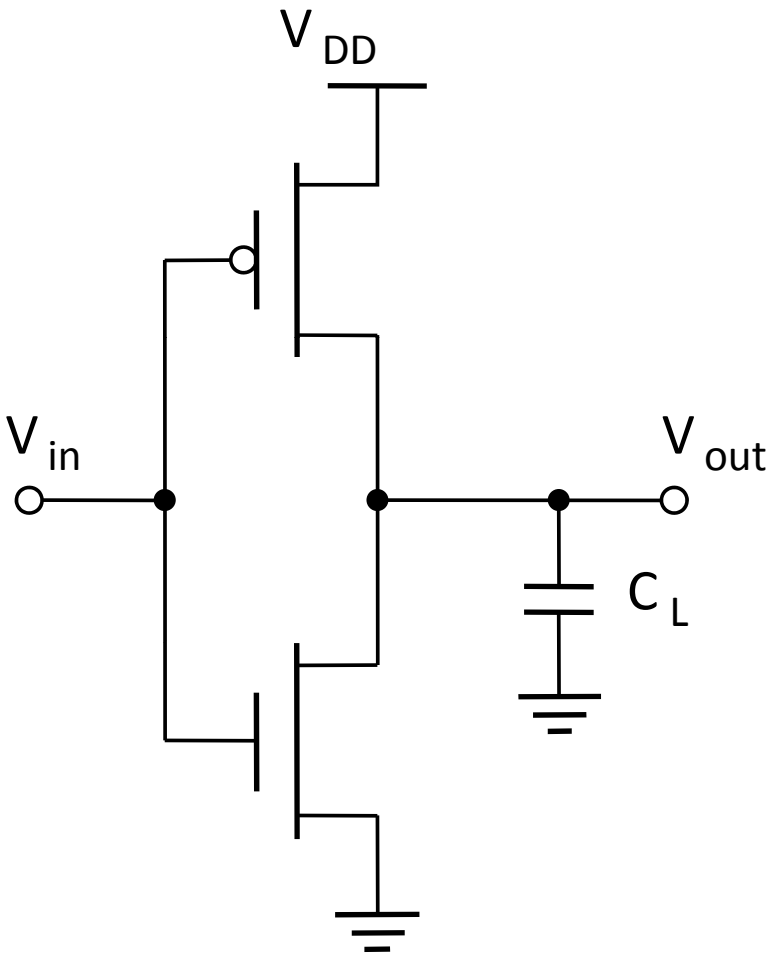
ECE680: Physical VLSI Design

Chapter III

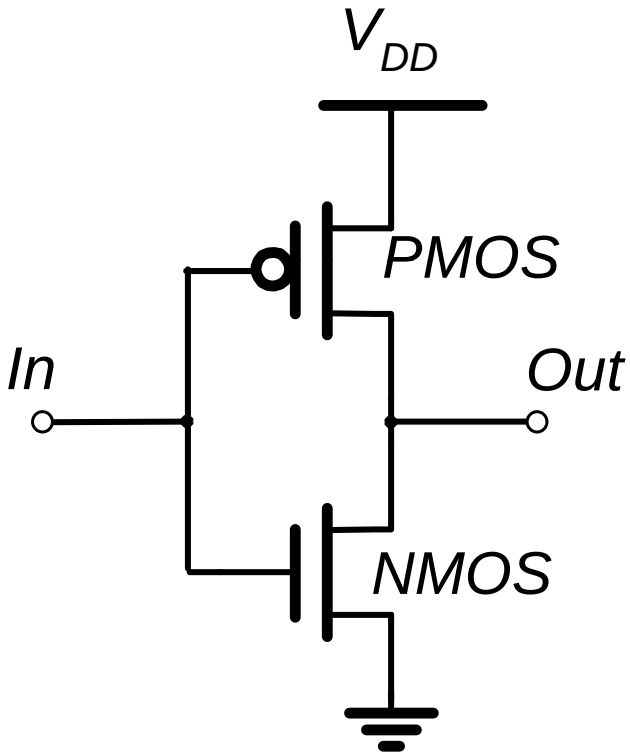
CMOS Device, Inverter, Combinational circuit Logic and Layout

