

VLSI Design

Lecture 7: Transfer Characteristics, Delay and Power

Shaahin Hessabi

Department of Computer Engineering

Sharif University of Technology

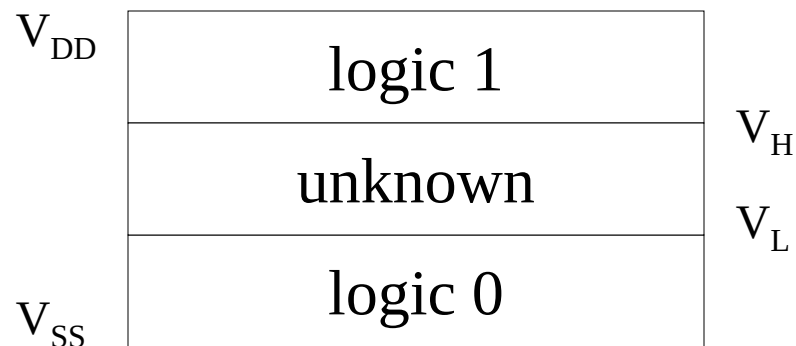
Adapted, with modifications, from lecture notes prepared by the author
(from Prentice Hall PTR)

Topics

- Electrical properties of static combinational gates:
 - transfer characteristics;
 - delay;
 - power.
- Effects of parasitics on gate.
- Driving large loads.

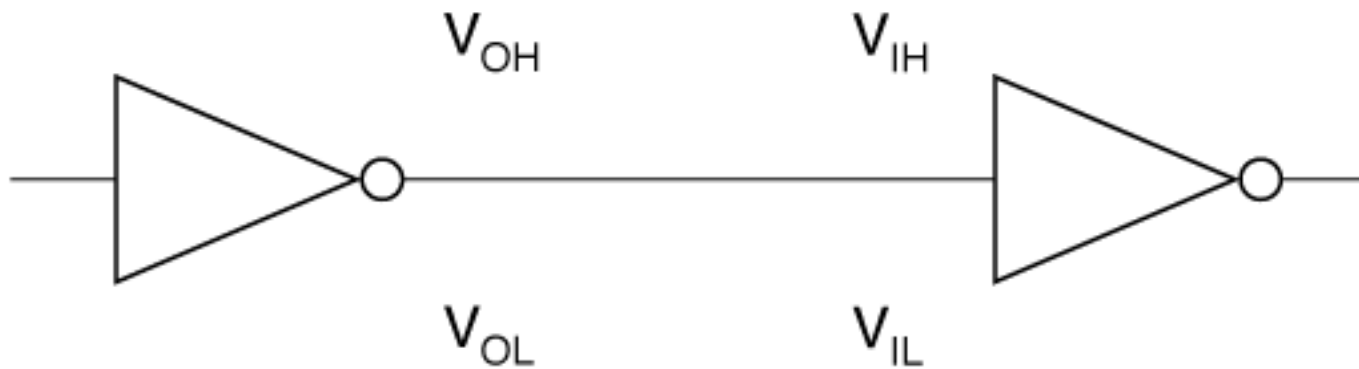
Logic levels

- Solid logic 0/1 defined by V_{SS}/V_{DD} .
- Inner bounds of logic values V_L/V_H are not directly determined by circuit properties, as in some other logic families.



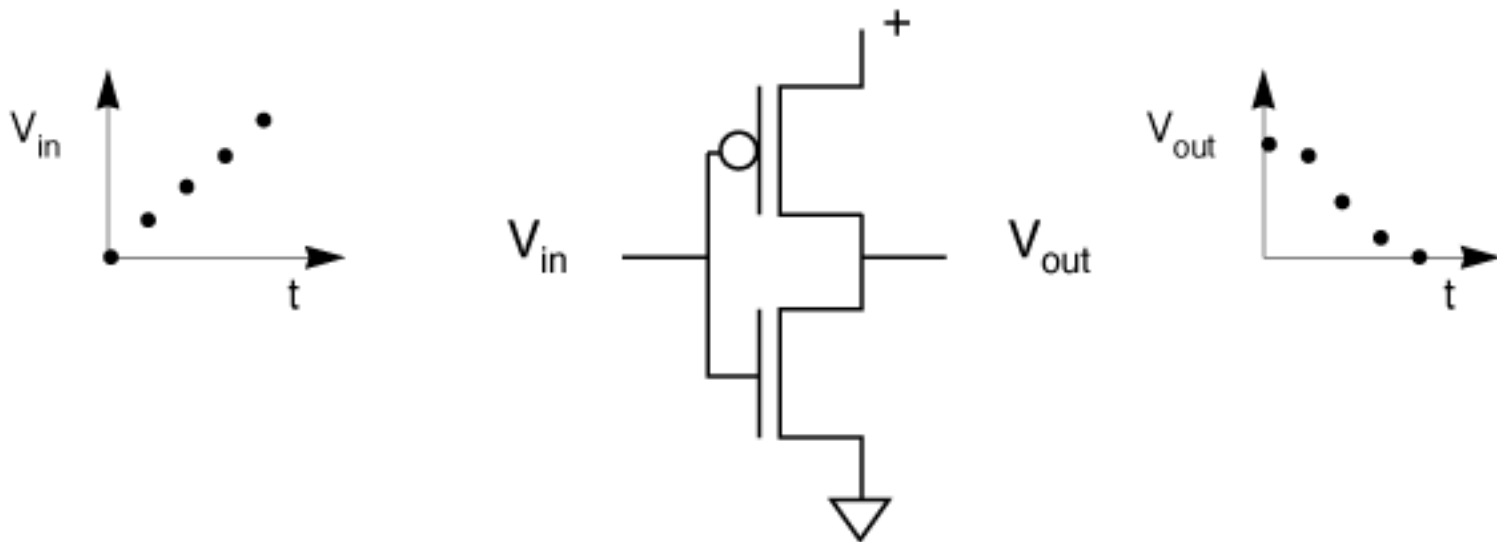
Logic level matching

- Levels at output of one gate must be sufficient to drive next gate.

