ECE 124A

VLSI Principles

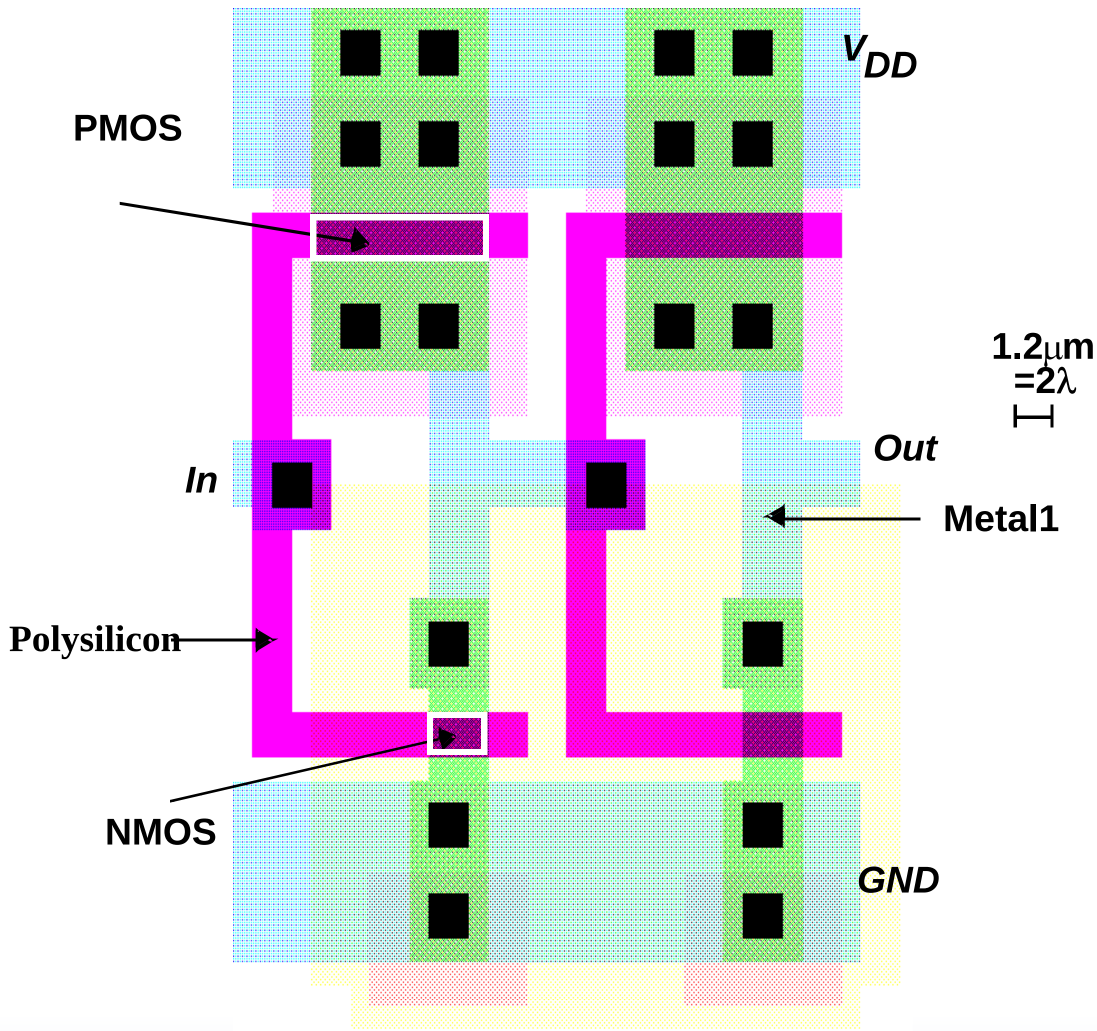
Lecture 10

CMOS Inverters: Delay, Power and Sizing

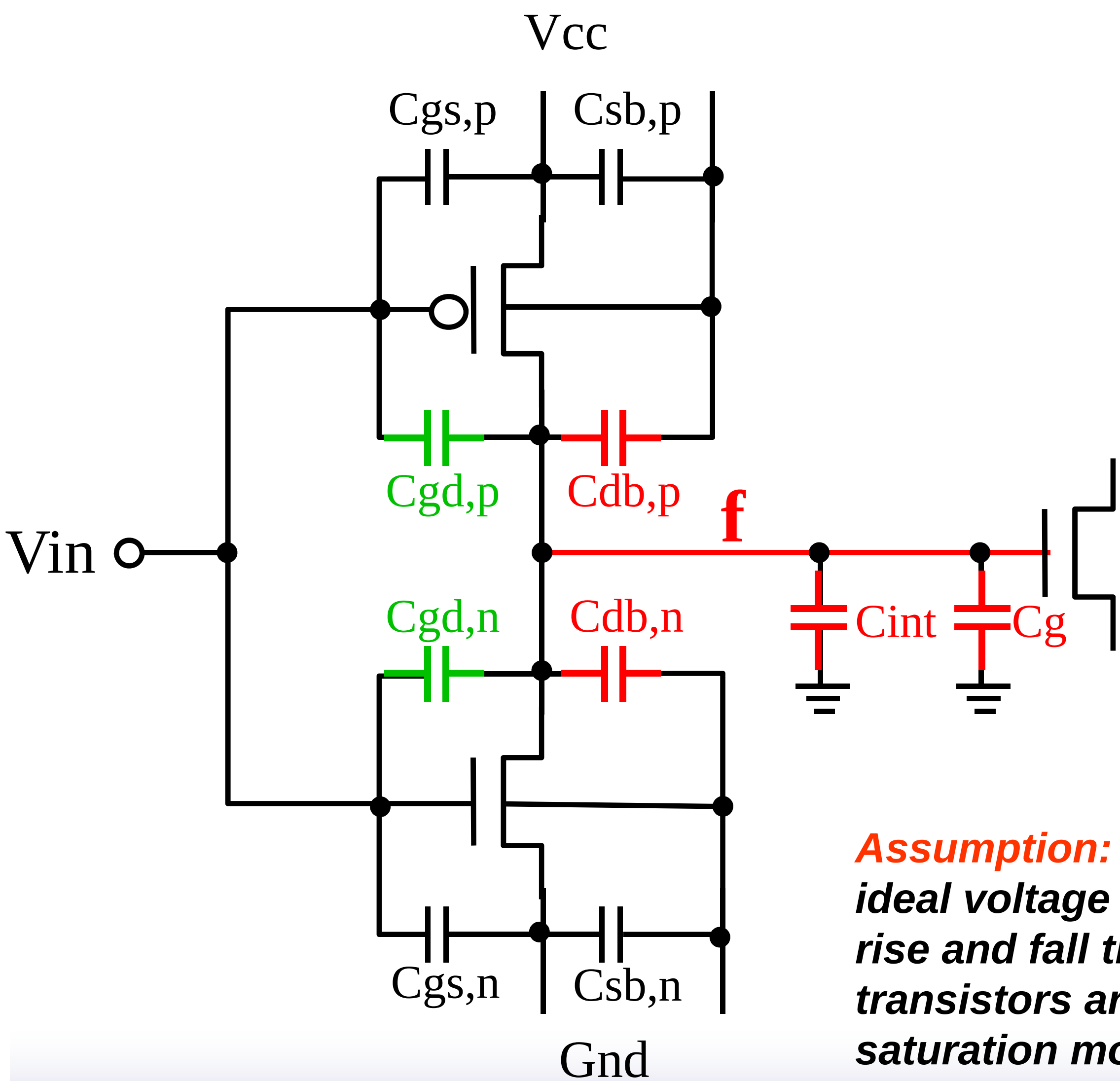
Prof. Kaustav Banerjee
Electrical and Computer Engineering
E-mail: kaustav@ece.ucsb.edu

