so the nmos are charging the intrinsic capacitance C_{int}.
Thus we can't use the usual value of the resistance but we have to calculate the value considering an initial with source at 0V an final condition of the mos with source at $V_{DD}/2$.



Considering the body effect in the initial condition we are in velocity saturation but at the end in pinch off saturatio ($V_{tn}^* \simeq 0.7 - 0.8$) so we can compute the equivalent resistance as

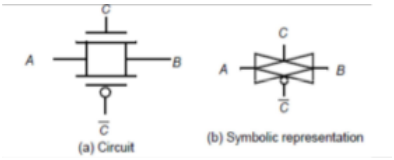
$$R_{eq,n} = \frac{1}{2}\left(\frac{V_{DD}}{I_{vsat}^{(i)}} + \frac{V_{DD}/2}{I_{po}^{(f)}}\right) \tag{7.3}$$

that depends on the equivalent reistance dimensions and it's the dominant term in the overall delay.
The high to low equivalent delay will be

$$\tau_{hl} = \ln(2) R_{eq,n} C_{int} + \tau_{phl}^{inv} = \ln(2) R_{eq,n} C_{int} + 18ps \tag{7.4}$$

7.3 Trasmission gate

Another solution to avoid the voltage drop problem is the use of trasmission gates that are one nmos and one pmos in parallel that are driven by complementary signals.
This device acts as a switch since turns on or off both devices. The advantage is that in this way we can always discharge/charge totally the load without stopping at a threshold voltage of distance.



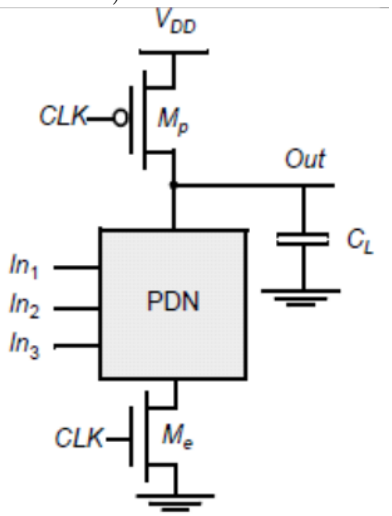
Chapter 8

Dynamic logic

Until now we have studied static logics in whitch the output node it's always connected to gnd or V_{DD} through a low impedance path.
Dynamic logics feature high impedance at the output and the information is stored in a capacitance.

8.1 Static and dynamic proprieties

This technology reduce the capacitances (both input and output) by using a single clocked pmos transistor (and a clocked nmos too)



This device has 2 phases: pre-charge and evaluation.
When CLK=0 we are in the pre-charge phase the pull-down network is disabled by the last nmos and the clocked pmos charge the output capacitance.
When CLK=1 we are in the evaluation phase the pull-down network is enabled and if the inputs are high than the output capacitance can be discharged. In order to the gate to work proprely during this phase the input signals can have only low to high transitions or they can remain stable (if they have a h-l commutation the output cannot change to high).

The logic function is implemented by the pull-down network that is the same of an FC-CMOS gate. The overall number of transistor is reduced at N+2 without losing the ratioless propriety. Reducing the number of transistors also the capacitance are reduced therefor we get an increase of speed.

Noise margins cannot be evaluated since we can't draw a characteristics but we can say that NMH$\simeq V_{DD} - V_{tp}$ and NML$\simeq V_{tn}$.

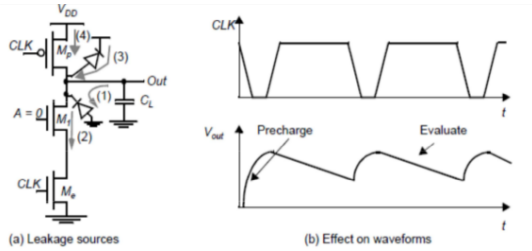
The positive aspects of this technology is that both input and output capacitance are smaller and we avoid any short circuit current with ideal clock signal.
However we have consider that we have another source of power dissipation that is the clock and that the switching activity is larger since p(1)=1 so

$$\alpha_{sw} = p(0) \tag{8.1}$$

8.2 Issues in dynamic logic

There are several issues that have to be take into account to verify that the circuit works proprely

8.2.1 Leakage current

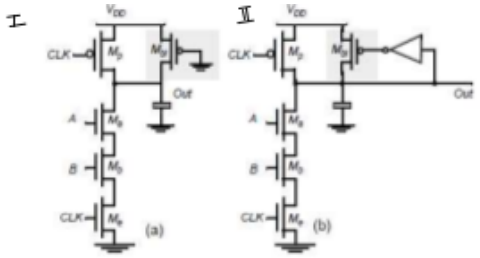


If the clock frequency is very low sketching the output voltage, we can observe that it does not remain at V_{DD} during the evaluation phase but slightly decreases. This effect is due to the leakage current. This is a problem only at low frequency of the clock when the voltage across the capacitance can decrease more than $V_{DD}/2$.

