

ECE 225

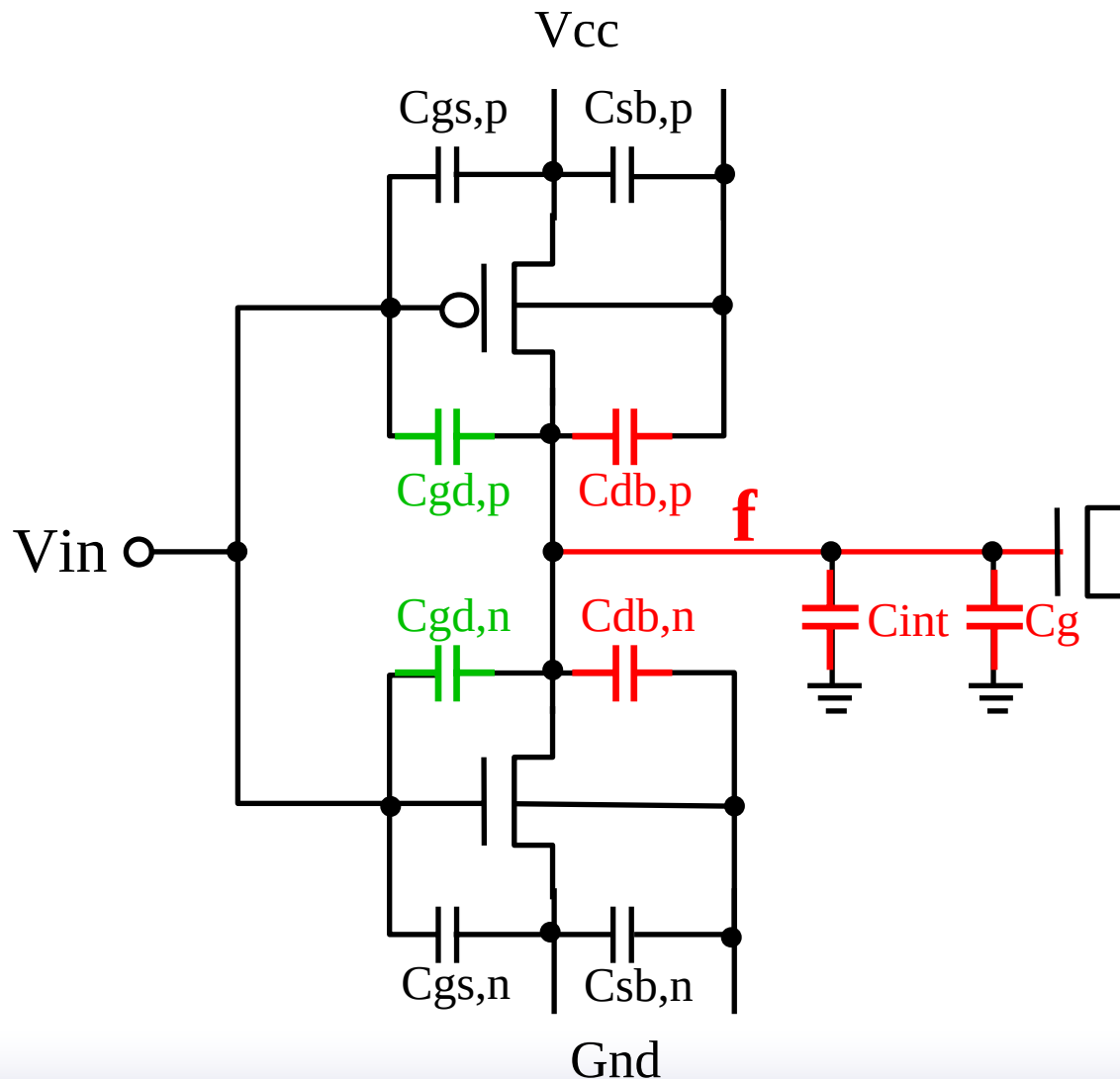
High-Speed Digital IC Design

Lecture 5

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

Propagation Delay

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

CMOS inverter capacitances

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

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

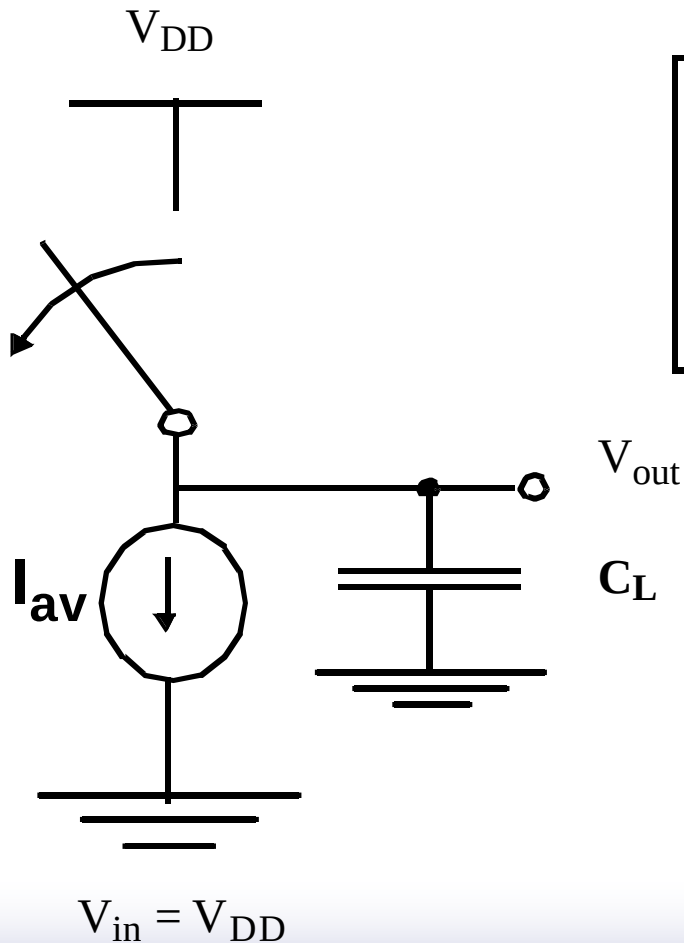
