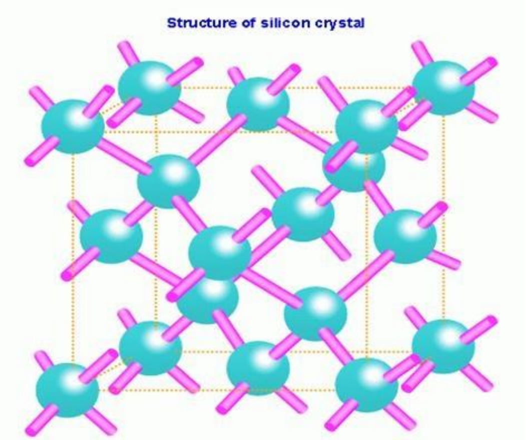
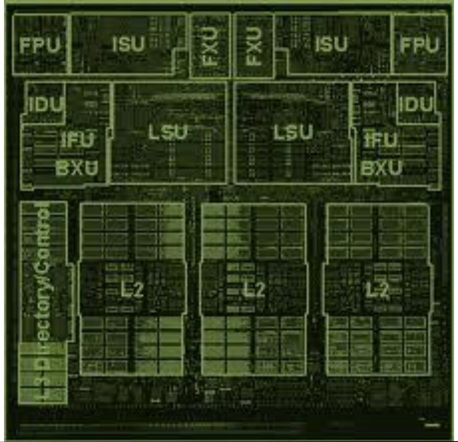
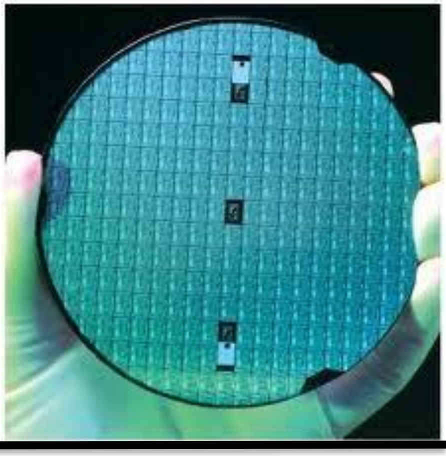






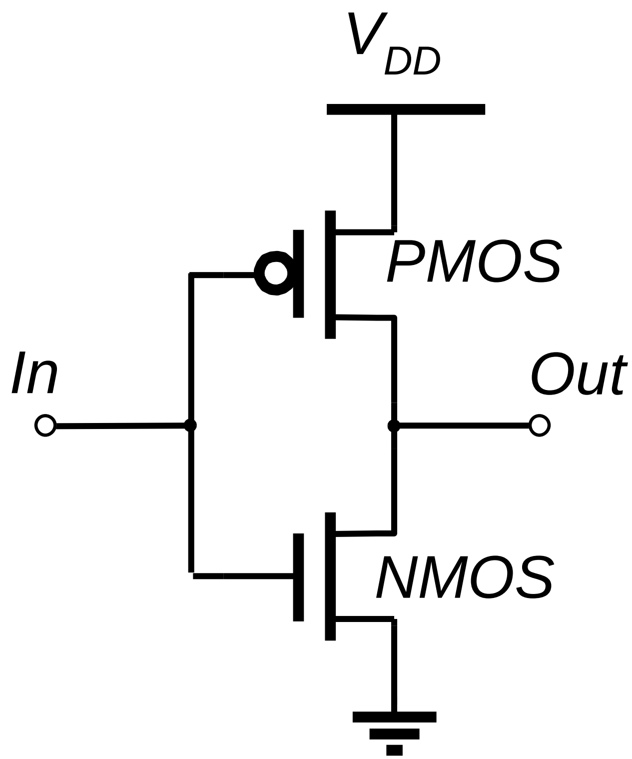
ECE 122A

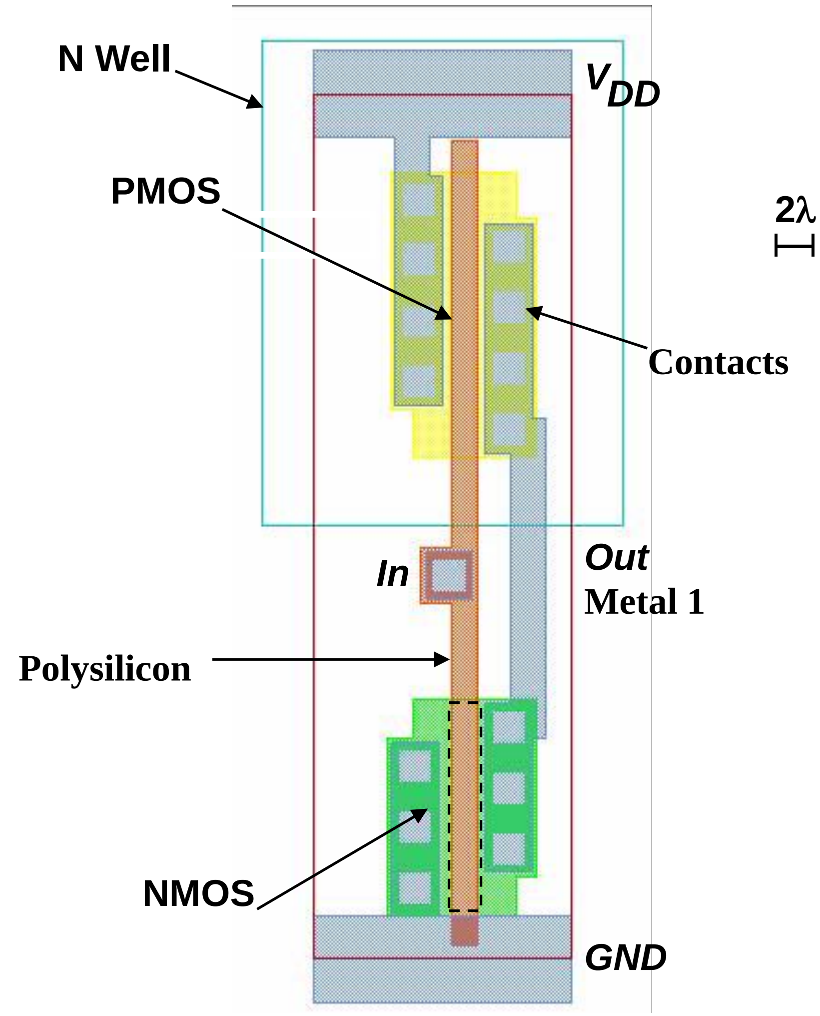
VLSI Principles

Lecture 9

Prof. Kaustav Banerjee
Electrical and Computer Engineering
University of California, Santa Barbara
E-mail: kaustav@ece.ucsb.edu