We can use two solutions to avoid this drawback.

I) Adopt a bleeder that keeps the output node at high level during the evaluation phase. But in this way we restore the ratioed problems so the transistor has to be correctly sized in order to be let the pull-down network dominate.

II) Use a pmos transistor in feedback configuration that operates only if the output voltage is high. This is still a ratioed solution since if the output is high and the pull-down network is enabled the pmos has to be enought weak to let the nmos prevail and discharge the capacitance



If we get a leakage current of I^* and no load applied to the output the drop rate of the ouput voltage is

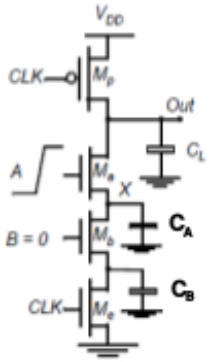
$$\frac{I^*}{C_{int}} = \gamma \tag{8.2}$$

and the minimum frequency at whitch we can go is (considering $V_{DD}/2$ as drop limit if the capacitance)

$$f = \frac{1}{2\frac{V_{DD}/2}{\gamma}} \tag{8.3}$$

8.2.2 Charge sharing

This is a problem of the inter-nodes parasitic capacitances that until now we've negleced.
Let's consider a 2 input NAND



If we are in the pre-charge phase and the input A toggles we have a charge sharing problem that varies the output voltage.
We can separate two cases large variation of the output or small variations.

For large variation of the output voltage ($\Delta V_{out} = V_{DD} - V_{out} \geq V_{tn}$) at the end the source of Ma will be the same as the output voltage at the end so we get that

$$\Delta V_{out} = V_{DD}\frac{C_A}{C_L + C_A} \tag{8.4}$$

For variations smaller than the nmos threshold the voltage at the source grows but cannot reach the output since the mos is shut off earlier.
In this case

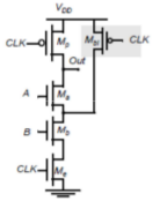
$$\Delta V_{out} = (V_{DD} - V_{tn})\frac{C_A}{C_L} \tag{8.5}$$

The case change with the value of the ratio of the two capacitance; equating the two equation to V_{tn} we get that the limit is (considering also the body effect

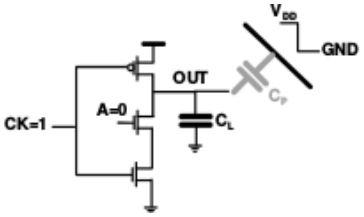
$$\frac{C_A}{C_L} = \frac{V_{tn}}{V_{DD} - V_{tn}} = 0.38 \tag{8.6}$$

This means that C_A has to be smaller than $0.38C_L$.

The solution is to precharge also the internal node as in figure



8.2.3 Capacitive coupling



High impedance node are very sensitive to interference and crosstalk so we have to be carefoul in the design of the overall circuit to do not put wires near the output of a dynamic gate in order to avoid interference and "parasitic" toggles due to a switch of the line. The resulting voltage drop at the output is

$$\Delta V_{out} = \frac{C_p}{C_L + C_p} V_{DD} \tag{8.7}$$

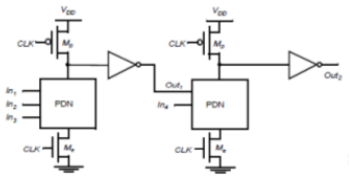
8.2.4 Clock feedthrough

It's a particular type of clock feedthrough that can cause latch-up problems.

8.3 Cascading dynamic gates

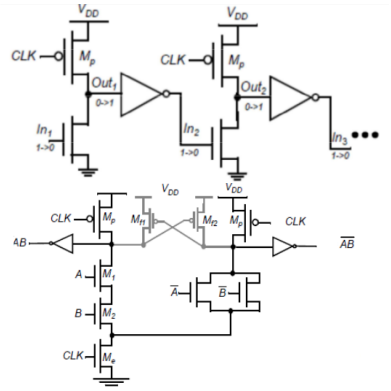
The problem in cascading dynamic gates is that only a low to high transition is permitted during the evaluation phase (or a stable signal).
We can cascade dynamic gates in two different ways.

8.3.1 Domino Logic

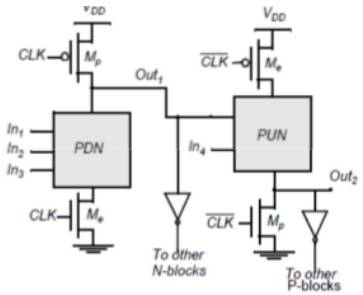


During the pre-charge phase, the output of the dynamic logic is charged up to V_{DD} while the inverter output is driven low. During the evaluation phase, the dynamic gate eventually discharges its output node and the output of the inverter transitions from 0 to 1. Otherwise, it remains low. This ensures that only transitions from 0 to 1 can happen in the evaluation phase. In this way sequential logic gates are driven by low impendance device increasing the reliability. The main drawback is that it's impossible to implement inverting functions in this way.