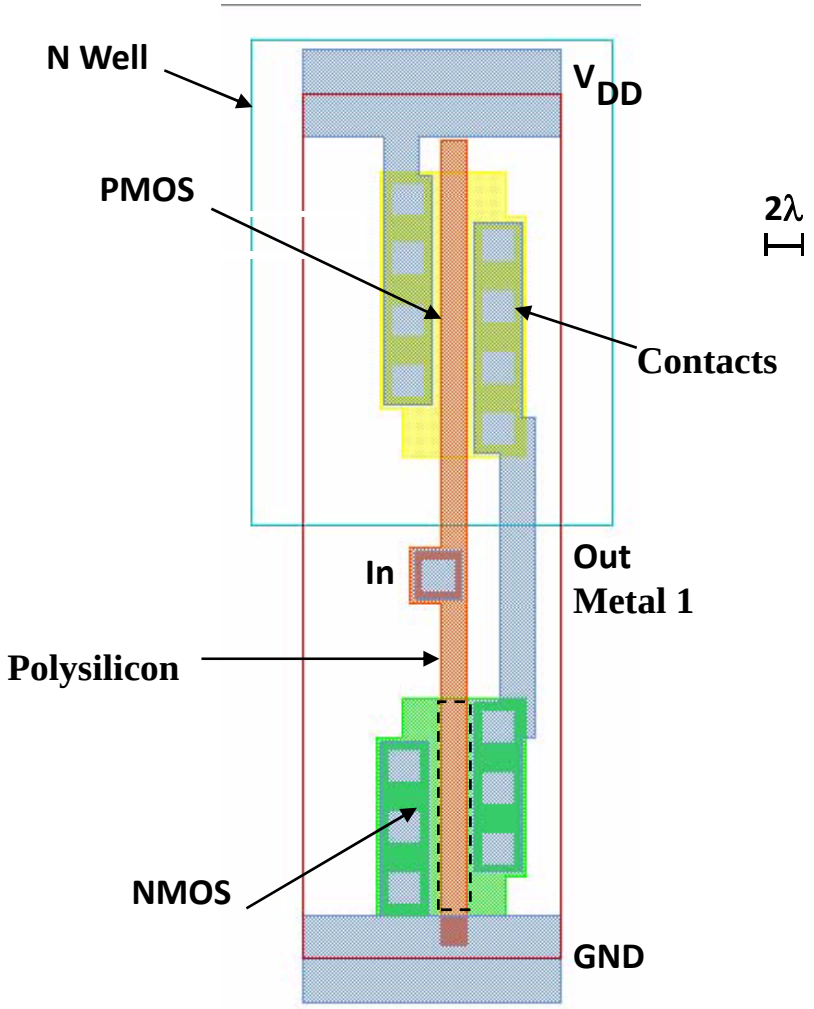
Part 2 CMOS Inverter

The CMOS Inverter: A First Glance



CMOS Inverter

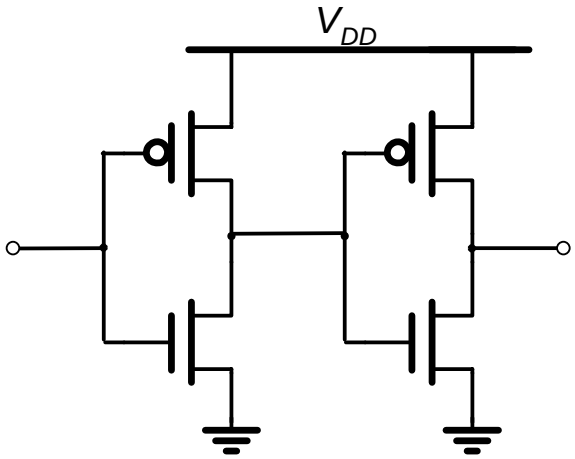


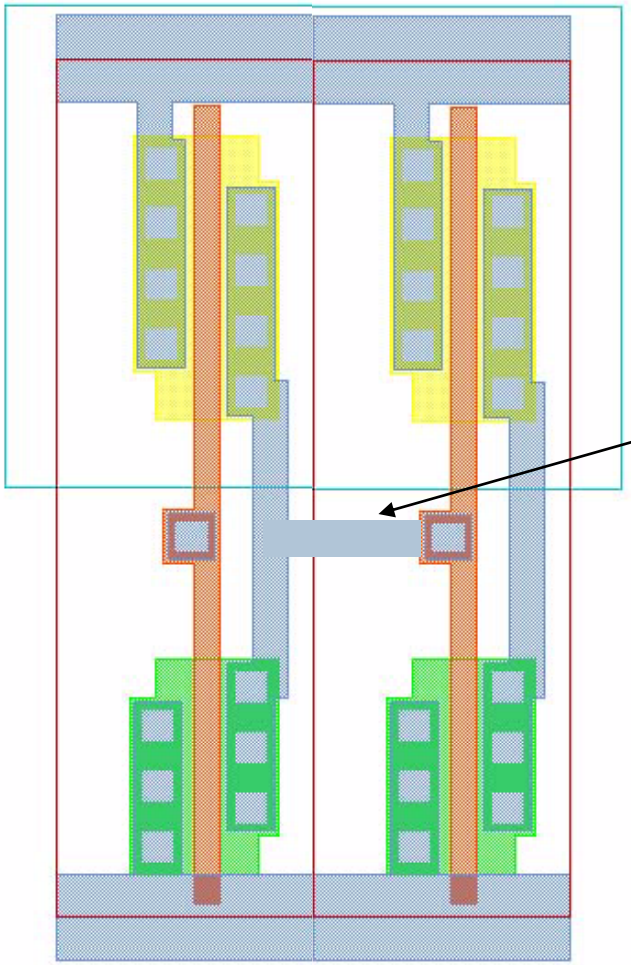


Two Inverters

Share power and ground

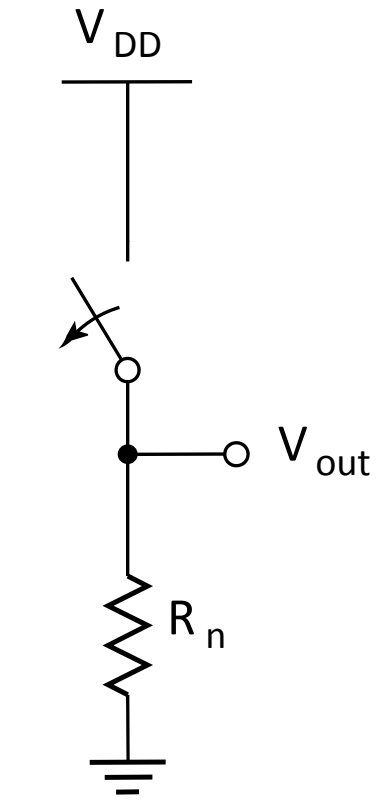
Abut cells



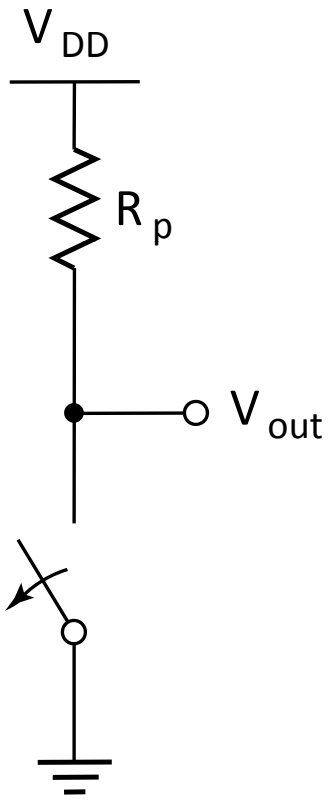


Connect in Metal

CMOS Inverter
First-Order DC Analysis



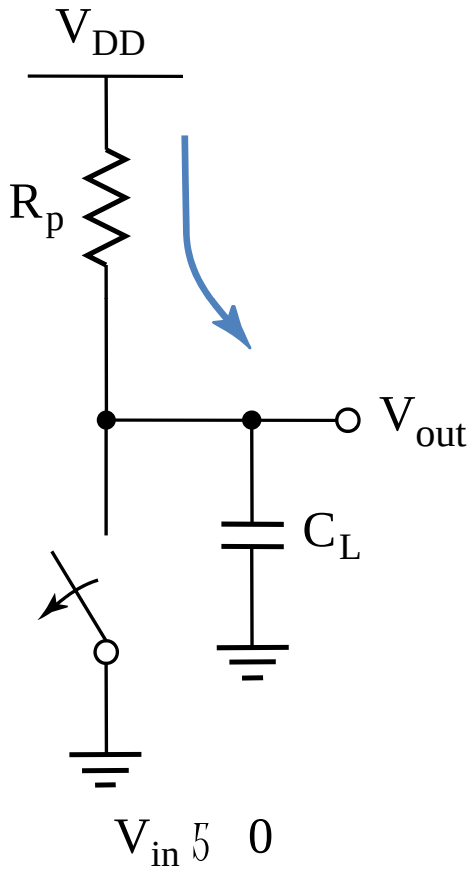
$V_{in} = V_{DD}$



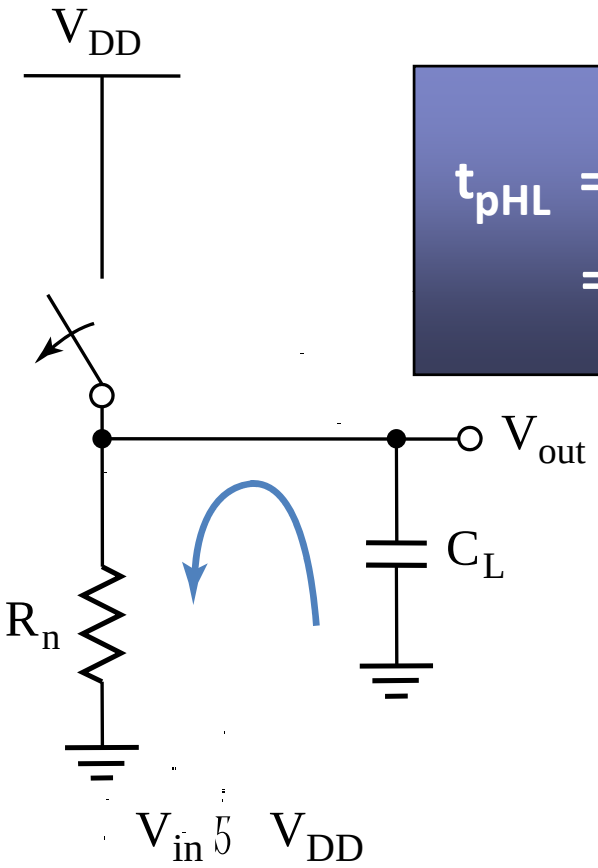
$V_{in} = 0$

$V_{OL} = 0$
$V_{OH} = V_{DD}$
$V_M = f(R_n, R_p)$

CMOS Inverter: Transient Response



(a) Low-to-high

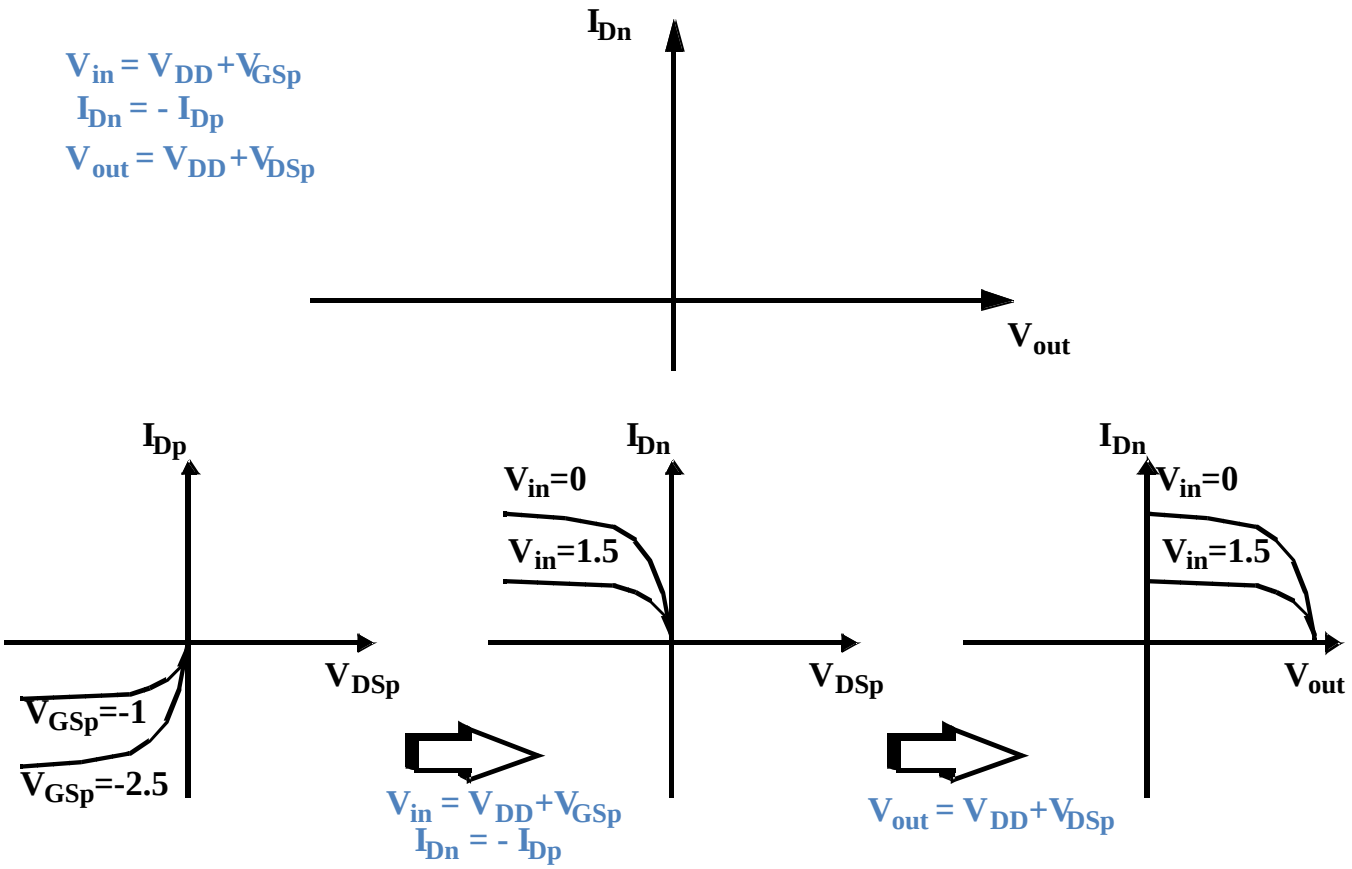


(b) High-to-low

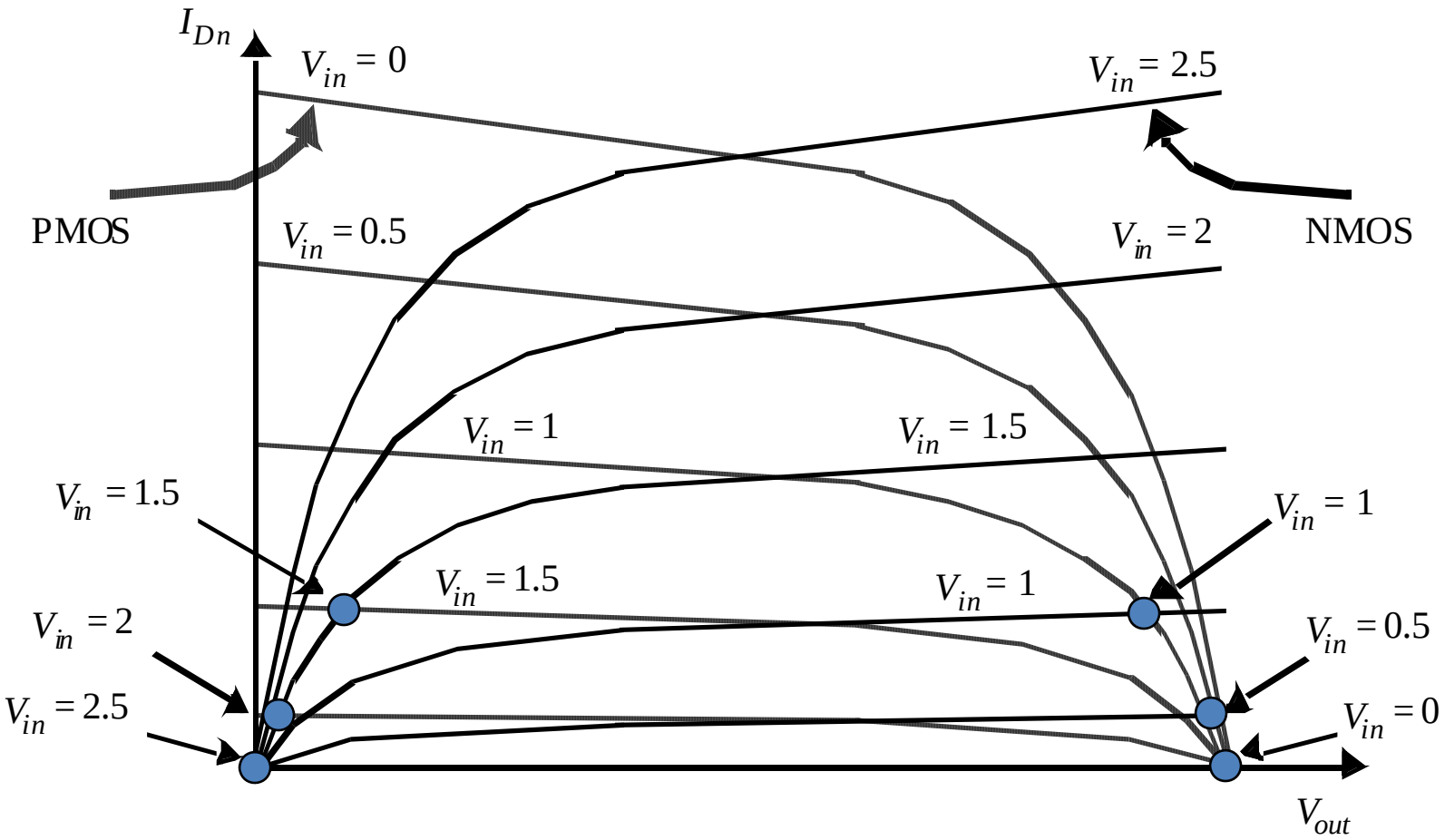
$t_{pHL} = f(R_{on}.C_L)$

$= 0.69\ R_{on} C_L$

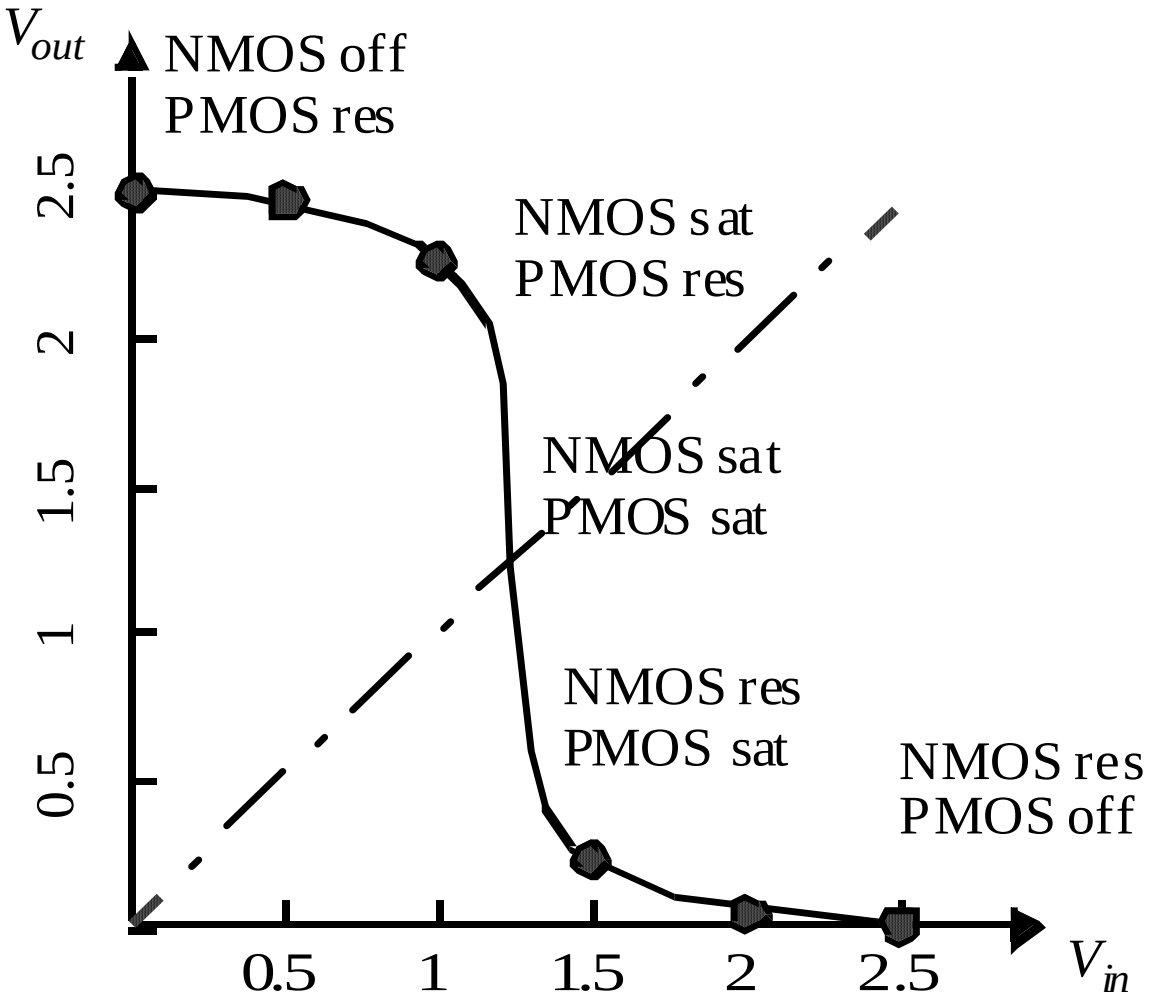
1. Voltage Transfer Characteristic
PMOS Load Lines