$$C_{gd,n}, C_{gd,p} = 2WL_D C_{ox}$$

↑
Miller effect

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

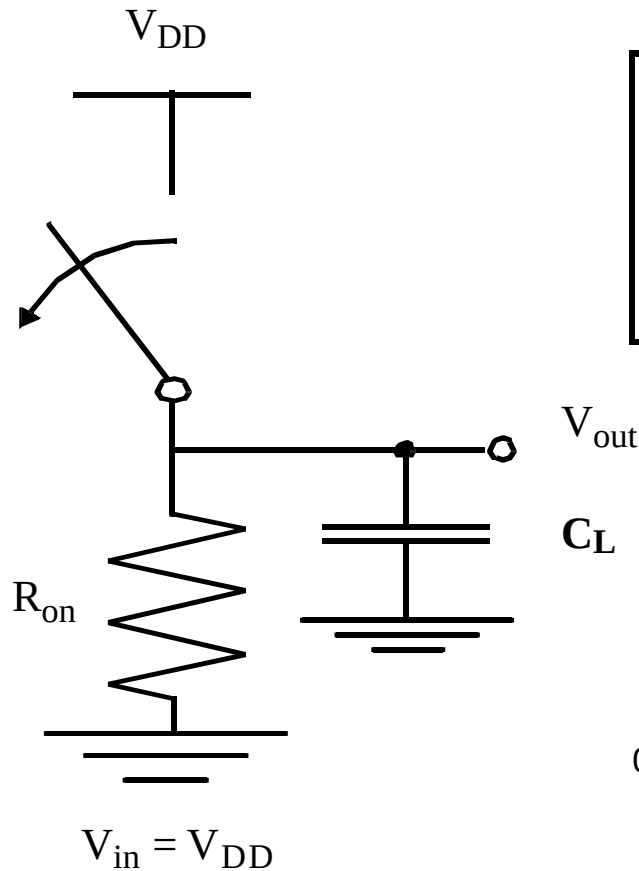
↑
For each gate

CMOS Inverter Propagation Delay Approach 1

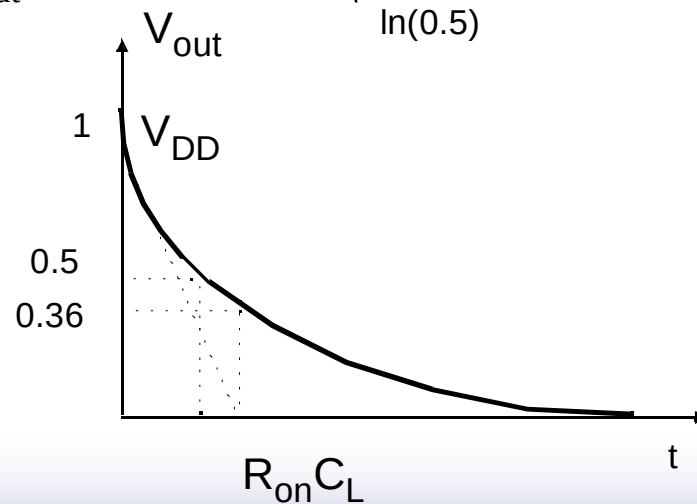


$$t_{pHL} = \frac{C_L V_{swing}/2}{I_{av}}$$

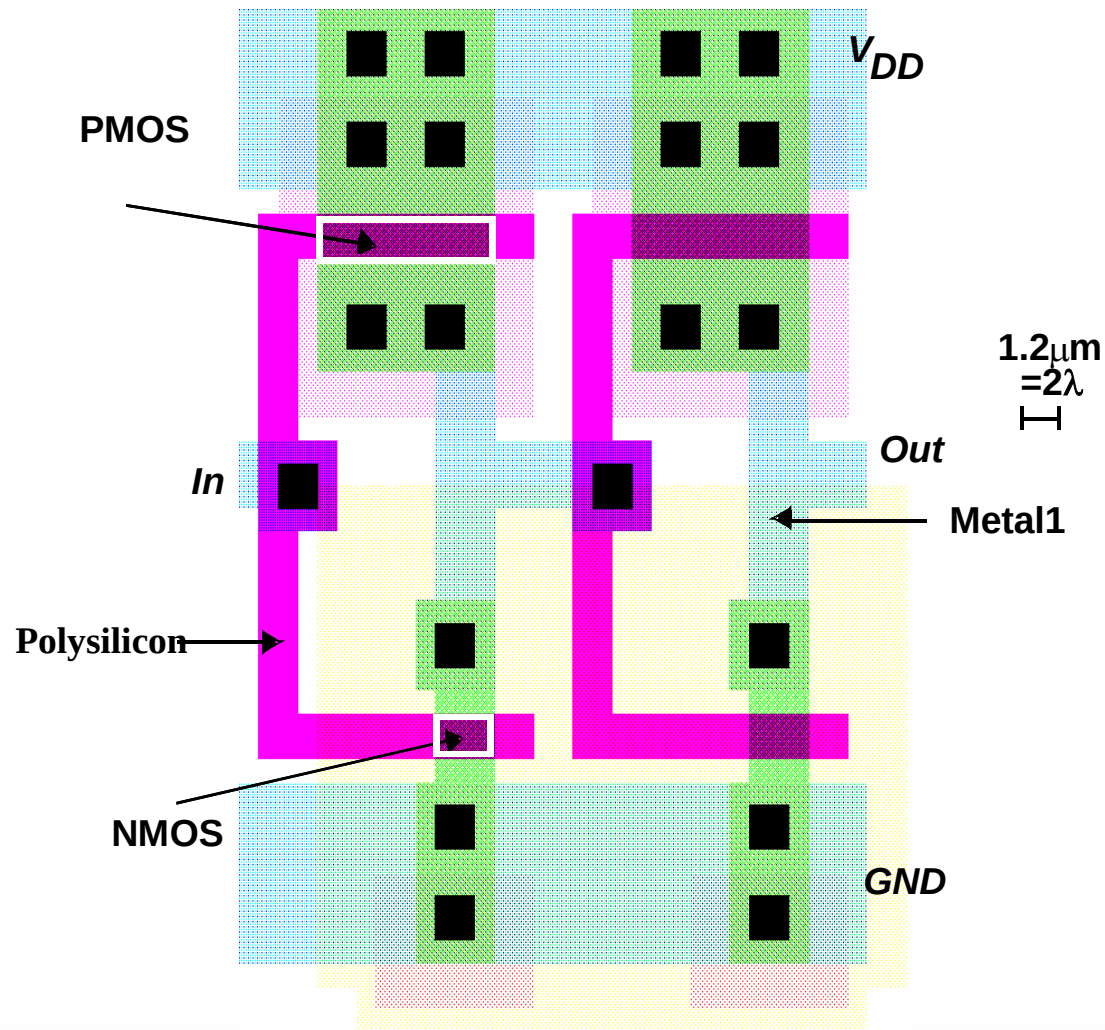
CMOS Inverter Propagation Delay Approach 2