Alternative way is to use the unfooted logic (wihout the clocked nmos) or the Dual Rail Domino logic.



8.3.2 NP-CMOS dynamic



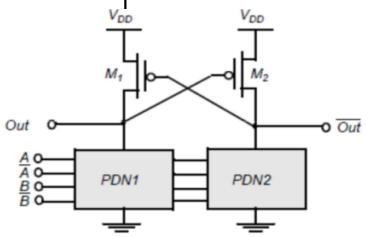
The idea is to cascade n-type dynamics gate to p-type in order to have always the correct commutation during the evaluation phase (note that both type have simultaniously evaluation and precharge phase).

Chapter 9

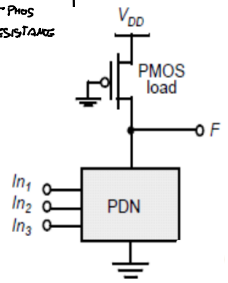
Scheme of all the logic families

LOGIC IMPLEMENTATIONS

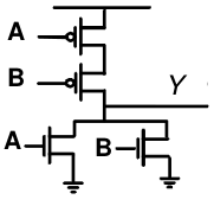
DCVSL
DIFFERENTIAL CASCODE VOLTAGE SWITCH LOGIC



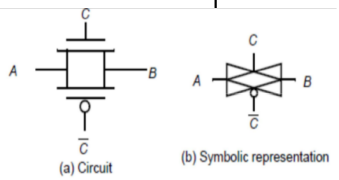
PSEUDO NMOS LOGIC
- $V_{OL} \neq 0$
- STATIC POWER WHEN PDN ACTIVE OF $P_{ST} = P(0) V_{DD} I_{PMOS}$
- PULL DOWN RESISTANCE FOR τ_P IS $R_{eqn} = \frac{R_n}{1 - \frac{R_n}{R_p}}$



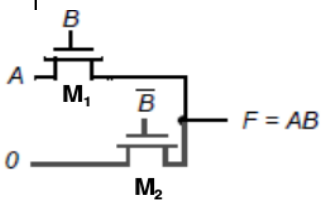
FULLY COMPLEMENTARY CMOS LOGIC
FC - CMOS



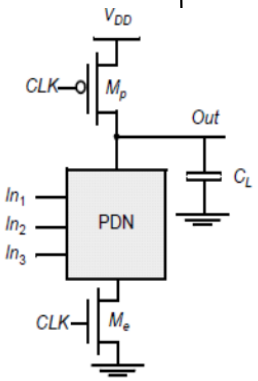
TRASMISSION GATE



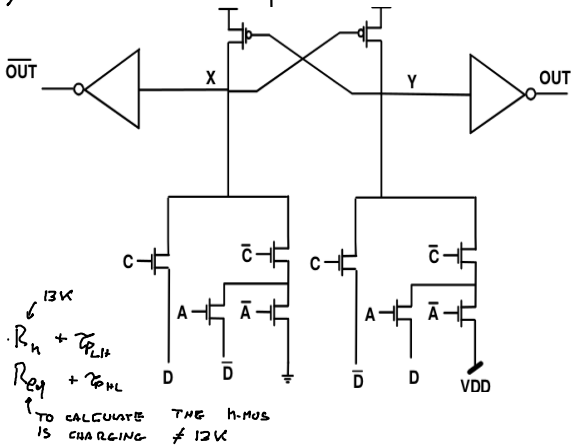
PASS TRANSISTOR LOGIC



DYNAMICS LOGIC



DIFFERENTIAL / COMPLEMENTARY PASS TRANSISTOR LOGIC
DPL / CPL



- $\tau_{LH} = \ln(2) C_{INT} \cdot R_n + \tau_{PLH}$ (13K)
- $\tau_{HL} = \ln(2) C_{INT} R_{eq} + \tau_{PHL}$

TO CALCULATE IS CHARGING THE N-MOS $\neq$ 13K

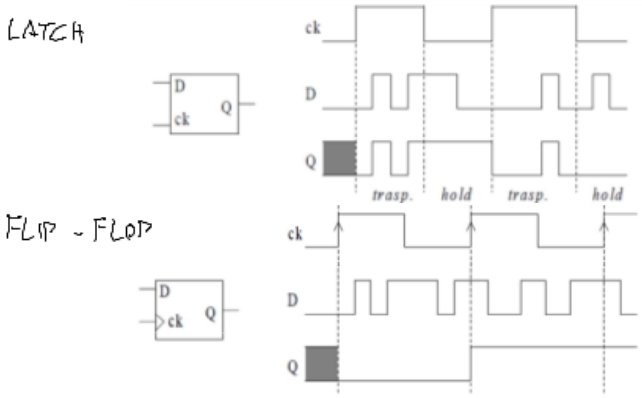
Chapter 10

Sequential circuits

Sequential logic circuits are such that the outputs depends not only on the present value of the input but also on theyr previous values. To achive this goal we need gates that have to remember the history of the input data; such circuits are said to have a state.
There are 2 basic memory elements: latches and flip-flops.