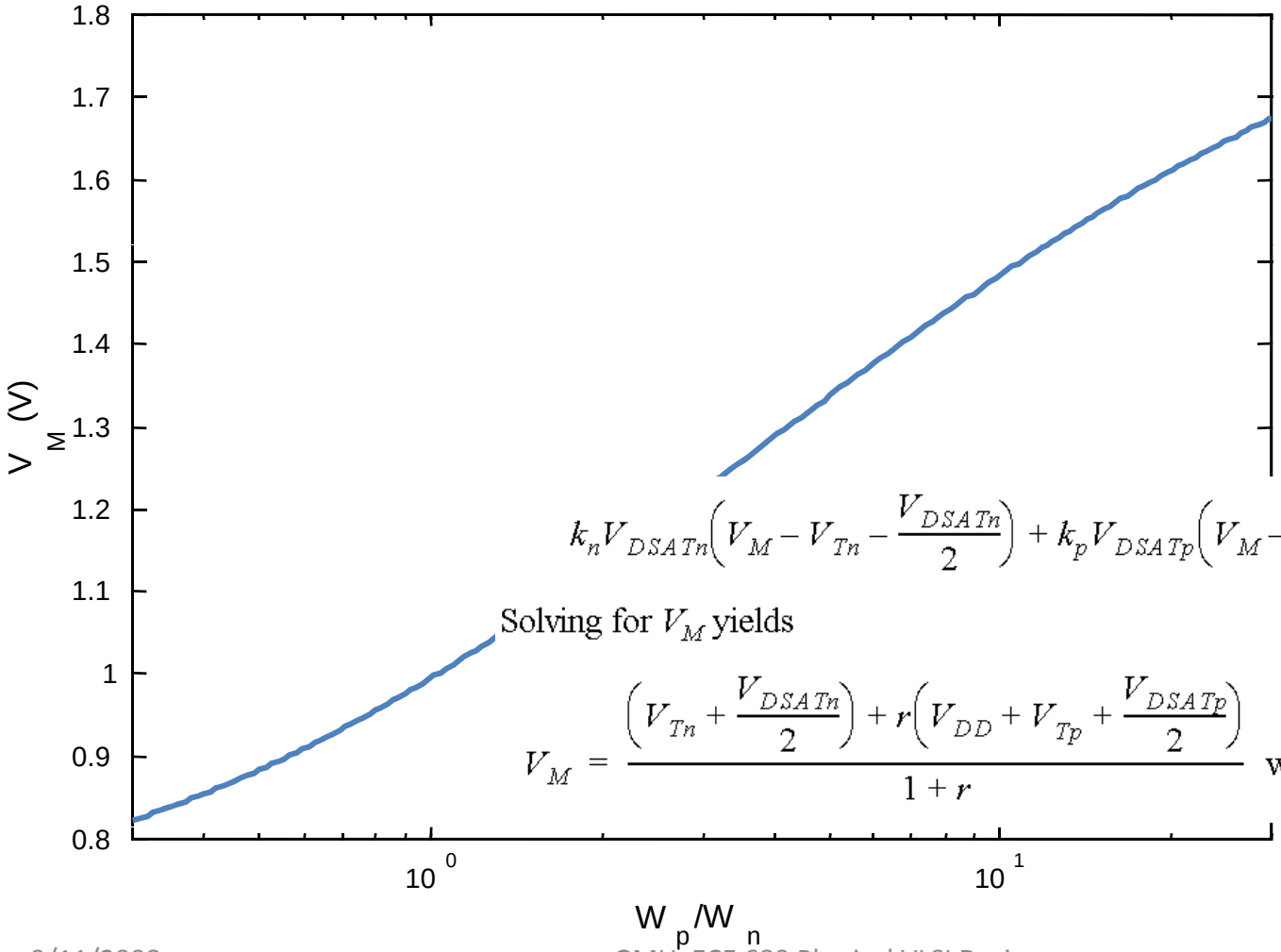
CMOS Inverter Load Characteristics



CMOS Inverter VTC



Switching Threshold as a function of Transistor Ratio

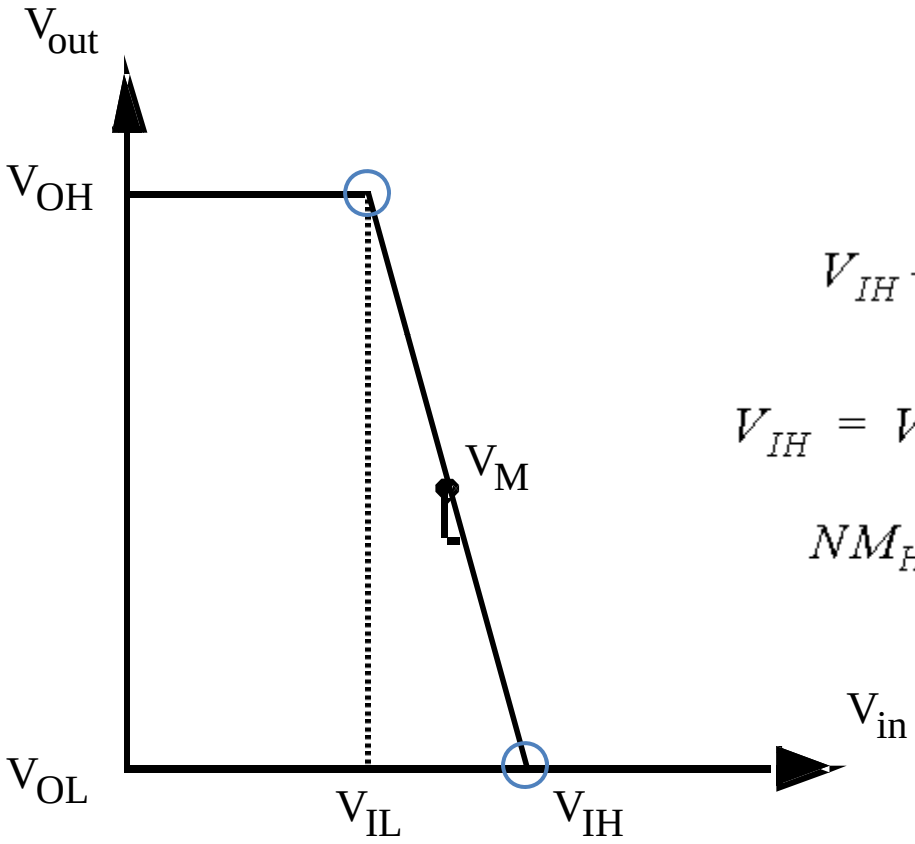


$$k_n V_{DSATn}\left(V_M - V_{Tn} - \frac{V_{DSATn}}{2}\right) + k_p V_{DSATp}\left(V_M - V_{DD} - V_{Tp} - \frac{V_{DSATp}}{2}\right) = 0$$

Solving for V_M yields

$$V_M = \frac{\left(V_{Tn} + \frac{V_{DSATn}}{2}\right) + r\left(V_{DD} + V_{Tp} + \frac{V_{DSATp}}{2}\right)}{1 + r} \quad \text{with} \quad r = \frac{k_p V_{DSATp}}{k_n V_{DSATn}} = \frac{\upsilon_{satp} W_p}{\upsilon_{satn} W_n}$$

Determining V_{IH} and V_{IL}



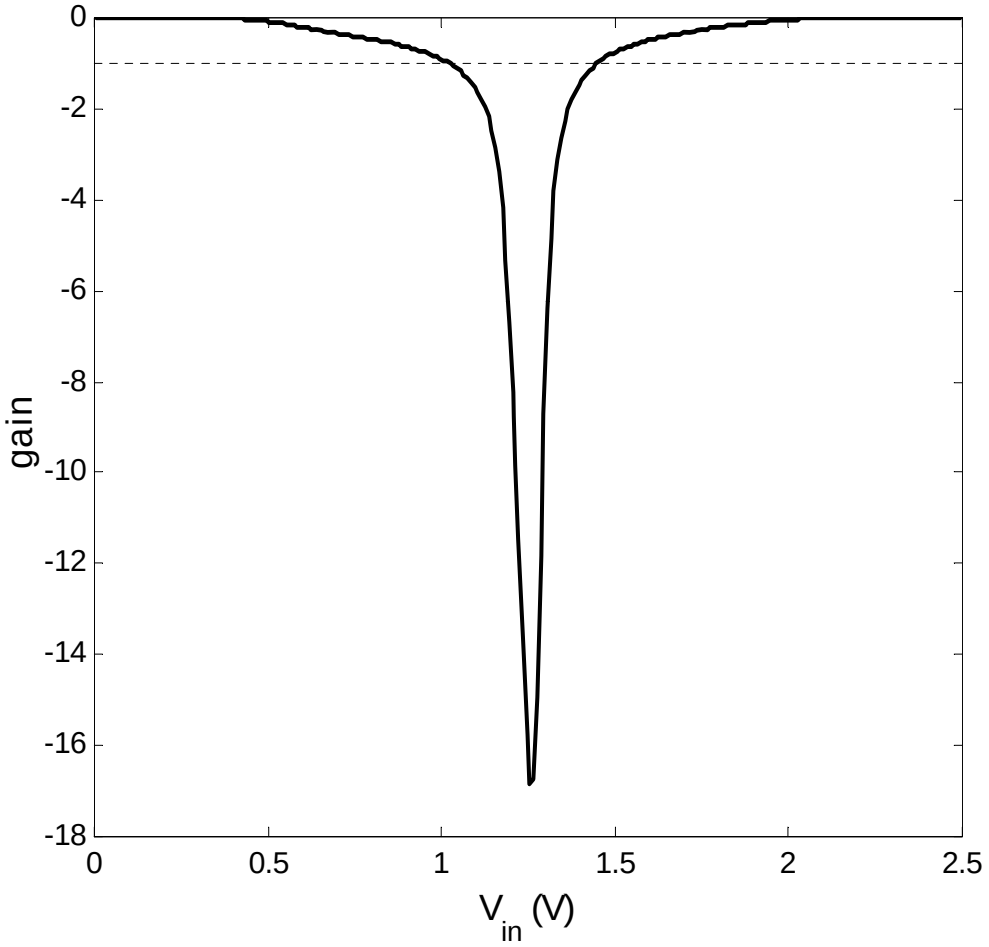
$$V_{IH} - V_{IL} = -\frac{(V_{OH} - V_{OL})}{g} = \frac{-V_{DD}}{g}$$

$$V_{IH} = V_M - \frac{V_M}{g} \qquad V_{IL} = V_M + \frac{V_{DD} - V_M}{g}$$

$$NM_H = V_{DD} - V_{IH} \qquad NM_L = V_{IL}$$

A simplified approach

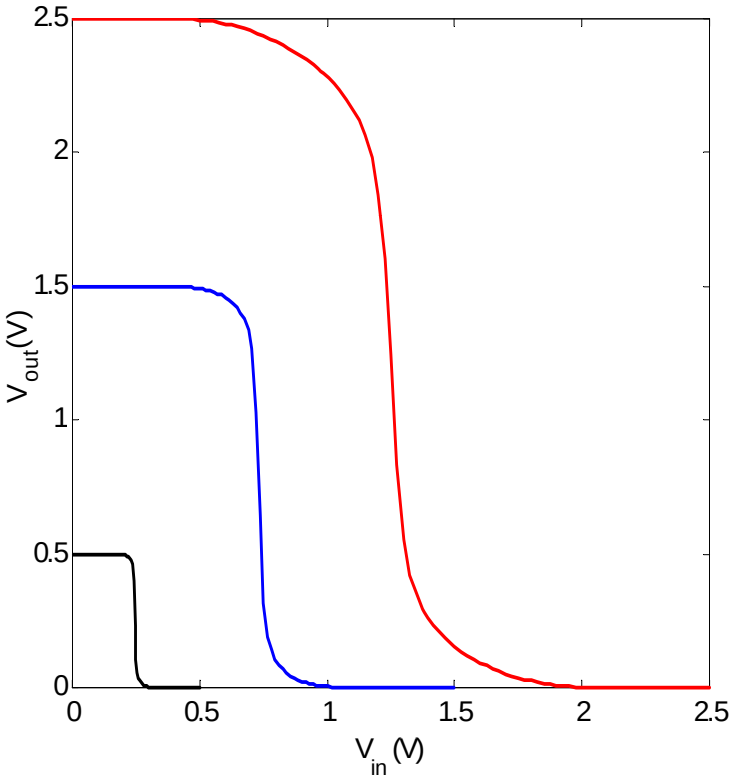
Inverter Gain

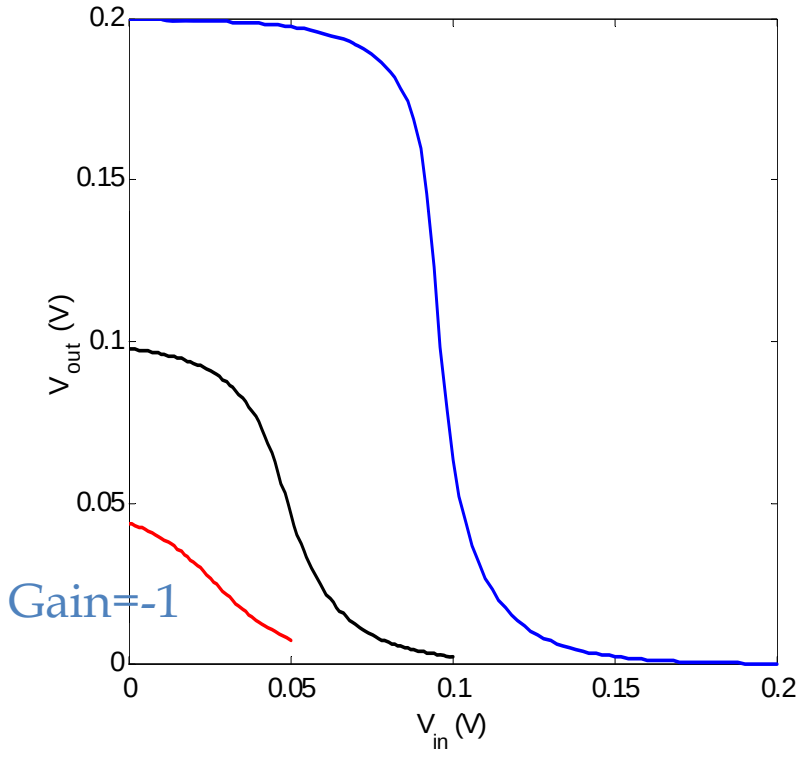


$$g = -\frac{1}{I_D(V_M)} \frac{k_n V_{DSATn} + k_p V_{DSATp}}{\lambda_n - \lambda_p}$$

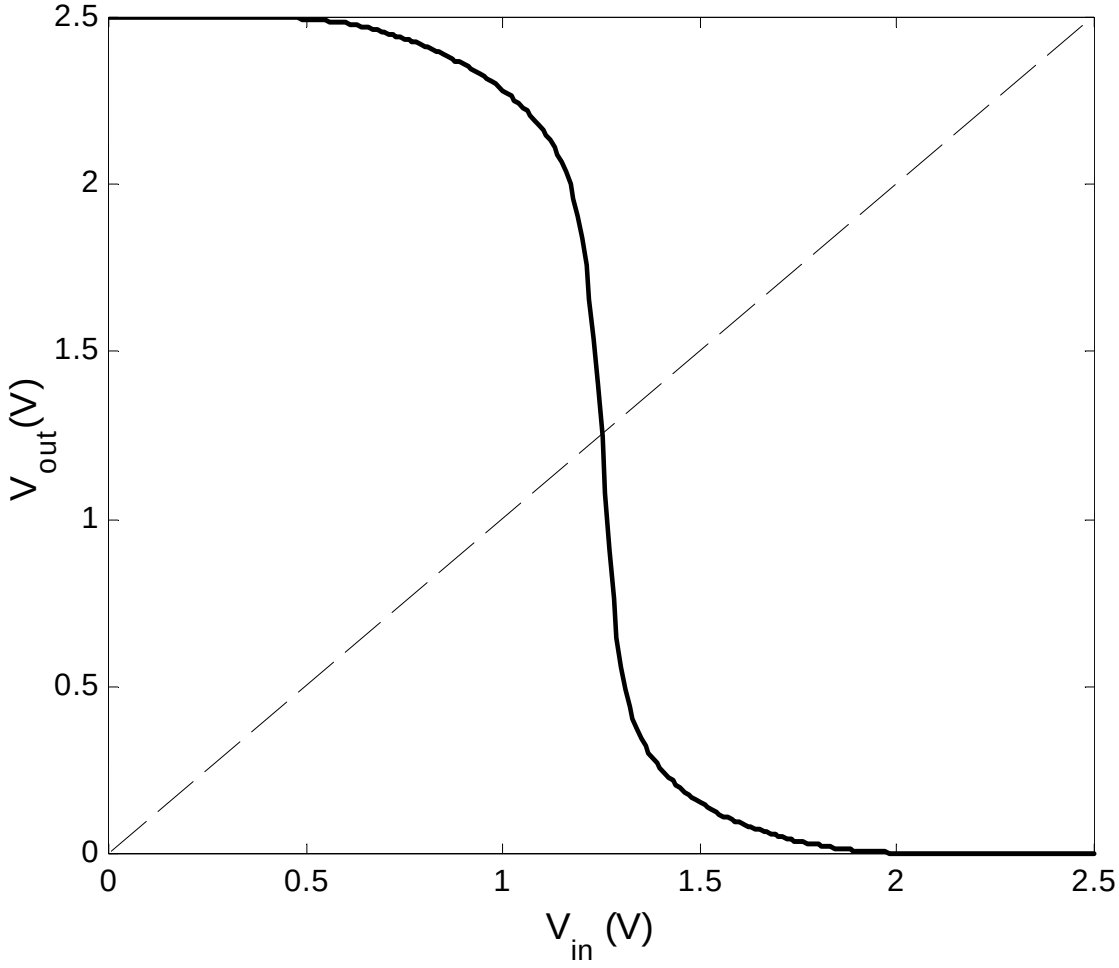
$$\approx \frac{1+r}{(V_M - V_{Tn} - V_{DSATn}/2)(\lambda_n - \lambda_p)}$$