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

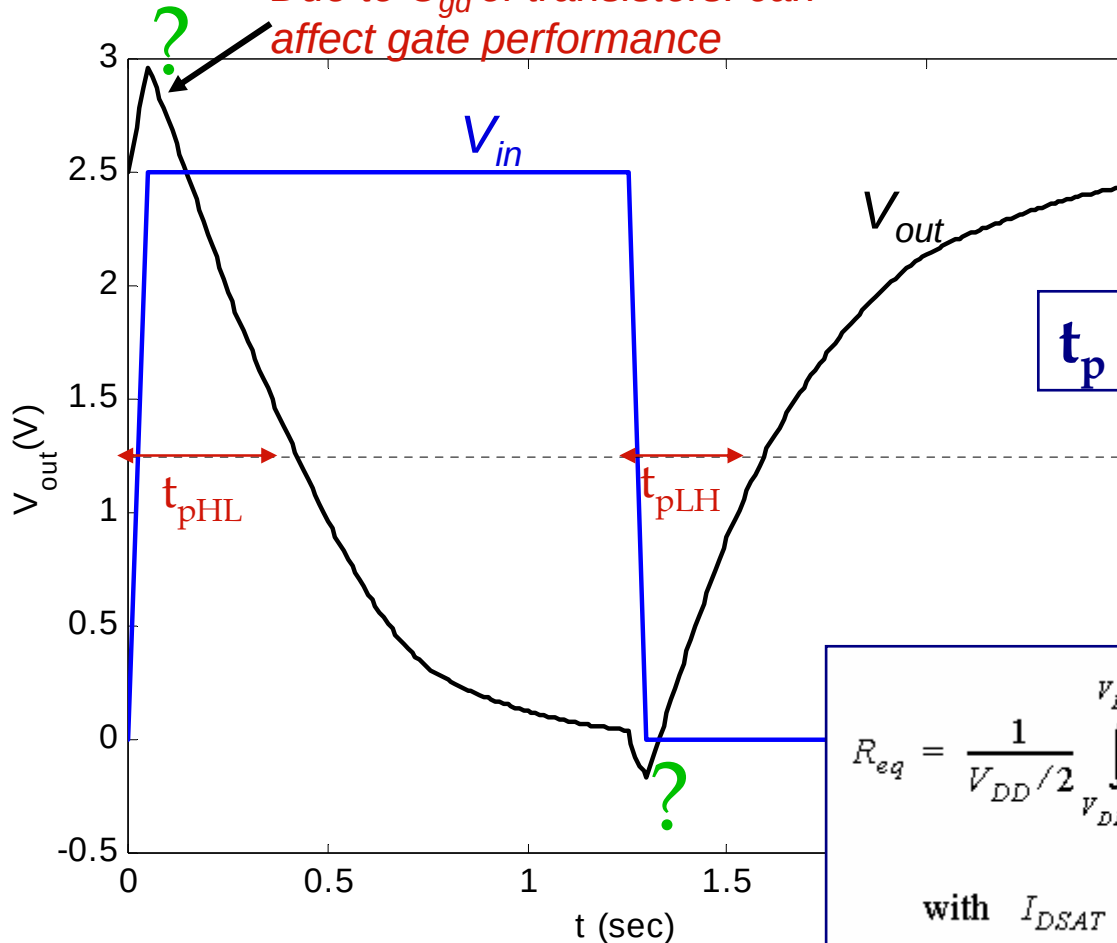


CMOS Inverters



Transient Response

Due to C_{gd} of transistors: 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 $I_{DSAT} = k' \frac{W}{L} \left((V_{DD} - V_T) V_{DSAT} - \frac{V_{DSAT}^2}{2} \right)$

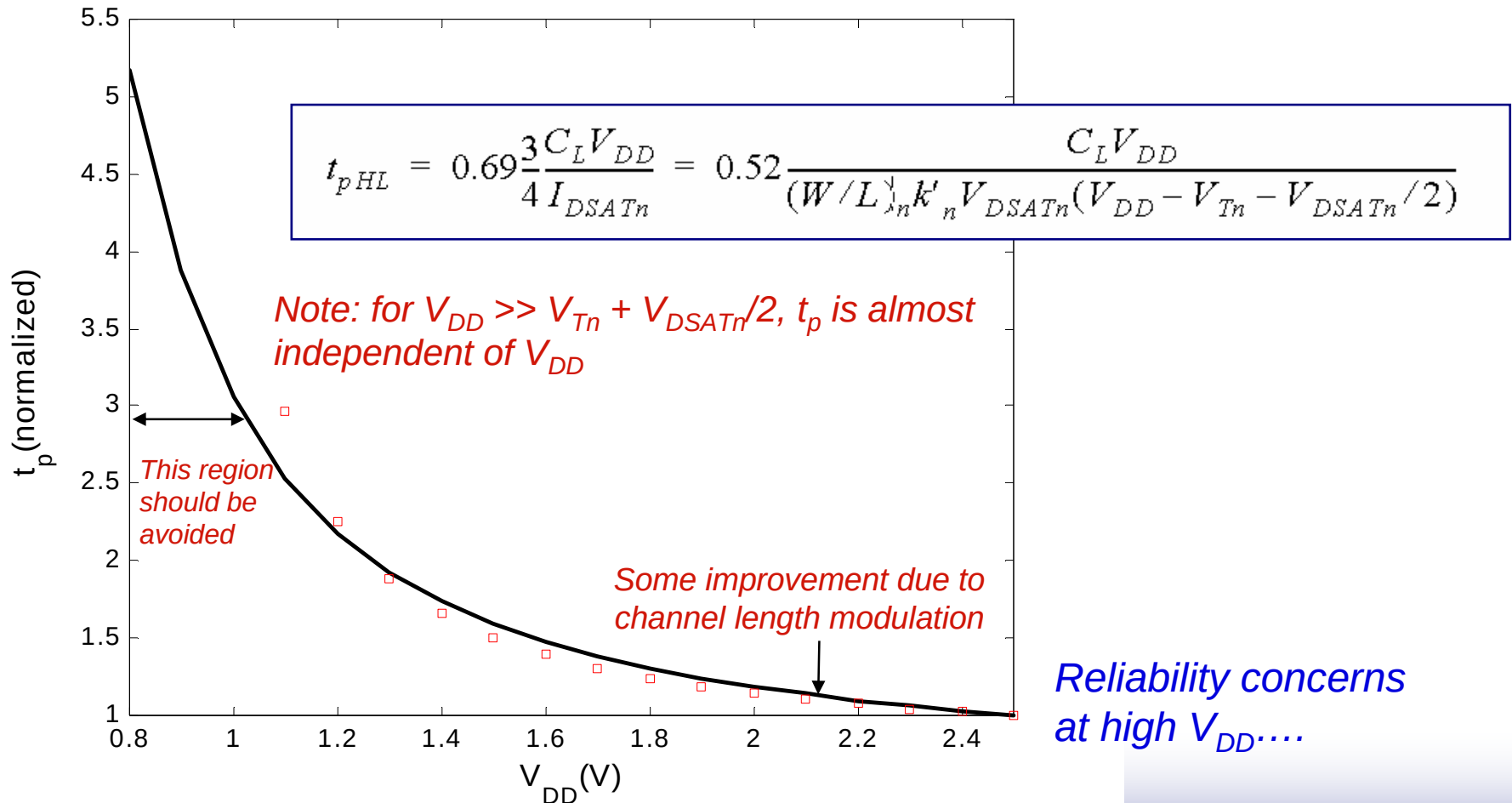
Design for Performance

- 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!