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

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

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

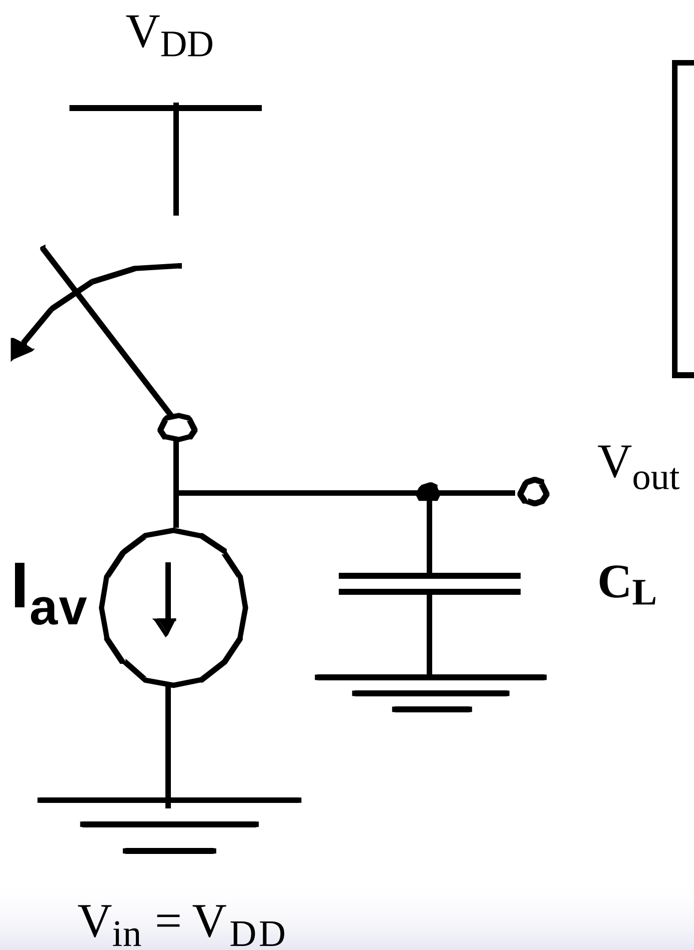
$$C_{gd,n}, C_{gd,p} = 2WX_D C_{ox}$$

Miller Effect

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

For each gate

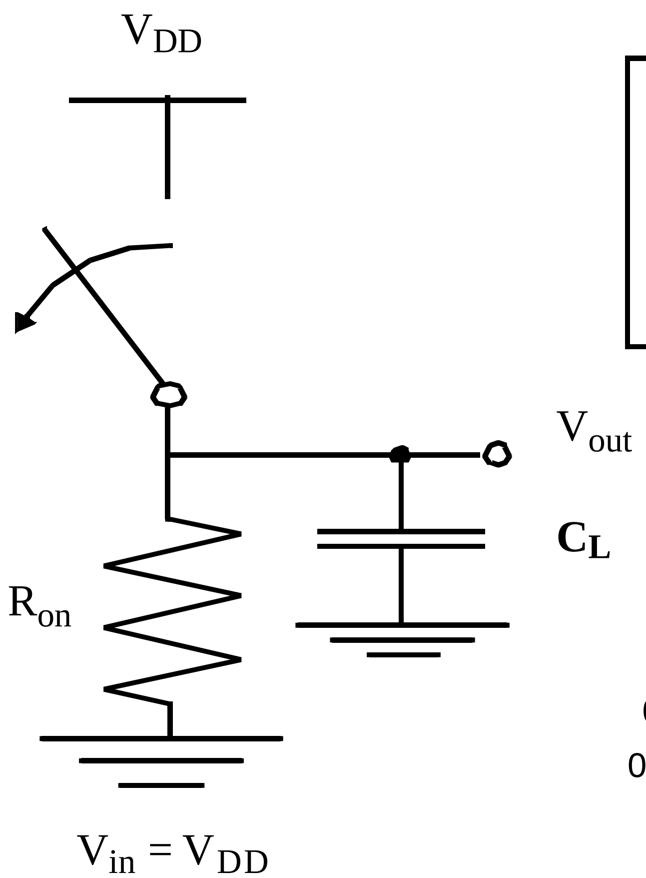
CMOS Inverter Propagation Delay Approach 1



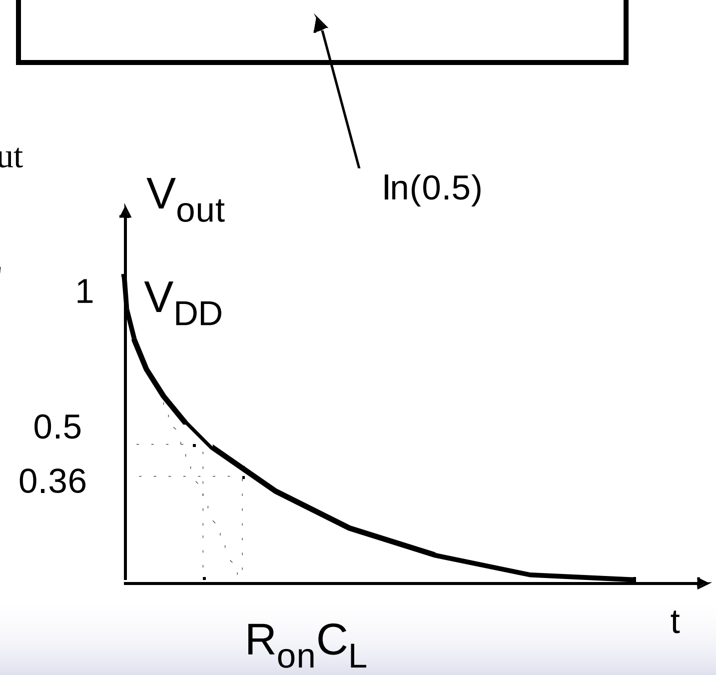
$$t_{pHL} = \frac{C_L V_{swing}/2}{I_{av}}$$