Gain as a function of VDD

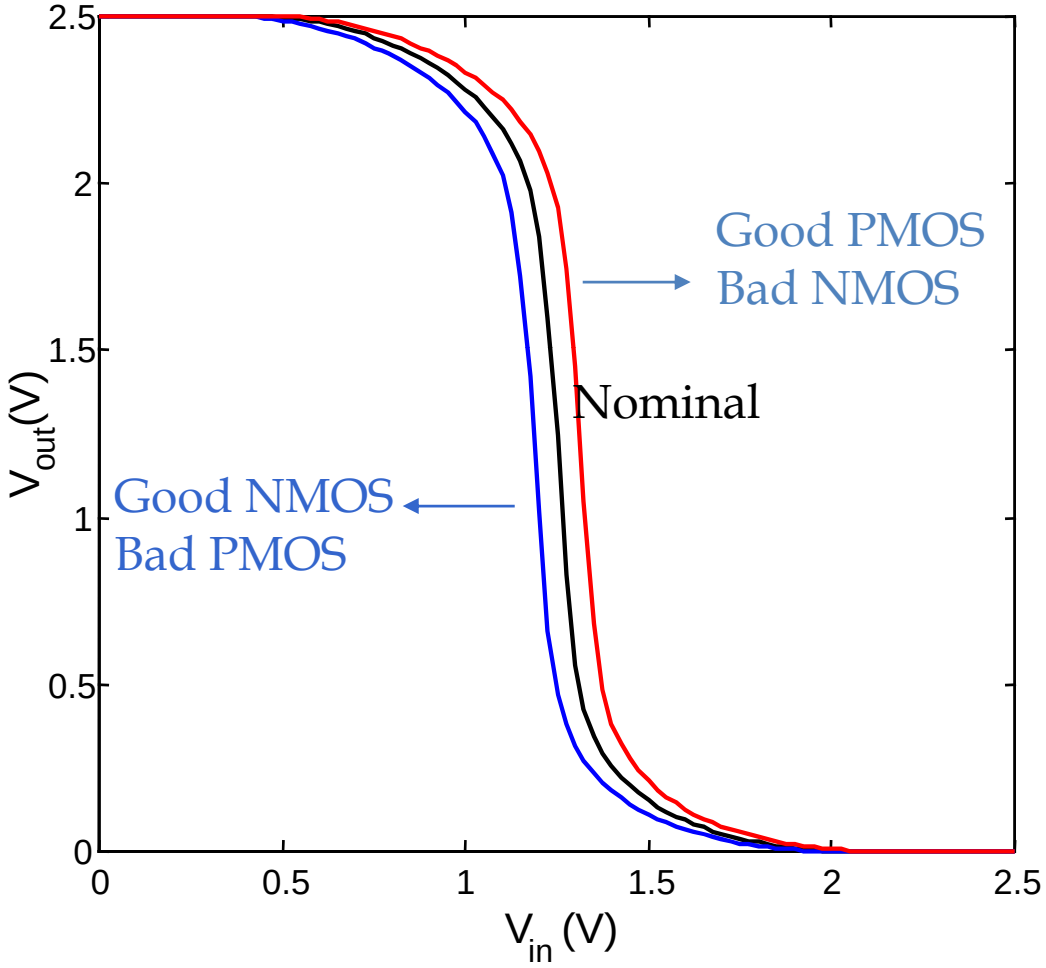




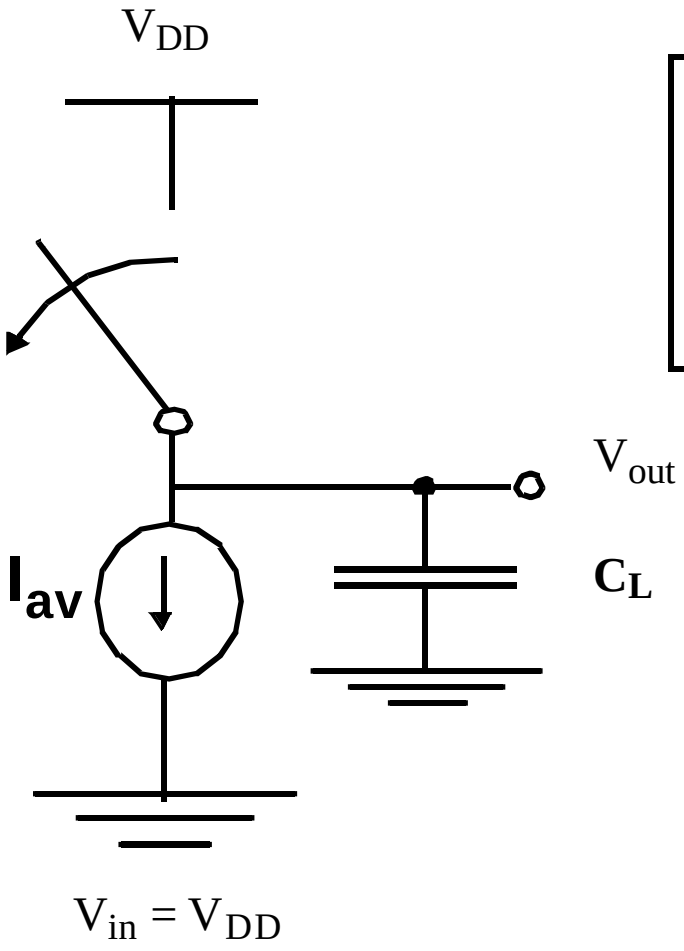
Simulated VTC



Impact of Process Variations



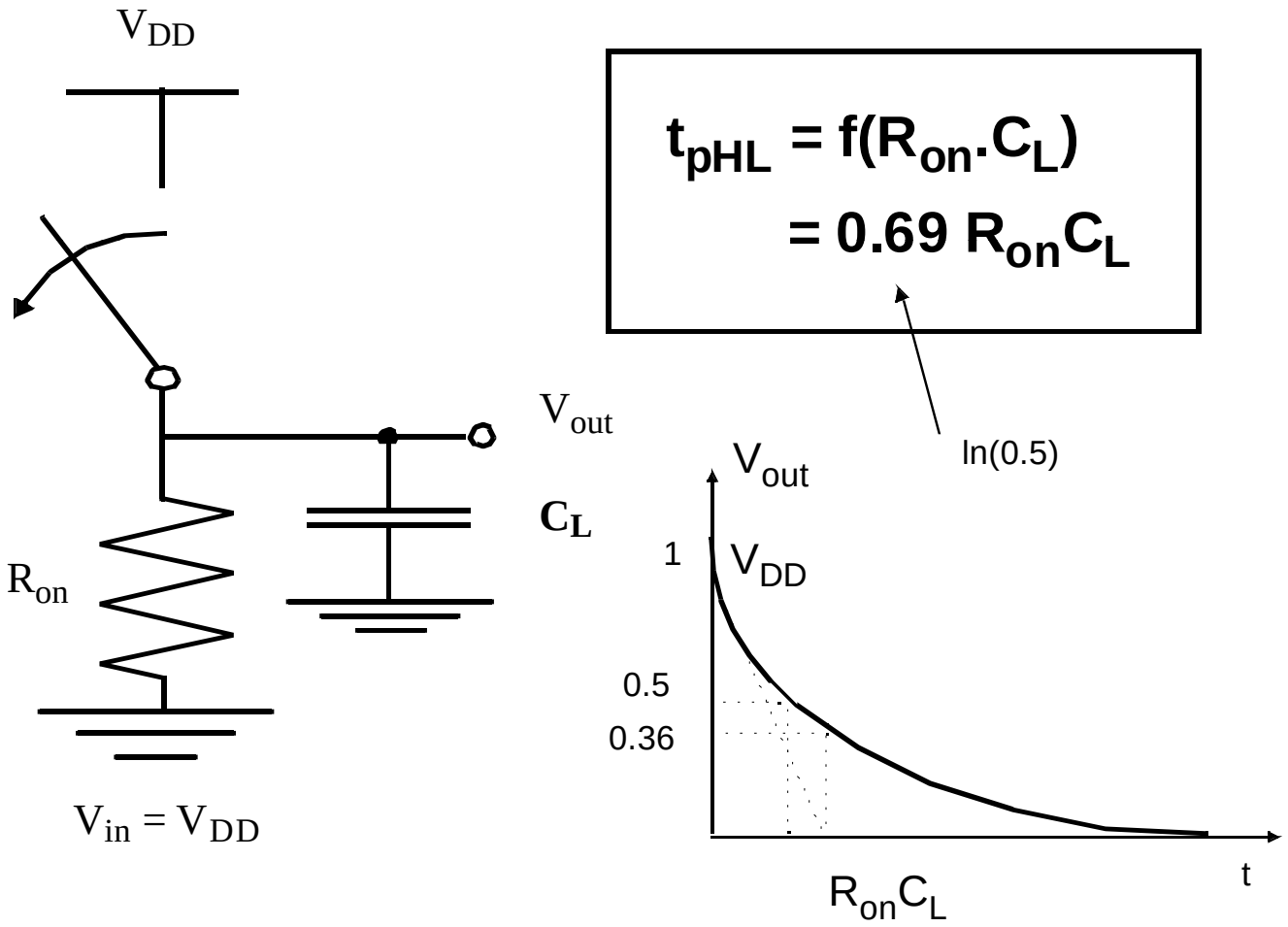
2. CMOS Inverter Propagation Delay Approach 1



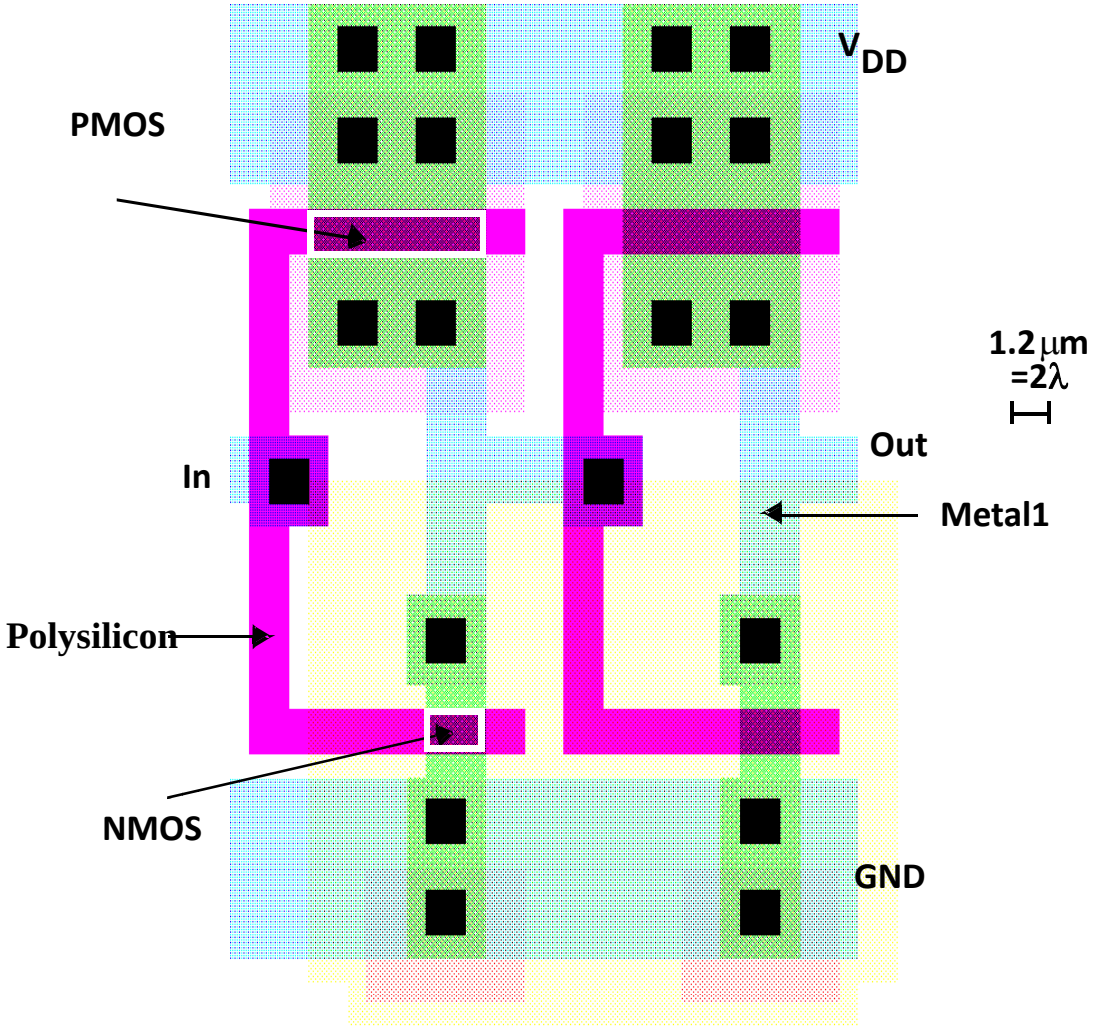
$$t_{pHL} = \frac{C_L \, V_{swing}/2}{I_{av}}$$

$$\sim \frac{C_L}{k_n \, V_{DD}}$$

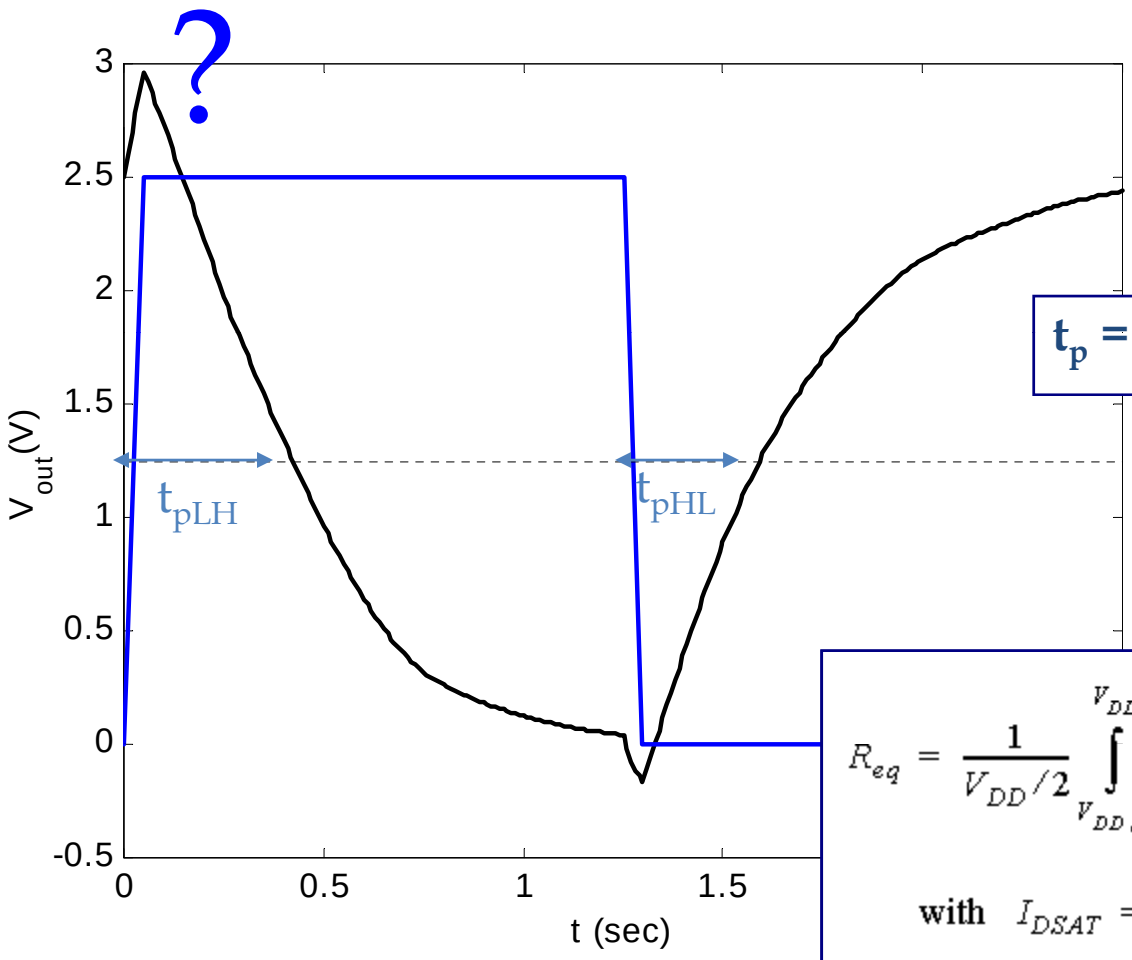
CMOS Inverter Propagation Delay Approach 2