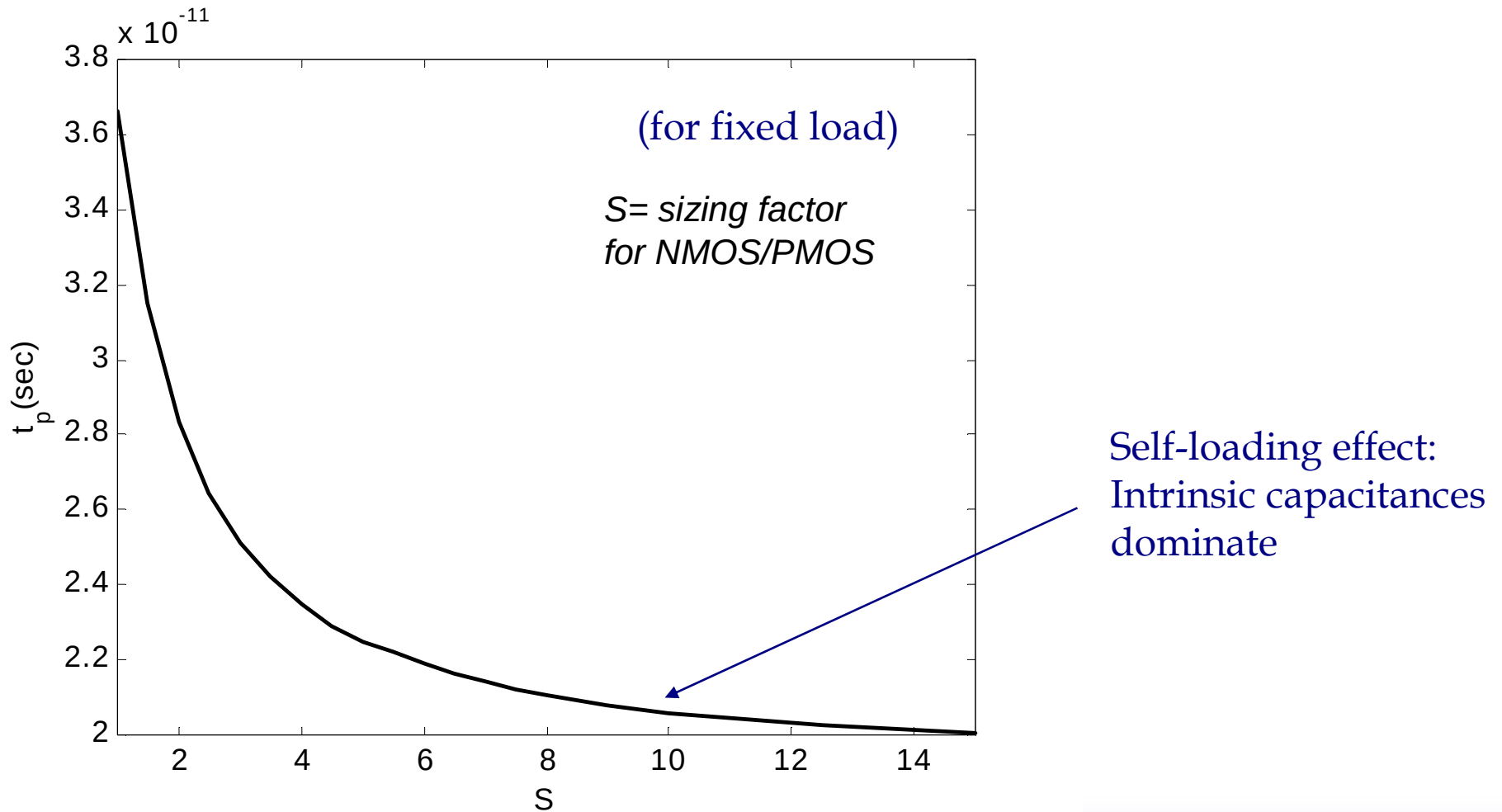
Delay as a function of V_{DD}

Same as the ON resistance of a transistor....

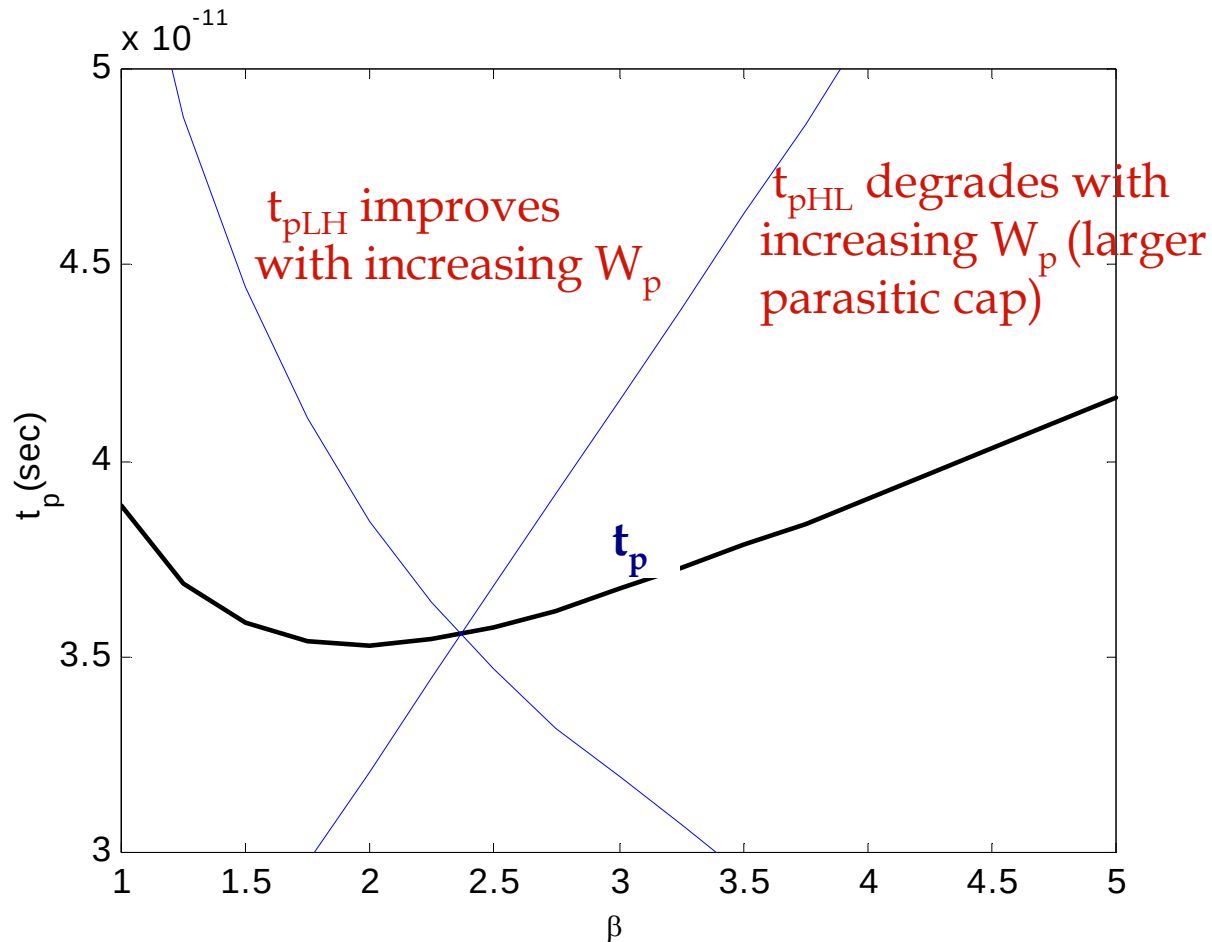
Trade off energy dissipation vs performance.....