Note: q = C V

CMOS Inverter Propagation Delay Approach 2

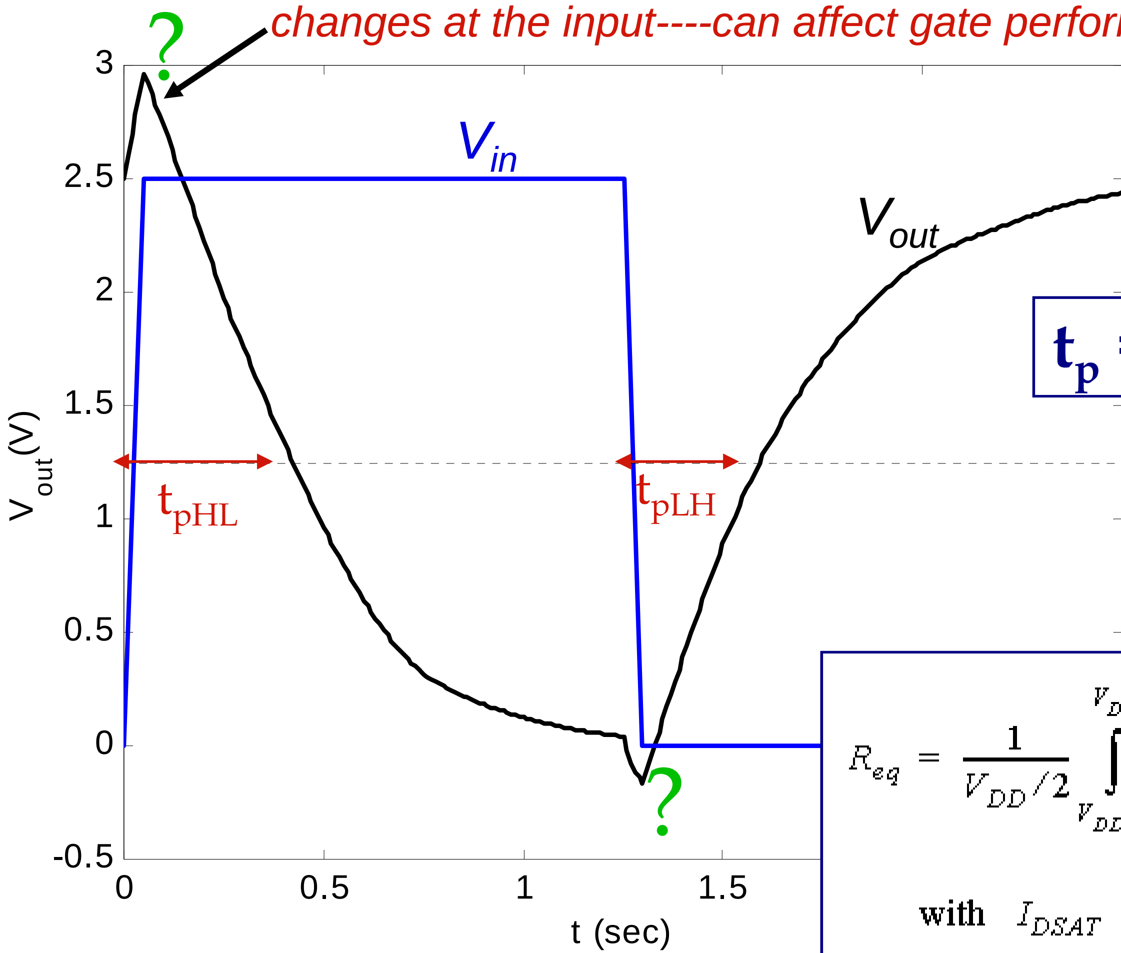


$$t_{pHL} = f(R_{on}.C_L)$$
$$= 0.69\, R_{on}C_L$$



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 $t_{pHL}=t_{pLH}$

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

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

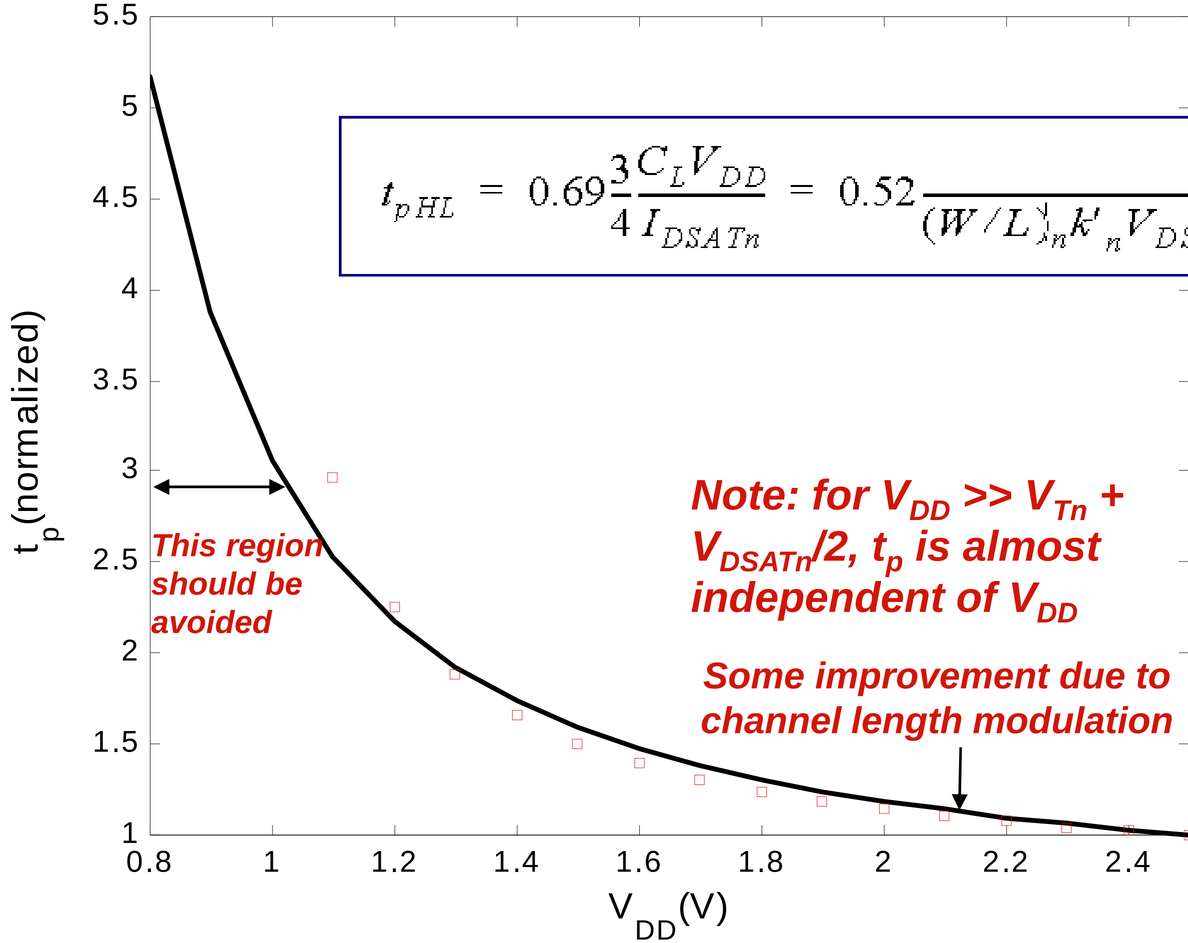
$$\text{with} \quad I_{DSAT} = k'\frac{W}{L}\left((V_{DD}-V_T)V_{DSAT}-\frac{V_{DSAT}^2}{2}\right)$$

Delay as a function of V_{DD}

Same as the ON resistance of a transistor….

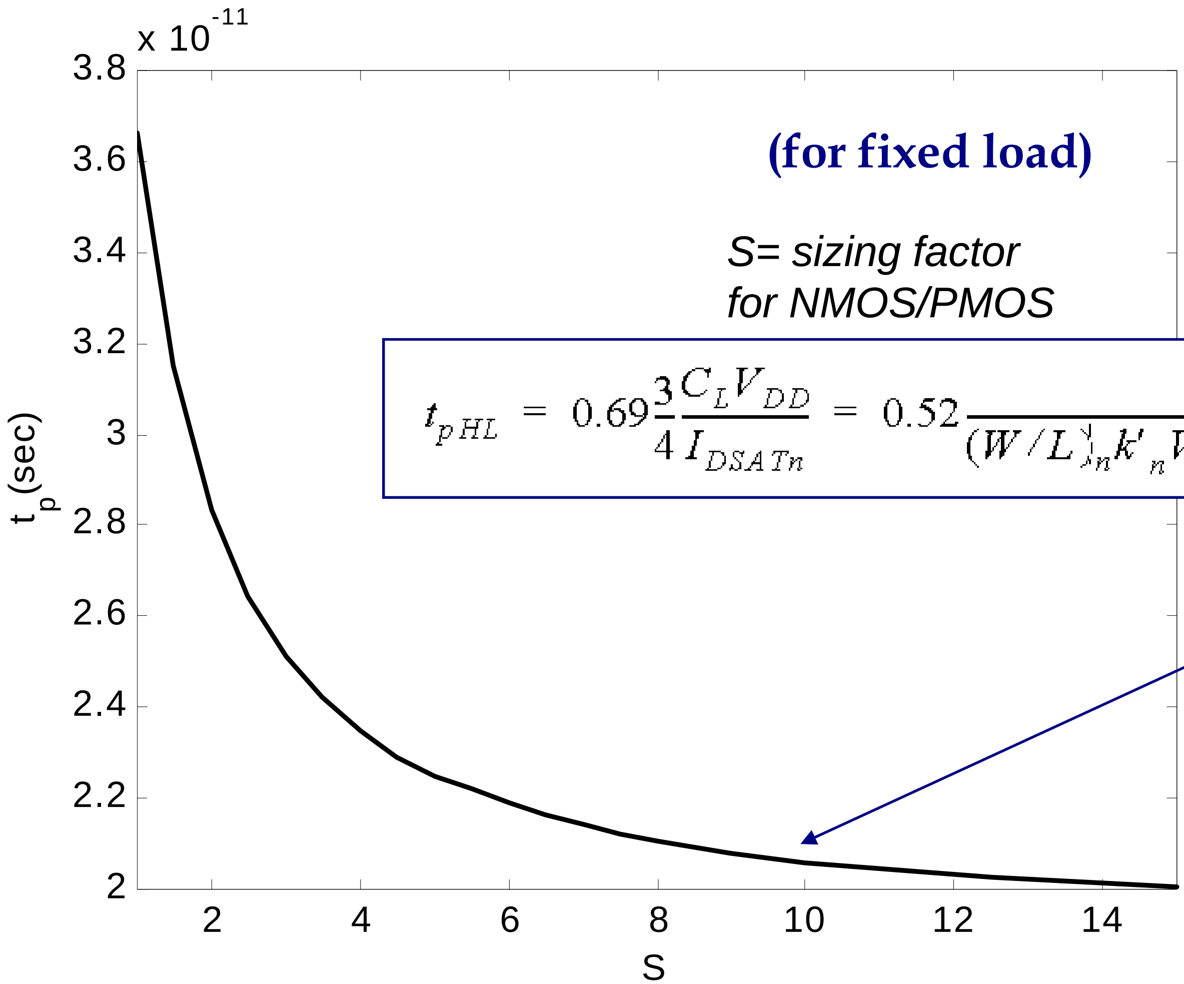
Trade off energy dissipation vs performance…..