CMOS Inverter

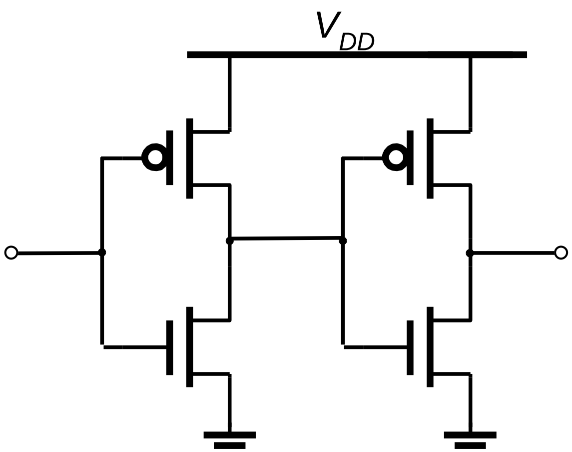


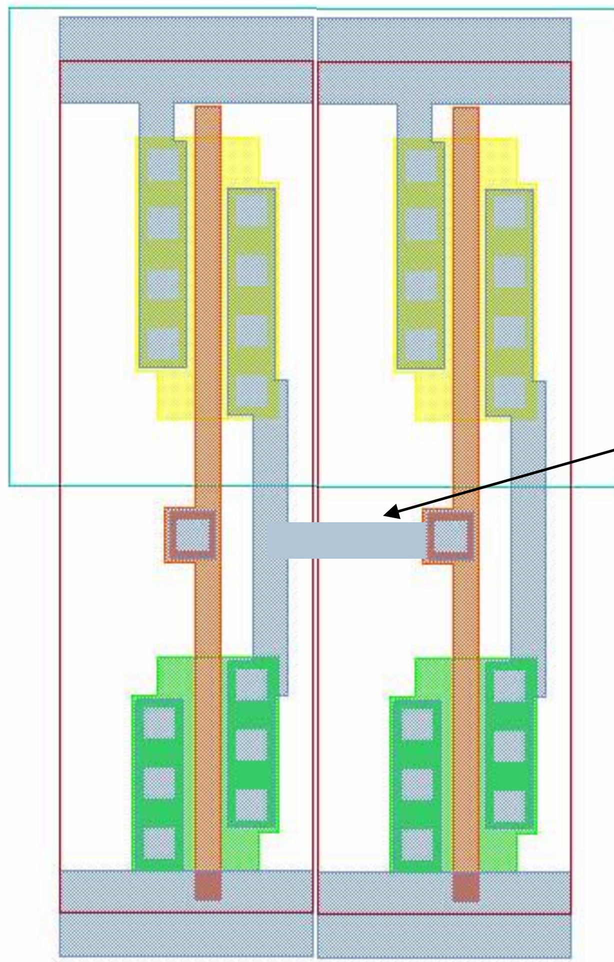


Two Inverters

Share power and ground

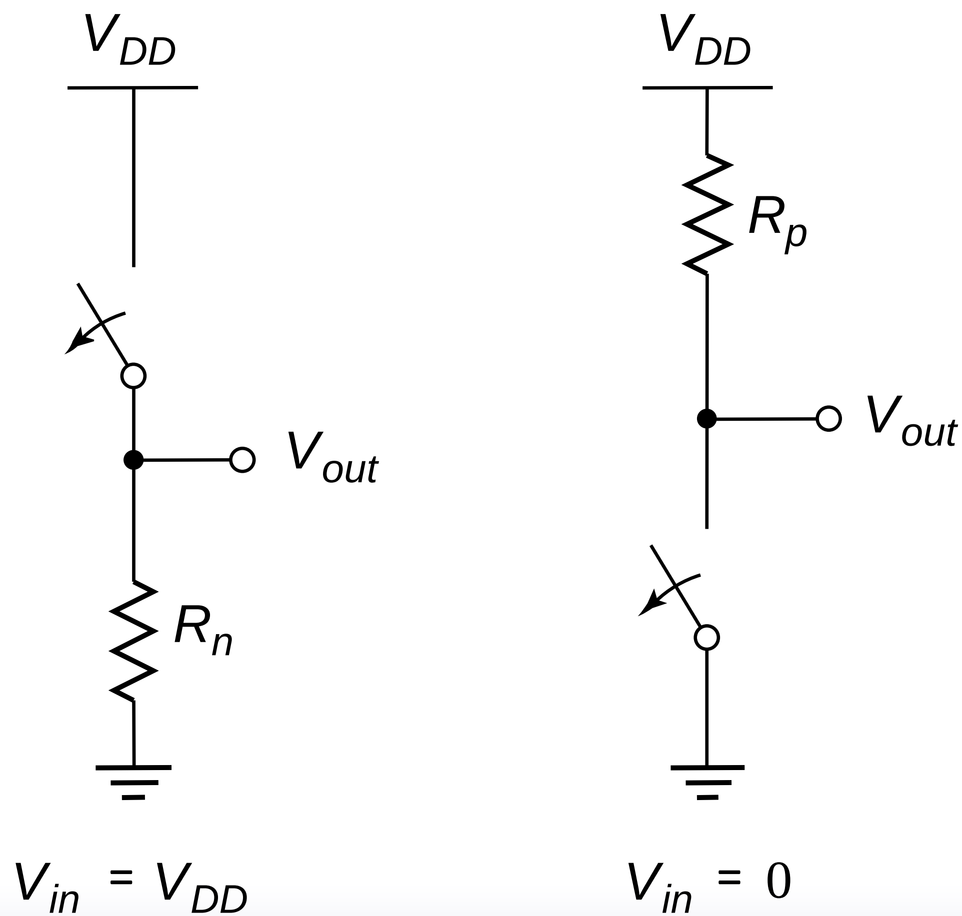
Abut cells





Connect in Metal

CMOS Inverter
First-Order DC Analysis

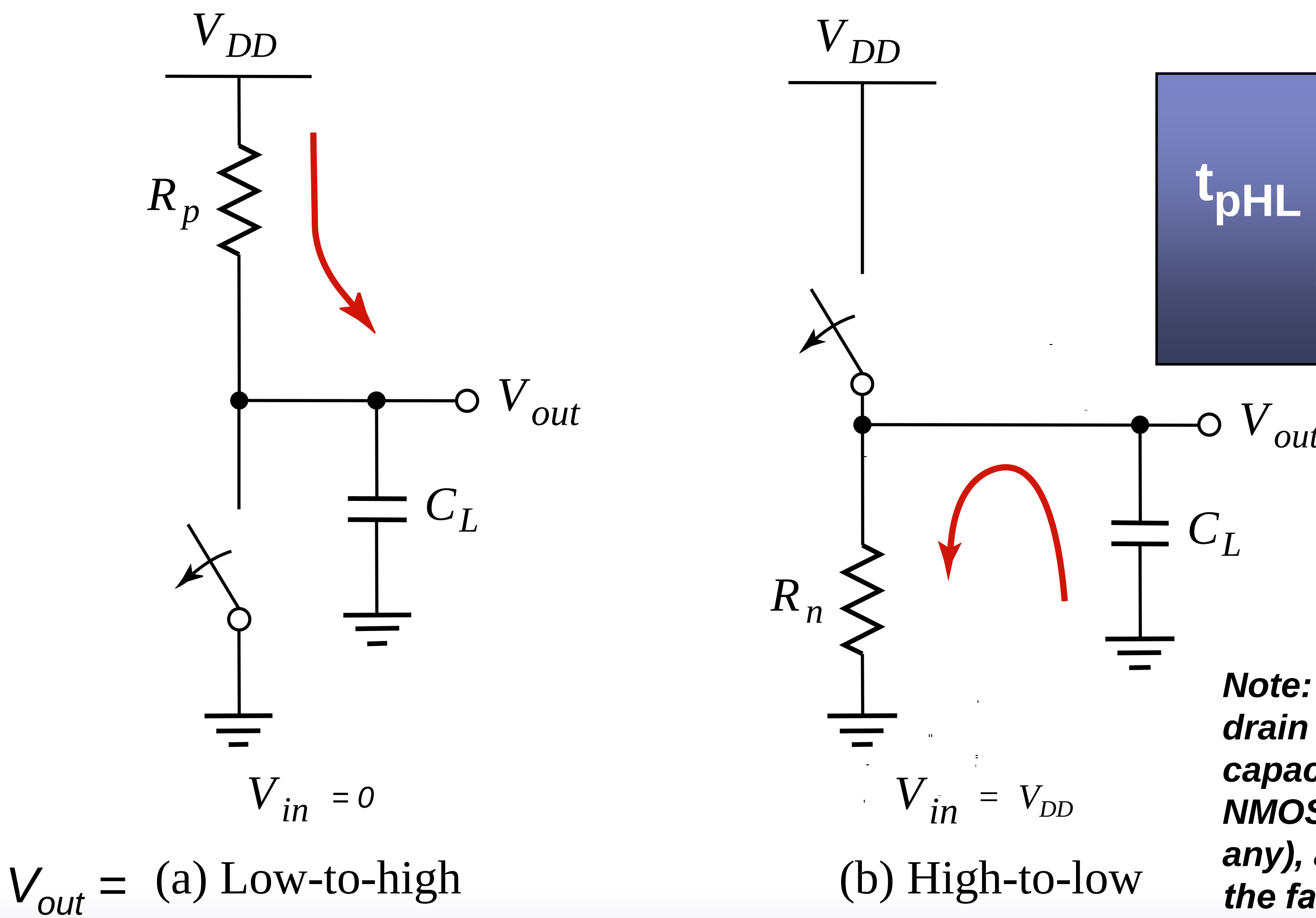


$V_{OL} = 0$

$V_{OH} = V_{DD}$

$V_M = f(R_n, R_p)$

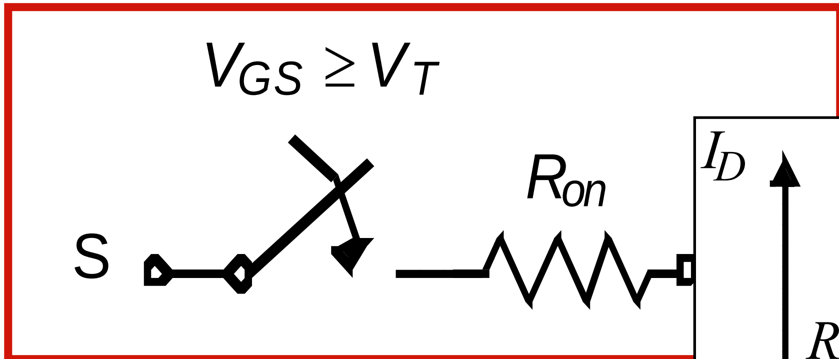
CMOS Inverter: Transient Response

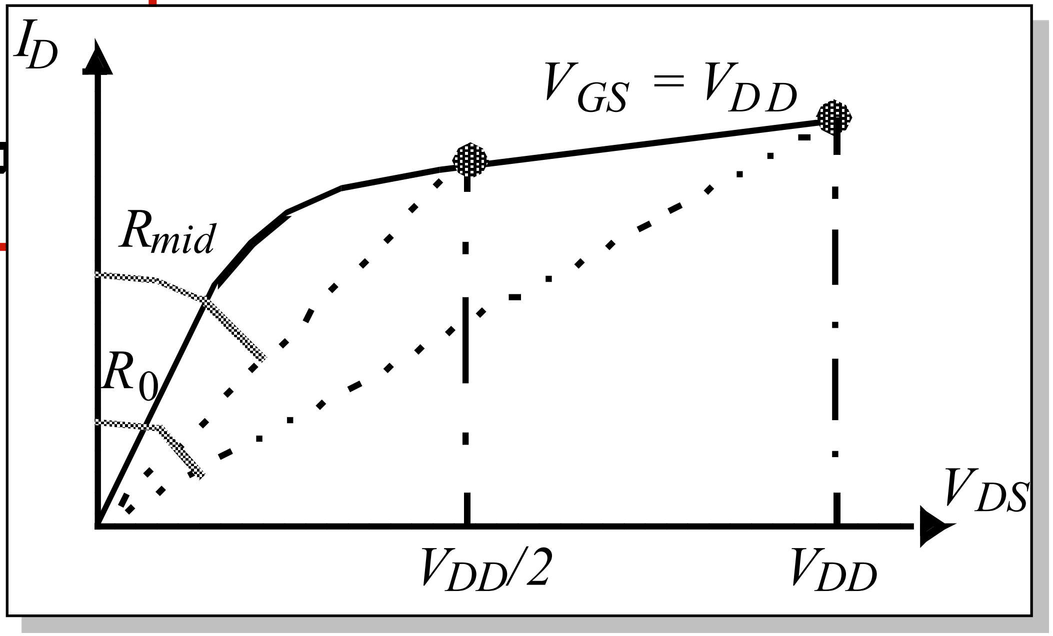


$t_{pHL} = f(R_{on}.C_L)$
$= 0.69\ R_{on}C_L$

Note: C_L is composed of drain diffusion capacitances of NMOS/PMOS, wire caps (if any), and the input cap. of the fanout gates…..

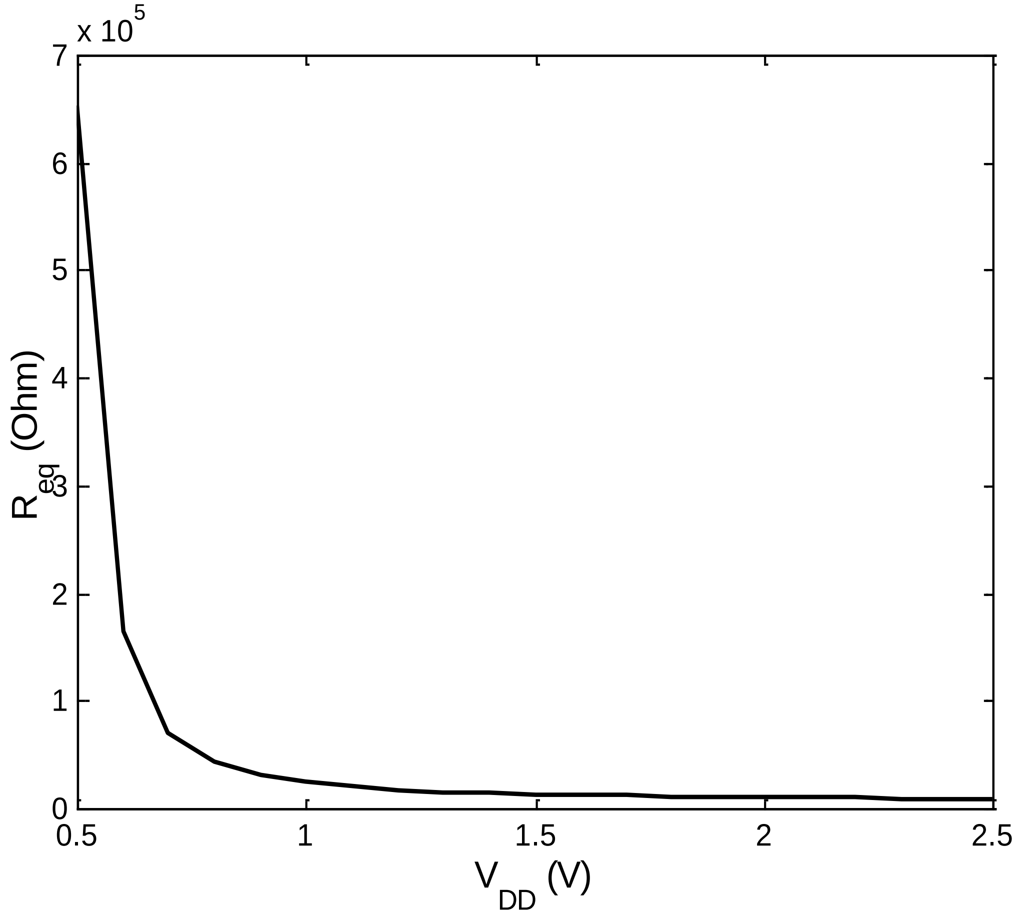
The Transistor as a Switch





$$R_{eq} = \frac{1}{2}\left(\frac{V_{DD}}{I_{DSAT}(1+\lambda V_{DD})} + \frac{V_{DD}/2}{I_{DSAT}(1+\lambda V_{DD}/2)}\right) \approx \frac{3}{4}\frac{V_{DD}}{I_{DSAT}}\left(1 - \frac{5}{6}\lambda V_{DD}\right)$$