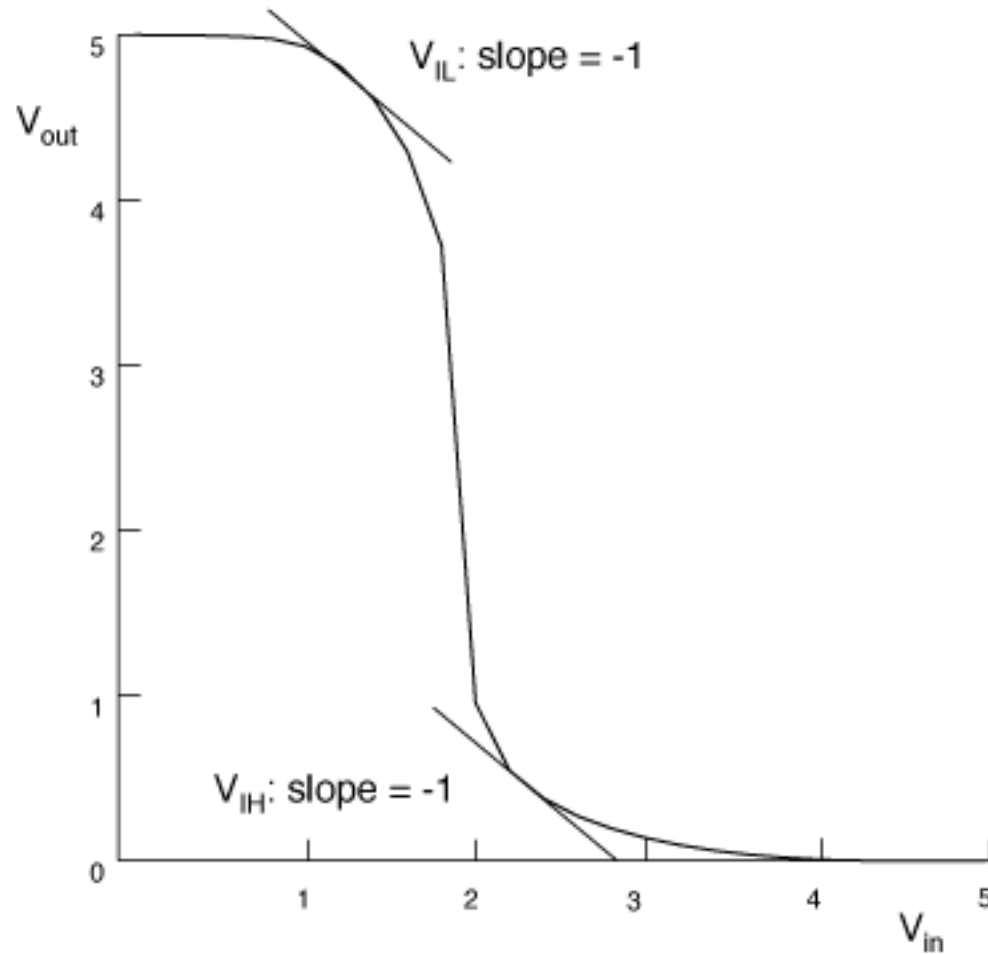


Transfer characteristics

- Transfer curve shows static input/output relationship; hold input voltage, measure output voltage.



Inverter transfer curve



Logic thresholds

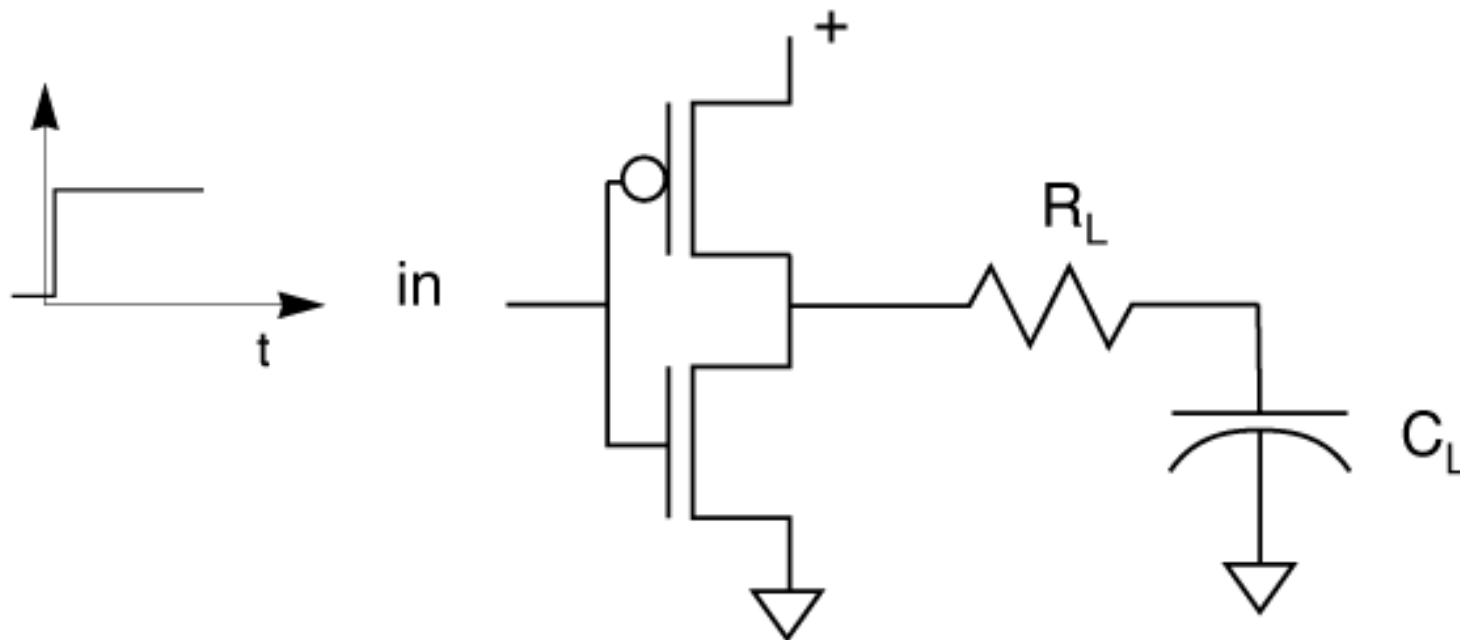
- Choose threshold voltages at points where slope of transfer curve = -1.
- Inverter has a high gain between V_{IL} and V_{IH} points, low gain at outer regions of transfer curve.
- Note that logic 0 and 1 regions are not equal sized; in this case, high pullup resistance leads to smaller logic 1 range.