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



Reliability concerns at high V_{DD}….

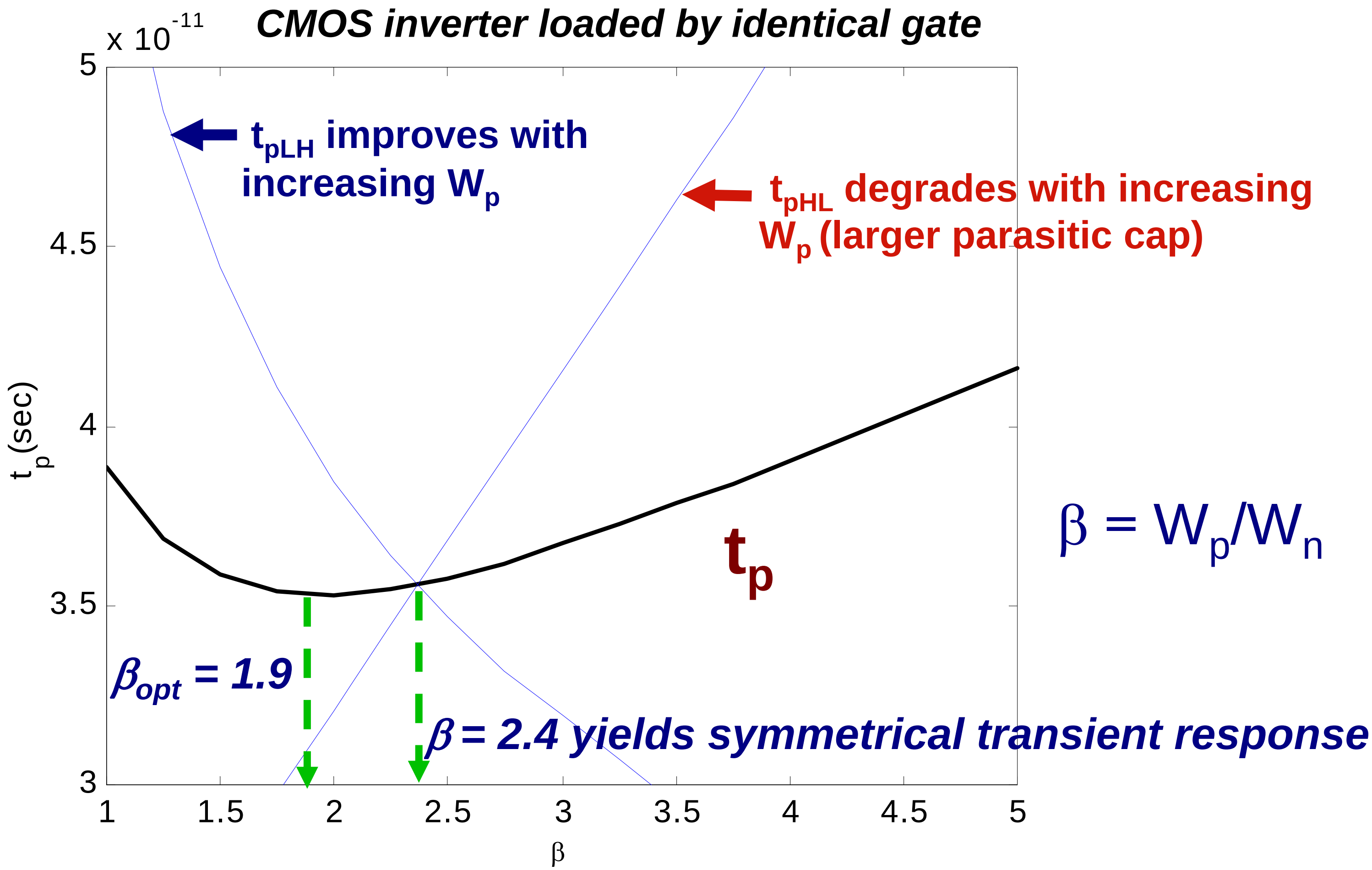
Effect of Device Sizing



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

Self-loading effect: Intrinsic (diffusion) capacitances dominate

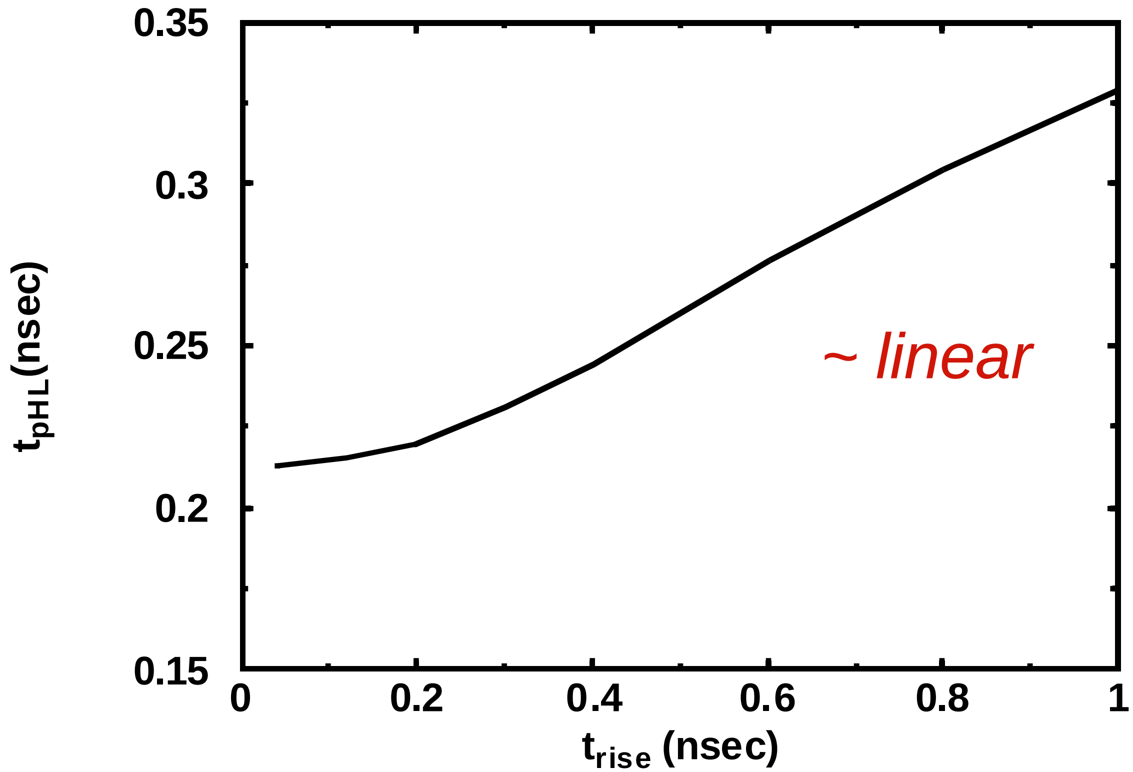
NMOS/PMOS ratio



$\beta = W_p/W_n$

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….

