Lecture #31

ANNOUNCEMENTS

- Prof. King's office hours this week are cancelled

OUTLINE

» Fan-out

» Propagation delay

» CMOS power consumption

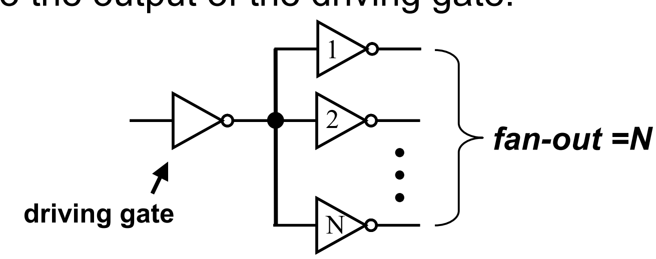
» Timing diagrams

Reading (Rabaey *et al.*)

- Chapter 1.3, pp. 21-22 & 24-28
- Chapter 5.2 & 5.5, pp. 148-149 & 173-184
- Chapter 6.2.1, pp. 204-207, 215-216

Fan-Out

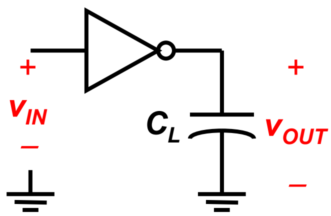
- Typically, the output of a logic gate is connected to the input(s) of one or more logic gates
- The ***fan-out*** is the number of gates that are connected to the output of the driving gate:

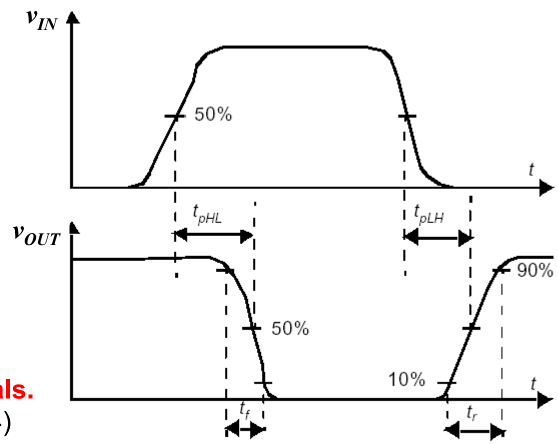


- Fanout leads to increased capacitive load on the driving gate, and therefore longer propagation delay
  - The input capacitances of the driven gates sum, and must be charged through the equivalent resistance of the driver

Effect of Capacitive Loading

- When an input signal of a logic gate is changed, there is a ***propagation delay*** before the output of the logic gate changes. This is due to capacitive loading at the output.





The propagation delay is measured between the 50% transition points of the input and output signals.
(Ref. Lecture 16, Slides 3&4)

Calculating the Propagation Delay

Model the MOSFET in the ON state as a resistive switch:

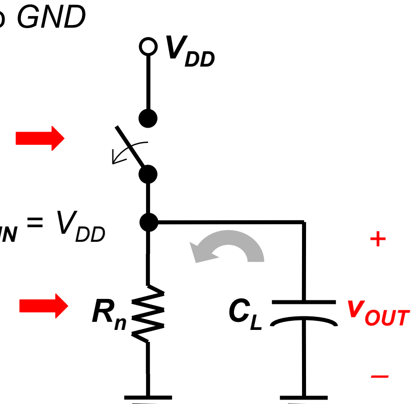
Case 1: V_{out} changing from High to Low
(input signal changed from Low to High)

- NMOSFET(s) connect V_{out} to GND

$t_{pHL} = 0.69 \times R_n C_L$

Pull-up network is modeled as an open switch

Pull-down network is modeled as a resistor



Calculating the Propagation Delay (cont'd)

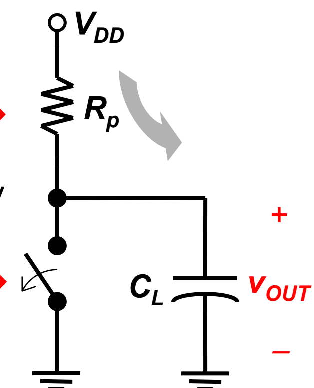
Case 2: V_{out} changing from Low to High
(input signal changed from High to Low)

- PMOSFET(s) connect V_{out} to V_{DD}

$t_{pLH} = 0.69 \times R_p C_L$

Pull-up network is modeled as a resistor

Pull-down network is modeled as an open switch



Output Capacitance of a Logic Gate

- The output capacitance of a logic gate is comprised of several components:

"intrinsic capacitance" →
- pn-junction and gate-drain capacitance
  - both NMOS and PMOS transistors