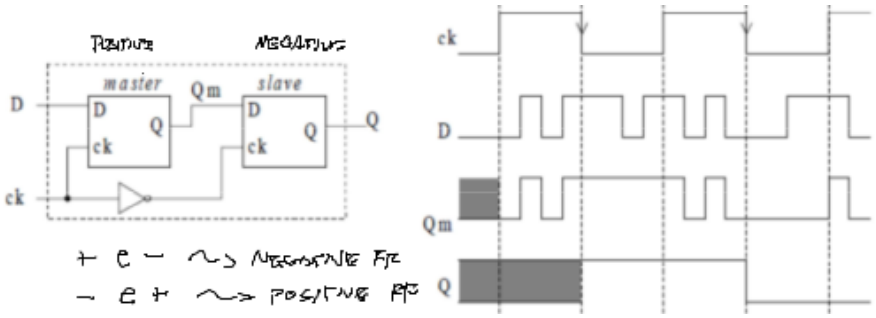
Latches are tipically used to build flip-flops in the master-slave configuration (that is the most adopted memory element). Both latches and flip-flop are in theyr basic configuration a 3 port devices with 2 inputs an 1 output: the two inputs are D data in and CLK clock , the output is Q (and an eventual $\bar{Q}$).

A latch is a memory element sensitive to the level of the clock, while the flip-flop is sensitive to the edge of the clock signal. We will refer to a latch as to an element level-triggered, and to the flip-flop as an element edge triggered.



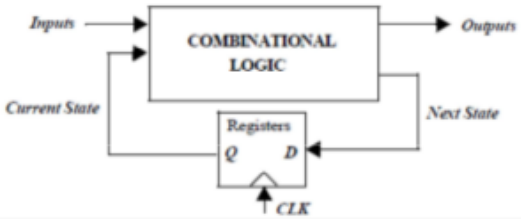
To build up a flip-flop we can use the master-slave structure using 2 latches one positive and

one negative connected as shown in figure



The main applications for this structrures are four: data sotrage in foreground memory (RAM is a background element), frequency divider, counters and finite state machines (FSM)

10.1 Finite state machines (FSM)



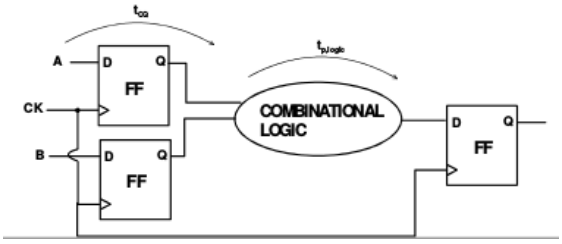
It's a compound of a logic circuit and one or more flip-flops adopted as memory elements to store data.

There are some importants times that has to be considered when we deal with memory elements

- The set-up time t_{su} is the time that the input data must be valid before the sensitive edge of the clock.
- The hold time t_{hold} is the time that the input data must remain stable after the edge of the clock (it can be also negative).
- If the previous times are verified the input data is copied at the output after a time t_{cq} that is form the edge of the clock to the cross of half supply at the output.

For a generic FSM we can underline two time constraints that have to be respected for the circuit to work.

10.1.1 Max-delay constraint



Denoting T as the clock period and $t_{p,logic}^{max}$ as the maximum delay of the logic, we have to ensure that

$$t_{cq} + t_{su} + t_{p,logic}^{max} \leq T \tag{10.1}$$

That can be translated in

$$t_{p,logic}^{max} \leq T - (t_{cq} + t_{su}) = T - t_{overhead} \tag{10.2}$$

The signal has to pass throught the flip-flop and the logic and be present at the input od the memory element a set-up time before the edge of the clock.