Noise margin

- Noise margin = voltage difference between output of one gate and input of next. Noise must exceed noise margin to make second gate produce wrong output.
- In static gates, $t=\infty$ voltages are V_{DD} and V_{SS} , so noise margins are $V_{DD}-V_{IH}$ and $V_{IL}-V_{SS}$.

Delay

- Assume ideal input (step), RC load.

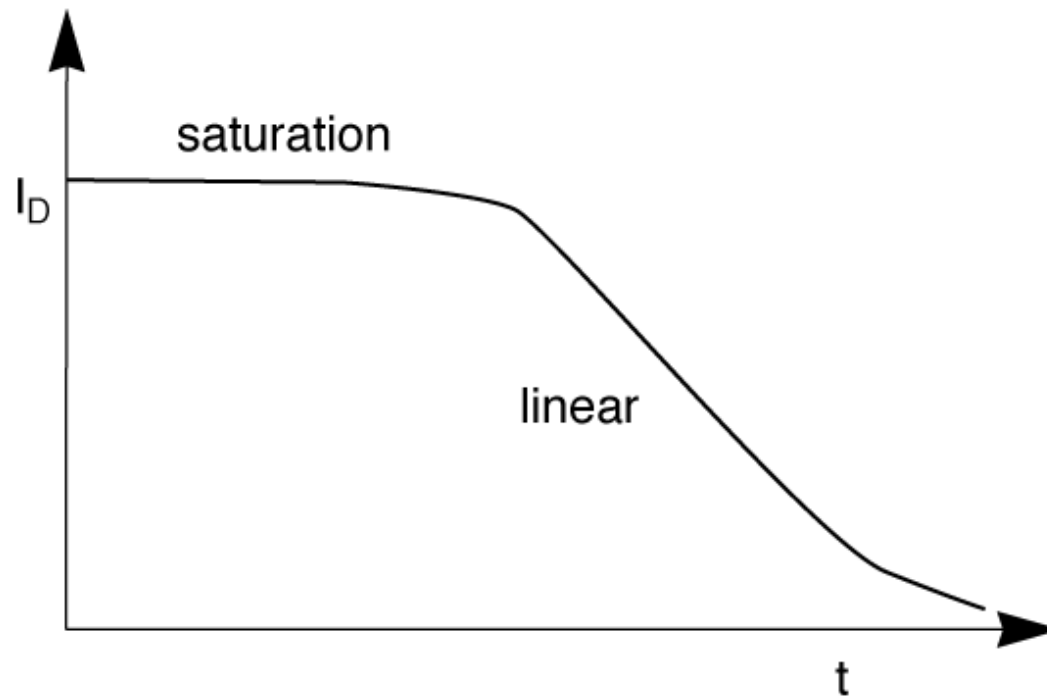


Delay assumptions

- Assume that only one transistor is ON at a time.
This gives two cases:
 - rise time, pullup on;
 - fall time, pullup off.
- Assume resistor model for transistor.
 - Ignores saturation region and mischaracterizes linear region, but results are acceptable!

Current through transistor

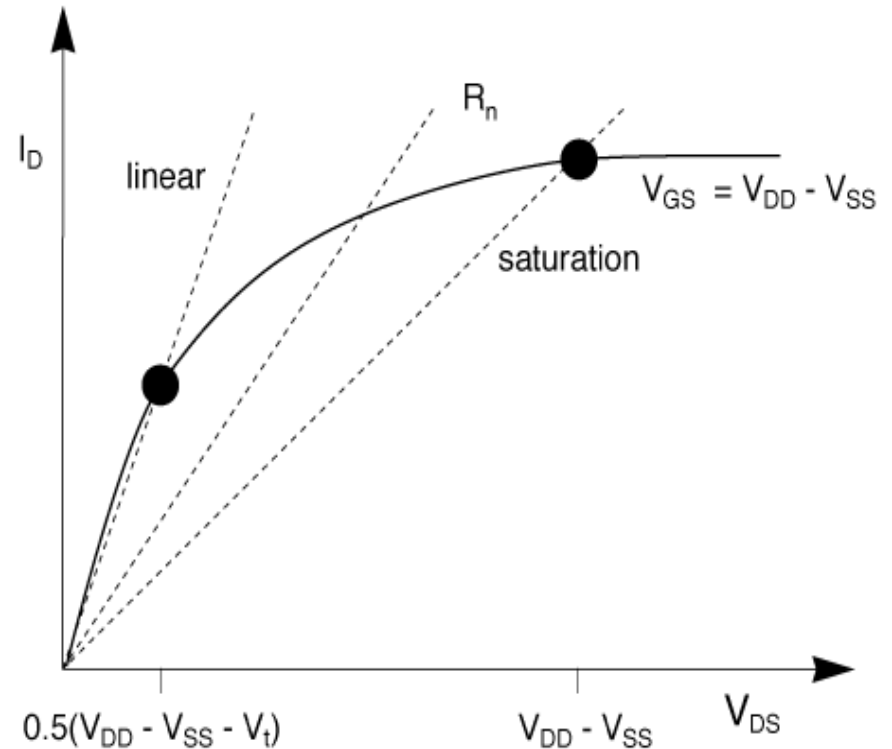
- Transistor starts in saturation region, then moves to linear region.