Device Sizing



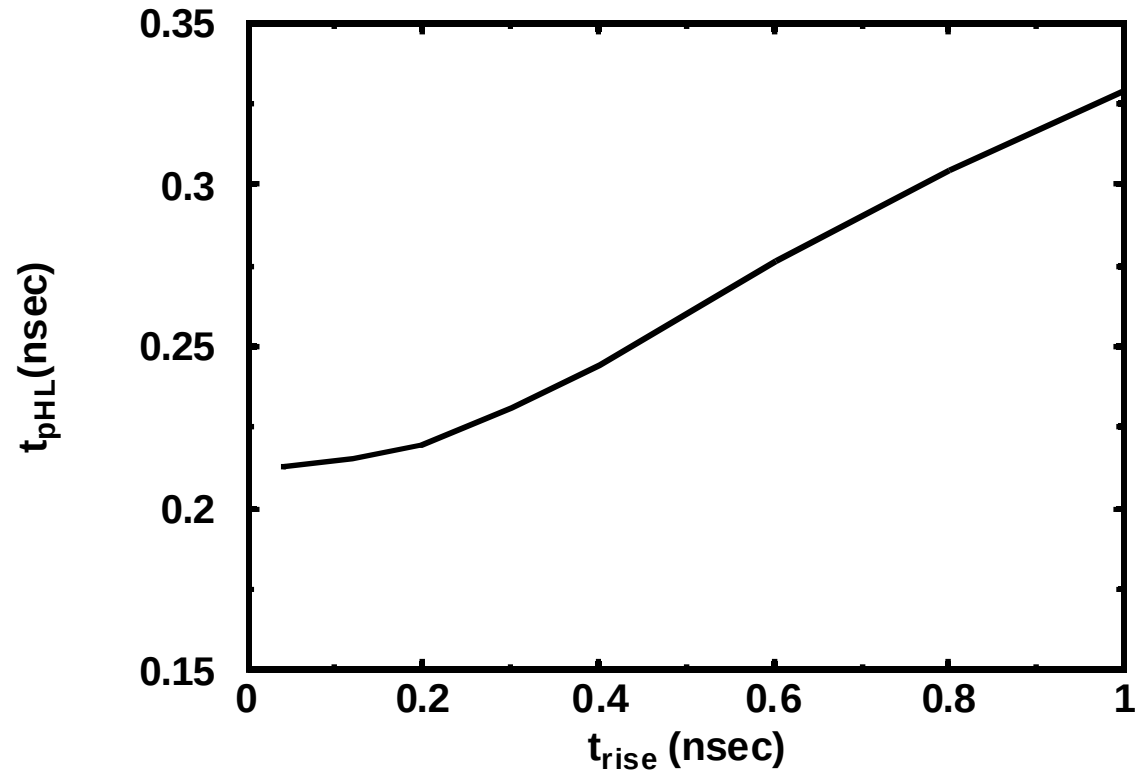
NMOS/PMOS ratio



$$\beta = W_p/W_n$$

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

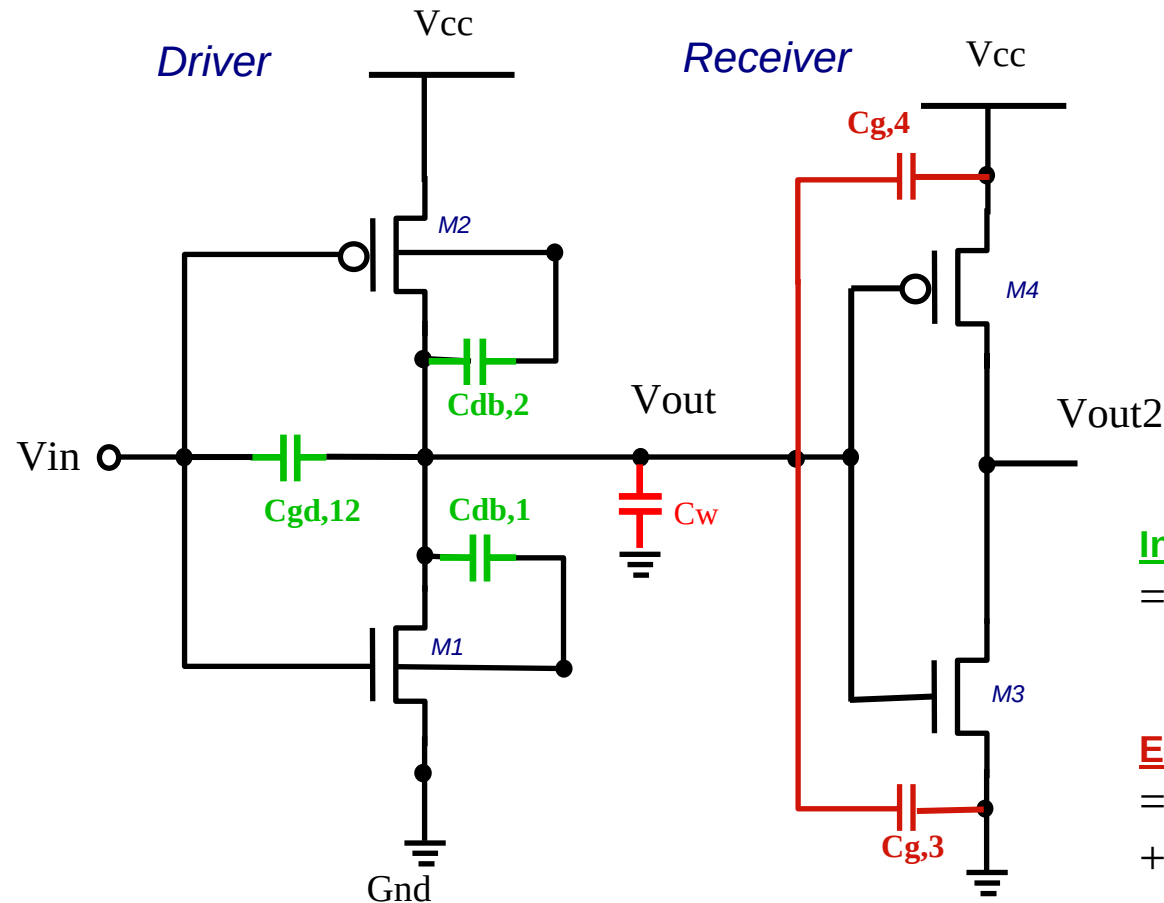
Impact of Rise Time on Delay



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

Inverter Sizing

Load capacitance



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

Internal Caps of Driver (C_{int}):