10.1.2 Min-delay constraint

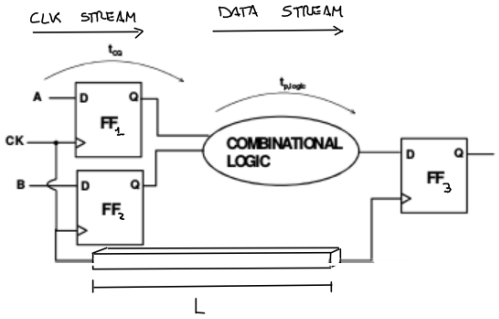
The hold constraint says that the signal cannot propagate throught the flip-flop and the logic too fast or it will remain at the input of the flip-flop for too little time

$$t_{cq} + t_{p,logic}^{min} \geq t_{hold} \tag{10.3}$$

10.1.3 Skew

If we have a non ideal line of the clock that connect the input group of flip-flops and the output group we encounter a phenomenon called skew that affects the delay constraints of our circuit. Using t_w as the propagation delay of the line we can consider two cases

Positive skew



In the case that the clok propagates in the same direction of the signal we are in the case of positive skew.
The max-delay constraint becomes

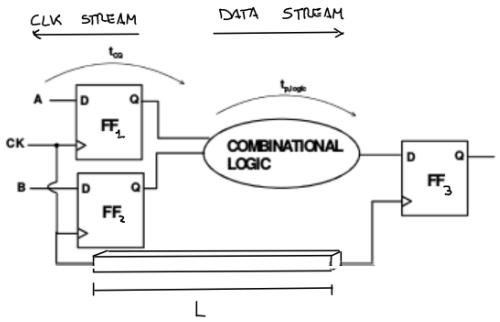
$$T_{max} = t_{cq} + t_{su} + t_{p,logic}^{max} - t_w \tag{10.4}$$

and the min-delay constraint

$$t_{hold} \leq t_{cq} + t_{p,logic}^{min} - t_w \tag{10.5}$$

We improve the maximum clock frequency but the hold constraint becomes more stringent.

Negative skew



In the case that the clok propagates in the opposite direction of the signal we are in the case of negative skew.
The max-delay constraint becomes

$$T_{max} = t_{cq} + t_{su} + t_{p,logic}^{max} + t_w \tag{10.6}$$

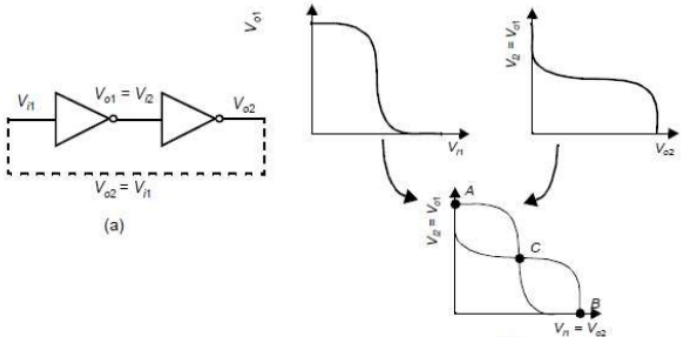
and the min-delay constraint

$$t_{hold} \leq t_{cq} + t_{p,logic}^{min} + t_w \tag{10.7}$$

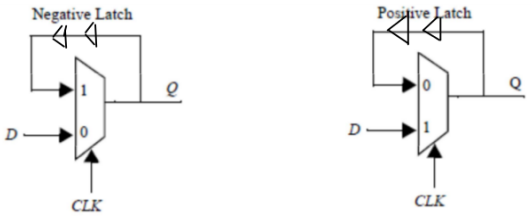
We decrease the maximum clock frequency but the hold constraint becomes less stringent.

10.2 Static memory devices

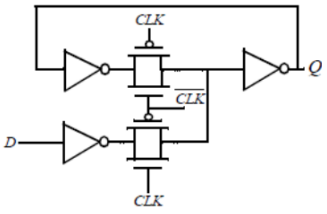
Typical static memory device is a couple of inverter connected in positive loop one another. This type of structure gives us (overlapping the 2 characteristics) has only 2 stable ponis (ground or V_{DD}) and one metastable point as shown in figure



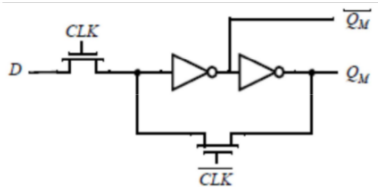
10.2.1 Multiplexer-based static latch



The loop can be opened to write the data or closed to store it. The mux is implemented with transimssion gates as in figure (positive latch in figure)



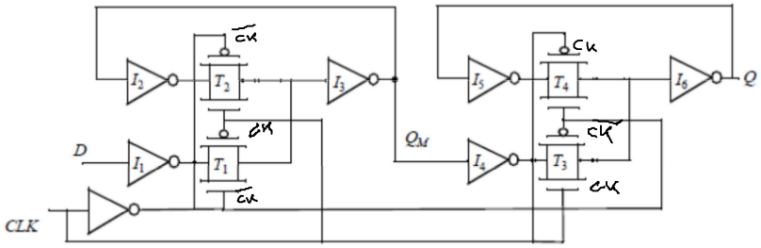
Since when implementing a clocked element the number of transistors connected to the clock signal play a foundamental role in power dissipation (since theyr $\alpha_{sw} = 1$) we can also use this structure



This structure has of course some drawbacks; the nmos passes a degraded high voltage that implies a larger propagation delay, reduced noise margins and static power consumptions of the inverters.

