Digital integrated circuit design
Notes

Professor:
Andrea Bonfanti

Student:
Matteo Baldo

March 29, 2017

Contents

Chapter 1

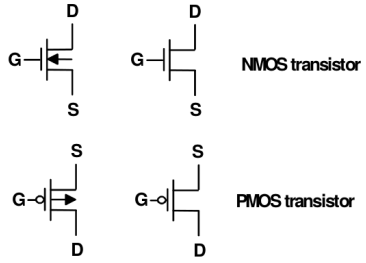
Figures of merit in a digital integrated circuit

Chapter 2

The MOS transistor

2.1 Symbols

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



2.2 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

2.2.1 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.

2.2.2 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.

2.3 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 can be easily neglected since is $\simeq 1$.
We can make three considerations:

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

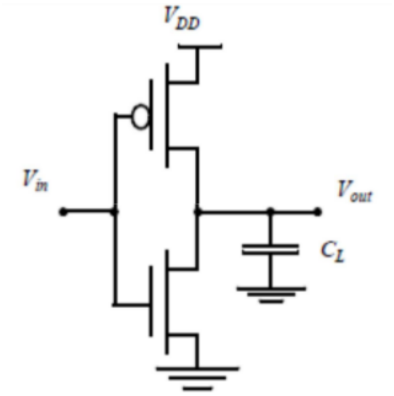
2)for $V_{DD} >> V_T + V_{DSAT}/2$, the resistance becomes virtually independent of the supply voltage

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 3

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.

3.1 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

$$V_{OH} = V_{DD} = 2.5V \tag{3.1}$$

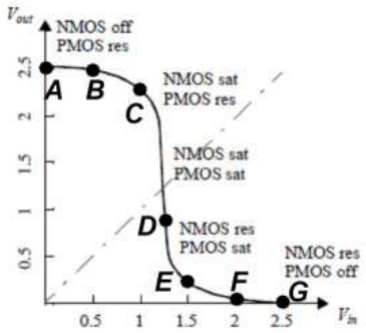
$$V_{OL} = GND = 0V \tag{3.2}$$

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

$$r_{on} = \frac{4.2k}{(W/L)_n} \qquad r_{on} = \frac{15.9k}{(W/L)_p} \tag{3.3}$$



3.1.1 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

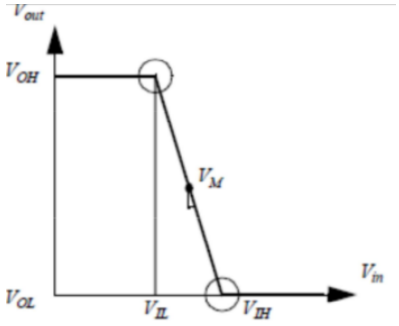
$$r = \frac{V_M - V_{TEn}}{V_{DD} - V_M - V_{TEP}}$$

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

$$\frac{(W/L)_p}{(W/L)_n} = 3.5 \simeq 3 \tag{3.4}$$

3.1.2 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

$$V_{OUT} = V_{IN}g + (1-g)V_M \qquad (g<0) \tag{3.5}$$

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

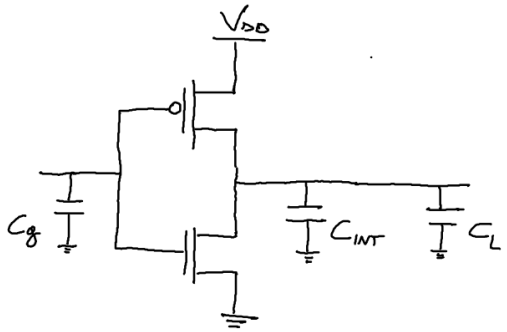
$$V_{IH} = V_M + \frac{V_{OL} - V_M}{g}$$