The Transistor as a Switch



As V_{DD} increases, drain current increases….for both NMOS and PMOS

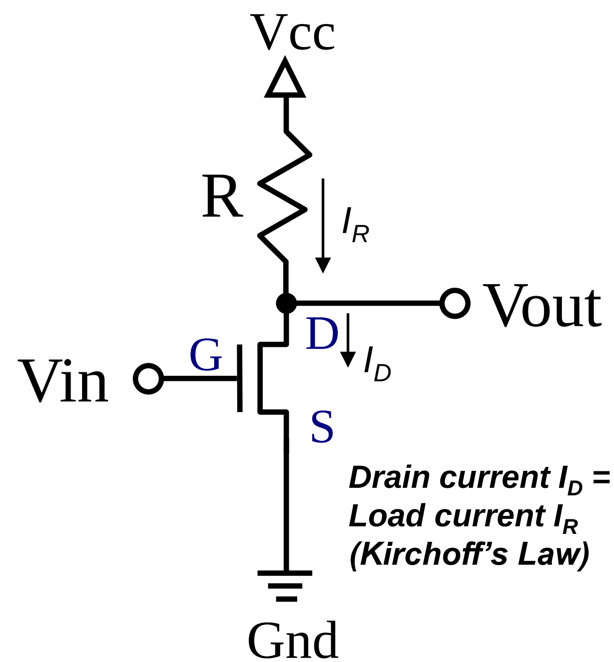
The Transistor as a Switch

Table 3.3 Equivalent resistance R_{eq} (W/L= 1) of NMOS and PMOS transistors in 0.25 μm CMOS process (with $L = L_{min}$). For larger devices, divide R_{eq} by W/L.

V_{DD} (V)	1	1.5	2	2.5
NMOS (kΩ)	35	19	15	13
PMOS (kΩ)	115	55	38	31

Resistive-load Inverter

- Requires only NMOS transistor and resistor
- When Vin = 0:
  - NMOS is OFF (V_{GS} = 0)
  - No current through NMOS
  - Vout ≈ Vcc
- When Vin = Vcc:
  - NMOS is ON (V_{GS} = Vcc)
  - NMOS ON resistance << R
  - Vout ≈ 0



Remember: if body terminal not shown, it is connected to gnd for NMOS, Vcc for PMOS

Not suitable for VLSI: large area of R, DC power dissipation.

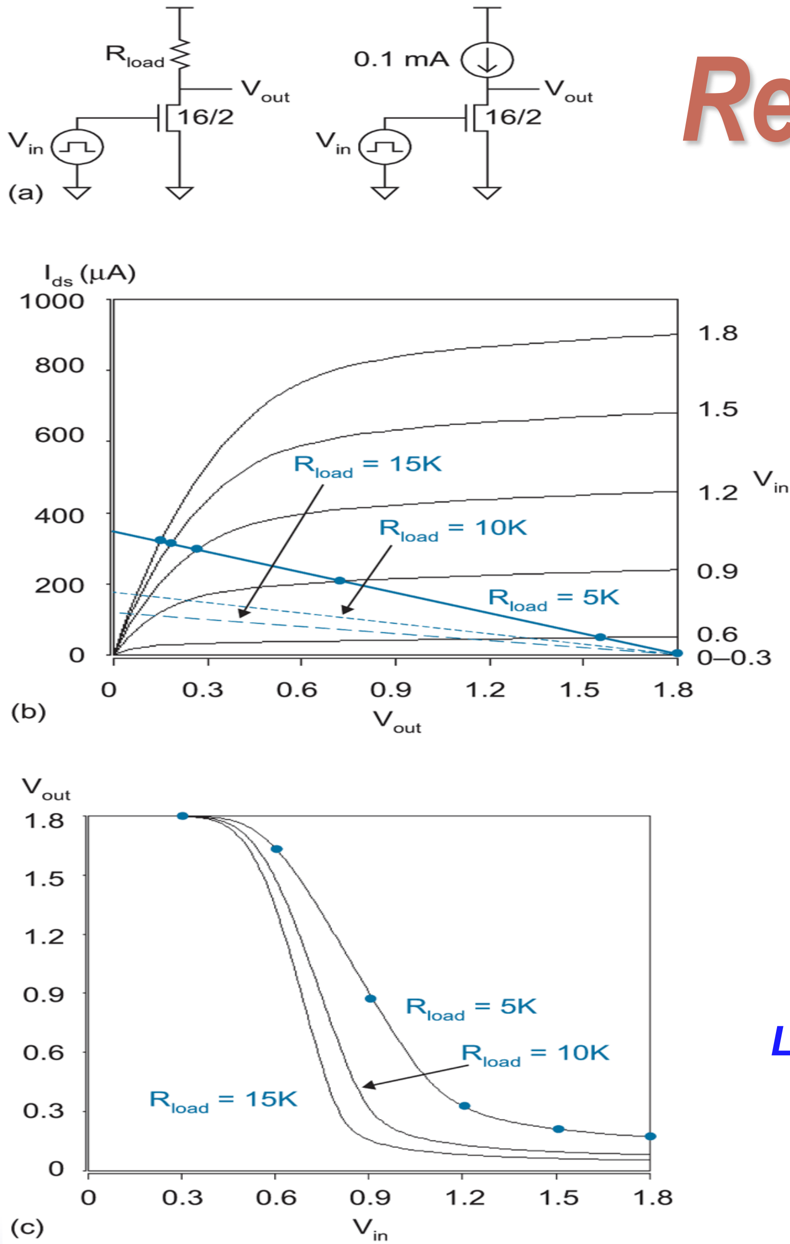


FIG 2.29 Generic nMOS inverters with resistive or constant current load

Resistive-load Inverter

$I_{ds} = V_{cc}/R_{load}$

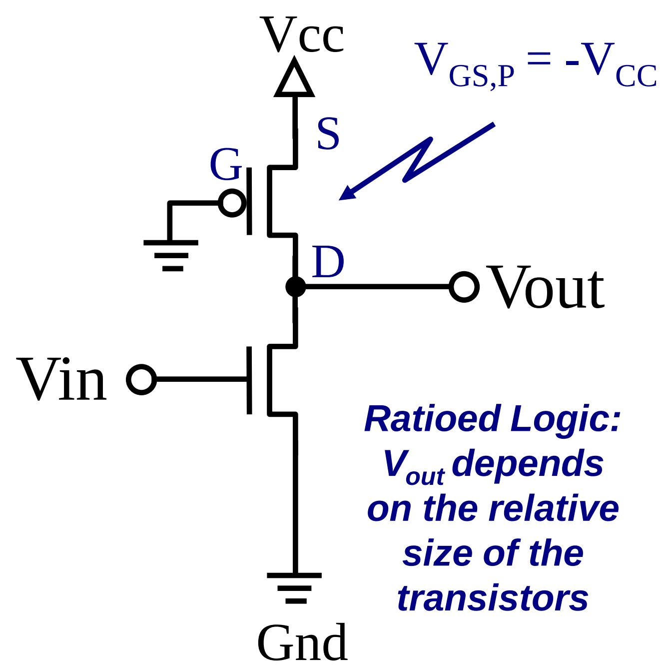
As R_{load} increases, V_{TC} becomes sharper!

Larger R_{load} = smaller PU transistor...LO skewed…

Kaustav Banerjee

Pseudo-NMOS Inverter

- Replace resistor with "always-on" PMOS transistor
- Easier to implement in standard process than large resistance value
- PMOS load transistor:
  - ON when $V_{GS} < V_T \rightarrow$
    $V_{GS} = -V_{CC}$: transistor always on
  - Linear when $V_{DS} > V_{GS} - V_T \rightarrow$
    $V_{out} - V_{cc} > -V_{cc} - V_T \rightarrow V_{out} > -V_T$
  - Saturated when $V_{DS} < V_{GS} - V_T \rightarrow$
    $V_{out} - V_{cc} < -V_{cc} - V_T \rightarrow V_{out} < -V_T$



Ratioed Logic: V_{out} depends on the relative size of the transistors

Remember:
$V_T(PMOS) < 0$

Pseudo-NMOS Inverter

Compare with resistive-load inverter!