CMOS Inverters



Transient Response



$$t_p = 0.69\, C_L\, (R_{eqn}+R_{eqp})/2$$

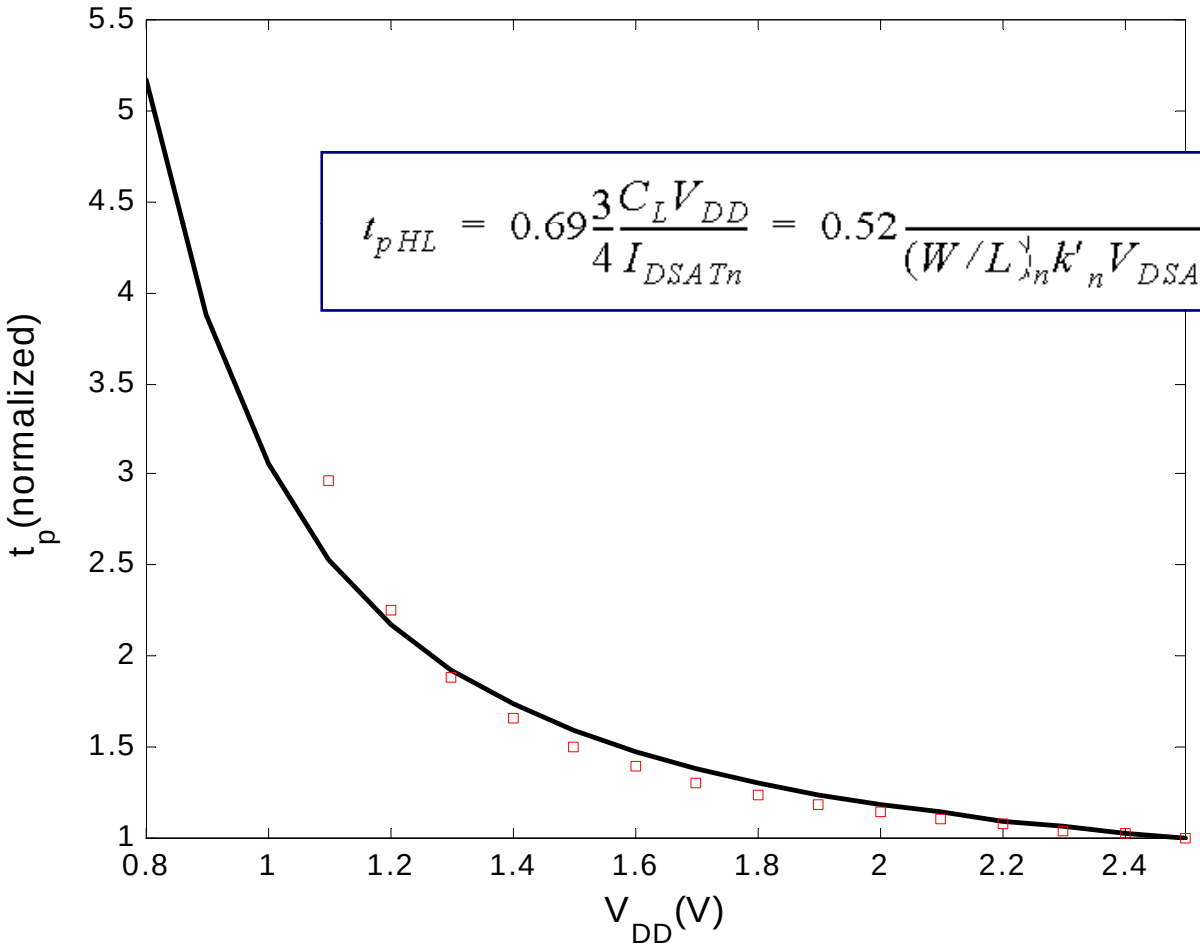
$$R_{eq} = \frac{1}{V_{DD}/2}\int_{V_{DD}/2}^{V_{DD}} \frac{V}{I_{DSAT}(1+\lambda V)}dV \approx \frac{3}{4}\frac{V_{DD}}{I_{DSAT}}\left(1-\frac{7}{9}\lambda V_{DD}\right)$$

$$\text{with}\quad I_{DSAT} = k'\frac{W}{L}\left((V_{DD}-V_T)V_{DSAT}-\frac{V_{DSAT}^2}{2}\right)$$

Design for Performance

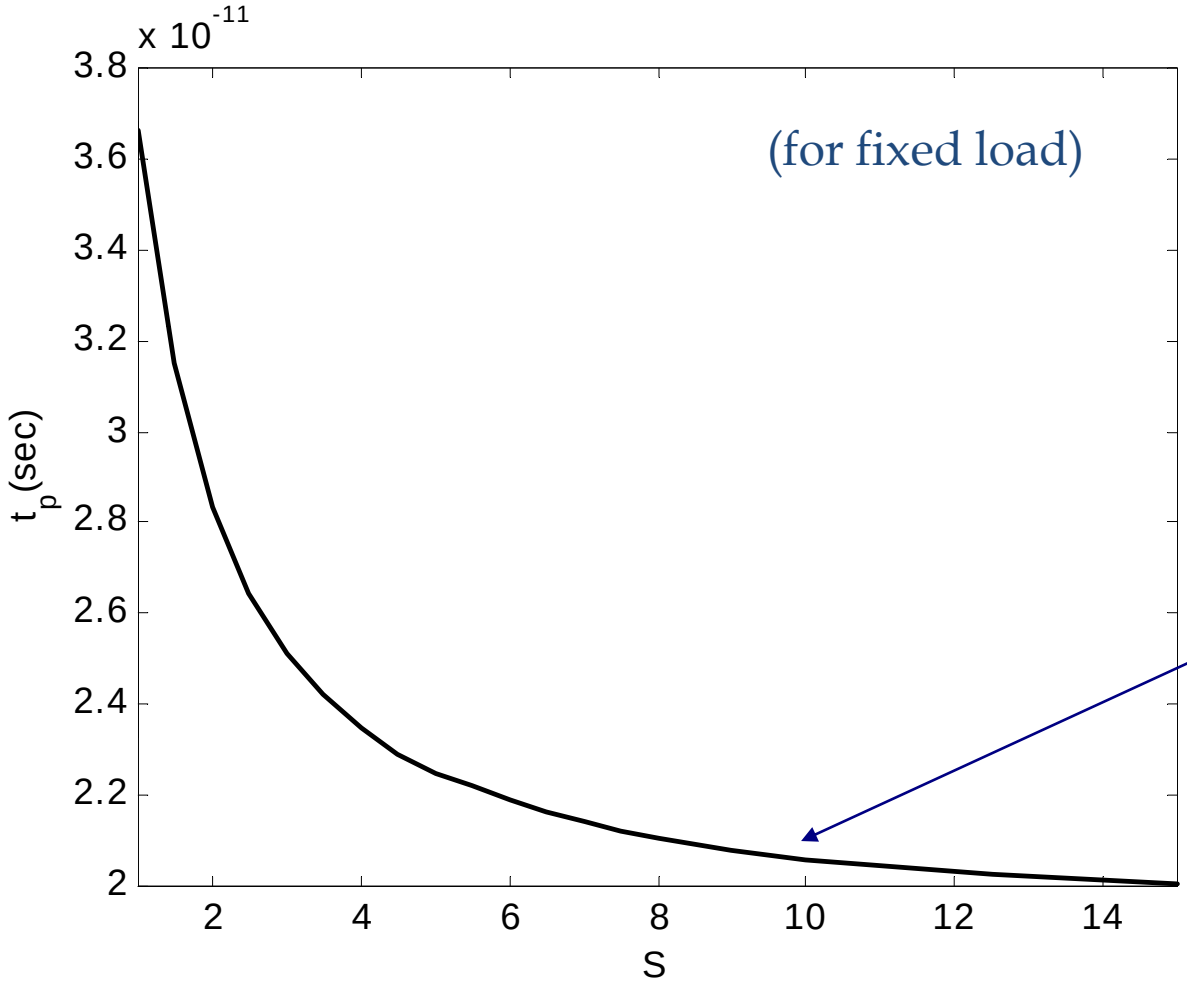
- Keep capacitances small
- Increase transistor sizes
  - watch out for self-loading!
- Increase V_{DD} (????)

Delay as a function of V_{DD}



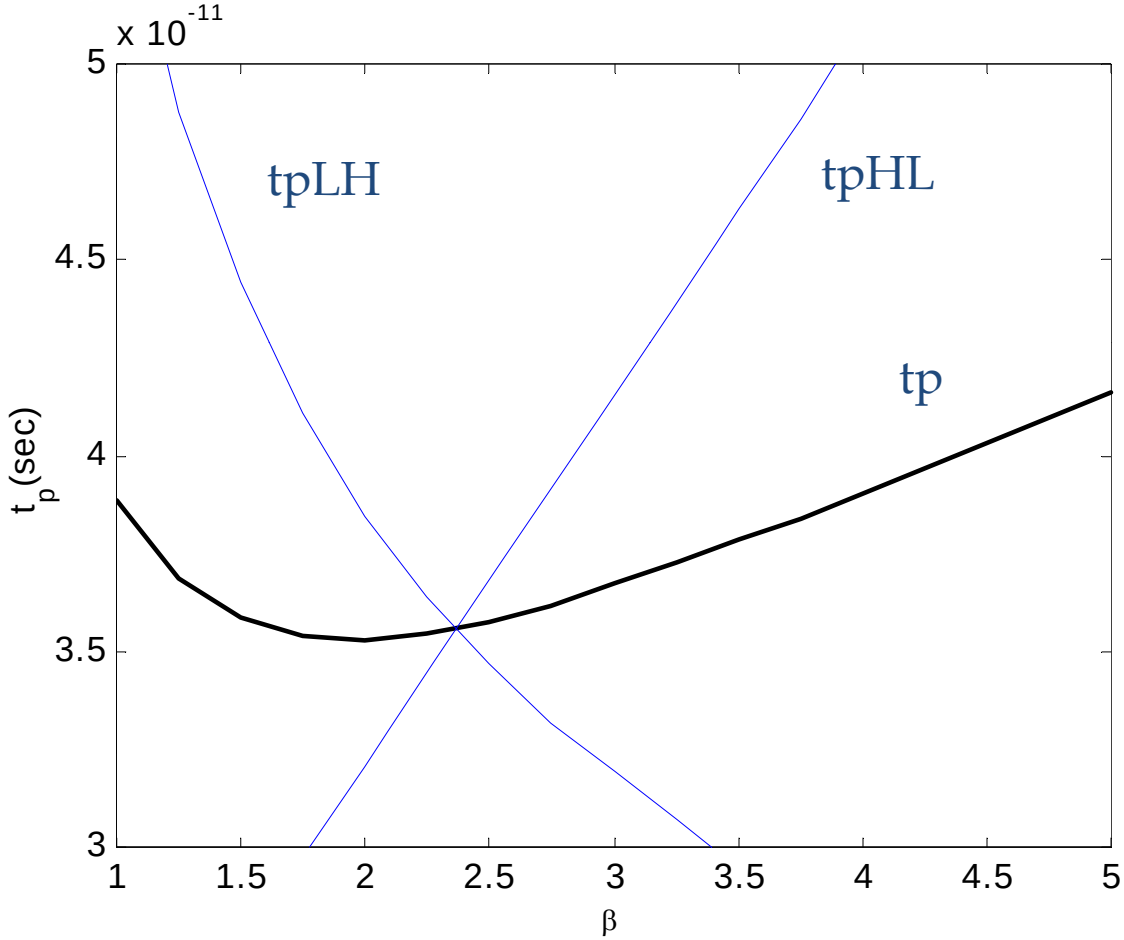
$$t_{pHL} = 0.69\frac{3}{4}\frac{C_L V_{DD}}{I_{DSATn}} = 0.52\frac{C_L V_{DD}}{(W/L)_n k'_n V_{DSATn}(V_{DD} - V_{Tn} - V_{DSATn}/2)}$$

Device Sizing



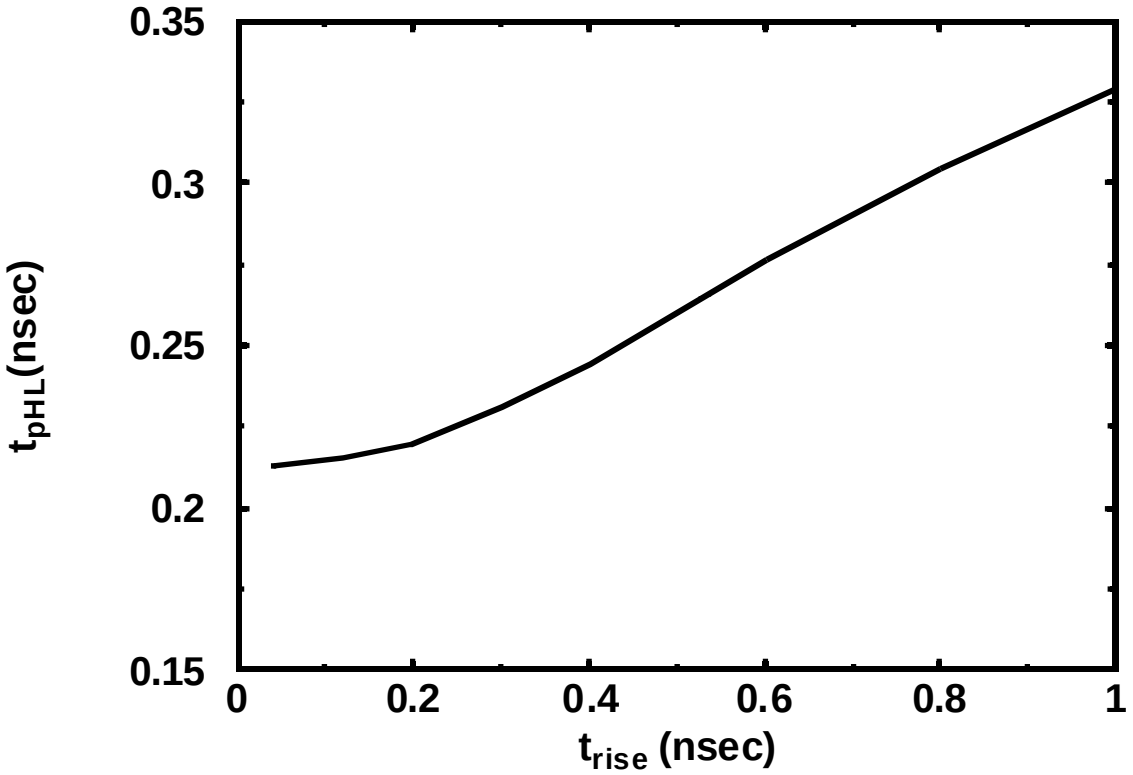
Self-loading effect:
Intrinsic capacitances dominate