= Junction caps: $C_{db,12} +$

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

External Caps (C_{ext}):

= 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:

$$\begin{aligned} t_p &= 0.69 R_{eq} (C_{int} + C_{ext}) \\ &= 0.69 R_{eq} C_{int} \left(1 + \frac{C_{ext}}{C_{int}} \right) \\ &= t_{p0} \left(1 + \frac{C_{ext}}{C_{int}} \right) \end{aligned}$$

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)

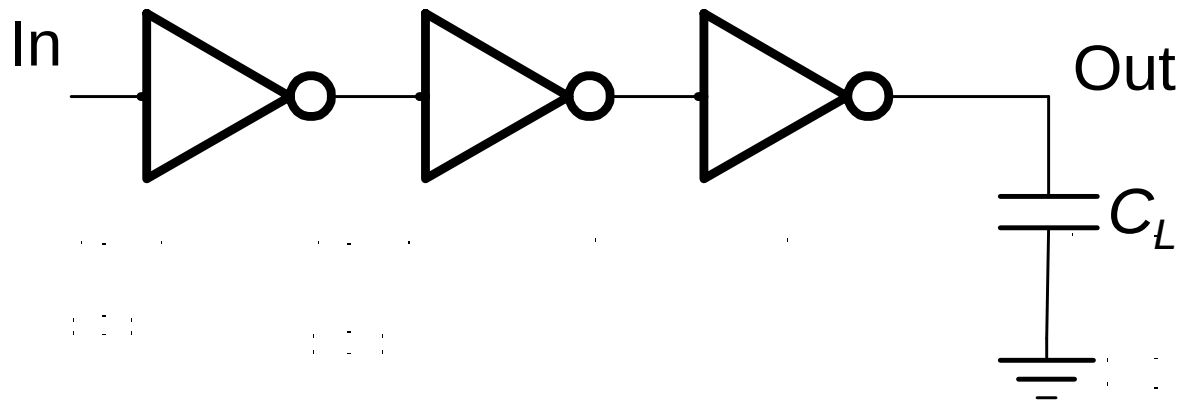
C_{iref} be the internal capacitance of the reference gate

$$\begin{aligned}C_{int} &= S C_{iref}, \quad R_{eq} = \frac{R_{ref}}{S} \\t_p &= 0.69 \left(\frac{R_{ref}}{S} \right) (S C_{iref}) \left(1 + \frac{C_{ext}}{S C_{iref}} \right) \\&= 0.69 R_{ref} C_{iref} \left(1 + \frac{C_{ext}}{S C_{iref}} \right) \\&= t_{p0} \left(1 + \frac{C_{ext}}{S C_{iref}} \right)\end{aligned}$$

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



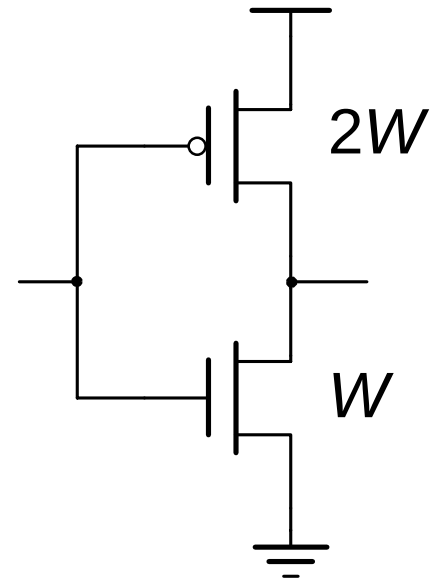
If C_L is given:

- How many stages are needed to minimize the delay?
- How to size the inverters?

May need some additional constraints.

Inverter Delay

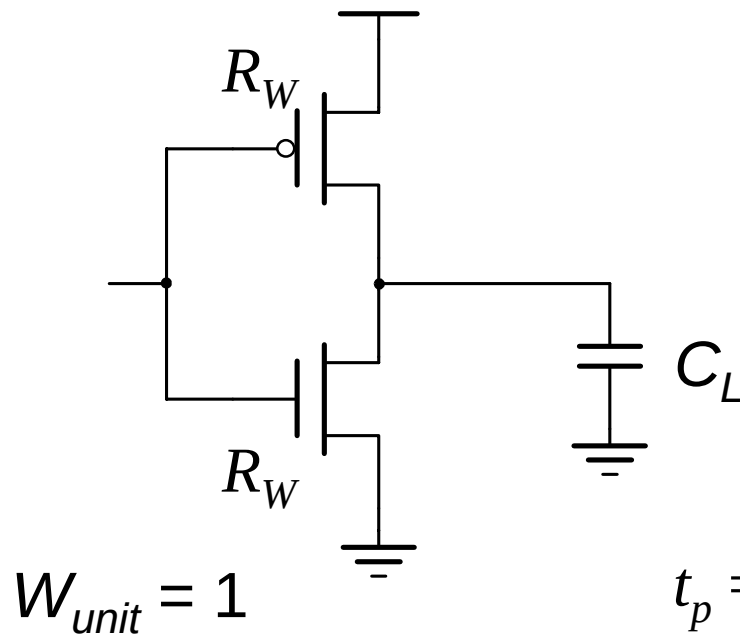
- Minimum length devices, $L=0.25\mu\text{m}$
- Assume that for $W_P = 2W_N = 2W$
 - same pull-up and pull-down currents
 - approx. equal resistances $R_N = R_P$
 - approx. equal rise t_{pLH} and fall t_{pHL} delays
- Analyze as an RC network



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

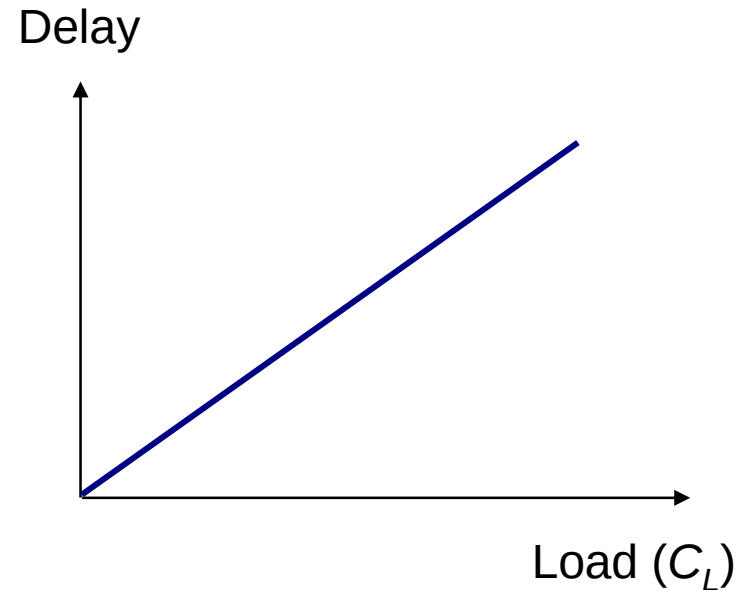
$$\text{Load for the next stage: } C_{gin} = 3 \frac{W}{W_{unit}} C_{unit}$$