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

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….

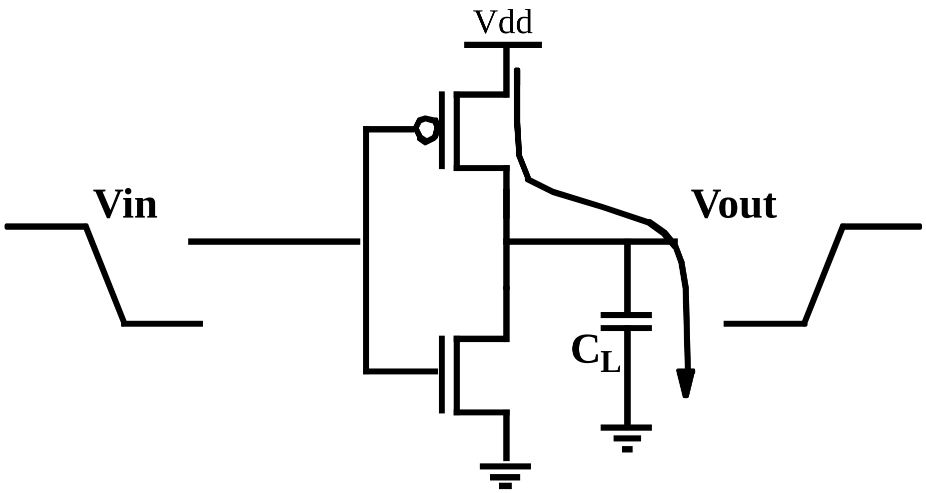
- **Leakage Power**

  - *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 = $C_L * V_{dd}^2$

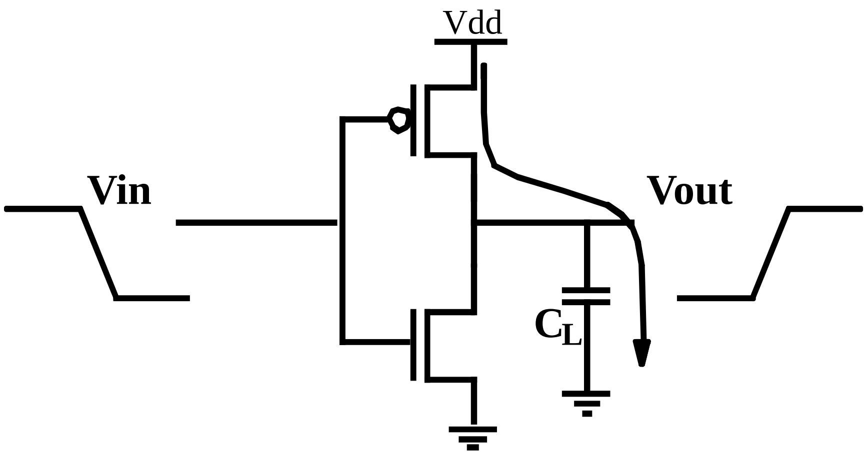
Power = Energy/transition * f = $C_L * V_{dd}^2 * f$

- **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:

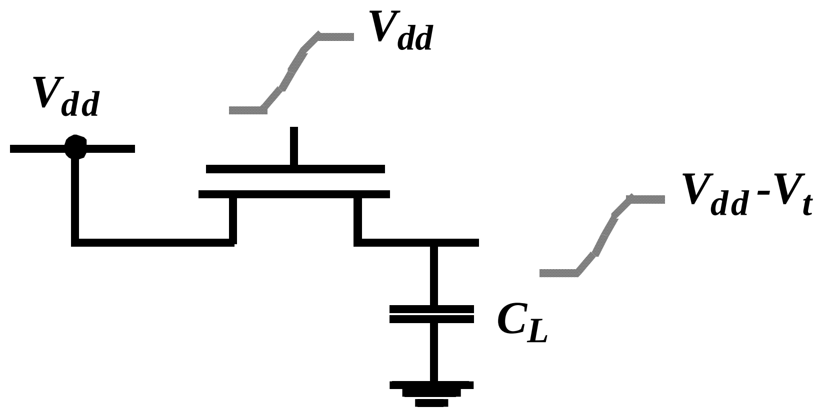
$$E_{V_{DD}} = \int_0^\infty i_{V_{DD}}(t) V_{DD}\, dt = V_{DD} \int_0^\infty C_L \frac{dv_{out}}{dt}\, dt = C_L V_{DD} \int_0^{V_{DD}} dv_{out} = C_L V_{DD}^2$$

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*

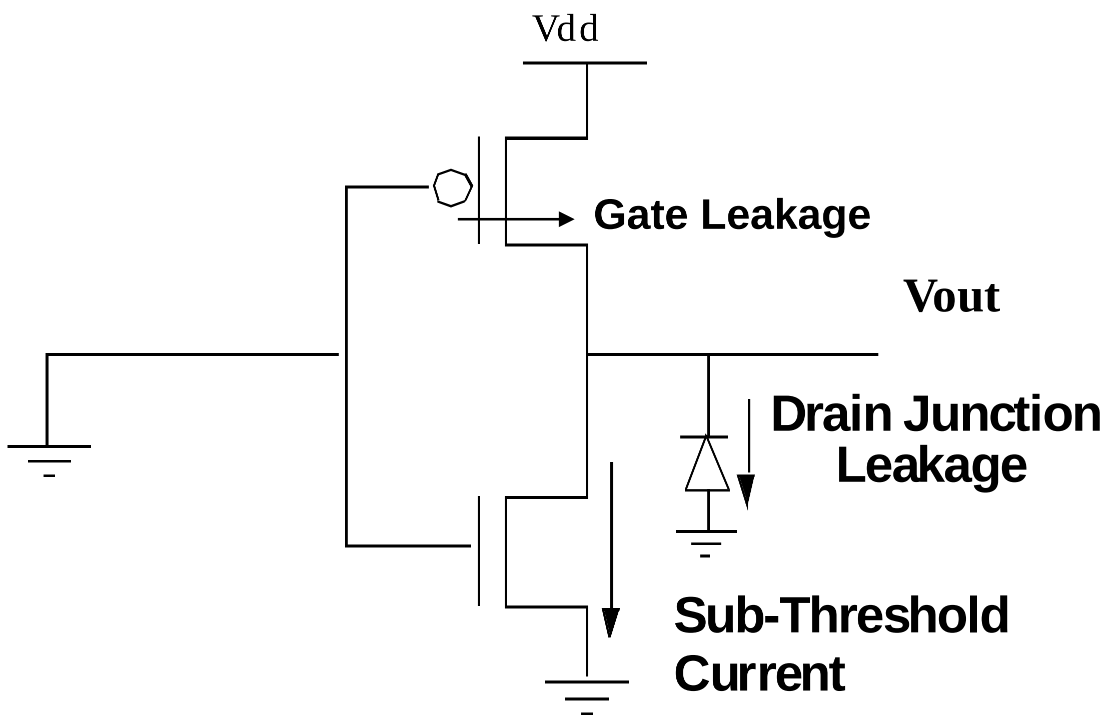
Modification for Circuits with Reduced Swing