NMOS/PMOS ratio



$\beta = W_p/W_n$

Impact of Rise Time on Delay



$$t_{pHL} = \sqrt{t_{pHL(step)}^2 + (t_r/2)^2}$$

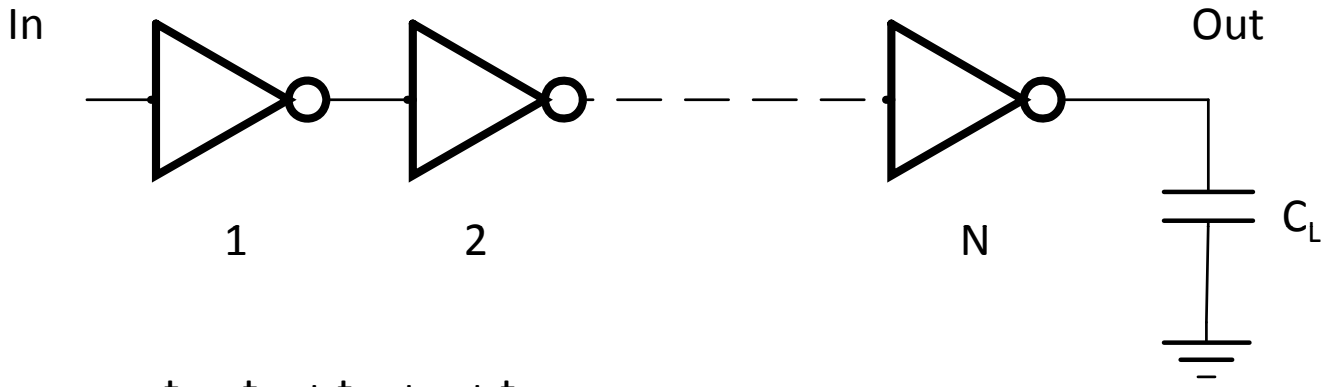
Delay Formula

$$Delay \sim R_W \left(C_{int} + C_L\right)$$

$$t_p = kR_W C_{\text{int}} \left(1 + C_L / C_{int}\right) = t_{p0}\left(1 + f / \gamma\right)$$

$C_{int} = \gamma C_{gin}$ with $\gamma \approx 1$
$f = C_L/C_{gin}$ - effective fanout
$R = R_{unit}/W$; $C_{int} = WC_{unit}$
$t_{p0} = 0.69R_{unit}C_{unit}$

3. Inverter Sizing
Apply to Inverter Chain



$t_p = t_{p1} + t_{p2} + \ldots + t_{pN}$

$$t_{pj} \sim R_{unit} C_{unit} \left(1 + \frac{C_{gin,j+1}}{\gamma C_{gin,j}} \right)$$

$$t_p = \sum_{j=1}^{N} t_{p,j} = t_{p0} \sum_{i=1}^{N} \left(1 + \frac{C_{gin,j+1}}{\gamma C_{gin,j}} \right), \; C_{gin,N+1} = C_L$$

Optimal Tapering for Given N

Delay equation has N - 1 unknowns, $C_{gin,2}$ – $C_{gin,N}$

Minimize the delay, find N - 1 partial derivatives

Result: $C_{gin,j+1}/C_{gin,j} = C_{gin,j}/C_{gin,j-1}$

Size of each stage is the geometric mean of two neighbors

- each stage has the same effective fanout (C_{out}/C_{in})
- each stage has the same delay

$$C_{gin,j} = \sqrt{C_{gin,j-1} C_{gin,j+1}}$$

Optimum Delay and Number of Stages

When each stage is sized by f and has same eff. fanout f:

$$f^N = F = C_L / C_{gin,1}$$

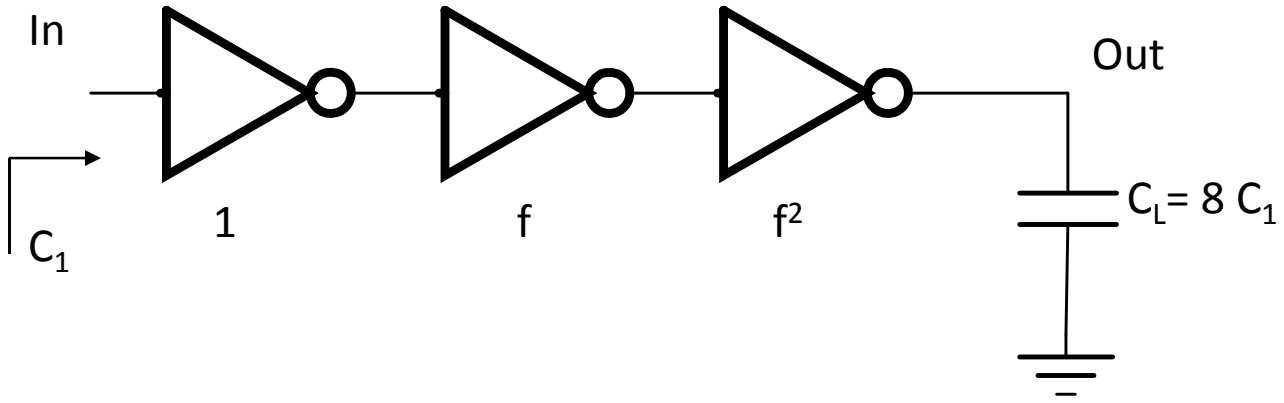
Effective fanout of each stage:

$$f = \sqrt[N]{F}$$

Minimum path delay

$$t_p = N t_{p0}\left(1 + \sqrt[N]{F} / \gamma\right)$$

Example



C_L/C_1 has to be evenly distributed across N = 3 stages:

$$f = \sqrt[3]{8} = 2$$

Optimum Number of Stages

For a given load, C_L and given input capacitance C_{in}
Find optimal sizing f

$$C_L = F \cdot C_{in} = f^N C_{in} \quad \text{with} \quad N = \frac{\ln F}{\ln f}$$

$$t_p = N t_{p0}\left(F^{1/N}/\gamma + 1\right) = \frac{t_{p0} \ln F}{\gamma}\left(\frac{f}{\ln f} + \frac{\gamma}{\ln f}\right)$$

$$\frac{\partial t_p}{\partial f} = \frac{t_{p0} \ln F}{\gamma} \cdot \frac{\ln f - 1 - \gamma / f}{\ln^2 f} = 0$$

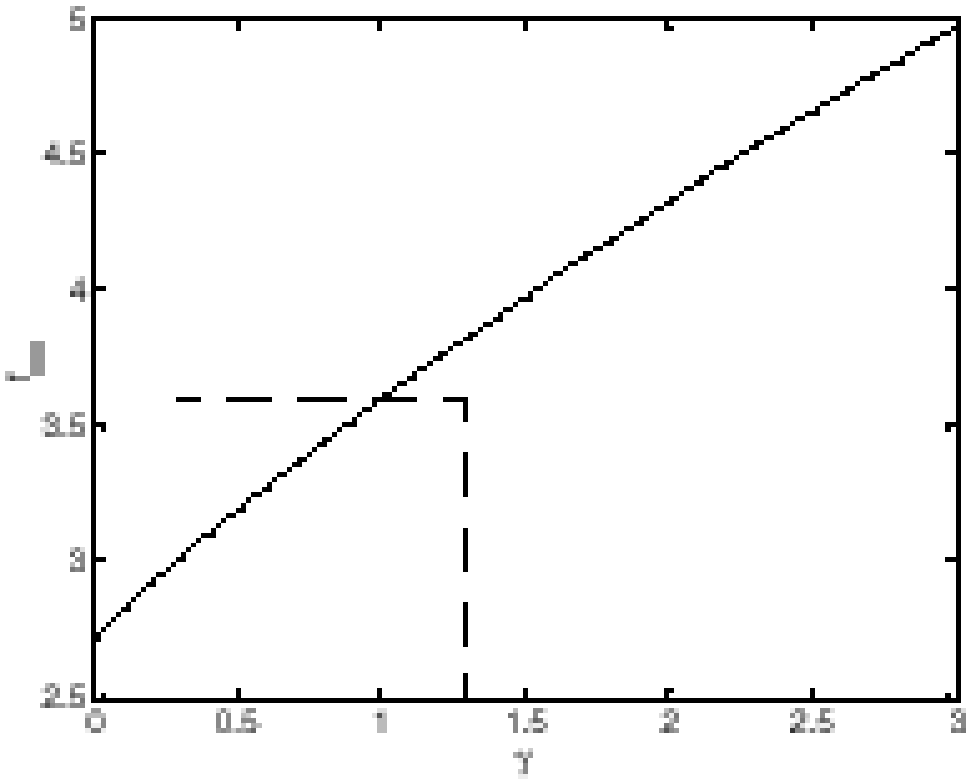
For $\gamma = 0$, f = e, N = lnF

$$f = \exp\left(1 + \gamma / f\right)$$

Optimum Effective Fanout f

Optimum f for given process defined by γ

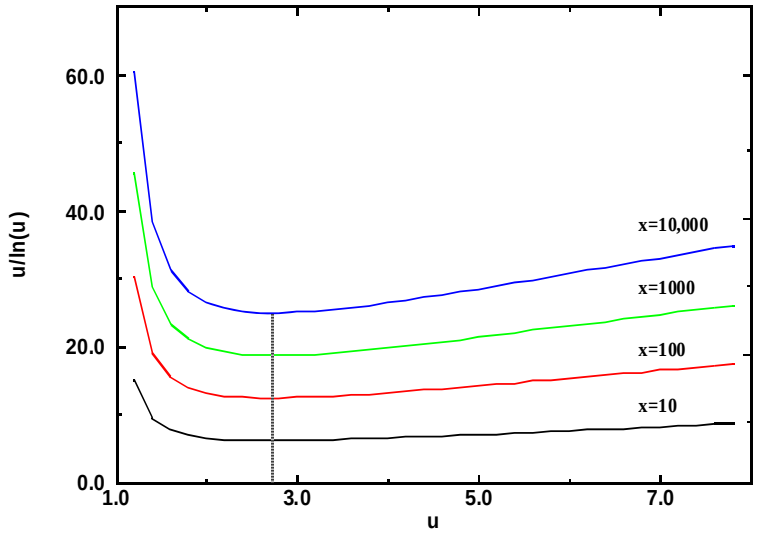
$$f = \exp(1 + \gamma / f)$$



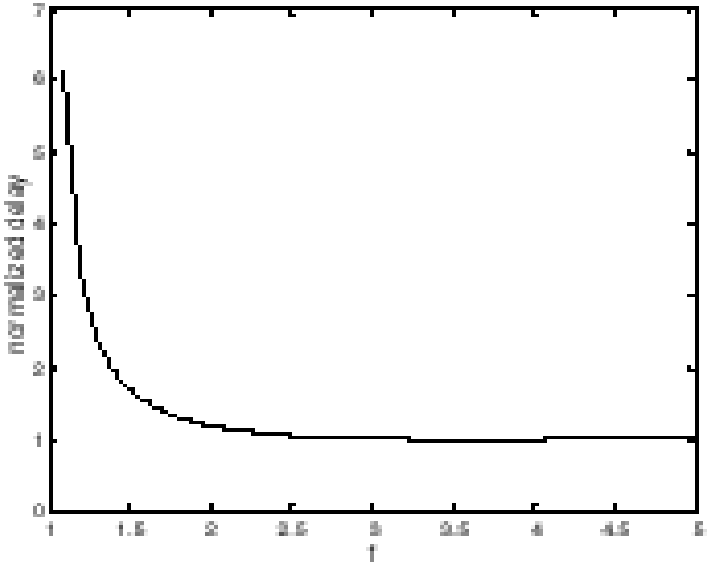
f_{opt} = 3.6
for γ=1

Impact of Self-Loading on tp

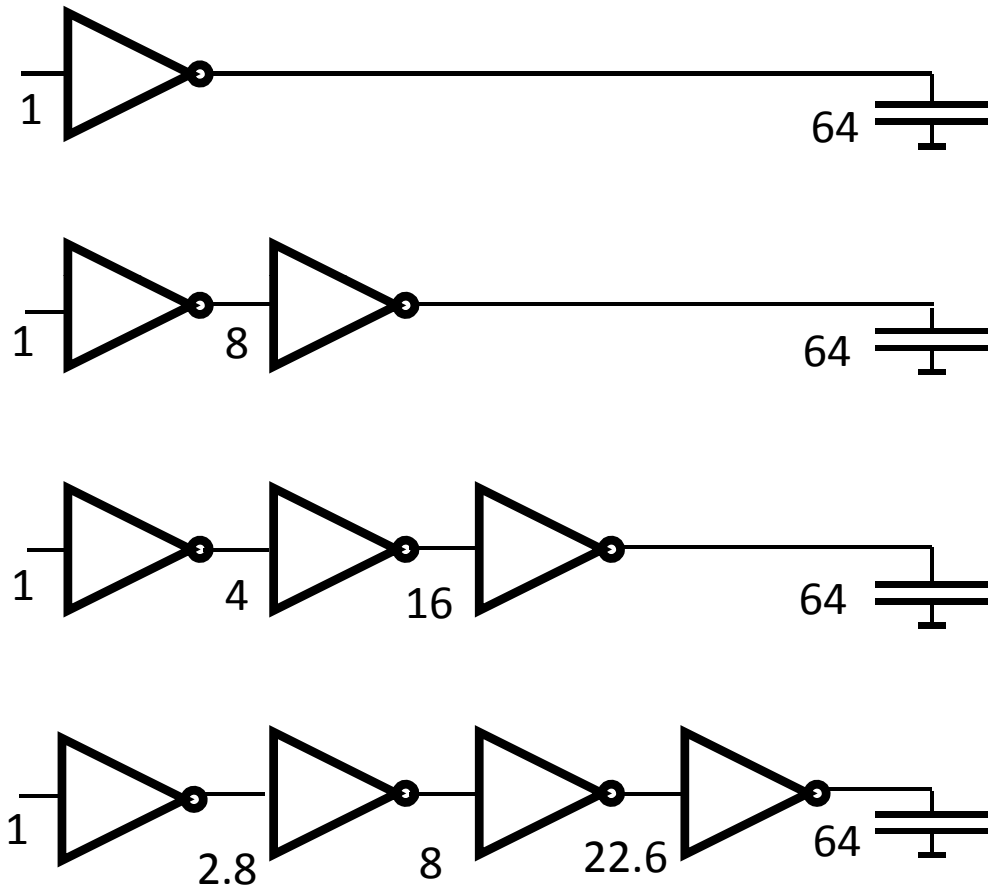
No Self-Loading, γ=0



With Self-Loading γ=1



Buffer Design



N	f	t_p
1	**64**	**65**
2	**8**	**18**
3	**4**	**15**
4	**2.8**	**15.3**

4. Power Dissipation

Where Does Power Go in CMOS?