10.2.2 Multiplexer-based static flip-flop

The multiplexer-based static flip-flop is implemented in the master-slave configuration as follows



We define t_{inv} as the propagation delay of the inverter and t_{tx} as the propagation delay of a trasmission gate; doing so we can hilight the different time constraint of this device.

- **Set-up time**
  We need that the input signal passes throught the first inverter I_1 the trasmission gate T_1 and the other 2 inverters I_2, I_3 otherwise there is the possibility to have conflicts or

incorrect values on the 2,3 inverters so

$$t_{su} = 3t_{inv} + t_{tx} \tag{10.8}$$

- **Propagation delay**
  Due to set-up time we already have the signal after the I_4 inverter. The signal have to be propagated throught the trasmission gate I_4 and the inverter I_6

$$t_{cq} = t_{inv} + t_{tx} \tag{10.9}$$

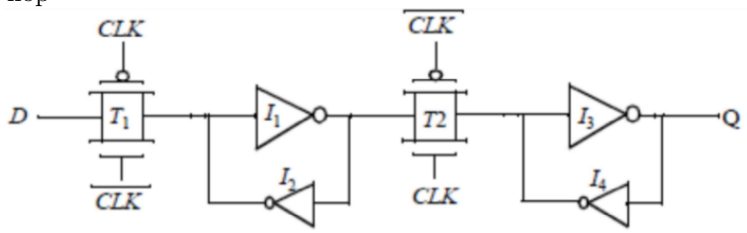
- **Hold time**
  The rising edge of the clock turns the transmission gate T_1 off, thus any change of the input signal does not cause a change of the flip-flop state. Since the inverter I_1 has a propagation delay, the input D can change also before the rising edge of the clock without being sampled by the transmission gate; the hold time is negative

$$t_{hold} = -t_{inv} \tag{10.10}$$

10.2.3 Time constraints rigorous definitions

10.2.4 Brute force filp-flop

To reduce the number of transistors involved in a flip-flop (and also the problem of high number of transistor feeded by the clock signal) we can use a ratioed solution like the brute-force flip-flop

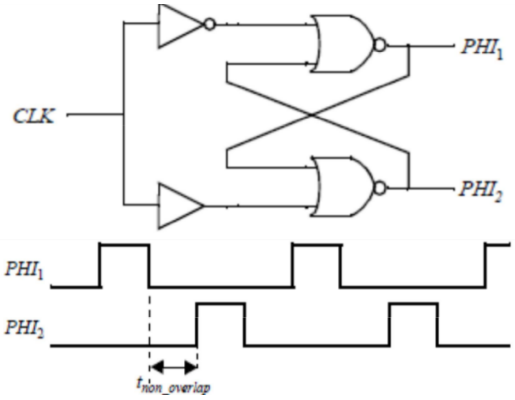


The first latch can be written only if the first trasmission gate and the driver circuit are stroger than the feedback inverter to change the state at the input of I_1.
Moreover there is a problem of "reverse conduction". When the clock is high the second trasmission gate is closed and there can be a conflict between the first and the second stage; I_4 has to be sized in a way that can't overwhelm I_1 (if this happen the first loop can change it's state).

10.2.5 Clock overlap

Due to different paths the clk and the $c\bar{l}k$ signals aren't sincronized but they have a period of overlap (0-0 or 1-1 overalap) that can cause wrong toggles (due to critical races that is that the output is connected to the input in 1-1 overlap) or undefined states (due to 0-0 overlap).

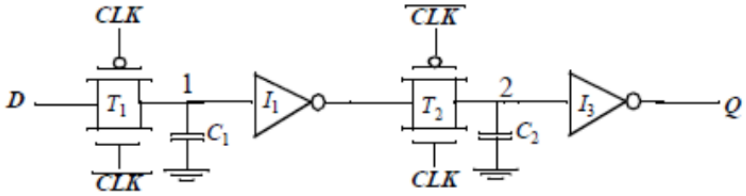
A solution to solve the 1-1 overlap is the adoption of the following circuit to generate the 2 clock signals



There is never a period when both signals are high but on the other hand there are a lot of 0-0 overlaps that can destroy the state if this condition last for too long.