Note: Smaller Width of PMOS (P) means larger resistance

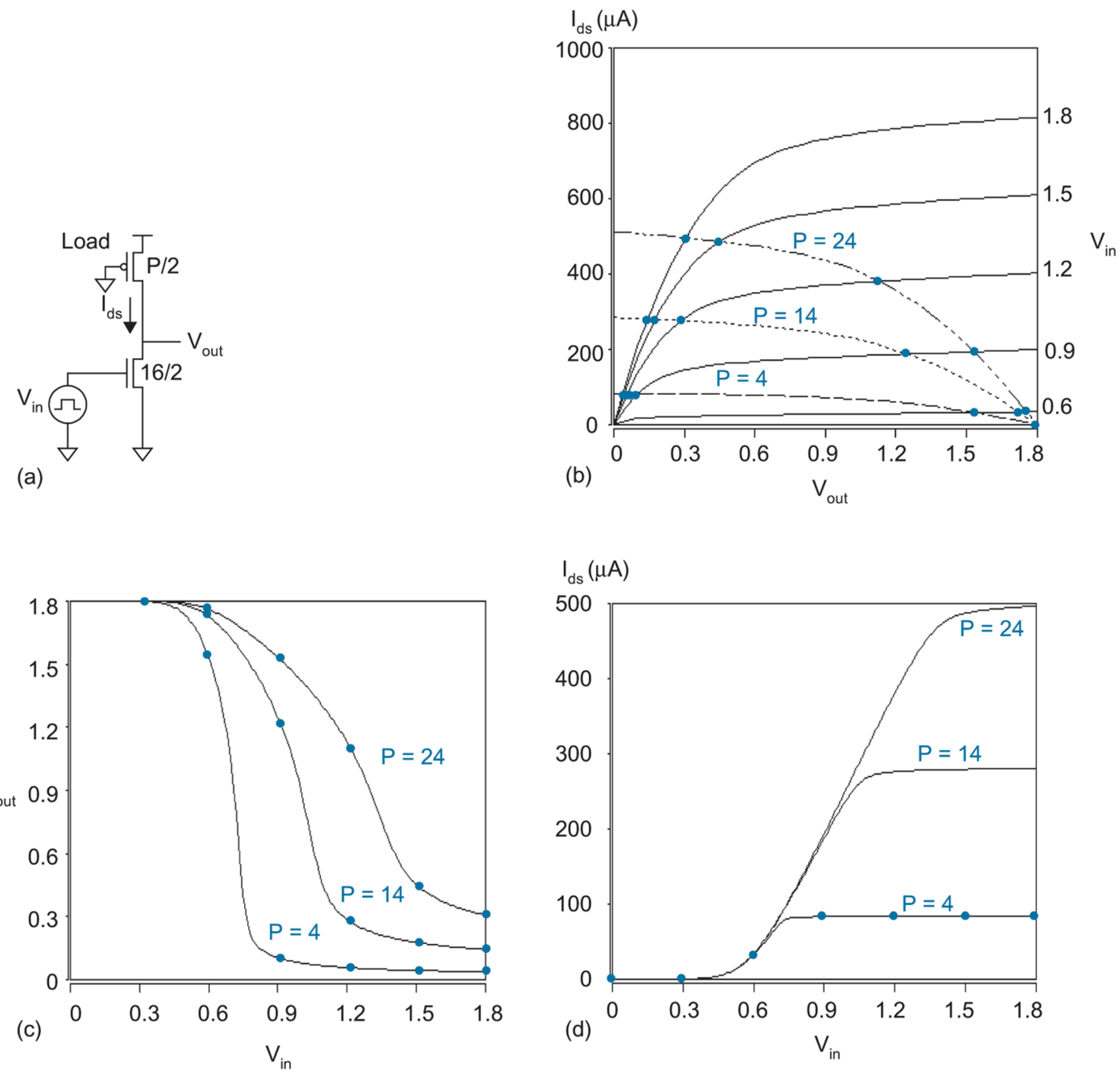
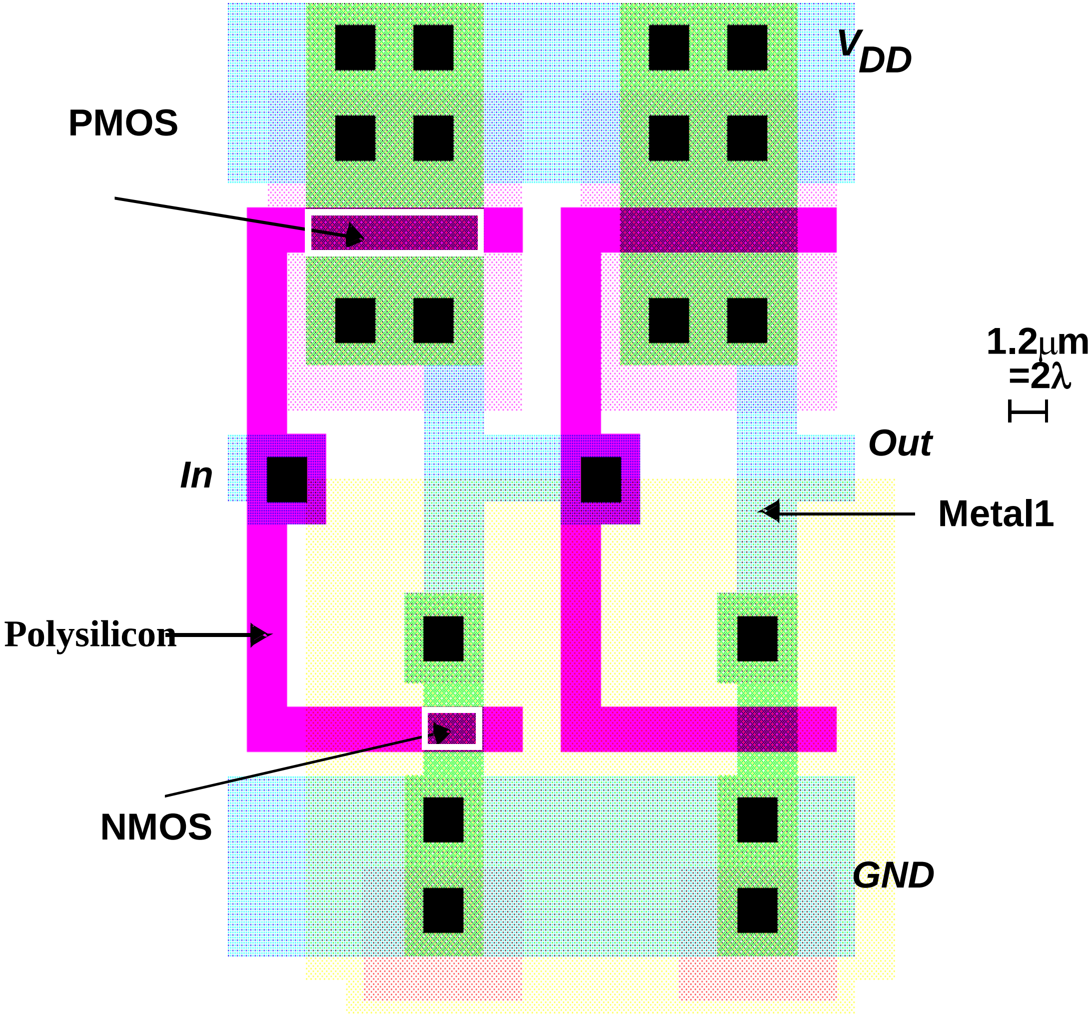


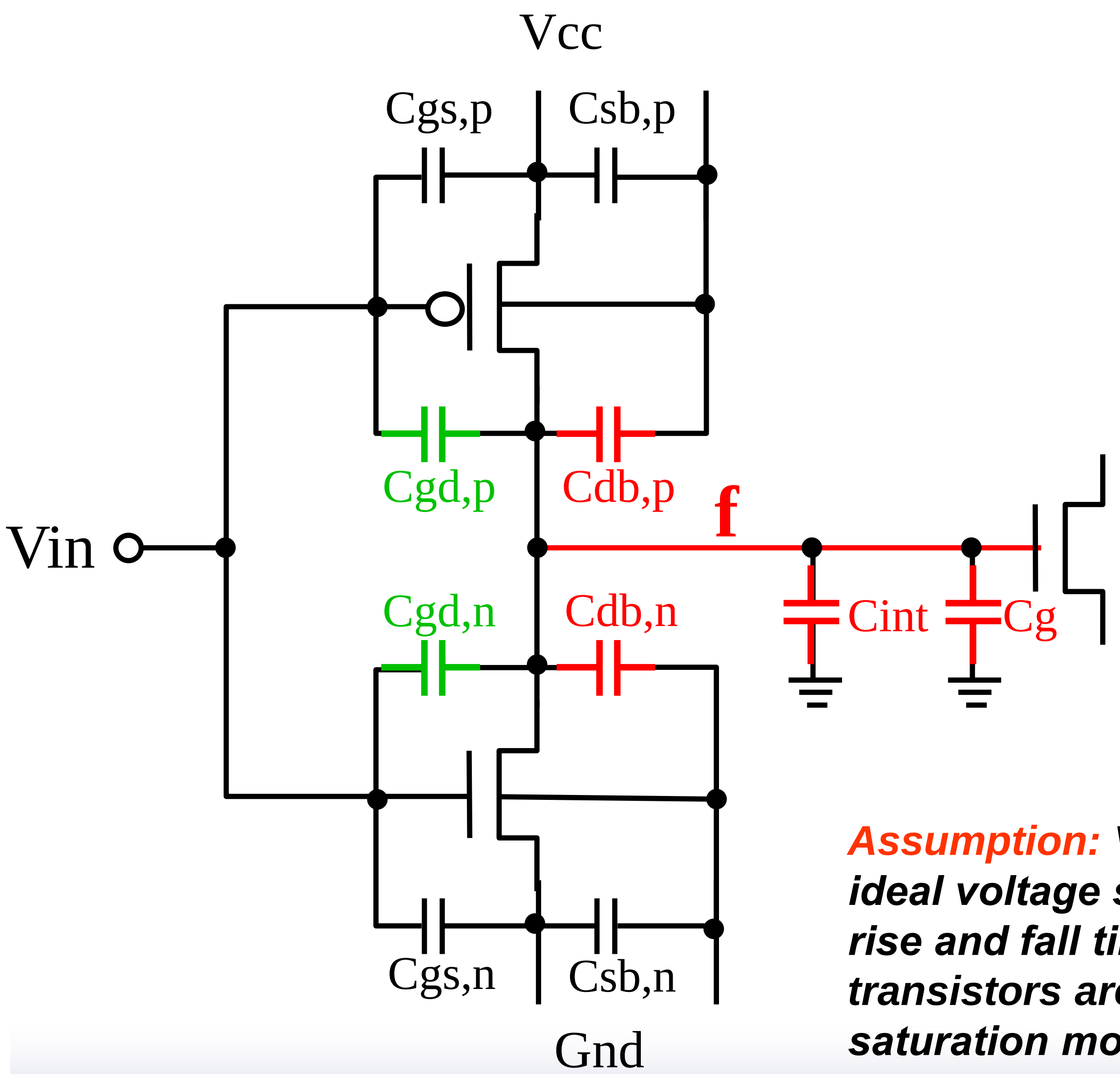
FIG 2.30 Pseudo-nMOS inverter and DC transfer characteristics

Propagation Delay

Two CMOS Inverters…



CMOS inverter capacitances



Cap on Node f:

- Junction cap: $C_{db,p}$ and $C_{db,n}$
- Gate (overlap) capacitance $C_{gd,p}$ and $C_{gd,n}$ (beware of Miller effect)
- Interconnect cap: C_{int}
- Receiver gate cap: C_g

***Assumption:* Vin is driven by an ideal voltage source….with zero rise and fall times….hence the transistors are either in cut-off or saturation mode…hence, no channel capacitance**