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

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

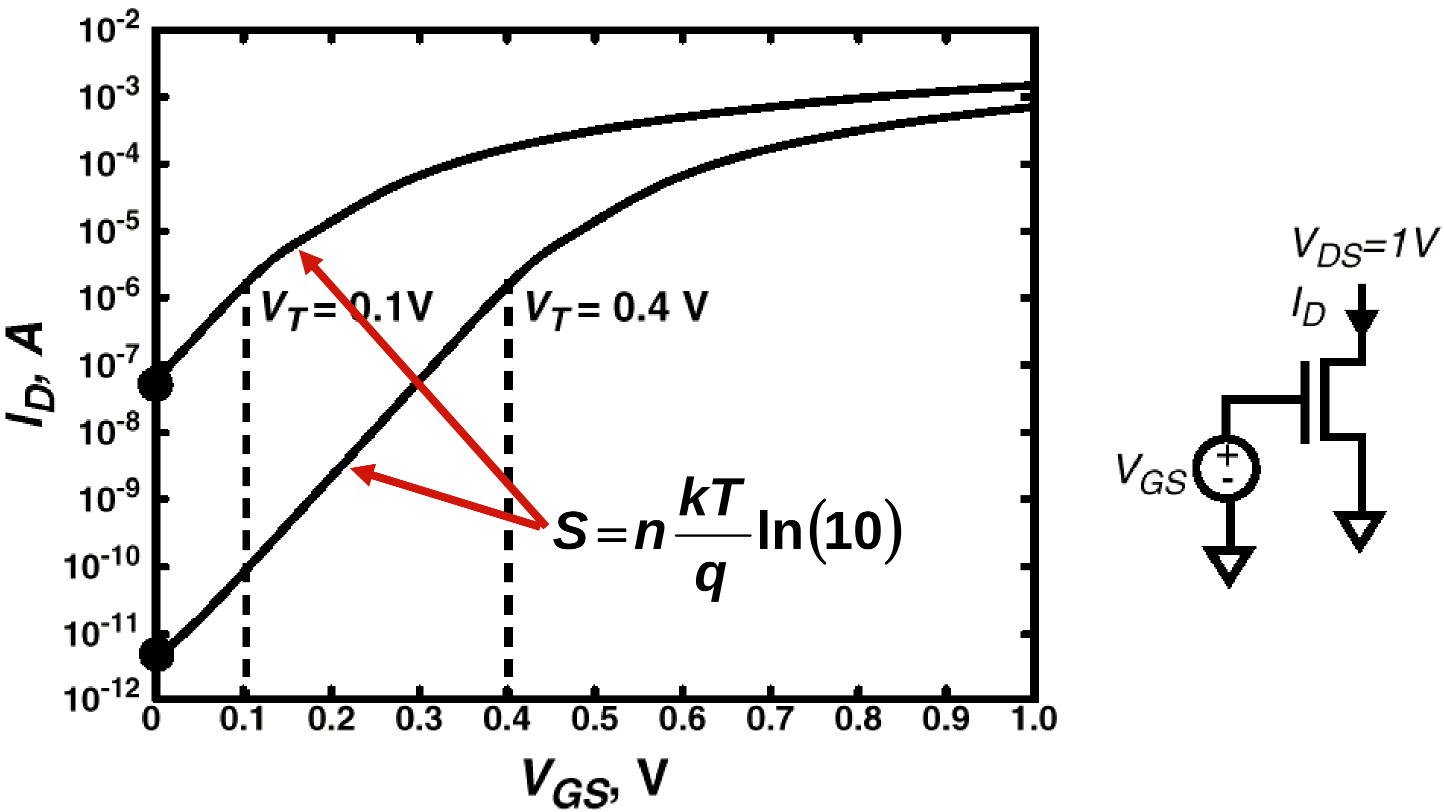
Primary Leakage Mechanisms



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

Subthreshold Leakage Component

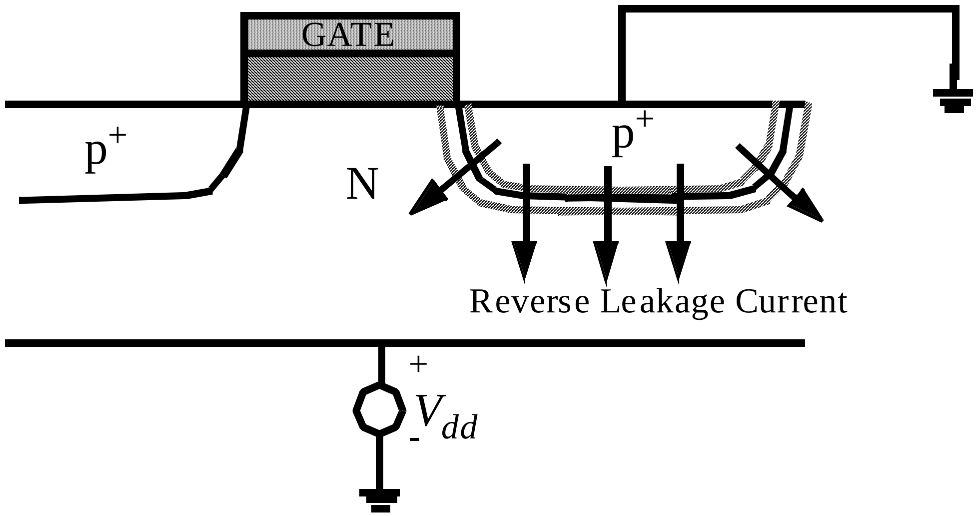
Subthreshold Slope = S = 60 mv/decade (for ideal transistor with n=1)



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

Kaustav Banerjee

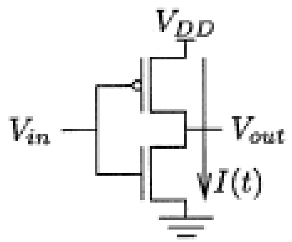
Reverse-Biased Diode Leakage



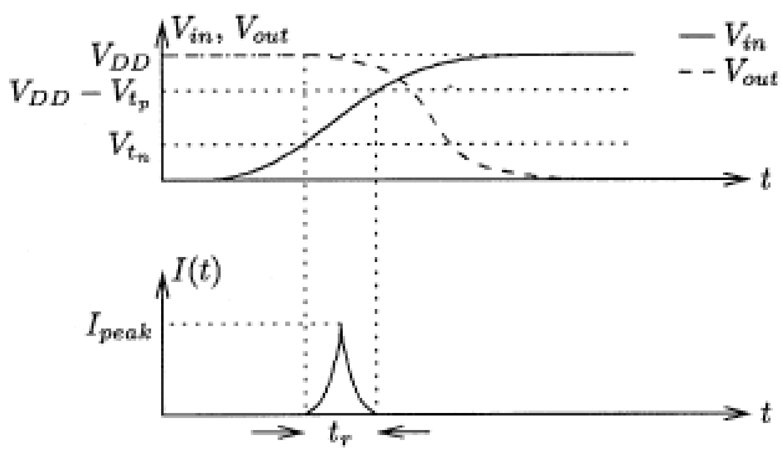
$$I_{DL} = J_S \times A$$

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.



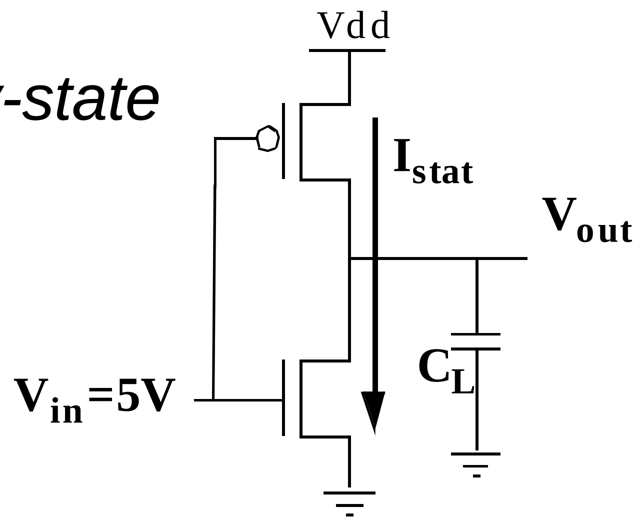
(b) Voltage and current waveforms.