- **Dynamic Power Consumption**

  Charging and Discharging Capacitors

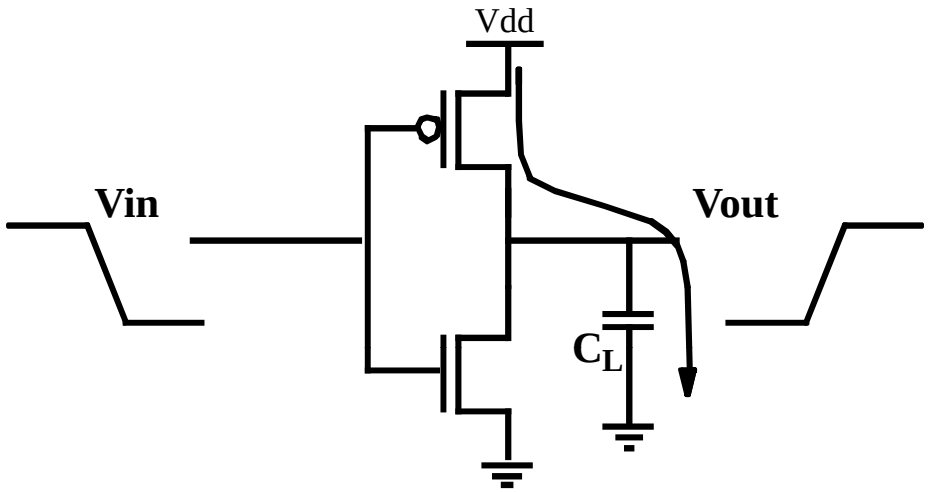
- **Short Circuit Currents**

  Short Circuit Path between Supply Rails during Switching

- **Leakage**

  Leaking diodes and transistors

Dynamic Power Dissipation

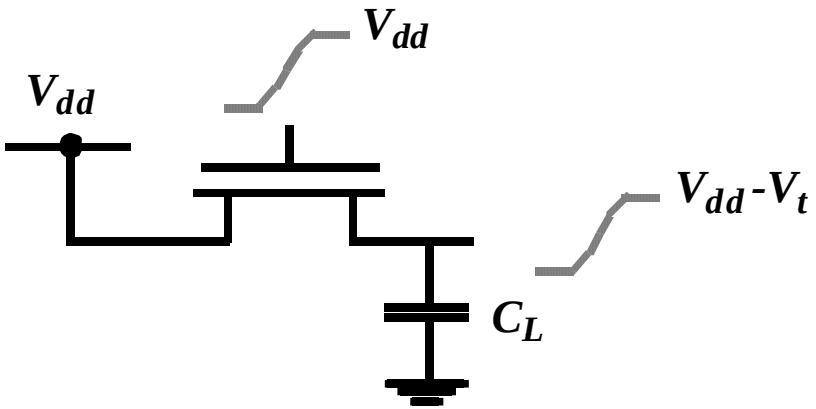


Energy/transition = $C_L * V_{dd}^2$

Power = Energy/transition * f = $C_L * V_{dd}^2 * f$

- **Not a function of transistor sizes!**
- **Need to reduce C_L, V_{dd}, and f to reduce power.**

Modification for Circuits with Reduced Swing

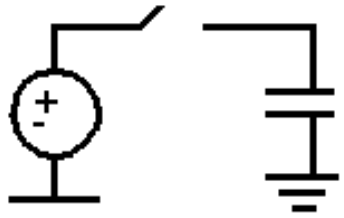


$$E_{0 \rightarrow 1} = C_L \bullet V_{dd} \bullet (V_{dd} - V_t)$$

- **Can exploit reduced swing to lower power (e.g., reduced bit-line swing in memory)**

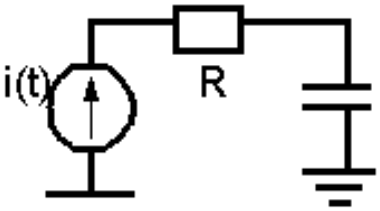
Adiabatic Charging

Charging a capacitor



$CV_{dd}^2/2$

Consider



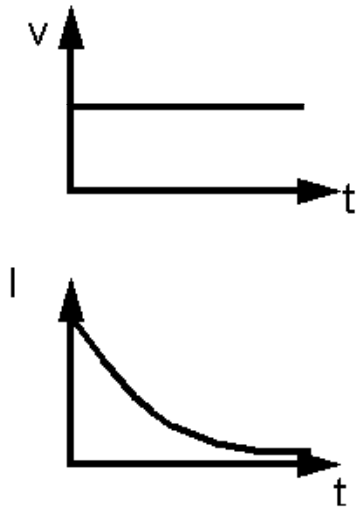
$$v_c = \frac{1}{C}\cdot\int_0^T i\,dt = \frac{1}{C}\cdot I_{av}\cdot T \qquad I_{av} = \frac{C\cdot v_c}{T}$$

$$E_{dis} = R\cdot\int_0^T i^2(t)\,dt \geq R\cdot\int_0^T I_{av}^2\,dt = R\cdot I_{av}^2\cdot T = \frac{RC}{T}\cdot C\cdot V_c^{\,2}$$

Adiabatic Charging

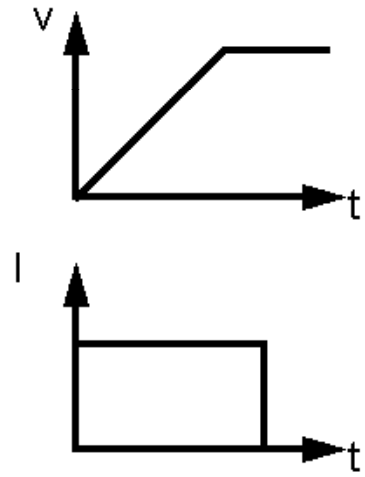
$$V_I = RI + V_c = RC\frac{dv_c}{dt} + V_c$$

V_I = cst -> Exponential current



$E_R = CV_c^2/2$

I = I_{av} -> Linear ramp on V_I



wins if T > 2RC

mimimal energy

$E_R = RC/T\ CV_c^2$

Node Transition Activity and Power

- **Consider switching a CMOS gate for *N* clock cycles**

$$E_N = C_L \bullet V_{dd}^2 \bullet n(N)$$

E_N : the energy consumed for *N* clock cycles

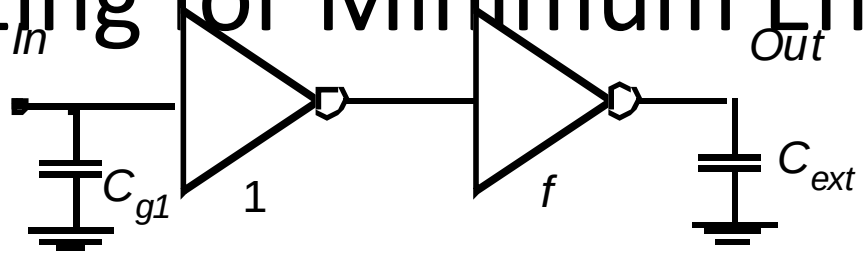
***n*(*N*): the number of 0->1 transition in *N* clock cycles**

$$P_{avg} = \lim_{N \to \infty} \frac{E_N}{N} \bullet f_{clk} = \left(\lim_{N \to \infty} \frac{n(N)}{N} \right) \bullet C_L \bullet V_{dd}^2 \bullet f_{clk}$$

$$\alpha_{0 \to 1} = \lim_{N \to \infty} \frac{n(N)}{N}$$

$$P_{avg} = \alpha_{0 \to 1} \bullet C_L \bullet V_{dd}^2 \bullet f_{clk}$$

Transistor Sizing for Minimum Energy



- Goal: Minimize Energy of whole circuit
  - Design parameters: *f* and V_{DD}
  - $tp \leq tpref$ of circuit with f=1 and $V_{DD} = V_{ref}$

$$t_p = t_{p0}\left(\left(1+\frac{f}{\gamma}\right)+\left(1+\frac{F}{f\gamma}\right)\right)$$

$$t_{p0} \propto \frac{V_{DD}}{V_{DD}-V_{TE}}$$

Transistor Sizing (2)

- Performance Constraint (γ=1)

$$\frac{t_p}{t_{pref}} = \frac{t_{p0}}{t_{p0ref}} \frac{\left(2 + f + \frac{F}{f}\right)}{(3+F)} = \frac{V_{DD}}{V_{ref}} \frac{V_{ref} - V_{TE}}{V_{DD} - V_{TE}} \frac{\left(2 + f + \frac{F}{f}\right)}{(3+F)} = 1$$

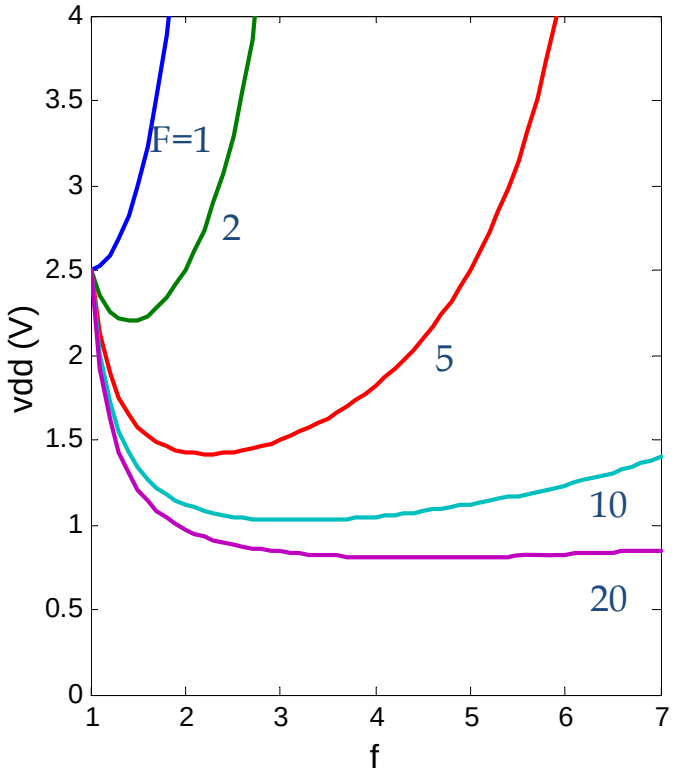
- Energy for single Transition

$$E = V_{DD}^2 C_{g1} \left[(1+\gamma)(1+f) + F\right]$$

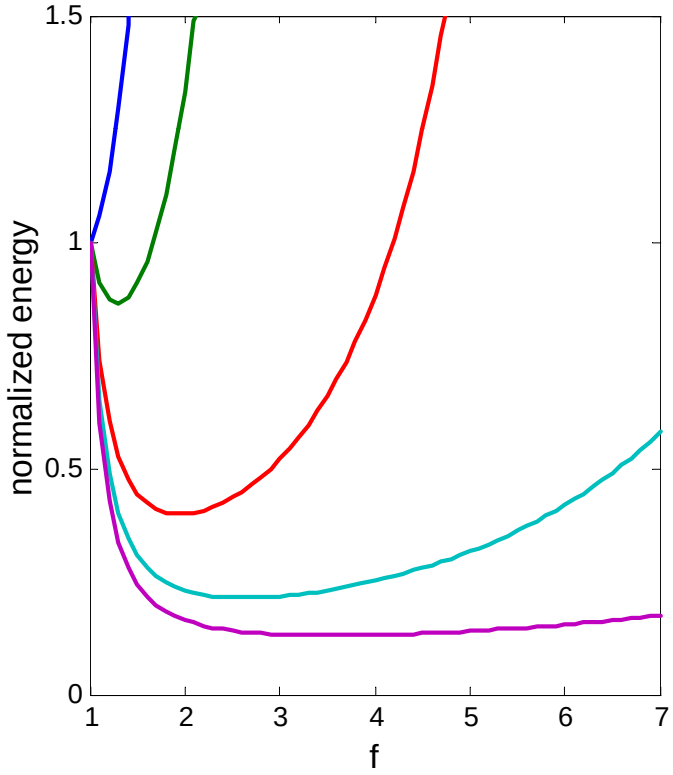
$$\frac{E}{E_{ref}} = \left(\frac{V_{DD}}{V_{ref}}\right)^2 \left(\frac{2 + 2f + F}{4 + F}\right)$$

Transistor Sizing (3)