Resistive model for transistor

■ Average V/I at two voltages:

- maximum output voltage
- middle of linear region



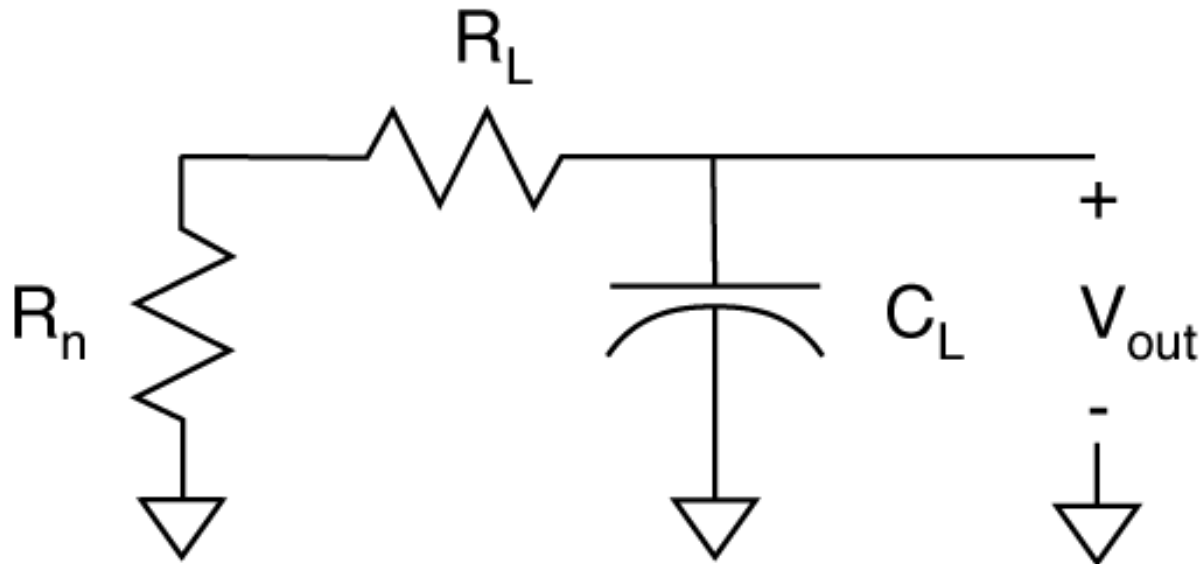
- ## ■ Voltage is V_{ds} , current is given I_d at that drain voltage. Step input means that $V_{gs} = V_{DD}$ always.

Ways of measuring gate delay

- Delay: time required for gate's output to reach 50% of final value.
- Transition time: time required for gate's output to reach 10% (logic 0) or 90% (logic 1) of final value.

Inverter delay circuit

- Load is resistor + capacitor, driver is resistor.



Inverter delay with τ model

- τ model: gate delay based on RC time constant τ .
- $V_{out}(t) = V_{DD} \cdot e^{-t/[(R_n + R_L) \cdot C_L]}$
- $t_f = 2.2 R C_L$
- For pullup time, use pullup resistance.

τ model inverter delay

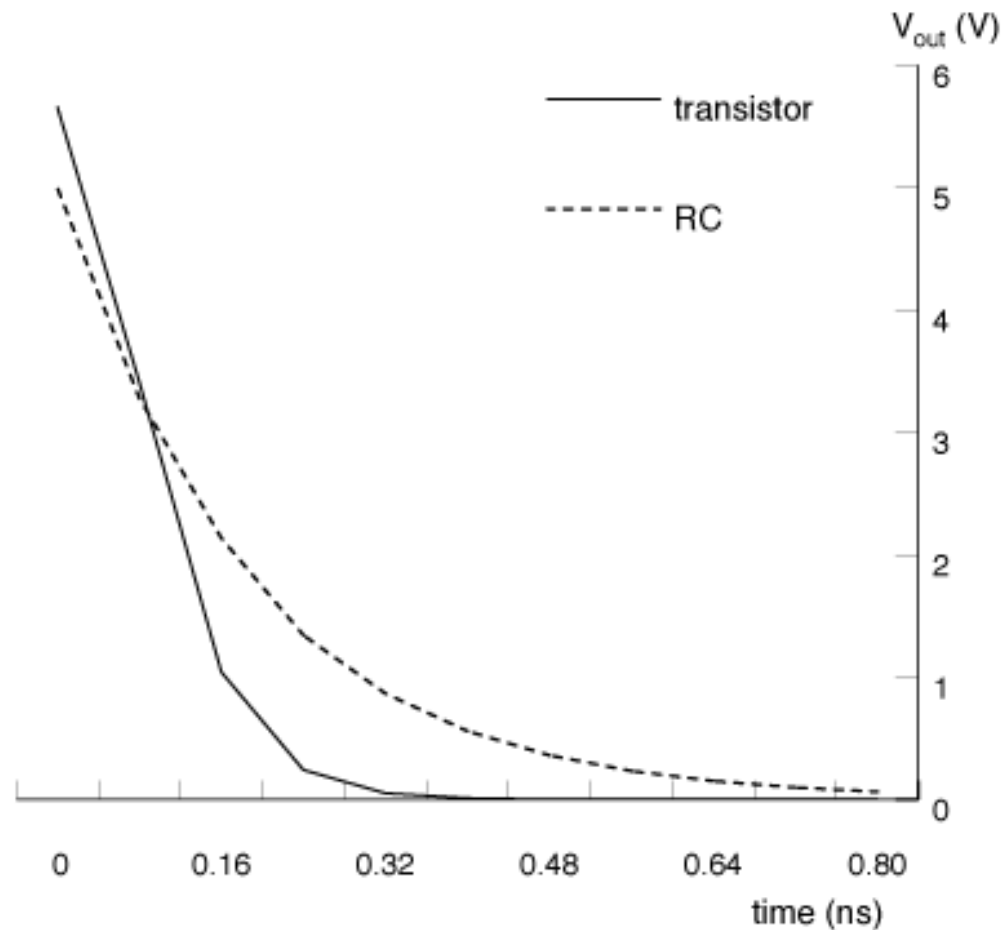
■ 0.5 micron process:

- $R_n = 3.9 \text{ k}\Omega$
- $C_l = 0.68 \text{ fF}$

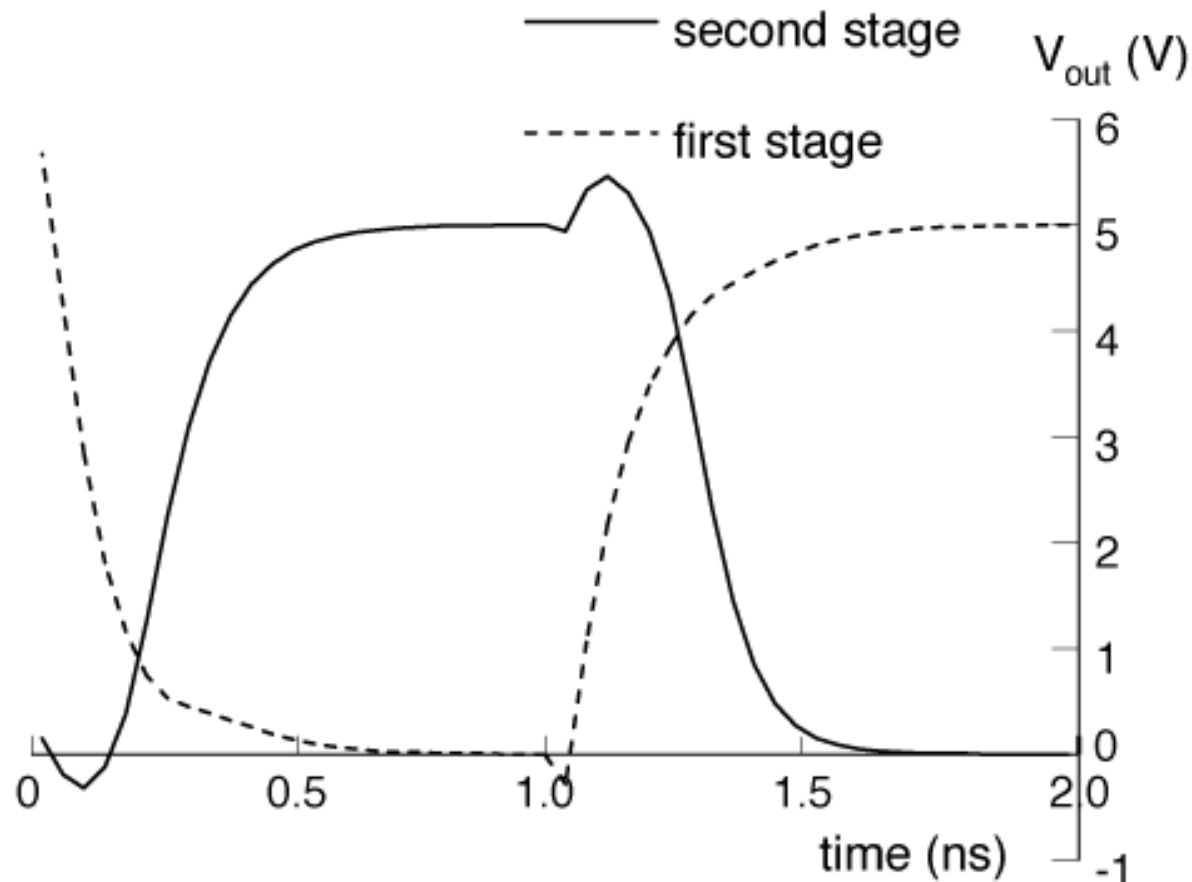
■ So

- $t_d = 0.69 \times 3.9 \times .68 = 1.8 \text{ ps.}$
- $t_f = 2.2 \times 3.9 \times .68 = 5.8 \text{ ps.}$