CMOS inverter capacitances

$$C_{load} = \underbrace{C_{db,n} + C_{db,p}}_{} + \underbrace{C_{gd,n} + C_{gd,p}}_{} + C_{int} + \underbrace{C_{gate}}_{}$$

$$C_{db,n}, C_{db,p} = AK_{eq}C_{j0} + PK_{eqsw}C_{jsw}$$

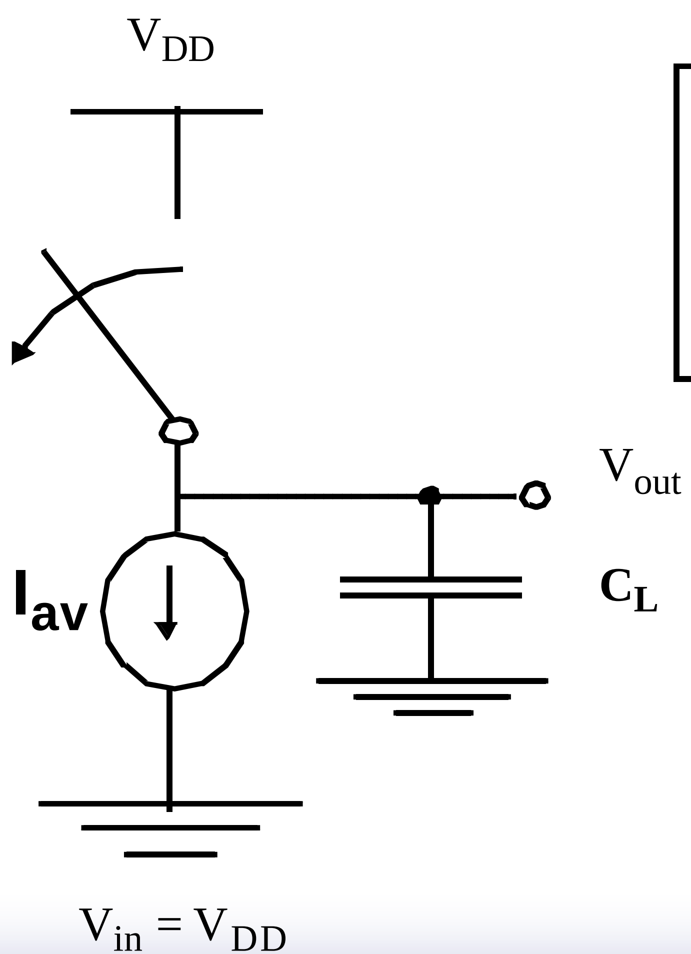
$$C_{gd,n}, C_{gd,p} = 2WX_{D}C_{ox}$$

Miller Effect

$$C_{gate} = WL_{drawn}C_{ox}$$

For each gate

CMOS Inverter Propagation Delay Approach 1



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

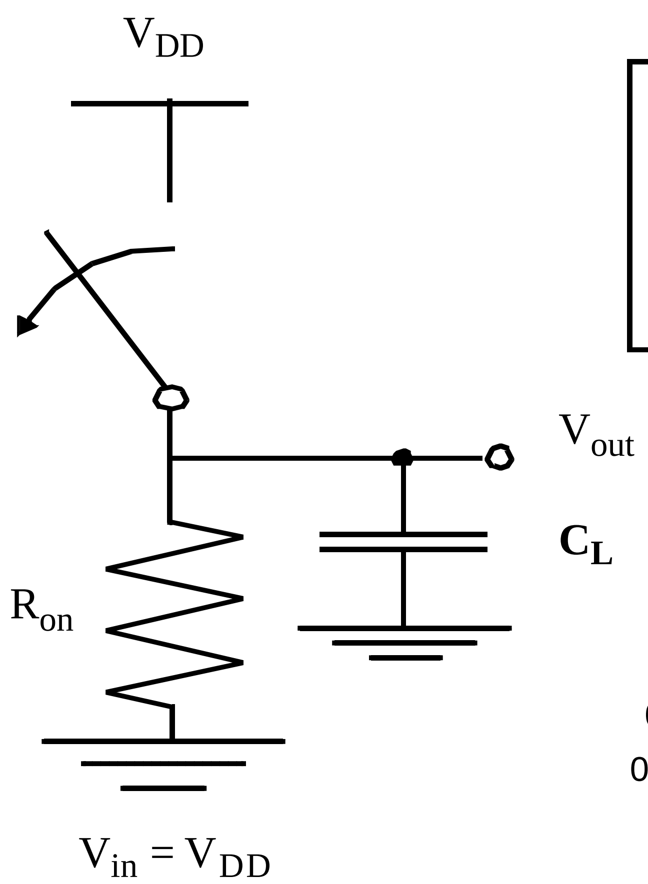
Note: $q = C\,V$

t_{pHL} is defined as time to discharge to 50% of the "High" value of the Vout… hence, Vswing/2

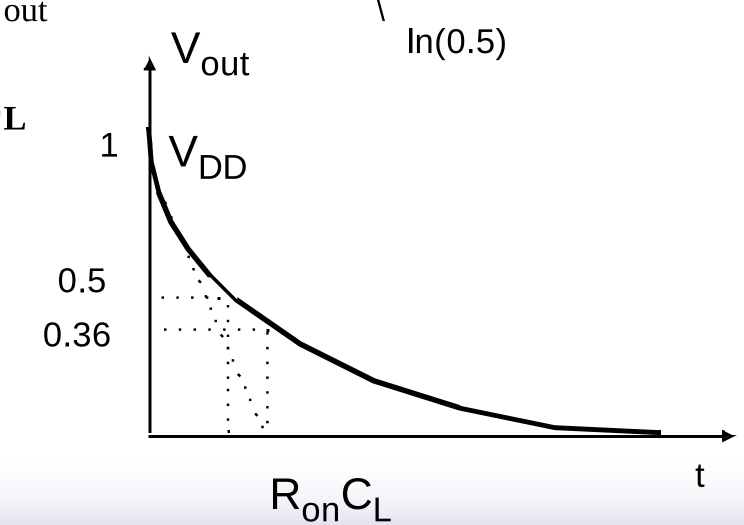
$C_L\, V_{swing}/2$ = charge
and charge/I_{av} = time

Kaustav Banerjee

CMOS Inverter Propagation Delay Approach 2

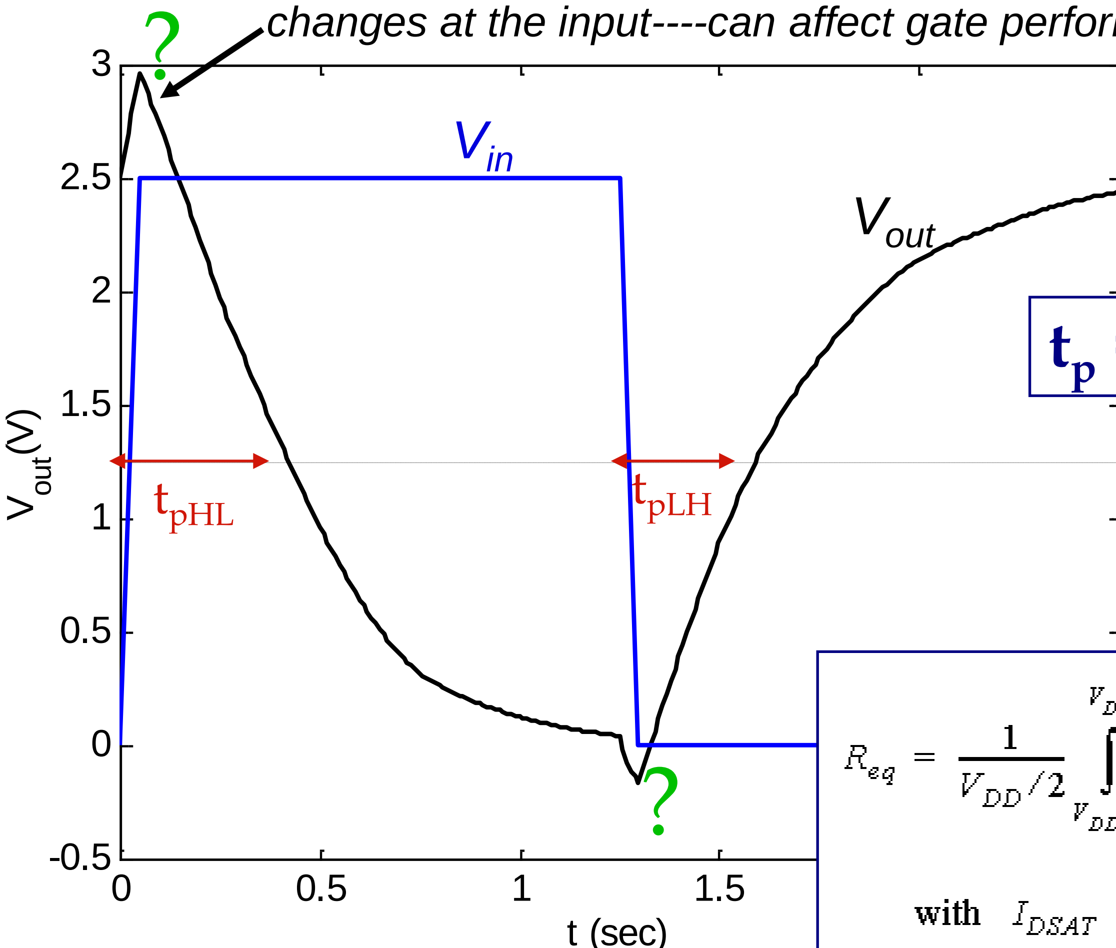


$$t_{pHL} = f(R_{on}.C_L)$$
$$= 0.69\ R_{on}C_L$$



Transient Response

***Due to C_{gd} of transistors**: directly couples voltage at input to output before the transistors can even start to react to changes at the input----can affect gate performance*



Symmetric inverter has $t_{pHL}=t_{pLH}$

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

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

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

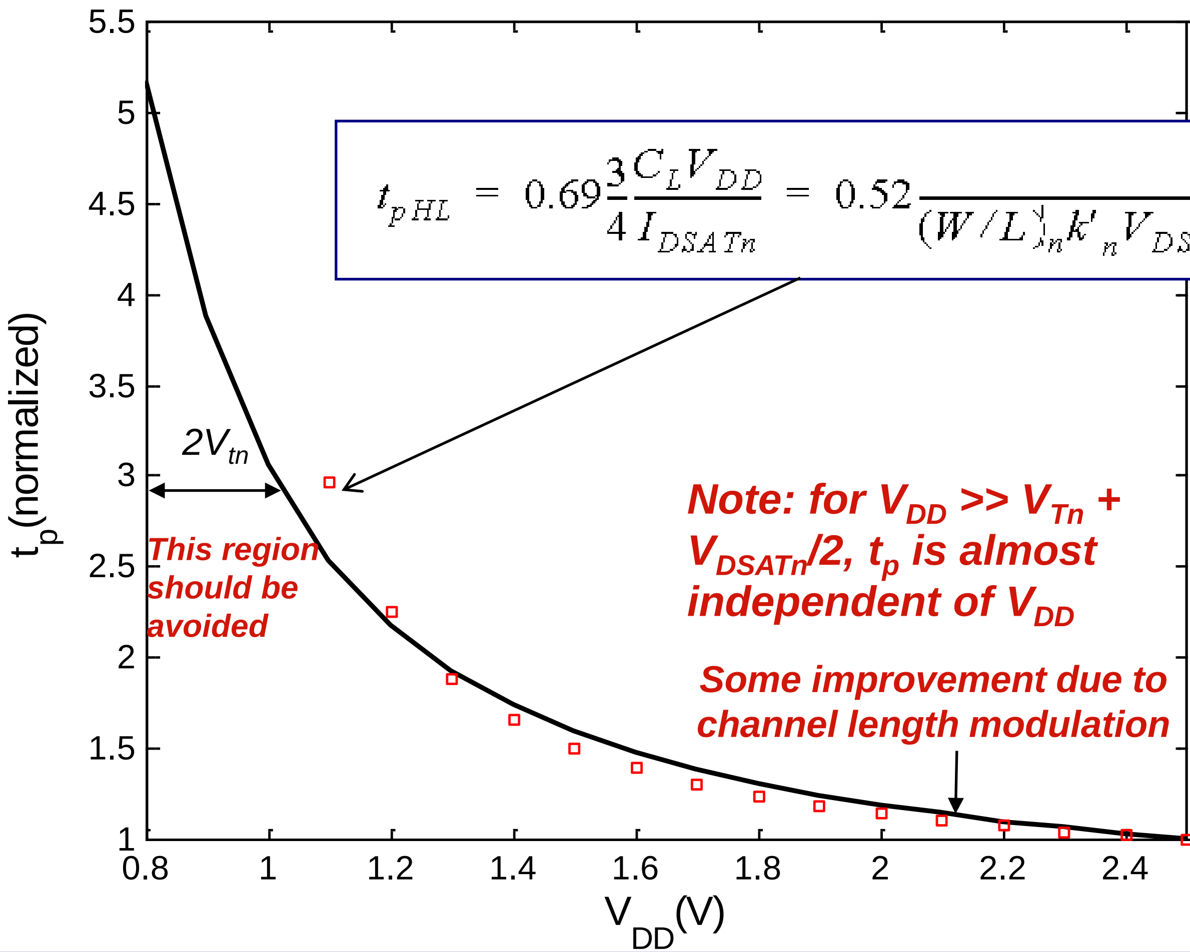
CMOS Inverter Delay as a function of V_{DD}