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

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

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

3.2 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

3.2.1 Intinsic capacitance C_{int}

$$C_{int} = C_{overlap} + C_{dbn} + C_{dbSWn} + C_{dbp}C_{dbSWp} \tag{3.6}$$

Overlap capacitance

$$C_{ovelap} = 2(C'_{ovp}W_p + C'_{ovn}W_n) \tag{3.7}$$

Drain-bulk n capacitances

$$C_{dbn} = \frac{C'_{jn}A_{dn}}{2}\left(\frac{1}{(1+\frac{V_{DD}}{\phi})^{0.5}} + \frac{1}{(1+\frac{V_{DD}}{2\phi})^{0.5}}\right) \tag{3.8}$$

$$C_{dbSWn} = \frac{C'_{jSWn}P_{dn}}{2}\left(\frac{1}{(1+\frac{V_{DD}}{\phi})^{0.44}} + \frac{1}{(1+\frac{V_{DD}}{2\phi})^{0.44}}\right) \tag{3.9}$$

Drain-bulk p capacitances

$$C_{dbp} = \frac{C'_{jp}A_{dp}}{2}\left(1 + \frac{1}{(1+\frac{V_{DD}}{2\phi})^{0.48}}\right) \tag{3.10}$$

$$C_{dbp} = \frac{C'_{jSWp}P_{dp}}{2}\left(1 + \frac{1}{(1+\frac{V_{DD}}{2\phi})^{0.32}}\right) \tag{3.11}$$

3.2.2 Gate capacitance C_g

$$C_g = C'_{ovn}W_n + C'_{ovp}W_p + C'_{ox}(W_nL_n + W_pL_p) \tag{3.12}$$

3.2.3 Self-loading factor

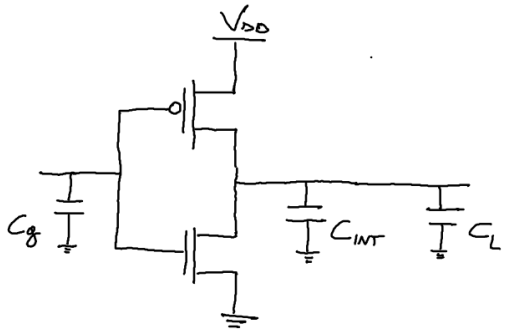
We can define the self-loading factor γ as

$$C_{int} = \gamma C_g \tag{3.13}$$

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

$$C_g^{(\alpha)} = \alpha C_g^{(1)} = \alpha \cdot 2fF \tag{3.14}$$

3.3 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 forom 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 $C_{ext} = 0$

$$\tau_{p0} = \frac{\tau_{HL} + \tau_{LH}}{2} \tag{3.15}$$

3.3.1 Fan-out

Considering an external capacitance we get the folowing expression

$$\tau_p = \ln(2) R_{eq}(C_{int} + C_{ext}) = \tau_{p0}\left(1 + \frac{C_{ext}}{C_{int}}\right) = \tau_{p0}\left(1 + \frac{C_{ext}}{\gamma C_g}\right) = \tau_{p0}\left(1 + \frac{f}{\gamma}\right) \tag{3.16}$$