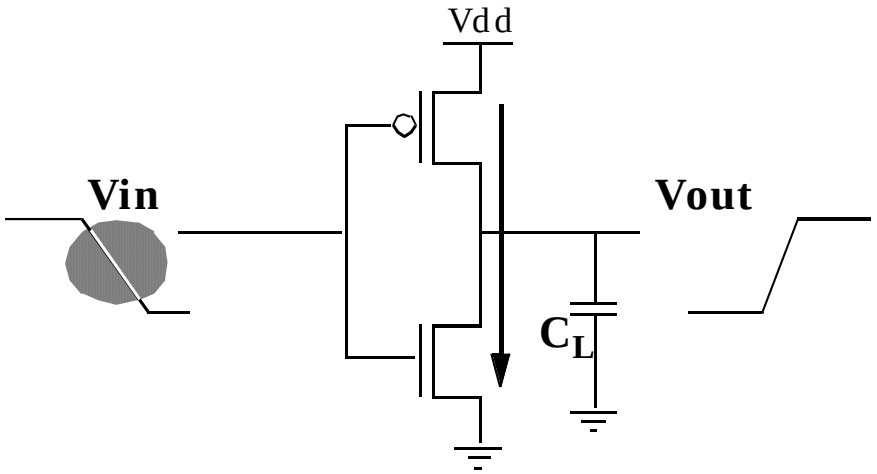
$V_{DD}=f(f)$

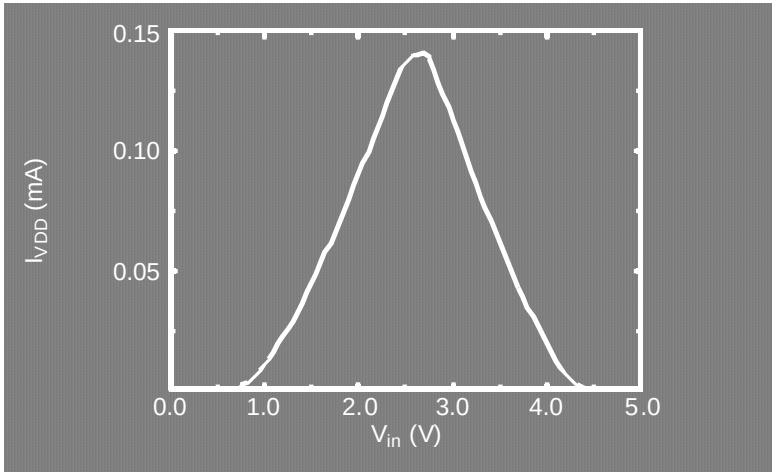


$E/E_{ref}=f(f)$

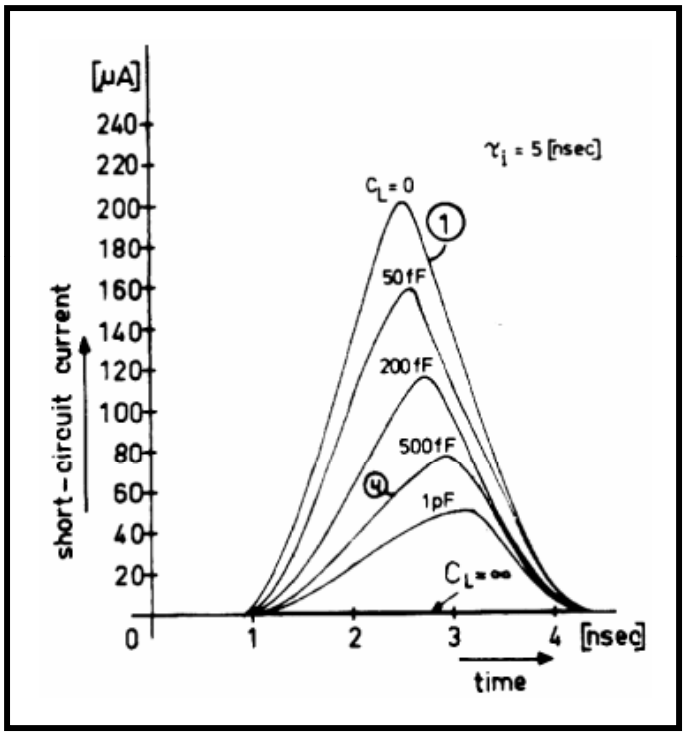


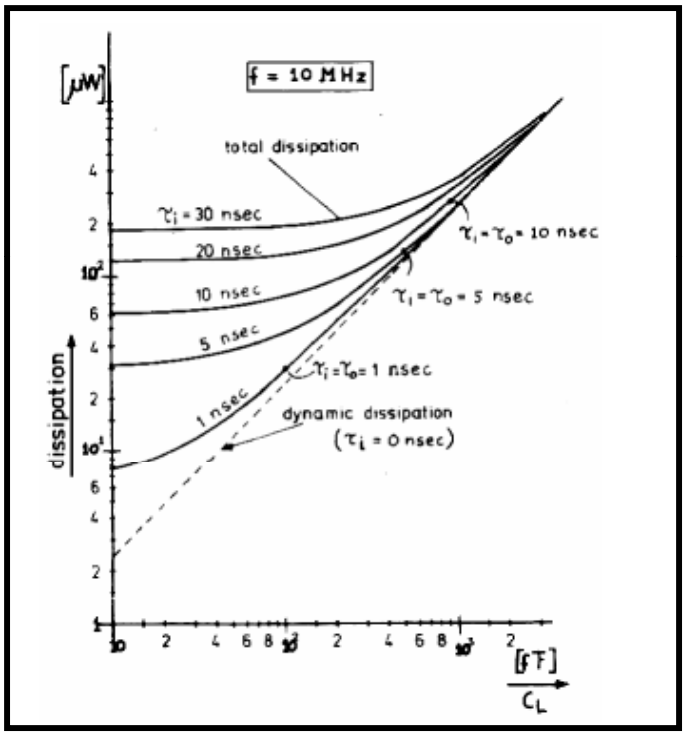
Short Circuit Currents





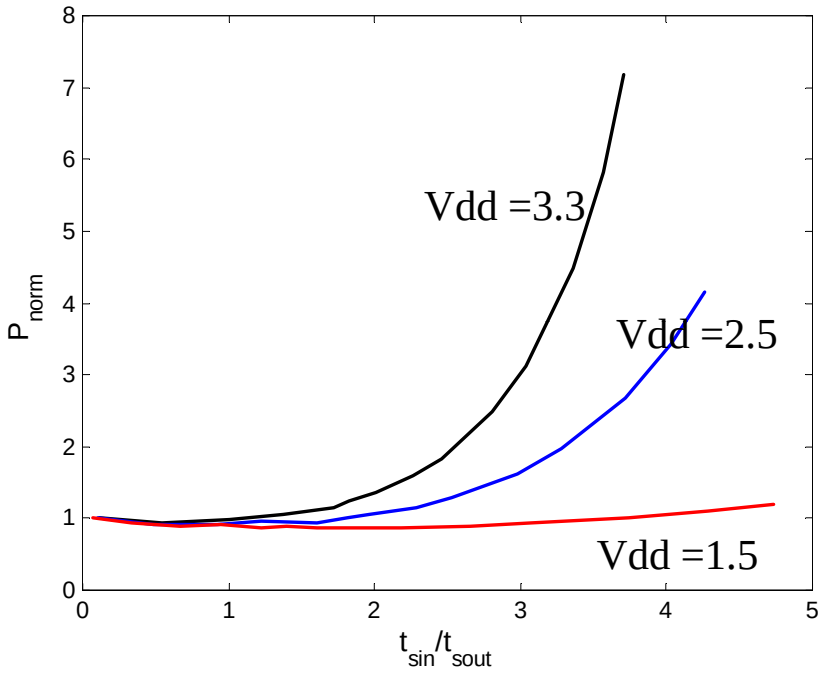
How to keep Short-Circuit Currents Low?





Short circuit current goes to zero if $t_{fall} >> t_{rise}$,
but can't do this for cascade logic, so ...

Minimizing Short-Circuit Power



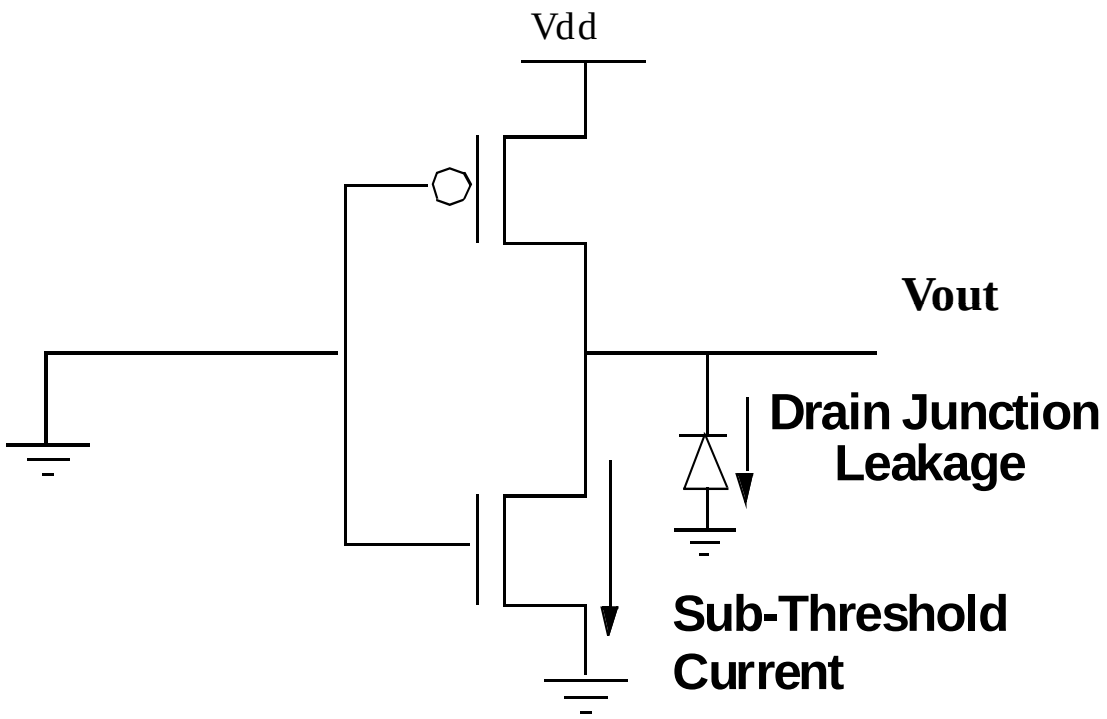
- Keep the input and output rise/fall times the same (< 10% of Total Consumption)

  from [Veendrick84]

  (*IEEE Journal of Solid-State Circuits*, August 1984)

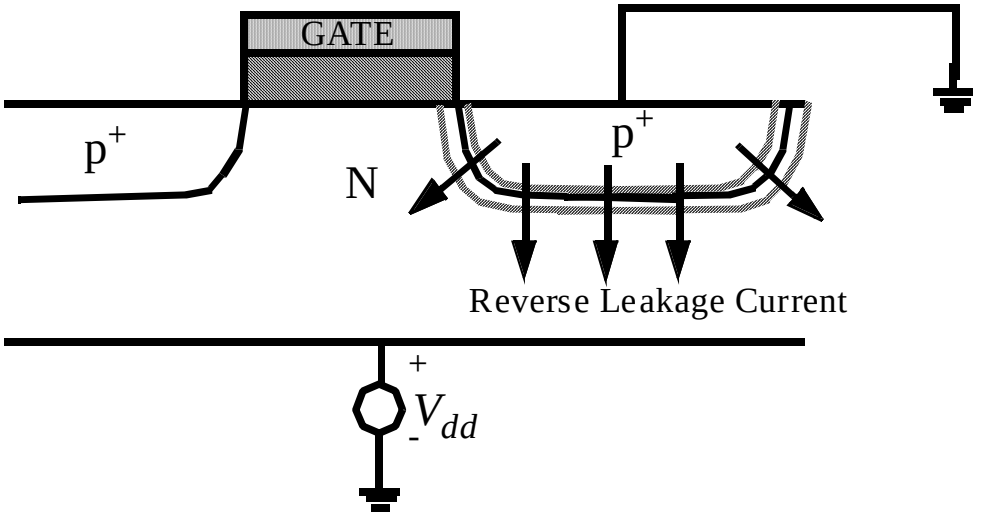
- If $V_{dd} < V_{tn} + |V_{tp}|$ then short-circuit power can be *eliminated*!

Leakage



Sub-threshold current one of most compelling issues in low-energy circuit design!

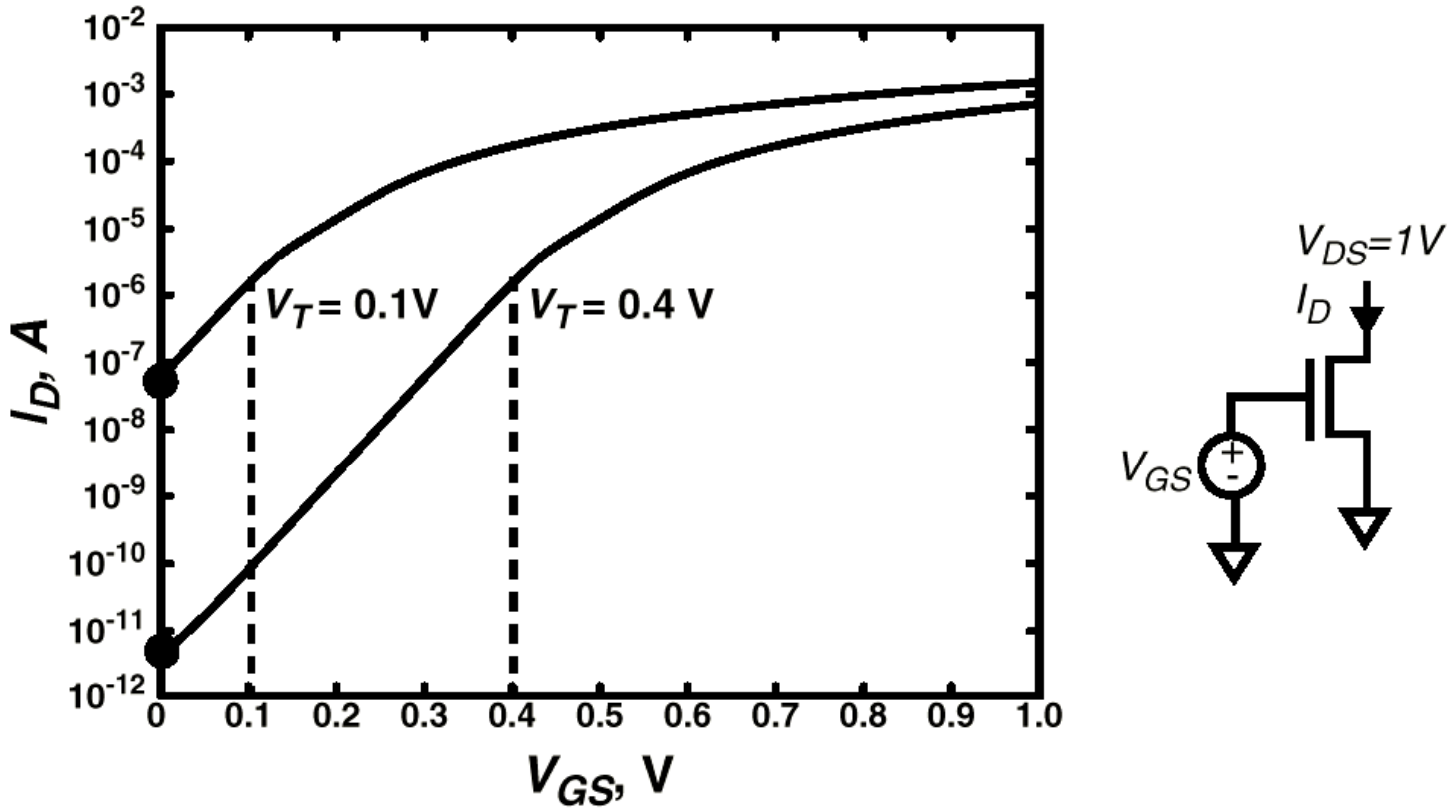
Reverse-Biased Diode Leakage



$$I_{DL} = J_S \times A$$

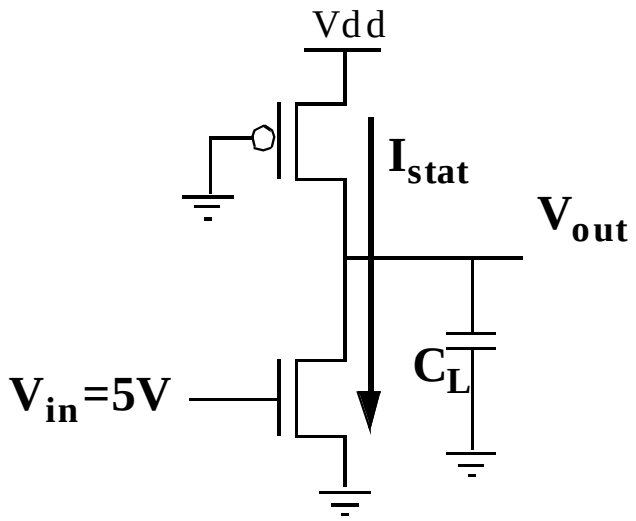
JS = 10-100 pA/μm2 at 25 deg C for 0.25μm CMOS
JS doubles for every 9 deg C!

Subthreshold Leakage Component



- **Leakage control is critical for low-voltage operation**

Static Power Consumption



$$P_{stat} = P_{(In=1)} . V_{dd} . I_{stat}$$

Wasted energy …
Should be avoided in almost all cases,
but could help reducing energy in others (e.g. sense amps)

Principles for Power Reduction

- Prime choice: Reduce voltage!
  - Recent years have seen an acceleration in supply voltage reduction
  - Design at very low voltages still open question (0.6 … 0.9 V by 2010!)
- Reduce switching activity
- Reduce physical capacitance
  - Device Sizing: for F=20
    - f_{opt}(energy)=3.53, f_{opt}(performance)=4.47

5. Impact of Technology Scaling

Goals of Technology Scaling

- Make things cheaper:
  - Want to sell more functions (transistors) per chip for the same money
  - Build same products cheaper, sell the same part for less money
  - Price of a transistor has to be reduced
- But also want to be faster, smaller, lower power

Technology Scaling

- Goals of scaling the dimensions by 30%:
  - Reduce gate delay by 30% (increase operating frequency by 43%)
  - Double transistor density
  - Reduce energy per transition by 65% (50% power savings @ 43% increase in frequency
- Die size used to increase by 14% per generation
- Technology generation spans 2-3 years