Inverter with Load



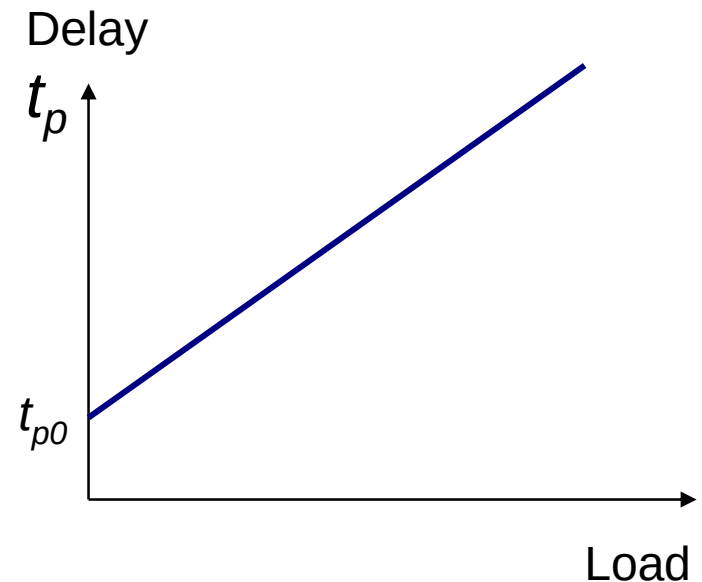
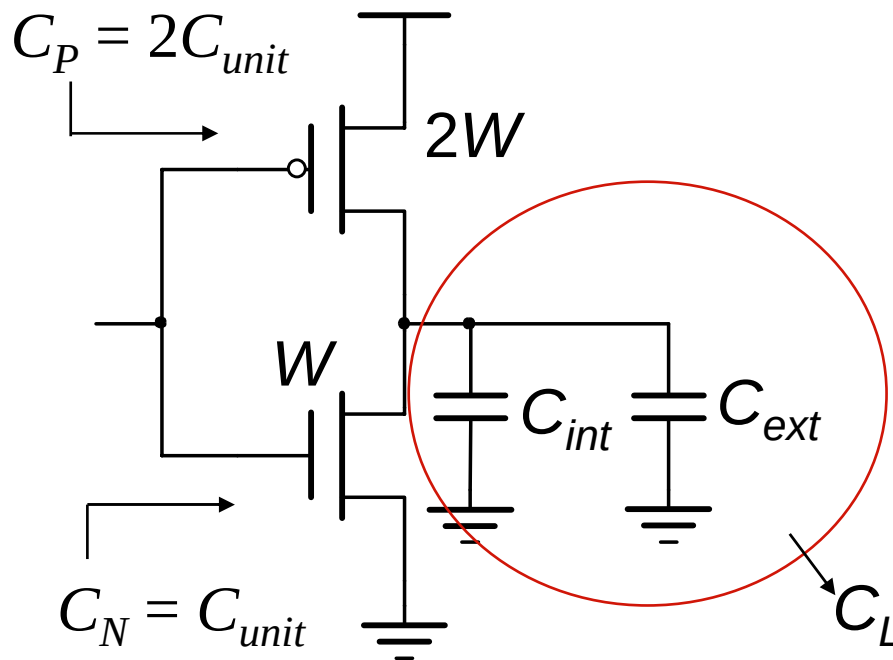
$$t_p = k R_W C_L$$

k is a constant, equal to 0.69



Assumptions: no load $\longrightarrow$ zero delay?

Inverter with Load



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

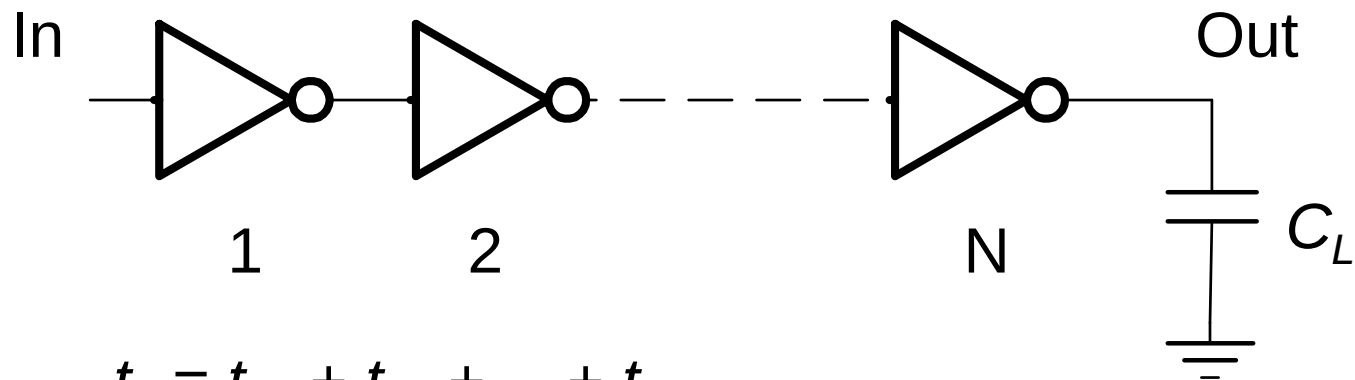
Delay Formula: inverter chain

$$\begin{aligned} \text{Let } C_{int} &= \gamma C_{gin} \text{ with } \gamma \approx 1 \\ f &= C_{ext}/C_{gin} - \text{effective fanout} \\ t_{p0} &= 0.69 R_{eq} C_{int} \end{aligned}$$

$$\text{Delay} \sim R_{eq} (C_{int} + C_{ext})$$

$$t_p = 0.69 R_{eq} C_{int} (1 + C_{ext} / C_{int}) = t_{p0} (1 + f / \gamma)$$

Apply to Inverter Chain



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

$$t_{p,j} = t_{p0} \left(1 + \frac{C_{gin,j+1}}{\gamma C_{gin,j}} \right)$$

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

Optimal Tapering for Given N

Delay equation has $N - 1$ unknowns, $C_{g,2} \dots C_{g,N}$

Minimize the delay, find $N - 1$ partial derivatives

Result: $C_{g,j+1}/C_{g,j} = C_{g,j}/C_{g,j-1}$ With $j = 2, \dots, N$

Size of each stage is the geometric mean of two neighbors

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

- each stage has the same effective fanout ($f_j = f = C_{ext}/C_{g,j}$)
- each stage has the same delay: $t_p = t_{p0} (1 + f/\gamma)$

