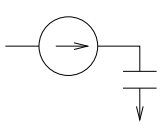
Delay of a Digital Gate

Fundamentally, In ALL cases: D = C * Vswing/ I



Unit Check:

C [Farad] = Q [coulomb] / V [volt]

I [Amp] = Q [coulomb] / s [sec]

$$D[sec] = CV/I = \frac{coulomb * volt * sec}{volt \quad coulomb}$$

Now then, if only we had a constant current source and a constant capacitance it would be trivial to calculate delay.

BUT: Current (I) is a nonlinear function of a bunch of variables.

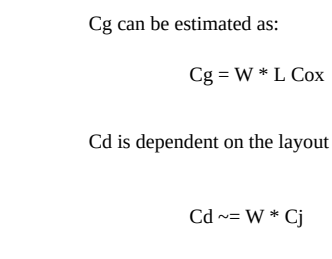
AND, All the capacitances are voltage dependent.

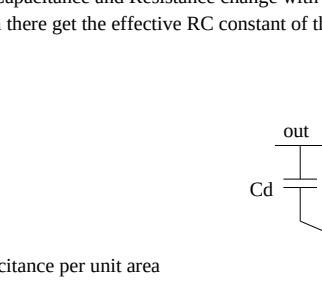
Cutoff Ids = 0

Linear $Ids = W/L * (ue_{ox} / t_{ox})[(Vgs - Vt)Vds - 0.5\ Vds^2]$

Saturation $Ids = W/L * (ue_{ox} / t_{ox})\ 0.5*[(Vgs - Vt)^2 (1-lambda*Vds)]$

Even these are highly simplified models...

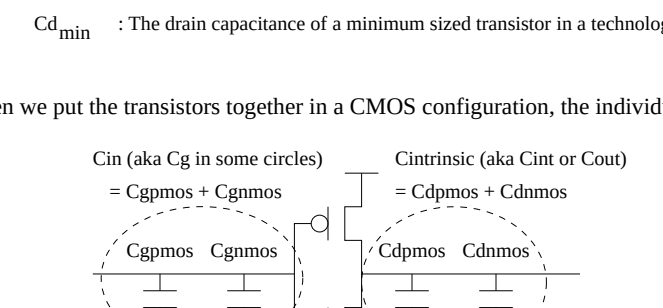




We also have even more complicated models for the voltage dependent capacitances!

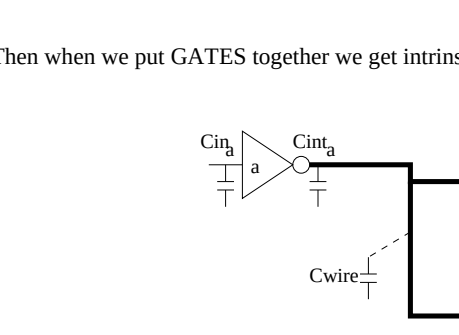
A SPICE simulator is great for solving all of those detailed equations. We want a simpler way!

In digital circuits, we are always doing the same thing... Vgs is a step function from 0 to VDD or VDD to 0. What does the output look like?



When we simulate it, we get lucky − the output LOOKS A LOT LIKE a resistor charging a constant capacitor!!!

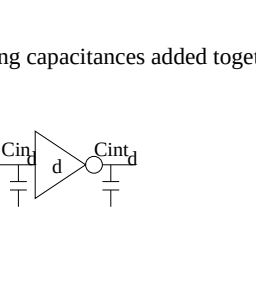
An RC Circuit Charging



From RC circuit theory we know that the delay to reach 1/2 Vout is:

D = 0.69 RC

A CMOS gate switching



In the digital world, this is called the propogation delay from low to how (tplh)

D = CV/I Recall: R = V/I

If we 'pretend' that there is a constant resistance and capacitance:

tplh = 0.69 * Ctotal * Reffective

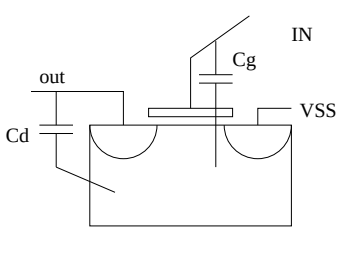
So, the question becomes, what are the values for Ctotal and Reffective?

Remember, we introduced these mythical terms to get rid of the complex models. If you REALLY felt the urge you could try to work out Ctotal and Reff but it would be a losing battle. You would need to know the operating mode of the transistors to estimate the instantaneous capacitance and resistance, but the operating modes swing wildly! That's what spice is for.

It is much easier to develop heuristics (rules of thumb) about how the Capacitance and Resistance change with common circuit parameters. You can run a simulation to get the delay (either tphl or tplh), and from there get the effective RC constant of the circuit.