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 (0.6 … 0.9 V by 2010!)
- Reduce switching activity
- Reduce physical capacitance
  - Device Sizing

Leakage Power Reduction

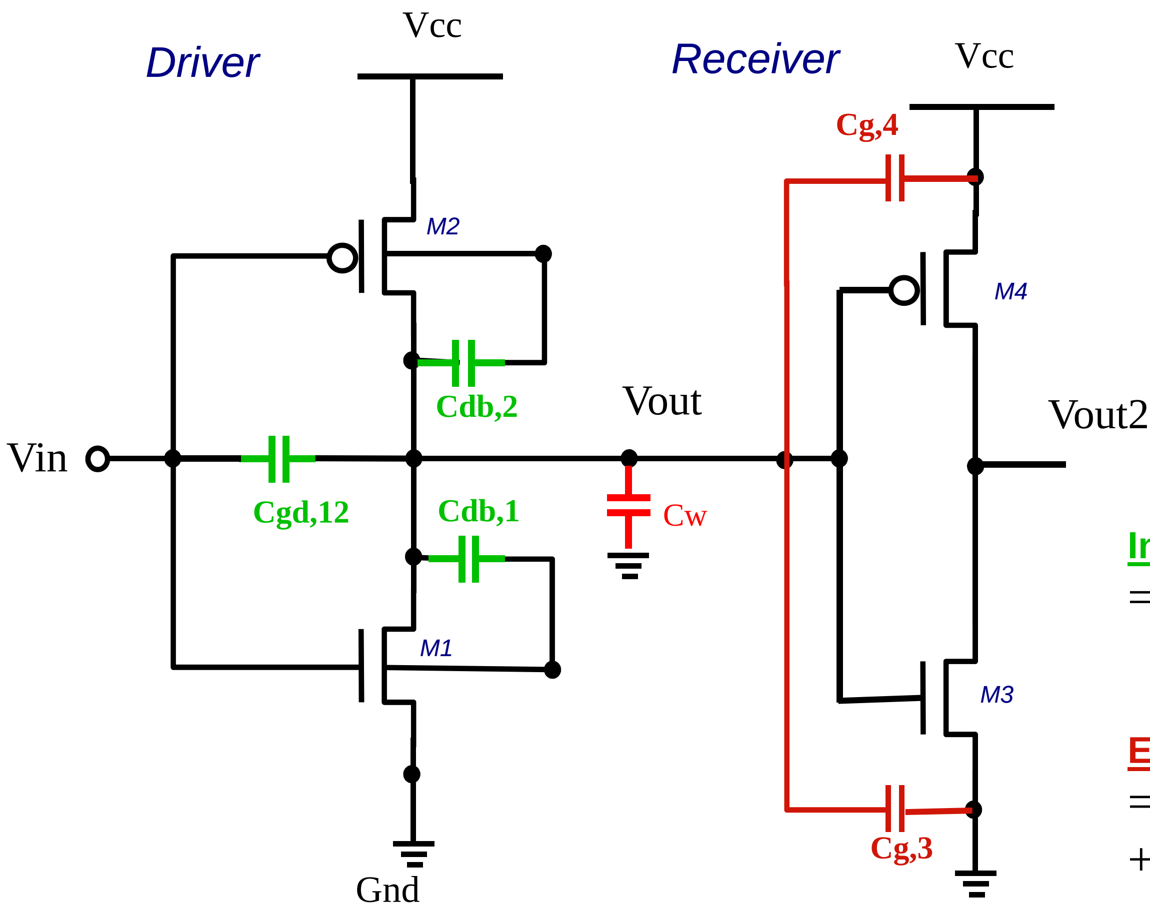
- Process scaling
  - V_T reduces with each new process (historically)
  - Leakage increases ~10X!
- Leakage vs. performance tradeoff:
  - For high-speed, need small V_T and L
  - For low leakage, need high V_T and large L
- One solution: dual-V_T process
  - Low-V_T transistors: use in critical paths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

Scaling Theory:

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

Inverter Sizing

Load capacitance



$$C_L = C_{int} + C_{ext}$$

Internal Caps of Driver (Cint):

= Junction caps: $C_{db,12}$ +
Gate caps: $C_{gd,12}$ (including Miller Caps.)

External Caps (Cext):

= Interconnect cap: C_w
+ Receiver gate caps: $C_{g,43}$

Intrinsic delay of CMOS inverter

Let R_{eq} be the equivalent resistance of the gate (inverter), then delay (t_p) is defined as:

$$t_p = 0.69\, R_{eq} \left(C_{\text{int}} + C_{ext} \right)$$

$$= 0.69\, R_{eq}\, C_{\text{int}} \left(1 + \frac{C_{ext}}{C_{\text{int}}} \right)$$

$$= t_{p0} \left(1 + \frac{C_{ext}}{C_{\text{int}}} \right)$$

t_{p0} is the intrinsic delay

Impact of sizing on gate delay

Let S be the sizing factor

R_{ref} be the resistance of a reference gate (usually a minimum size gate)

C_{iref} be the internal capacitance of the reference gate

$$C_{\text{int}} = S\,C_{iref}\ , \quad R_{eq} = \frac{R_{ref}}{S}$$

$$t_p = 0.69\left(\frac{R_{ref}}{S}\right)\left(S\,C_{iref}\right)\left(1+\frac{C_{ext}}{SC_{iref}}\right)$$

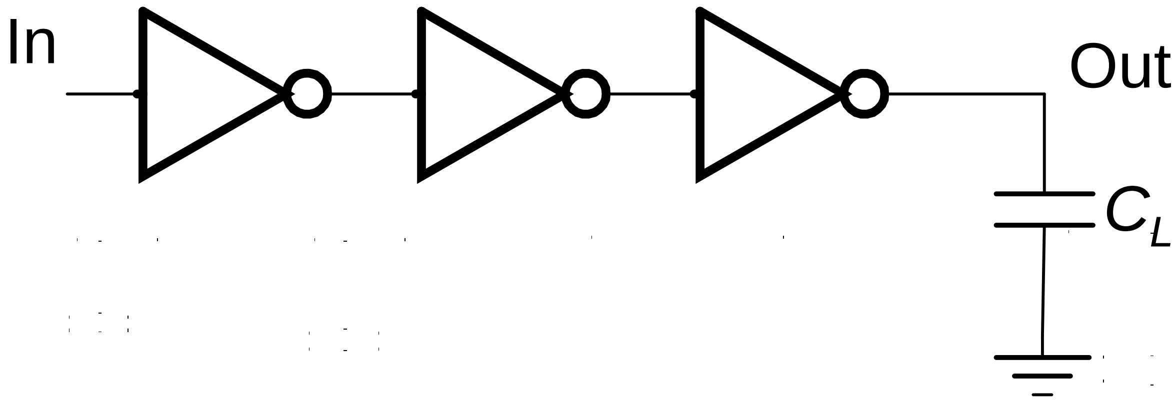
$$= 0.69\,R_{ref}\ C_{iref}\left(1+\frac{C_{ext}}{SC_{iref}}\right)$$

$$= t_{p0}\left(1+\frac{C_{ext}}{SC_{iref}}\right)$$

Hence:

1. ***Intrinsic delay is independent of gate sizing, and is determined only by technology and inverter layout***
2. ***If S is made very large, gate delay approaches the intrinsic value but increases the area significantly***

Inverter Chain

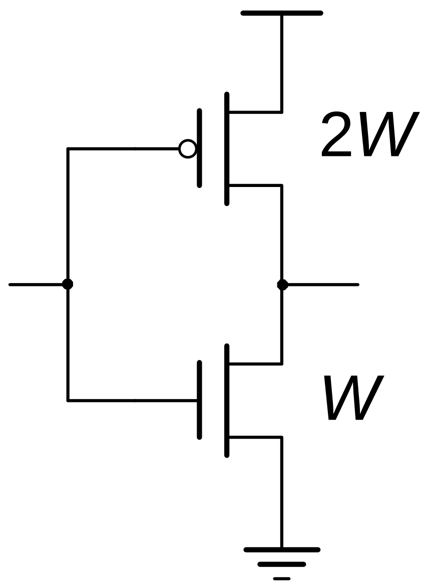


If C_L is given:
- How many stages are needed to minimize the delay?
- How to size the inverters?