Same trend as the ON resistance of a transistor….

Trade off energy dissipation vs performance…..

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

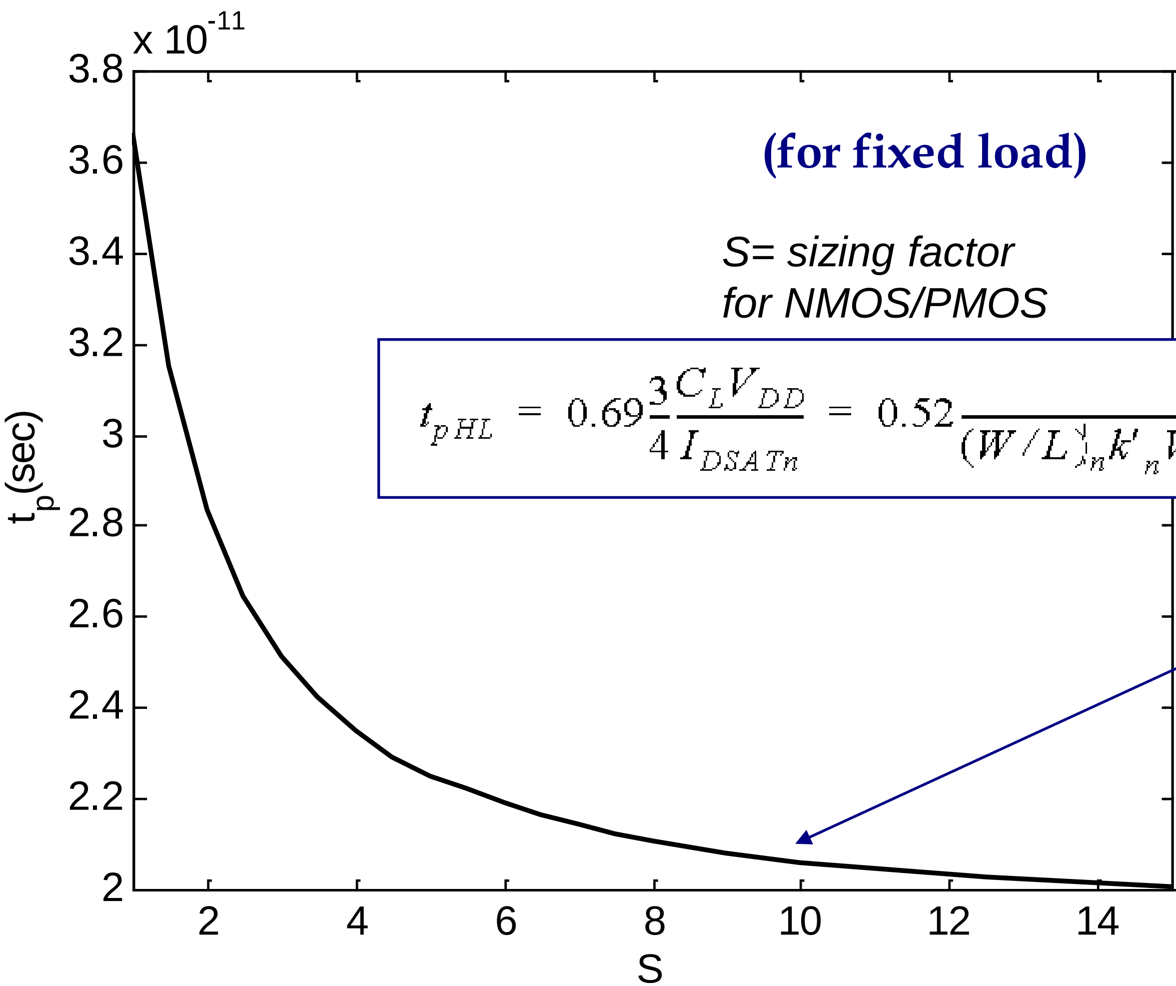
Only valid for velocity saturated devices…..



Reliability concerns at high V_{DD}….