Estimating Average Capacitances in CMOS circuits

For a particular transistor there are 2 capacitances of interest, the gate (Cg) and drain (Cd).



Cg can be estimated as:

$Cg = W * L\ Cox$ where Cox = the capacitance per unit area

Cd is dependent on the layout of the drain but is also proportional to the width.

$Cd \sim= W * Cj$ where Cj = the juniction capacitance per unit width of the drain

Note: This is simplified from what you might see elsewhere.

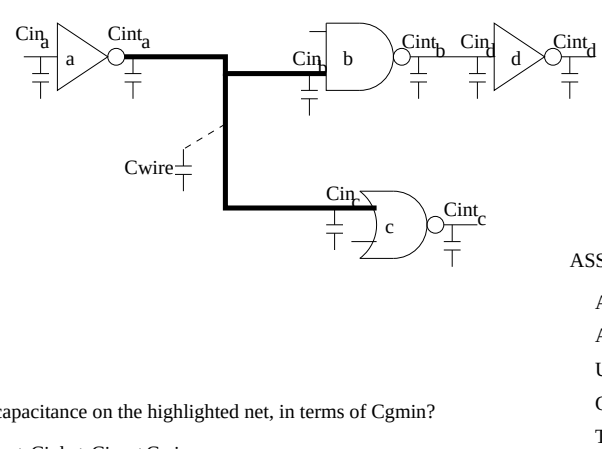
In many situations it is usefull to define some base values of Capacitances (and Resistances) relative to the smallest transistor available.

Cg_{min} : The gate capacitance of a minimum sized transistor in a technology (eg. W=220um, L=180um in 0.18um TSMC)

Cd_{min} : The drain capacitance of a minimum sized transistor in a technology (eg. W=220um in 0.18um TSMC)

Inside a gate, when we put the transistors together in a CMOS configuration, the individual capacitances can add together.

Then when we put GATES together we get intrinsic, input AND wiring capacitances added together



ASSUME:

All NMOS are minimum sized.

All PMOS are 2x min width, min length.

Unit wiring capacitance is 0.2*Cgmin/um

Cdmin = Cgmin / 4

The highlighted net is 4 um long.

Example:

What is the total capacitance on the highlighted net, in terms of Cgmin?

Cnet = Cinta + Cinb + Cinc +Cwire

= (Cdmin + 2Cdmin) + (Cgmin + 2Cgmin) + (Cgmin + 2Cgmin) + 4*0.2*Cgmin

= 0.75 Cgmin + 6 Cgmin + 0.8 Cgmin

= 7.55 Cgmin

Example:

The data sheets for a standard cell inverter specify that it has an input capacitance of 6 fF and output capacitance of 3fF. From some clever reverse engineering you've figured out that the PMOS is 2.200um/0.180um and NMOS is 1.180um/0.180 um. Minimum gate length is 0.18 um, width is 0.22um.

What is Cgmin and Cdmin?

Cin = Cgpmos + Cgnmos

6fF = 10xCgmin + (1180/220) Cgmin

6fF = 18.18 Cgmin

Cgmin = 0.330 fF

Cout = Cint = Cdpmos + Cdnmos

3 fF = 10 Cdmin + (1180/220) Cdmin

Cdmin = 0.165 fF

Conclusion:

When referring to digital gates, we simplify capacitances in terms of an input (Cin) and output/intrinsic capacitance (Cint).

Given relative transistor sizing in a system, you should be able to determine the capacitance on each net in terms of Cgmin and Cdmin.

Estimating Effective Resistances in CMOS circuits

Let's simplify a transistor by converting it into an effective resistance (Reff)

Try to coax it into an expression that looks like V=IR or I = V/R and isolate terms to find R

But in charging or discharging a transistor, we go from cutoff −> saturation −> linear −> cutoff

Cutoff Ids = 0

Linear $Ids = W/L * (ue_{ox} / t_{ox})[(Vgs - Vt)Vds - 0.5\ Vds^2]$ Often insignificant