10.3 Dynamic memory devices

This memory elements store the data as chage across a capacitance as the dynamics gates. The main drawback is that charge may be lost if the data is not refreshed for a long time.
The classical master-slave implementation of a flip-flop reduces to 8 transistors as shown in figure (the capacitance are the parasitic of the trasmission gate and of the inverter)



The sensitive times related to this implementation are

- **Set-up time**
  The time needed to the trasmission gate to pass the data to the node "1"

$$t_{su} = t_{tx} \tag{10.11}$$

- **Propagation delay**
  Is the time that the data use to pass form the node "1" to the output (the transition throught the inverter has not yet passed)

$$t_{cq} = t_{inv} + 2t_{tx} \tag{10.12}$$

- **Hold time**
  The hold time is 0

During both clock overlaps in this gate we get a direct connection between in and out.
The 0-0 overlap can be avoided making the overlap period small enough that the signal does not reach the output

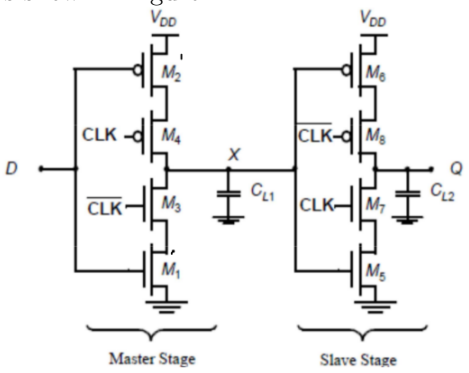
$$t_{overlap-00} \leq t_{inv} + 2t_{tx} \tag{10.13}$$

The 1-1 overlap can be avoided establishing a hold time large enough

$$t_{overlap-11} \leq t_{hold} \tag{10.14}$$

10.3.1 C^2MOS

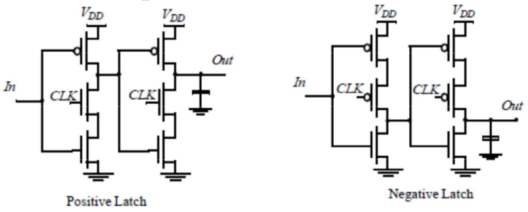
This is a dynamic solution insensitive to clock overlap problems. C^2 stand for clocked cmos and the implementation is shown in figure



The two latches are tristate inverters : the output node can be high low or at high impedance state depending on the inputs. The figure represents a positive edge triggered ff.
The only problem that arise is just after a 1-1 overlap as soon as $c\bar{l}k$ transit back to 0 the Q node is turned up to V_{DD}. This can be solved only imposing a hold time constraint.

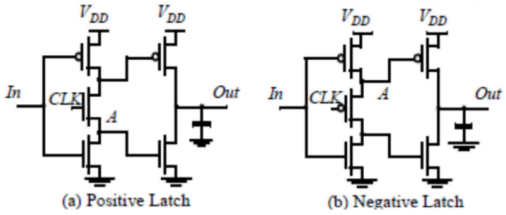
10.3.2 True single phase clock TSPC flip-flops

This implementation avoid the clock overlap problem by using only the clok signal avoiding it's complement relying on the fact that pmos and nmos are on for different voltages.



When clk is high the circuit is a cascade of two inverters and when the clock is low it's an open circuit in the positive latch and vice versa for the negative so that no direct path can be formed.

To spare transistors we can use the so called "split out" scheme



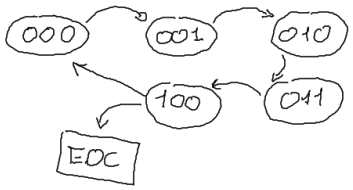
The drawback in this case is that the node A cannot feature a full swing leading to worst performance in terms of propagation delay.

10.4 Counters

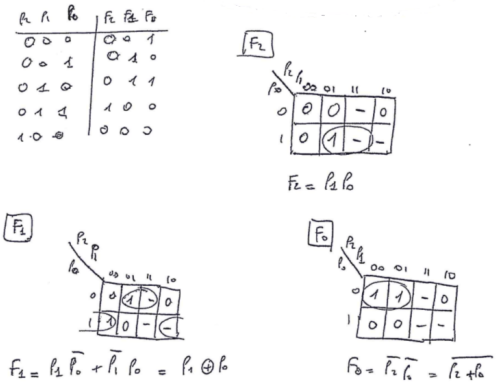
Conunters are finite states machines that counts up or down until a certain value.
The necessary flip-flops are equal to the number of bits that are necessay to code the decimal number in binary.
We can distinguish in two categories of counters: synchronous and asynchronous.

10.4.1 Synchronous counters

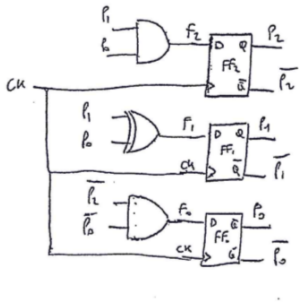
In synchronous counters all flip-flops are driven by the same clock signal. Taking in example a modulus 5 up-counter (that means it conts 000 001 010 011 100) adopting positive edge triggered flip-flops we can design the state diagram of our circuit as



Than we can represent in a table the corrispondance of the next states to the present ones and derive the expressions for the next states like

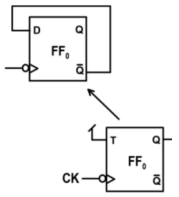


We have used the don't care states to simplify the expressions.
After this passage we have to properly connect the 3 flip-flops considering that the D of the i-th element is the i-th next state and the Q is the present state.