Kaustav Banerjee

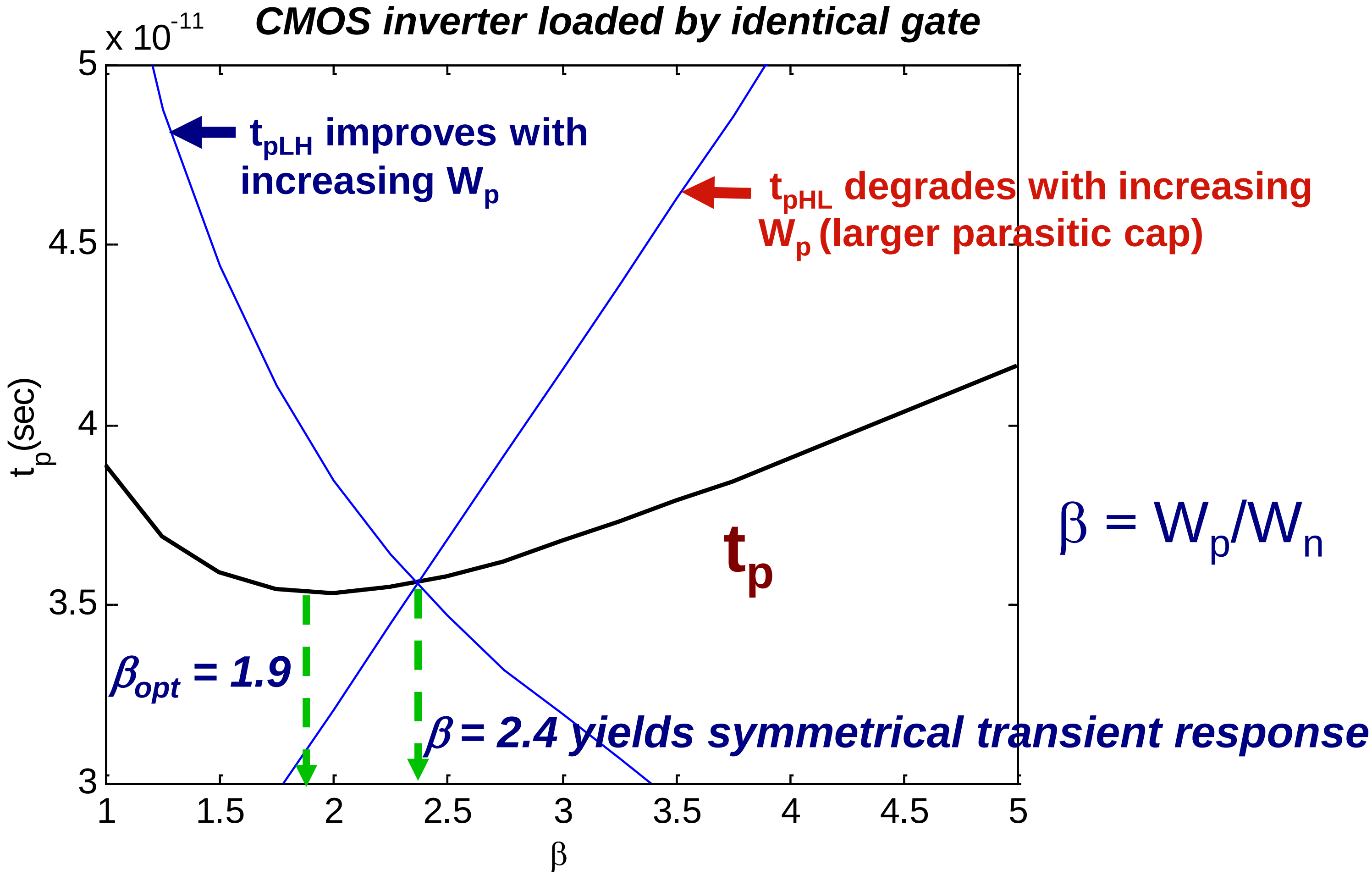
Effect of Device Sizing



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

Self-loading effect: Intrinsic (diffusion) capacitances dominate

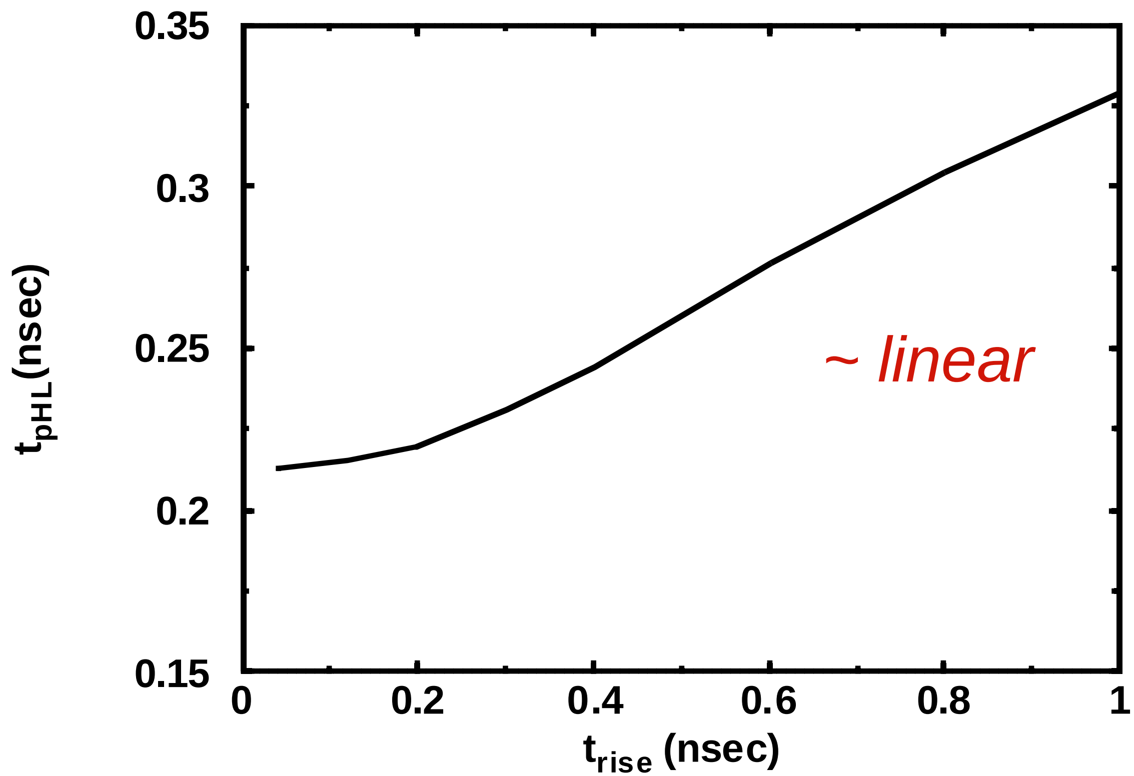
NMOS/PMOS ratio



$\beta = W_p/W_n$

If symmetry and noise margins are not of prime concern, inverter delay can be reduced by reducing the width of PMOS….

Impact of Rise Time on Delay



As the input signal changes gradually, both PMOS and NMOS conduct simultaneously….

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

Design for Performance

- Keep load capacitances (C_L) small
  - Recall that three major components contribute to the load cap.
    - internal diffusion + overlap caps
    - interconnect cap.
    - fan-out (gate cap)
- Increase transistor sizes
  - watch out for self-loading!
- Increase V_{DD} (??)
  - watch out for reliability issues!

Power Dissipation

Where Does Power Go in CMOS?

- **Dynamic Power**

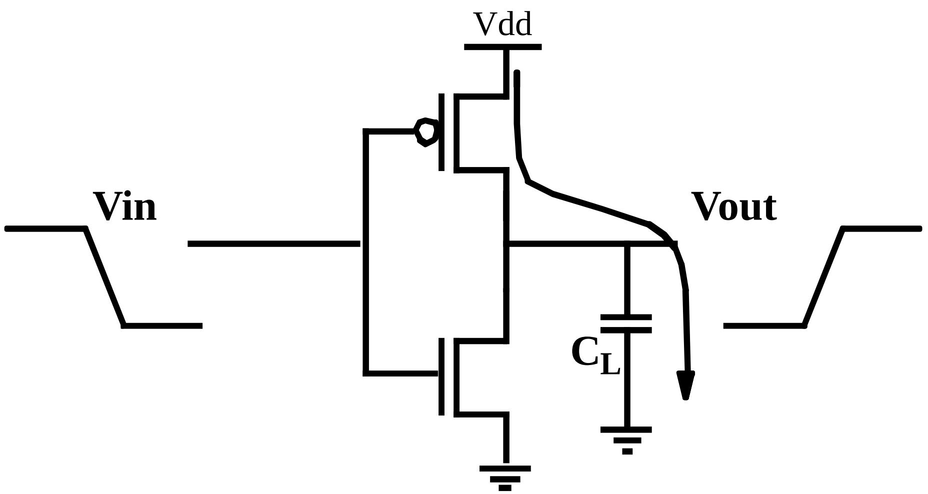
  Due to charging/discharging of capacitors….

- **Leakage Power**
  - *Subthreshold*
  - *Gate*
  - *Junction*

- **Short-Circuit Power**

  When NMOS and PMOS are both turned ON….

Dynamic Power Dissipation



Note: Here C_L is an external capacitor….

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

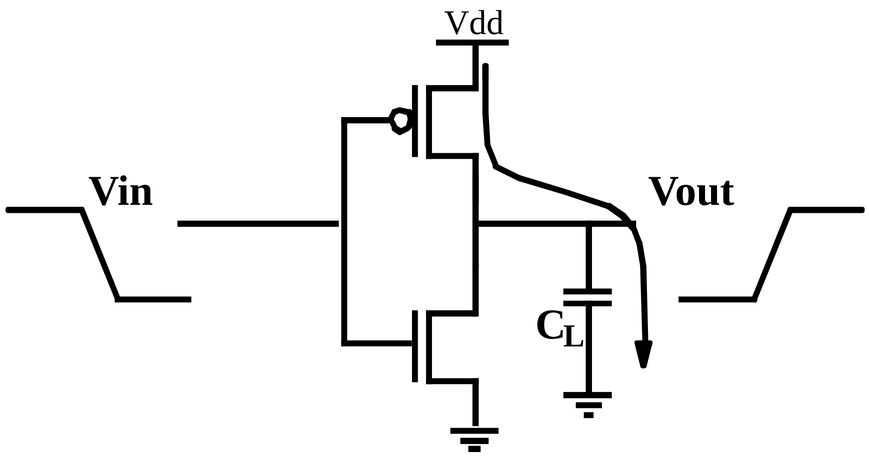
Power = Energy/transition * f = $C_L * V_{dd}^2 * f$

- **Energy/transition is not a function of transistor sizes!**
- **Need to reduce C_L, V_{dd}, and f to reduce power.**

Energy taken from supply during transition:

$$E_{V_{DD}} = \int_0^{\infty} i_{V_{DD}}(t) V_{DD}\, dt = V_{DD} \int_0^{\infty} C_L \frac{dv_{out}}{dt}\, dt = C_L V_{DD} \int_0^{V_{DD}} dv_{out} = C_L V_{DD}^2$$

Dynamic Power Dissipation



Note: during the discharge phase, charge is removed from C_L and its energy is dissipated in the NMOS

Energy taken from supply during transition:

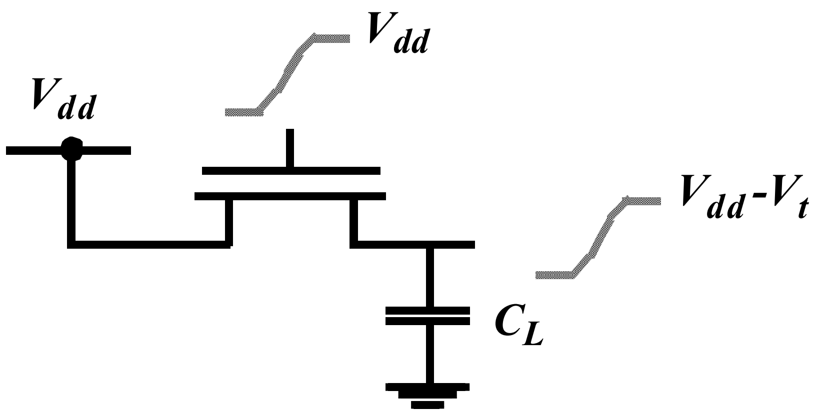
$$E_{V_{DD}} = \int_0^{\infty} i_{V_{DD}}(t) V_{DD}\, dt = V_{DD} \int_0^{\infty} C_L \frac{dv_{out}}{dt} dt = C_L V_{DD} \int_0^{V_{DD}} dv_{out} = C_L V_{DD}^2$$

Energy stored in the capacitor:

$$E_C = \int_0^{\infty} i_{V_{DD}}(t) v_{out}\, dt = \int_0^{\infty} C_L \frac{dv_{out}}{dt} v_{out}\, dt = C_L \int_0^{V_{DD}} v_{out}\, dv_{out} = \frac{C_L V_{DD}^2}{2}$$

Where is the other half of the energy? ….Dissipated by the PMOS

Modification for Circuits with Reduced Swing

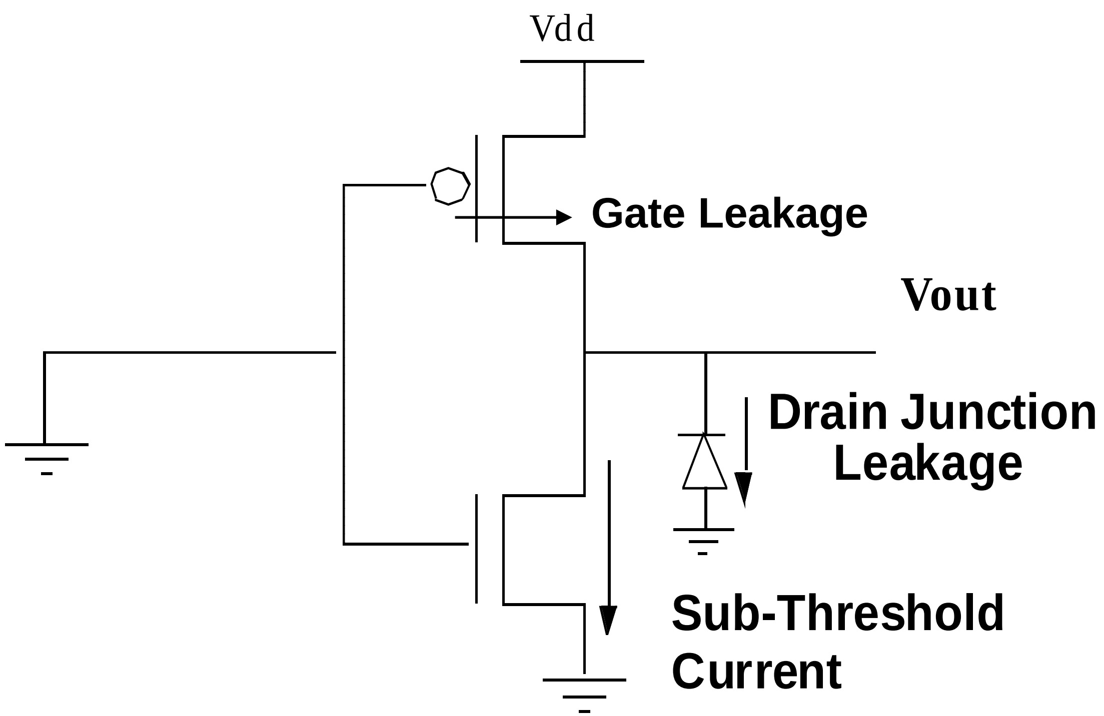


$$E_{0 \rightarrow 1} = C_L \bullet V_{dd} \bullet (V_{dd} - V_t)$$

Same as previous equation for energy drawn from supply voltage ($C_L\ V_{dd}^2$)…. integration of V_{out} yields (V_{dd}-V_t)