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

$$f = \frac{C_{ext}}{C_g} \qquad C_{int} = \gamma C_g \tag{3.17}$$

3.4 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

$$F = \frac{C_L}{C_{g,1}} \tag{3.18}$$

And from this we calculate the optimum fan-out

$$f_{opt} = \sqrt[N]{F} \tag{3.19}$$

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

$$s_n = s_1 \cdot (f_{opt})^n \quad \forall n = 1, ..., N \tag{3.20}$$

and the propagation delay is

$$\tau_p = N\tau_{p0}\left(1 + \frac{f_{opt}}{\gamma}\right) \tag{3.21}$$

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

$$N_{opt} = \frac{\ln(F)}{\ln(3.6)} \tag{3.22}$$

beacuse the best fan-out that we can have is

$$f_{opt}^{ideal} = 3.6 \tag{3.23}$$

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

Chapter 4

Inverter power consumption

4.1 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

$$P_{dyn} = C_L V_{DD}^2 f_{1\to 0} \tag{4.1}$$

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

$$f_{1\to 0} = f_{clk} \cdot \alpha_{SW} = f_{clk} \cdot \frac{1}{2} \tag{4.2}$$

In case of a randoom signal

$$f_{1\to 0} = f_{clk} \cdot \alpha_{SW} = f_{clk} \cdot \frac{1}{4} \tag{4.3}$$

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

$$P_{dyn} = C_L V_{DD}^2 f_{clk} \cdot \alpha_{SW} \tag{4.4}$$

$$E_{dyn} = C_L V_{DD}^2 \alpha_{SW} \tag{4.5}$$

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

Chapter 5

The wires

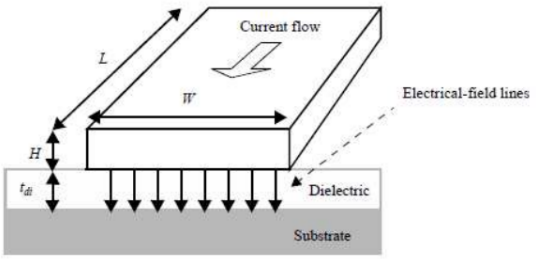
To simplify the analysis of the wires parasitics effects we introduce 3 simple assumption;

→ 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.

5.1 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

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

5.2 Resistance

Resistance of a wire can be assested as

$$R = \rho \frac{L}{WH} = R_{sh} \frac{L}{W} \tag{5.1}$$

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

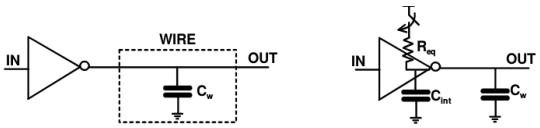
$$\sigma = \sqrt{\frac{\rho}{\pi f \mu}} \tag{5.2}$$

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.

5.3 Models for wires

5.3.1 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.

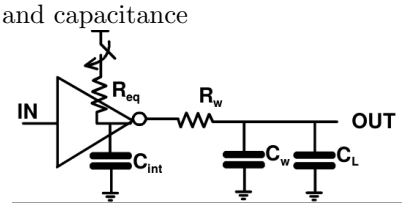


The wire capacitance has to be added to the intrinsic capacitance of the inverter in order to correctly estimate the propagation delay that in this case is

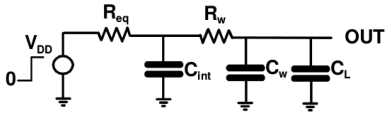
$$\tau = \ln(2) R_{eq}(C_{int} + C_w) \tag{5.3}$$

5.3.2 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

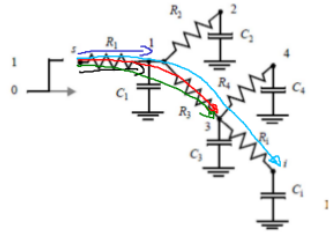


To estimate the overall propagation we can model the circuit as follows and use the Elmore theorem



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.

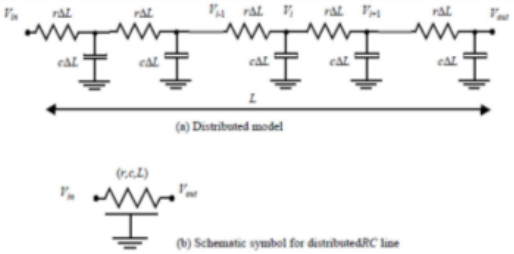


$$\tau = C_1 R_1 + C_2 R_1 + C_3(R_1 + R_3) + C_4(R_1 + R_3) + C_j(R_1 + R_2 + R_j) \tag{5.4}$$

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.

5.3.3 Distributed model

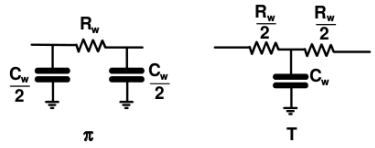
With the distributed model we divide the wire into lot of wires of length ΔL with $\Delta L \to 0$.



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

$$\tau_w \simeq \ln(2)\frac{R_w C_w}{2} \tag{5.5}$$

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



5.4 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

$$L \geq \sqrt{\frac{16 C_{int} R_{eq}}{cr}} \tag{5.6}$$

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

$$N = \sqrt{\frac{R_w C_w}{4 C_{int} R_{eq}}} = L\sqrt{\frac{rc}{4 C_{int} R_{eq}}} \tag{5.7}$$