May need some additional constraints.

Inverter Delay

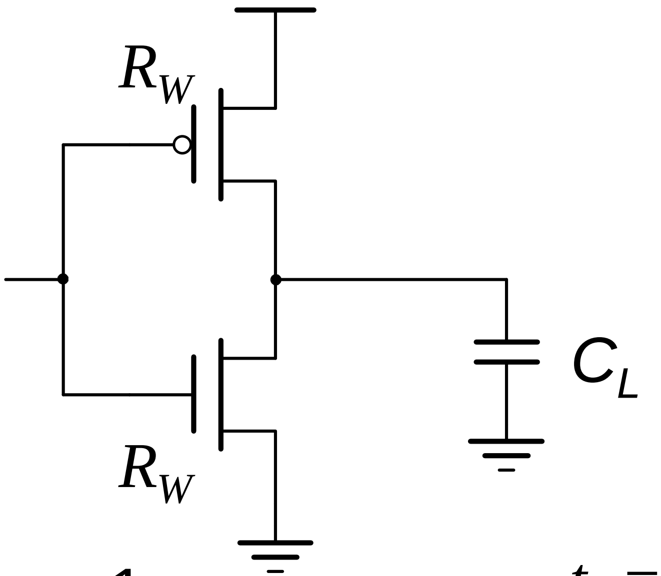
- Minimum length devices, L=0.25μm
- Assume that for $W_P = 2W_N = 2W$
  - same pull-up and pull-down currents
  - approx. equal resistances $R_N = R_P$
  - approx. equal rise t_{pLH} and fall t_{pHL} delays
- Analyze as an RC network



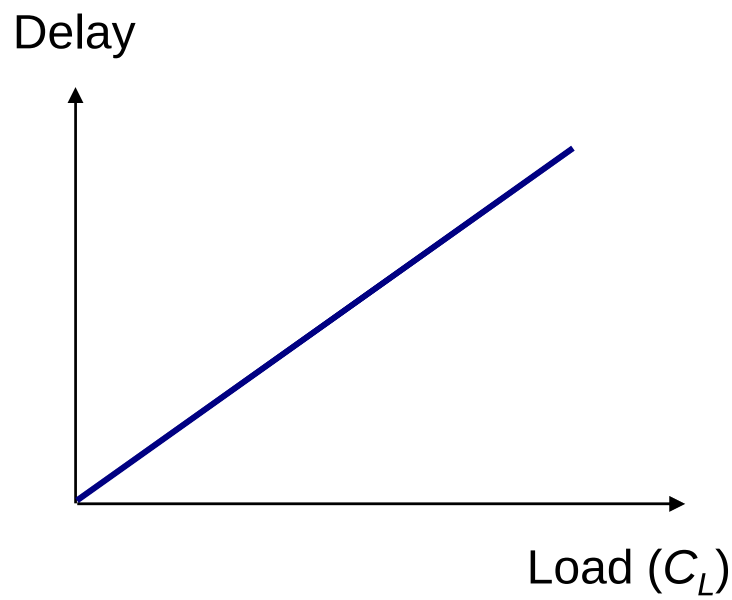
Delay (*D*): $t_{pHL} = (\ln 2)\, R_N C_L$ $\qquad t_{pLH} = (\ln 2)\, R_P C_L$

Load for the next stage: $C_{gin} = 3\dfrac{W}{W_{unit}} C_{unit}$

Inverter with Load



$W_{unit} = 1$

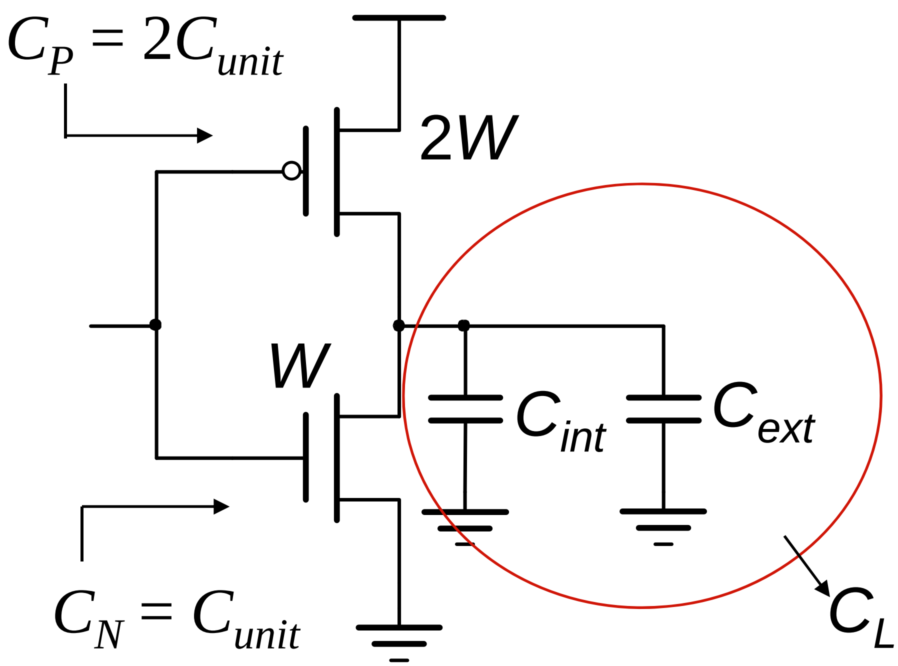


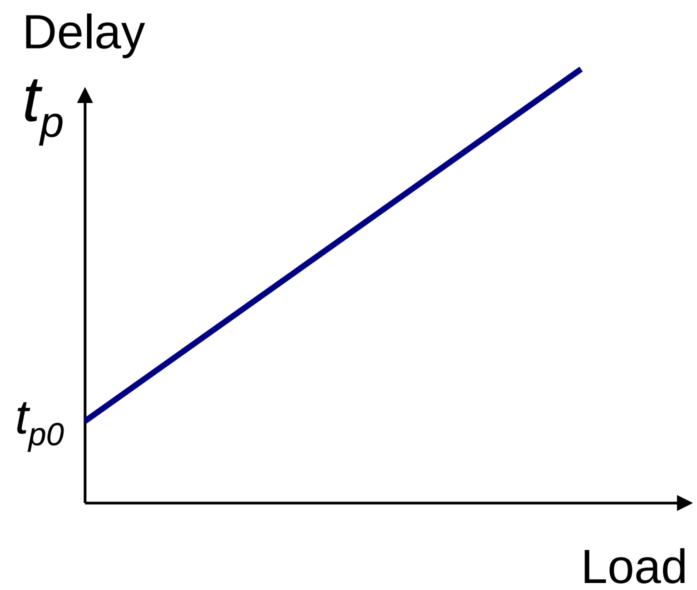
$$t_p = k\, R_W C_L$$

k is a constant, equal to 0.69

Assumptions: no load ⟶ zero delay?

Inverter with Load





Delay (t_p) = $kR_W(C_{int} + C_{ext}) = kR_W C_{int} + kR_W C_{ext} = \underbrace{kR_W C_{int}}_{t_{p0}}(1 + C_{ext}/C_{int})$

t_{p0} *(intrinsic delay)*