"extrinsic capacitance" {
- capacitance of connecting wires
- input capacitances of the fan-out gates

Impact of gate-drain capacitance

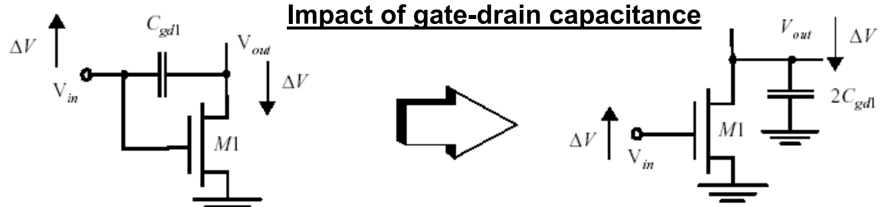
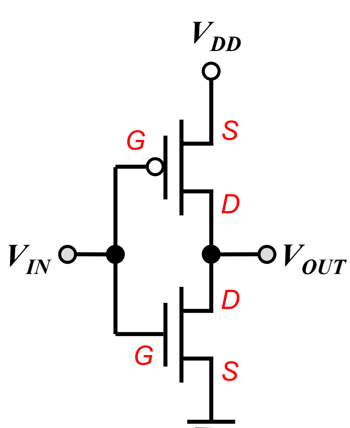


Figure 5.14 The Miller effect—A capacitor experiencing identical but opposite voltage swings at both its terminals can be replaced by a capacitor to ground, whose value is two times the original value.

Minimizing Propagation Delay

- A fast gate is built by
  1. **Keeping the output capacitance C_L small**
     – Minimize the area of drain pn junctions.
     – Lay out devices to minimize interconnect capacitance.
     – Avoid large fan-out.
  2. **Decreasing the equivalent resistance of the transistors**
     – Decrease *L*
     – Increase *W*
     … but this increases pn junction area and hence C_L
  3. **Increasing V_{DD}**
     → trade-off with power consumption & reliability

Transistor Sizing for Performance

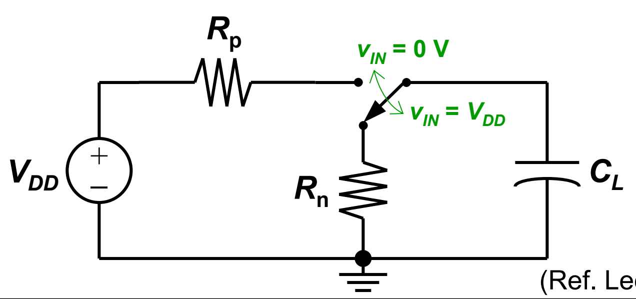


- Widening the transistors reduces resistance, but increases capacitance
- In order to have the on-state resistance of the PMOS transistor match that of the NMOS transistor (*e.g.* to achieve a symmetric voltage transfer curve), its *W*/*L* ratio must be larger by a factor of ~3. To achieve minimum propagation delay, however, the optimum factor is ~2.

CMOS Energy Consumption (Review)

- The energy delivered by the voltage source in charging the load capacitance is $C_L V_{DD}^2$
  - Half of this is stored in C_L; the other half is absorbed by the resistance through which C_L is charged.

→ In one complete cycle (charging and discharging), the total energy delivered by the voltage source is $C_L V_{DD}^2$



(Ref. Lecture 16, Slides 5&6)

CMOS Power Consumption

- The total power consumed by a CMOS circuit is comprised of several components:

  1. Dynamic power consumption due to charging and discharging capacitances*:

$$P_{dyn} = C_L V_{DD}^2 f_{0\rightarrow 1} = C_{EFF} V_{DD}^2 f$$

$f_{0\rightarrow 1}$ = frequency of 0→1 transitions ("switching activity")

f = clock rate (maximum possible event rate)

Effective capacitance C_{EFF} = average capacitance charged every clock cycle

*** This is typically by far the dominant component!**

CMOS Power Consumption (cont'd)

2. **Dynamic power consumption due to direct-path currents during switching** (Ref. Lecture 27, Slides 3&4)

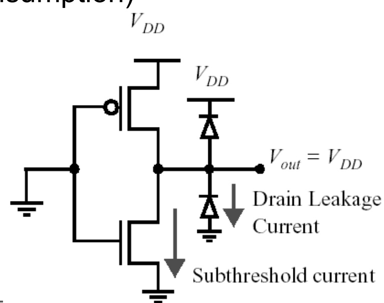
$$P_{dp} = C_{sc} V_{DD}^2 f$$

$C_{sc} = t_{sc} I_{peak} / V_{DD}$ is the equivalent capacitance charged every clock cycle due to "short-circuits" between V_{DD} & GND