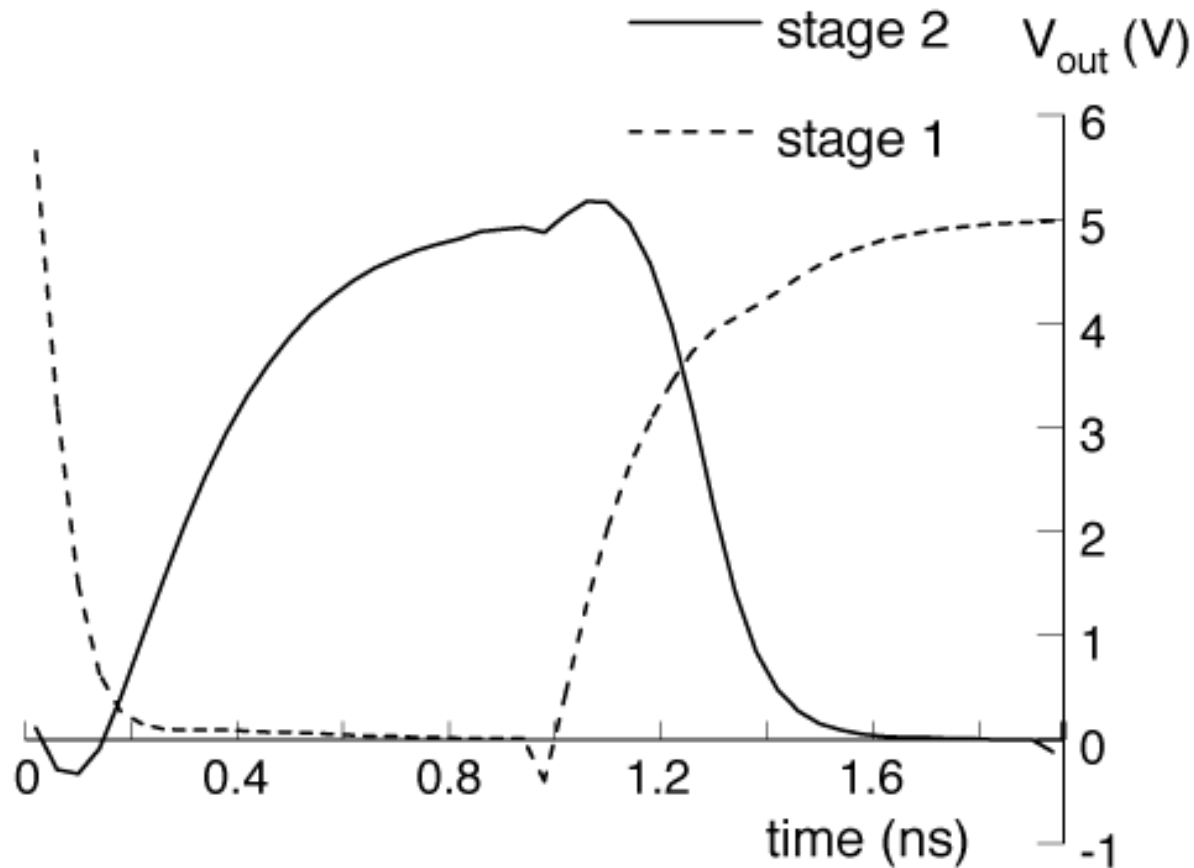
Quality of RC approximation



Quality of step input approximation



Results of using small pullup

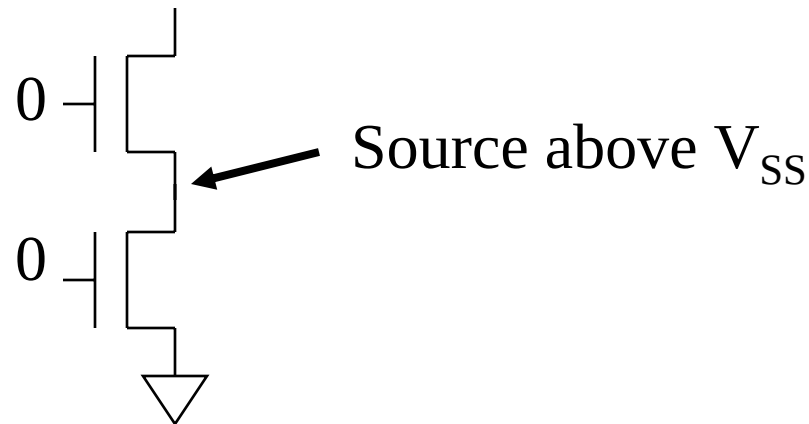


Other models

- Current source model (used in power/delay studies):
 - $t_f = C_L (V_{DD} - V_{SS}) / I_d$
 - $= C_L (V_{DD} - V_{SS}) / 0.5 k' (W/L) (V_{DD} - V_{SS} - V_t)^2$
- Fitted model: fit curve to measured circuit characteristics.

Body effect and gates

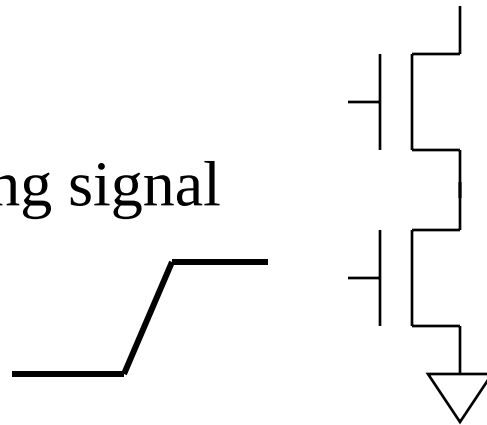
- Difference between source and substrate voltages causes body effect.
- Source for gates in middle of network may not equal substrate:



Body effect and gate input ordering

- To minimize body effect, put early arriving signals at transistors closest to power supply:

Early arriving signal

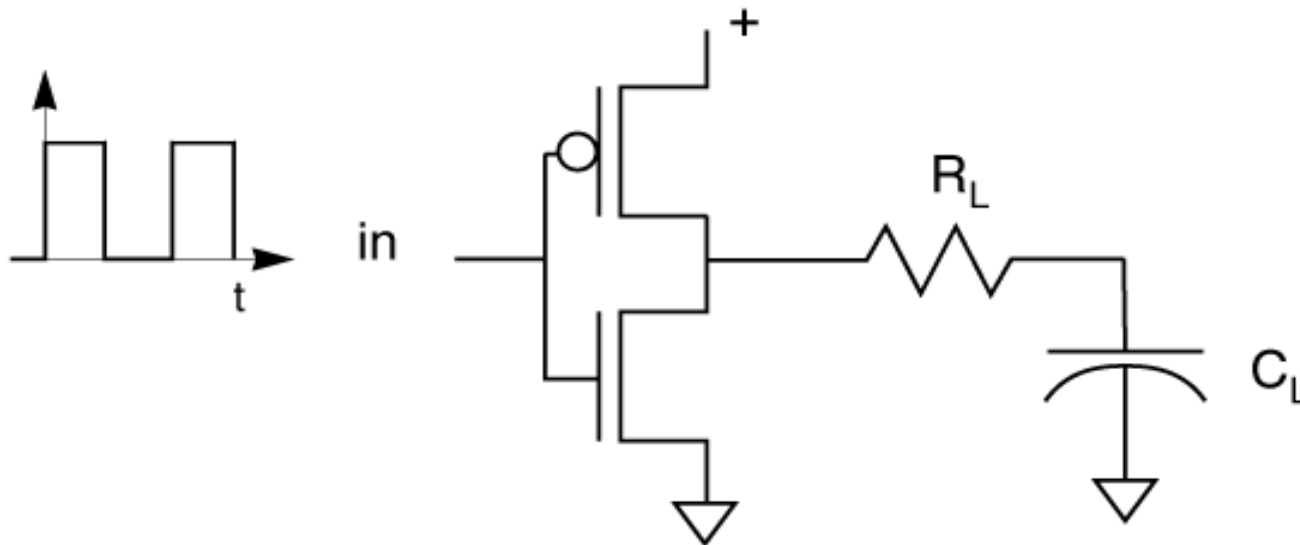


Power consumption analysis

- Almost all power consumption comes from switching behavior.
- Static power dissipation comes from leakage currents (few pA),
 - This small power is important in some very low power devices, such as implantable medical electronics, telephone chips that obtain power through the phone line,
- Surprising result: power consumption is independent of the sizes of the pullups and pulldowns.

Power consumption circuit

- Input is square wave, hence, only one transistor is ON at a time (the other has a current of a few pico Amps).



Power consumption

- A single cycle requires one charge and one discharge of capacitor: $E = C_L(V_{DD} - V_{SS})^2$.
 - Where is this energy dissipated?
- Clock frequency $f = 1/t$.
- Resistance of pullup/pulldown drops out of energy calculation.
 - It's logical, since it is *dynamic* power.
 - Size of transistors has indirect effect on power, since it determines C_L .

Observations on Power consumption

- The current through transistors (dependent on the size) determines the maximum speed at which the circuit can run, which indirectly determines power consumption.
- $\text{Power} = E f = f C_L (V_{DD} - V_{SS})^2.$
- In CMOS, power consumption depends on the frequency at which the circuit is operating.
 - in most other technologies (e.g., NMOS or bipolar) most of the power is consumed while the circuit is idle.
 - Slower-running circuits use less power (but not less energy to perform the same computation).

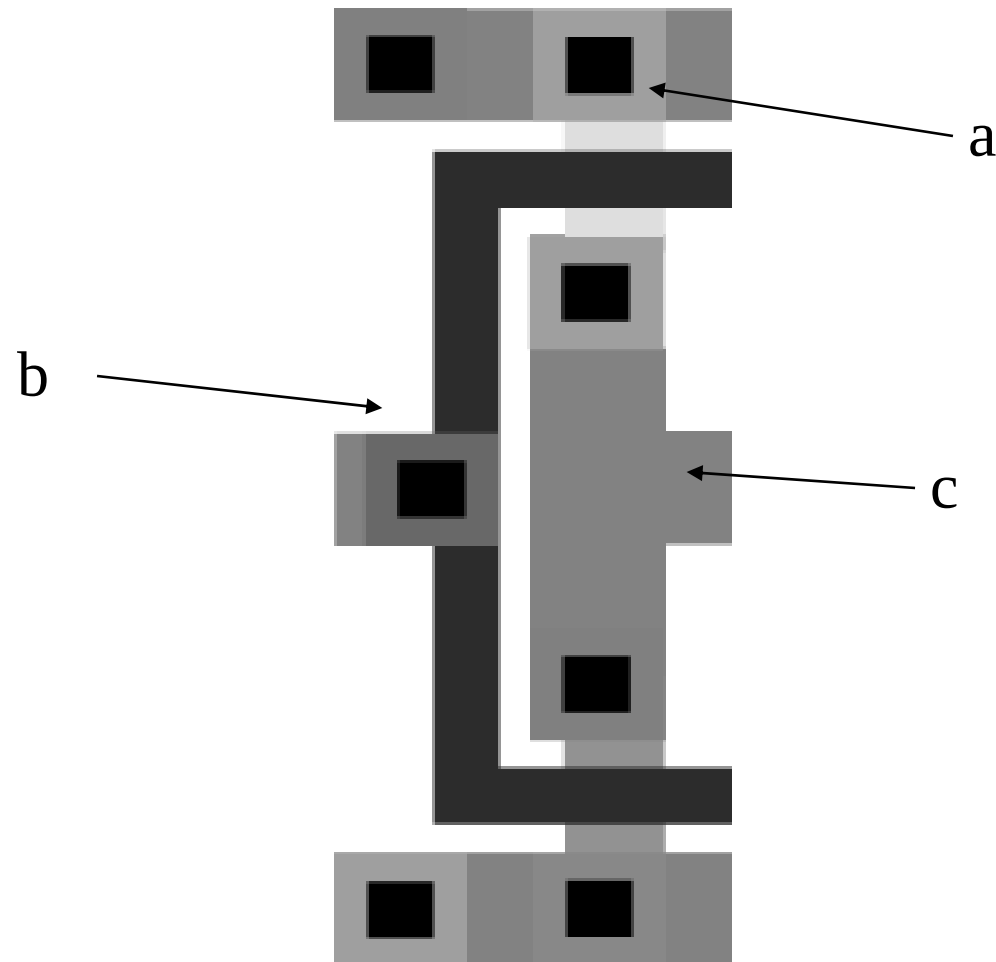
Speed-power product

- Also known as power-delay product.
- Helps measure quality of a logic family.
- For static CMOS:
 - $SP = P/f = CV^2$.
- Static CMOS speed-power product is independent of operating frequency.
 - Voltage scaling depends on this fact.

Speed-power product (cont'd)

- If we scale the supply voltage from V to V' :
 - P'/P is proportional to $(V'/V)^2$
 - t'_r/t_r is proportional to (V/V') [since R depends on V]
 - By reducing supply voltage, power consumption goes down faster than does speed.