$Ids = W/L * K\ Vds$ Where $K = (VDD - Vt)\ u\ e_{ox} / t_{ox}$

$Ids = 1/R\ (Vds)$

$R = 1/K * L/W$ depending on whether charging or discharging

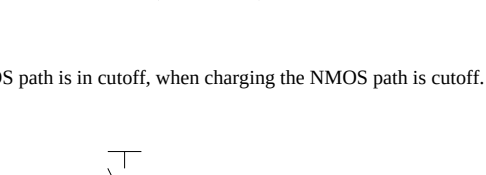
Electron or Hole Mobility

Saturation $Ids = W/L * (ue_{ox} / t_{ox})\ 0.5*[(Vgs - Vt)^2 (1-lambda*Vds)]$

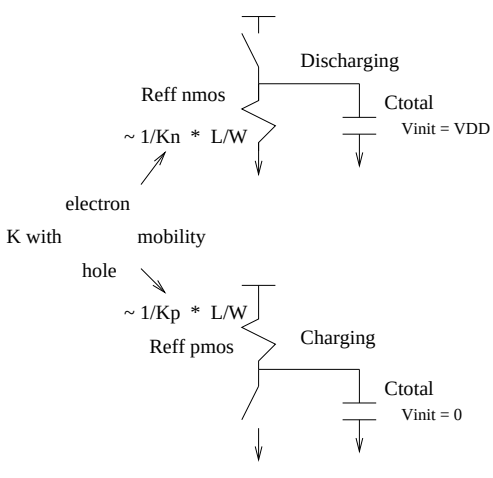
Note that an R does not make as much sense in saturation.
In Saturation, there is a relatively high I, but it is independent of V.

Since Rsat = dV/dI it has a very HIGH instantaneous resistance.

Rsat >> Rlinear

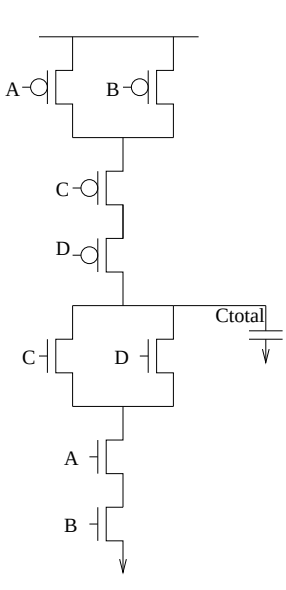


When discharging, the PMOS path is in cutoff, when charging the NMOS path is cutoff.



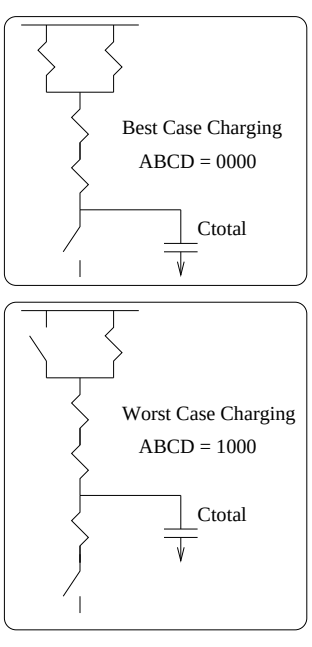
Modelling Complex CMOS gates

Replace each transistor with its equivalent resistance and simplify the resistor network down to one Reff resistance.



WARNING

Reff is 'pattern dependent'



* Assumes all transistors are the same size.

As with the capacitances, it is usefull to define a baseline Rnmin and Rpmin

Rpmin : Resistance of a minimum W/L PMOS in a technology

Rnmin : Resistance of a minimum W/L NMOS in a technology

Beware: These are MAXIMUM channel resistances.

Increasing the width beyond the minimum will lower resistance.

Example: For a given inverter, you know the total load Ctotal=4fF.

Detailed SPICE simulations reveal that tphl = 60ps and tplh = 118ps.
The W/L ratio of the NMOS is 220nm/180nm and is minimum size.
The W/L ratio of the PMOS is 360nm/180nm

Determine the Reff of the PMOS and NMOS. Also determine Rpmin and Rnmin

tplh = 0.69 Rpmos Ctotal

Rpmos = 118ps / 0.69*4f
= 42.75k Ohm

Minimum width is 220nm

360/220 = Rpmin/42.75k
Rpmin = 69.95k

tphl = 0.69 Rnmos Ctotal

Rnmos = 60ps / 0.69*4fF
= 21.74k Ohm

Since this is already minimum sized,

Rnmin = 21.74kOhm

Things we are not taking into account

1) In complex gates there are all of the intermediate drain capacitances which may need to be charged and discharged.

We do account for the Cd capacitances directly on the output node by including them in Ctotal

The others are hard to account for in most cases.
See Elmore's delay model for a good approximation.

2) The approzimation that D = 0.69 ReffCtotal assumes a step input.
For slightly sloped inputs (tr or tf < 1/2 D) it is still a good approximation.

3) In gates other than an inverter, we don't have full Vgs = VDD or −VDD

4) For the same reason as above, since the bulk to source voltage != 0, the effective threshold voltage (Vth) changes

5) We use first order model approximations to get Reff. This does not account for the non−linearities.

Conclusion

Given an arbitrty CMOS circuit with arbitrary transistor sizes,
you should be able to determine the simplified RC delay in terms of Cgmin, Cdmin, Rpmin, Rnmin

Given either technology constants or simulation results, you should bee able to determine Cgmin, Cdmin, Rpmin and Rnmin.