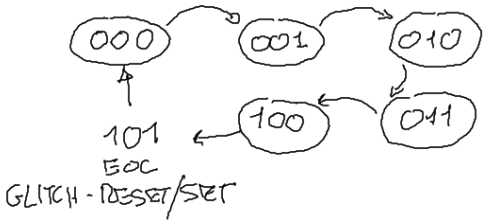


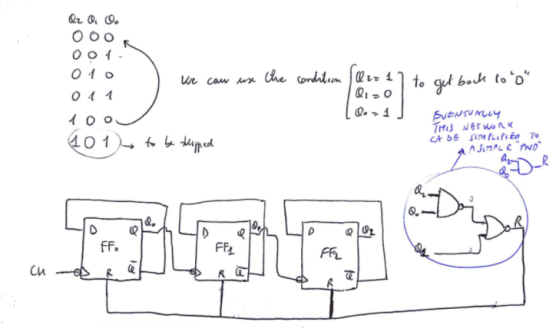
10.4.2 Asynchronous counters

In asynchronos counters we use a particular configuration of flip flops with the $\bar{Q}$ signal attached to D so that the Q output have the same shape of the clock of the flip-flop but with half the frequency (frequency divider).



Using negative edge flip-flops we can make upcounters cascading the flip-flops so that the Q signal of the n-1 ff is the clock signal of the n-th ff. To restart the count we have to use the set/reset inputs of the flip-flops controlled by an appropriate logic following the state diagram. Here in example a mod-5 up counter



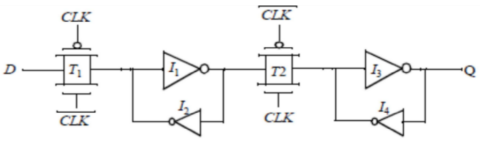


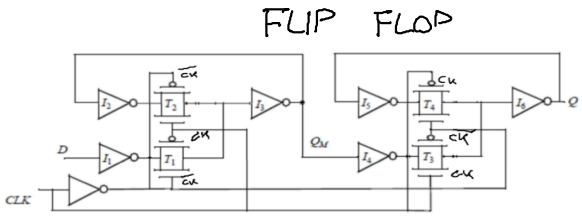
To make up-counters we have to use negative edge triggered flip-flops and to make down-counters positive edge triggered flip-flops.

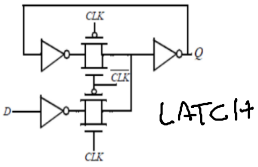
10.5 Scheme of all flip-flops implementations

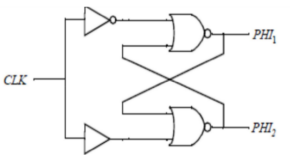
Flip Flops
Type of Implementation

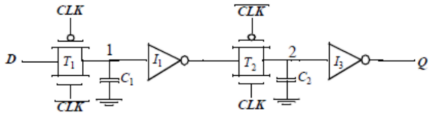
- Static
  - Brute Force (Ratioed)
  - Flip Flop
    - Multiplexer Based
  - Latch
  - CLK Overlap 1-1 Solver
- Dynamic
  - Classic
  - TSP
    - TSP "Split-Out"
  - C² MOS

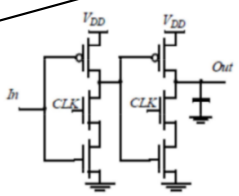




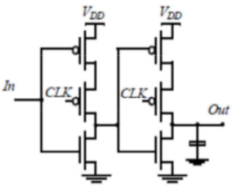




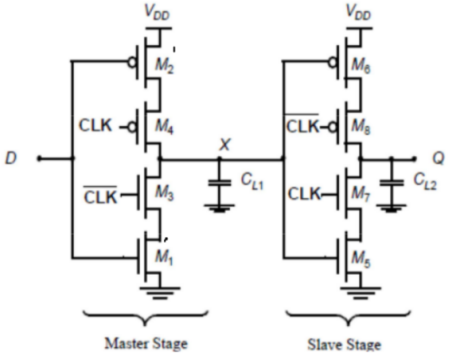


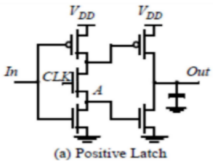


Positive Latch

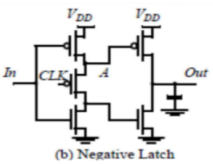


Negative Latch





(a) Positive Latch



(b) Negative Latch