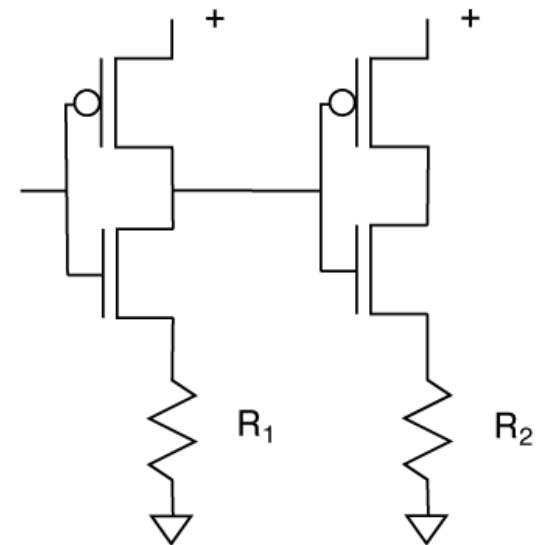
Parasitics and performance



Effect of parasitics

■ Node a:

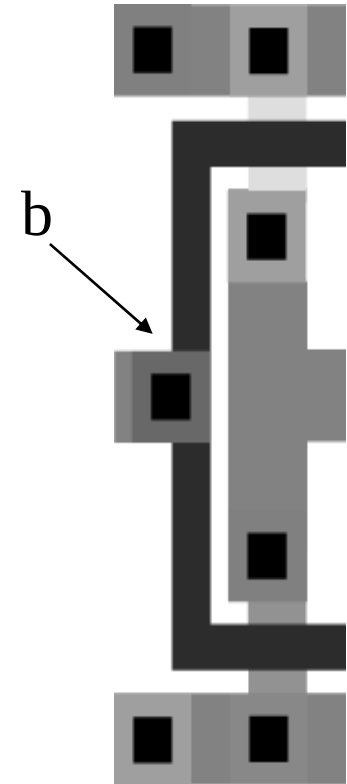
- Capacitance on power supply is not bad, can be good in absence of inductance.
- Resistance slows down static gates, may cause pseudo-nMOS circuits to fail.
- Layout around point *a* should be designed to minimize R:
 - power lines should be kept in metal as long as possible.
 - If the diffusion wire is wider than a via, use several parallel vias to connect the metal and diffusion lines.



Effects of parasitics, cont'd

■ Node b:

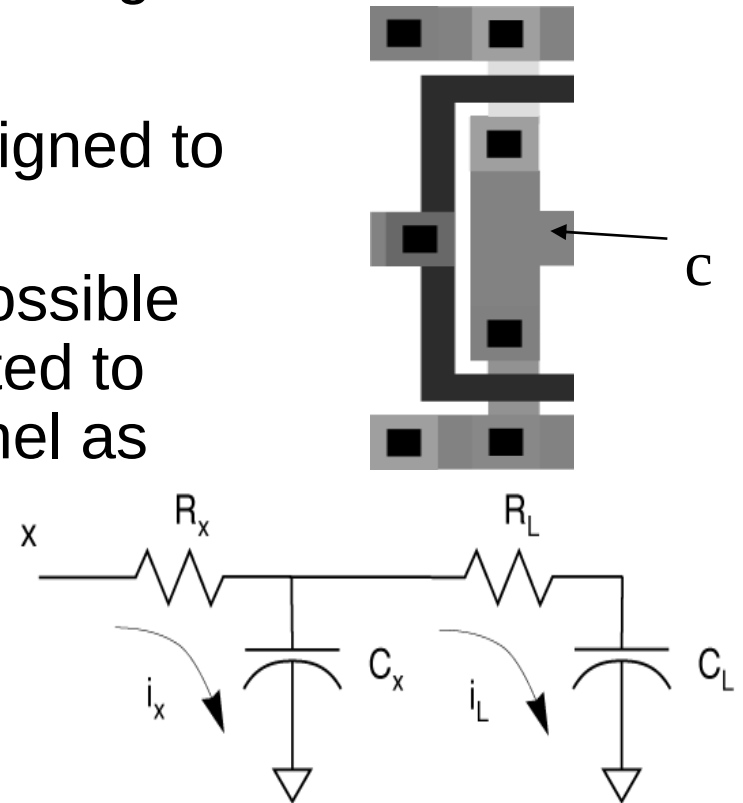
- Increasing capacitance/resistance reduces input slope (adds to the load of the driver).
- Gate capacitances are much larger than the capacitance added by this short wire.
- Avoid making big mistakes by using large sections of diffusion wire or a single via to connect high-current wires.



Effects of parasitics, cont'd

■ Node c:

- Similar to parasitics at b, but resistance near source is more damaging, since it must charge more capacitance.
- Layout around c should be designed to minimize resistance:
 - Using as little diffusion as possible (diffusion should be connected to metal/poly as close to channel as possible).
 - Using parallel vias at the diffusion/metal interface to minimize resistance.



Driving large loads

- Sometimes, large loads must be driven:
 - off-chip;
 - long wires on-chip (bus).
 - clock wires which go to many points.
- This increases the delay.
- Sizing up the driver transistors (to increase current) only pushes back the problem; driver now presents larger capacitance to earlier stage.

Cascaded driver circuit

- Use a chain of inverters, each stage has transistors α times larger than previous stage.
- The aim of optimization is to minimize the total delay through the inverter chain.
- Each inverter can produce α times more current than the previous stage: $\alpha = (C_{\text{big}}/C_g)^{1/n}$
- $t_{\text{total}} = n (C_{\text{big}}/C_g)^{1/n} t_{\text{min}}$; t_{min} : time to drive a min-size load.
- Optimal number of stages is found by differentiating:
 $n_{\text{opt}} = \ln(C_{\text{big}}/C_g)$. Choose an integer value larger than n_{opt}
- The optimum value is at $\alpha=e$. Driver sizes are exponentially tapered with size ratio α .