(typically <10% of total power consumption)

3. **Static power consumption due to transistor leakage and pn-junction leakage**

$$P_{stat} = I_{stat} V_{DD}$$



Low-Power Design Techniques

1. **Reduce V_{DD}**
   → quadratic effect on P_{dyn}

   Example: Reducing V_{DD} from 2.5 V to 1.25 V reduces power dissipation by factor of 4

   – **Lower bound is set by V_T: V_{DD} should be >$2V_T$**

2. **Reduce load capacitance**
   → Use minimum-sized transistors whenever possible

3. **Reduce the switching activity**
   – involves design considerations at the architecture level (beyond the scope of this class!)

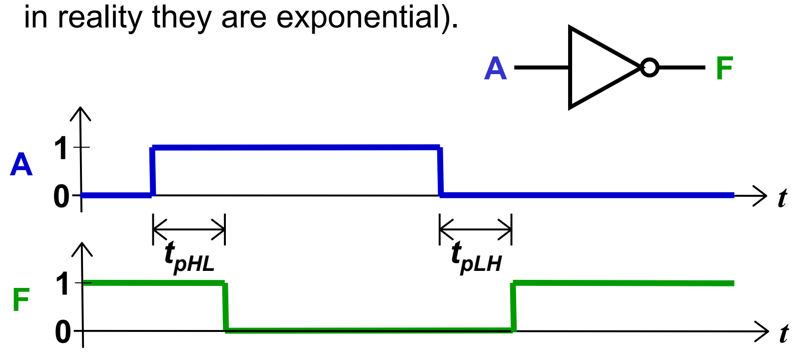
NAND Gates *vs.* NOR Gates

- In order for a 2-input NAND gate to have the same pull-down delay (t_{pHL}) as an inverter, the NMOS devices in the NAND gate must be made twice as wide.
  - This first-order analysis neglects the increase in capacitance which results from widening the transistors.
  - Note: The delay depends on the input signal pattern.
- In order for a 2-input NOR gate to have the same pull-up delay (t_{pLH}) as an inverter, the PMOS devices in the NOR gate must be made twice as wide.
  - Since hole mobility is lower than electron mobility (so that larger W / L ratios are needed for PMOS devices as compared with NMOS devices), stacking PMOS devices in series (as is done in a NOR gate) should be avoided as much as possible.

  → **NAND gates are preferred for implementing logic!**

Propagation Delay in Timing Diagrams

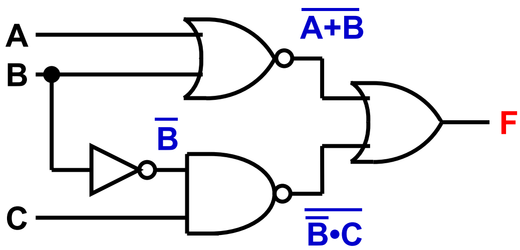
- To simplify the drawing of timing diagrams, we can approximate the signal transitions to be abrupt (though in reality they are exponential).

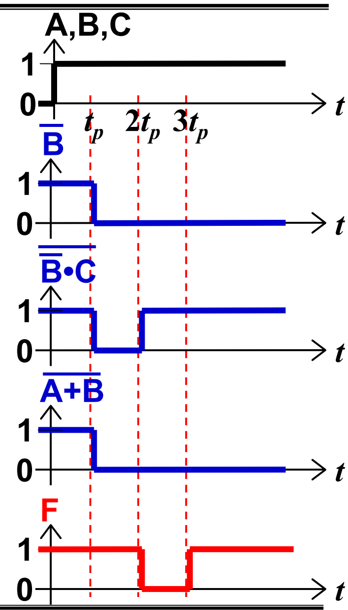


To further simplify timing analysis, we can define the propagation delay as $t_p = (t_{pHL} + t_{pLH})/2$

Glitching Transitions

The propagation delay from one logic gate to the next can cause spurious transitions, called ***glitches***, to occur. (A node can exhibit multiple transitions before settling to the correct logic level.)

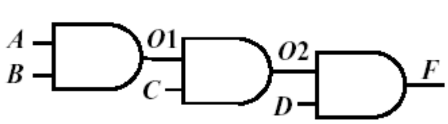




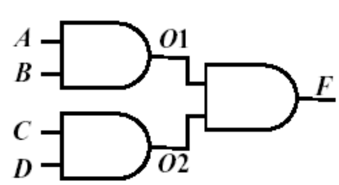
Glitch Reduction

- Spurious transitions can be minimized by balancing signal paths

<u>Example</u>: **F = A•B•C•D**



Chain structure



Tree structure