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

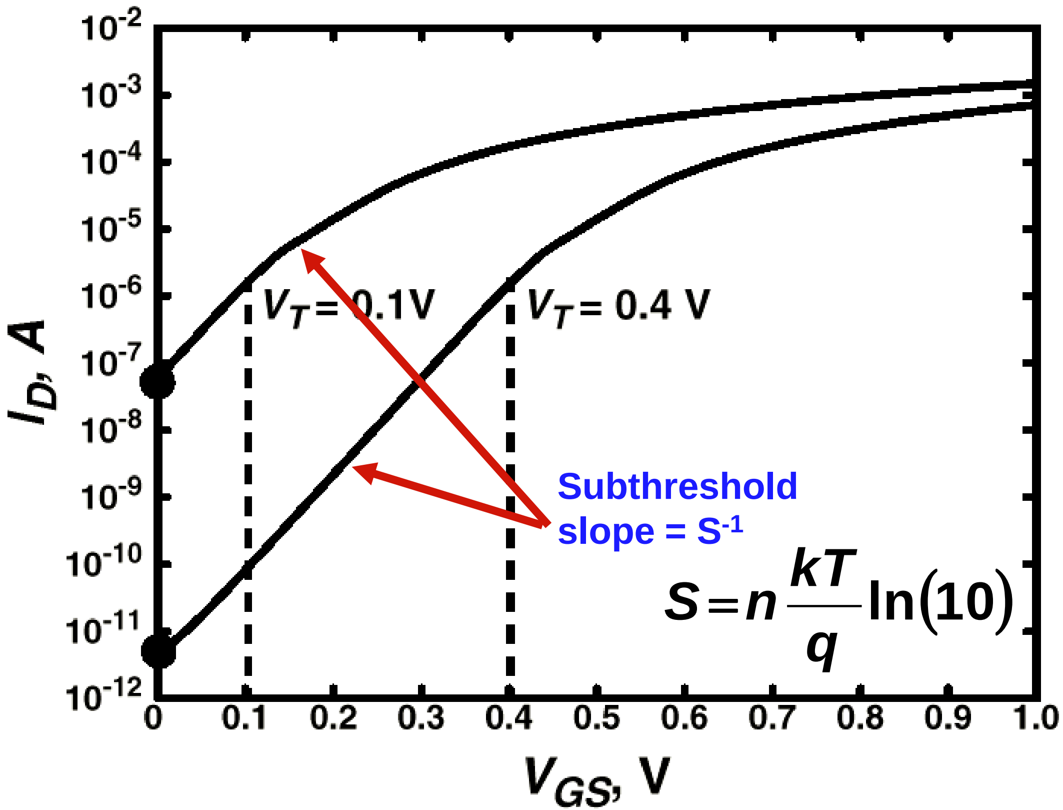
Primary Leakage Mechanisms

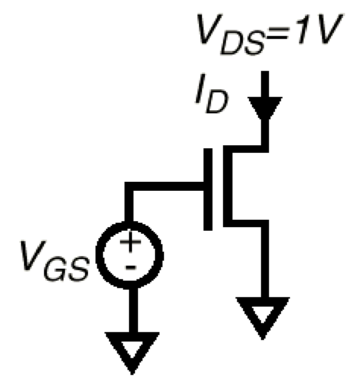


Sub-threshold current one of most compelling issues in low-energy circuit design!

Subthreshold Leakage Component

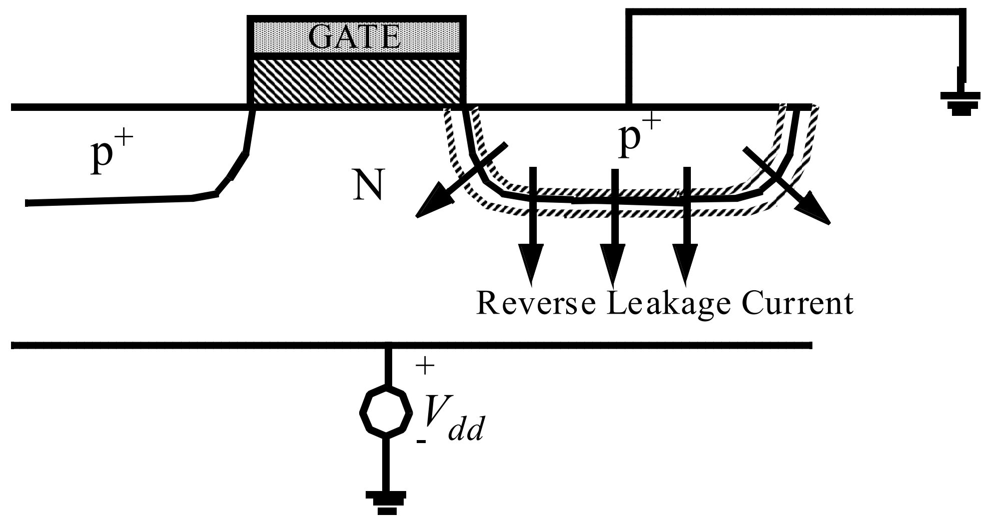
Subthreshold Swing = S = 60 mV/decade at room temperature (for ideal transistor with n=1)





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

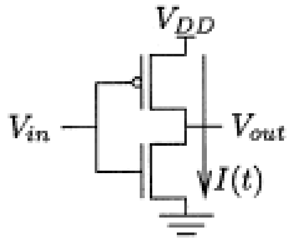
Reverse-Biased Diode Leakage



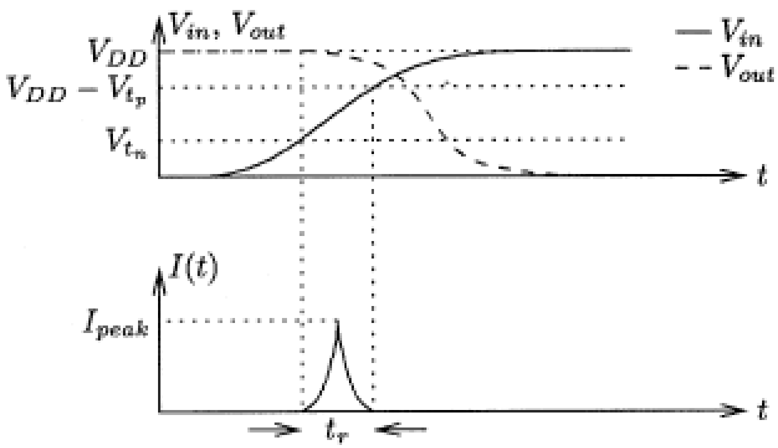
$$I_{DL} = J_S \times A$$

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

Short Circuit Currents



(a) CMOS inverter.



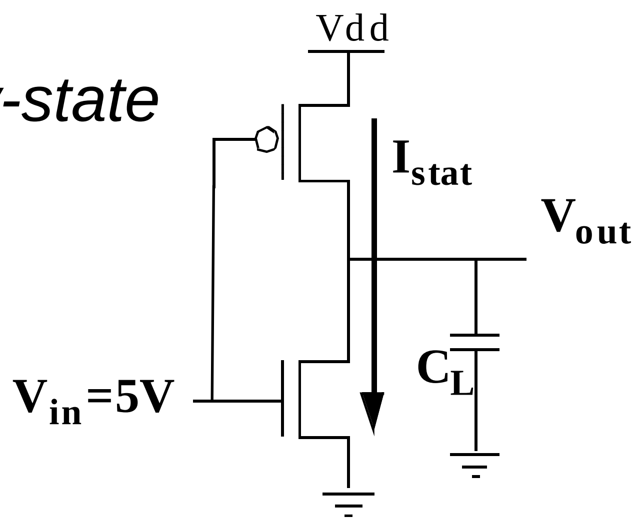
(b) Voltage and current waveforms.

Fig. 5. Voltage and current waveforms of a CMOS inverter.

K. Banerjee and A. Mehrotra, IEEE Transactions on Electron Devices, Vol. 49, No. 11, 2002.

Static Power Consumption

In the absence of switching….steady-state operation



Sources:

Due to all previously mentioned leakage currents……

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

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

Leakage

- Effect of leakage current
  - "Wasted" power: power consumed even when circuit is inactive
  - Leakage power raises temperature of chip
  - Can cause functionality problem in some circuits: memory, dynamic logic, etc.
- Reducing transistor leakage
  - Long-channel devices
  - Small drain voltage
  - Large threshold voltage V_T

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 (~0.3-0.4V)
- Reduce switching activity
- Reduce physical capacitance
  - Device Sizing

Leakage Power Reduction

- Process scaling
  - V_T reduces with each new process (historically)
  - Leakage increases ~10X!
- Leakage vs. performance tradeoff:
  - For high-speed, need small V_T and L
  - For low leakage, need high V_T and large L
- One solution: dual-V_T process
  - Low-V_T transistors: use in critical paths for high speed
  - High-V_T transistors: use to reduce power

Scaling Theory:

<table>
<tr><th colspan="2">Parameters</th><th>Constant Vdd scaling</th><th>Constant E scaling</th></tr>
<tr><td rowspan="4">Dimentions</td><td>Width = w</td><td colspan="2">0.7</td></tr>
<tr><td>Length = L</td><td colspan="2">0.7</td></tr>
<tr><td>Oxide thickness t_{ox}</td><td colspan="2">0.7</td></tr>
<tr><td>Junction depth X_j</td><td colspan="2">0.7</td></tr>
<tr><td colspan="2">Die Area</td><td colspan="2">$(0.7)^2$</td></tr>
<tr><td colspan="2">Gate capacitance per unit area ($C_{gox} = \varepsilon_{ox} / t_{ox}$)</td><td colspan="2">$\frac{1}{0.7}$</td></tr>
<tr><td colspan="2">Gate capacitance ($C_g = wLC_{gox}$)</td><td colspan="2">0.7</td></tr>
<tr><td colspan="2">Total Capacitance (C)</td><td colspan="2">0.7</td></tr>
<tr><td colspan="2">Supply Voltage (V_{DD})</td><td>1</td><td>0.7</td></tr>
<tr><td colspan="2">Current per device ($I_{DS} \propto \frac{w}{L} \frac{\varepsilon_{ox}}{t_{ox}} (V_{gs} - V_{th}) V_{DD}$)</td><td>1</td><td>0.7</td></tr>
<tr><td colspan="2">Intrinsic Gate Delay ($\tau = \frac{C_g \Delta V}{I_{AV}}$)</td><td>0.7</td><td>0.7</td></tr>
<tr><td colspan="2">Frequency ($f \propto \frac{1}{\tau}$)</td><td colspan="2">$\frac{1}{0.7}$</td></tr>
<tr><td colspan="2">Active Power Dissipation ($P_{active} = CV_{DD}^2 f$)</td><td>1</td><td>$(0.7)^2$</td></tr>
<tr><td colspan="2">Energy-Delay Product ($CV_{DD}^2 \tau$)</td><td>$(0.7)^2 = 0.49$</td><td>$(0.7)^4 = 0.2401$</td></tr>
<tr><td colspan="2">Power Dissipation density ($P_{active} / Area$)</td><td>$\frac{1}{(0.7)^2} \approx 2$</td><td>$\frac{(0.7)^2}{(0.7)^2} \approx 1$</td></tr>
</table>