Optimum Delay and Number of Stages

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

$$\frac{C_L}{C_{g,N}} = \frac{C_{g,N}}{C_{g,N-1}} = \dots = \frac{C_{g,2}}{C_{g,1}} = f$$

$$\text{Hence, } f^N = C_L / C_{g,1} = F$$

F is the overall effective fanout of the circuit

Effective fanout of each stage: $f = \sqrt[N]{F}$

Minimum path delay:

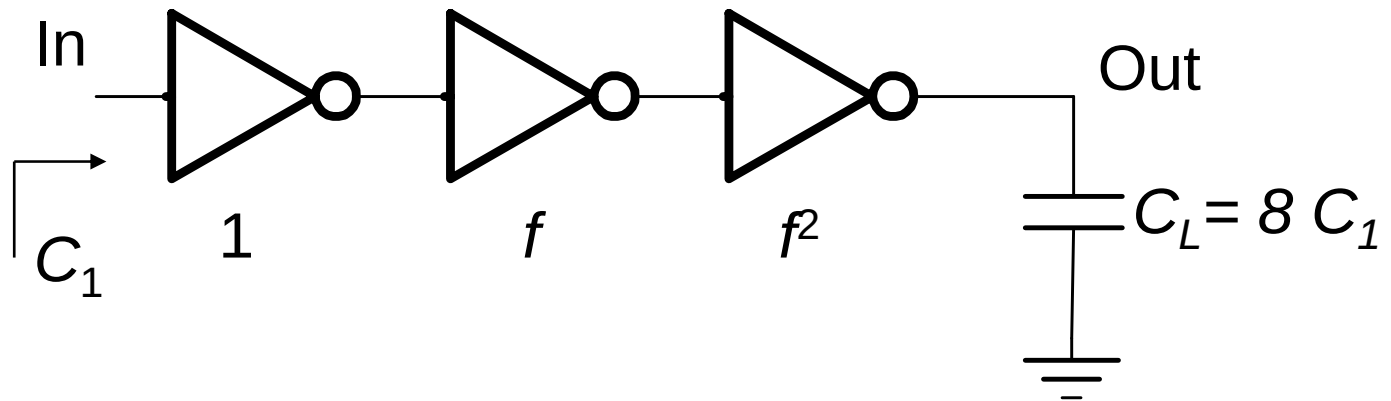
$$t_p = Nt_{p0} \left(1 + \sqrt[N]{F} / \gamma \right)$$

If N is too large, intrinsic delay of stages dominate, while if N is small, effective fanout of each stage (f) is large and the second term dominates

How to choose N?

Example

If N is given....



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

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

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

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

If self-loading is ignored....

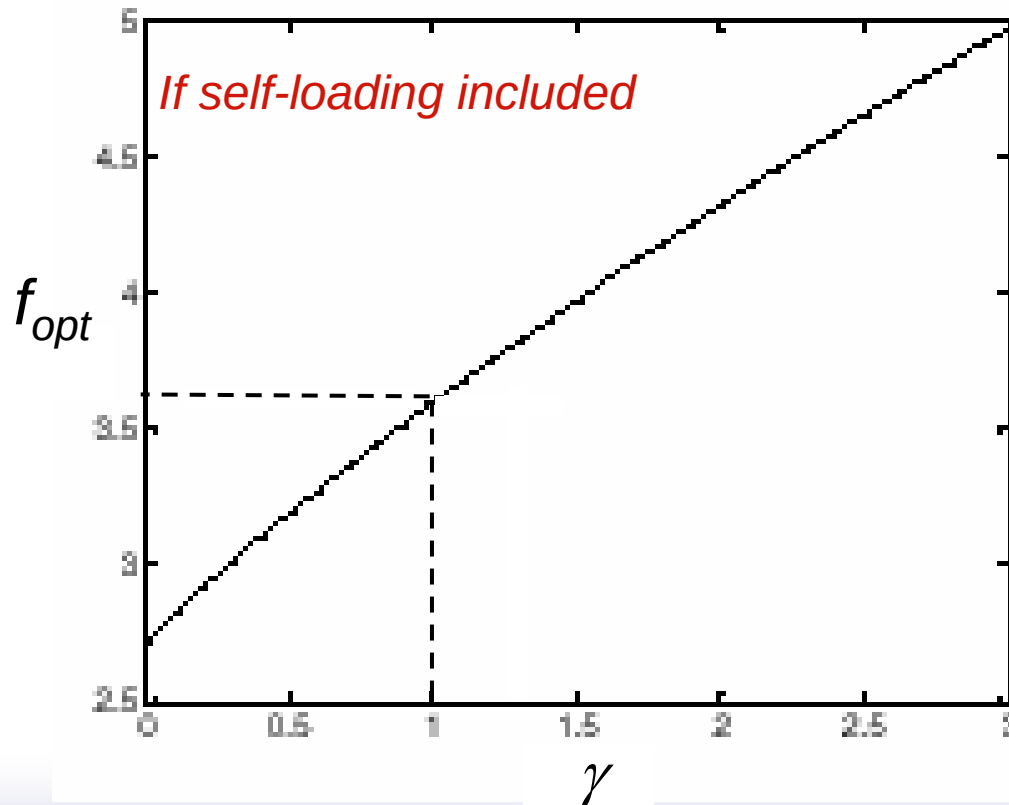
For $\gamma = 0$, $f = e$, $N = \ln F$

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

Optimum Effective Fanout f

Optimum f for given process defined by γ

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



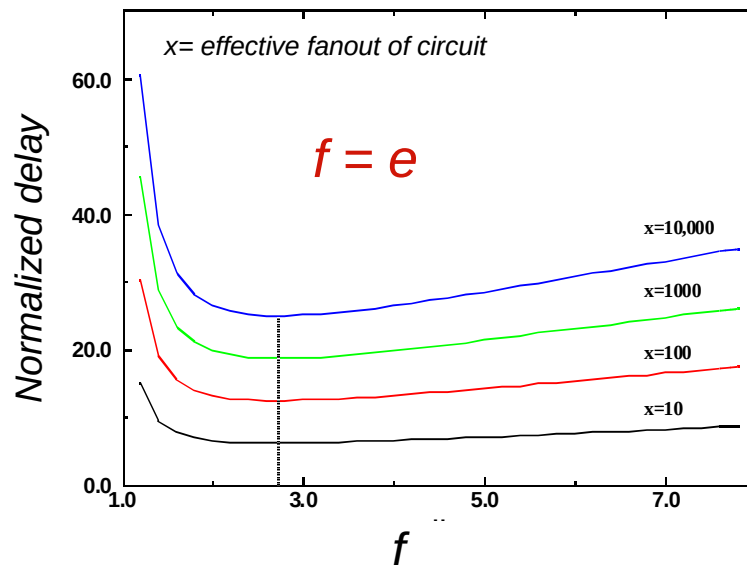
Optimum tapering factor:

$$f_{opt} = 3.6$$

for $\gamma=1$

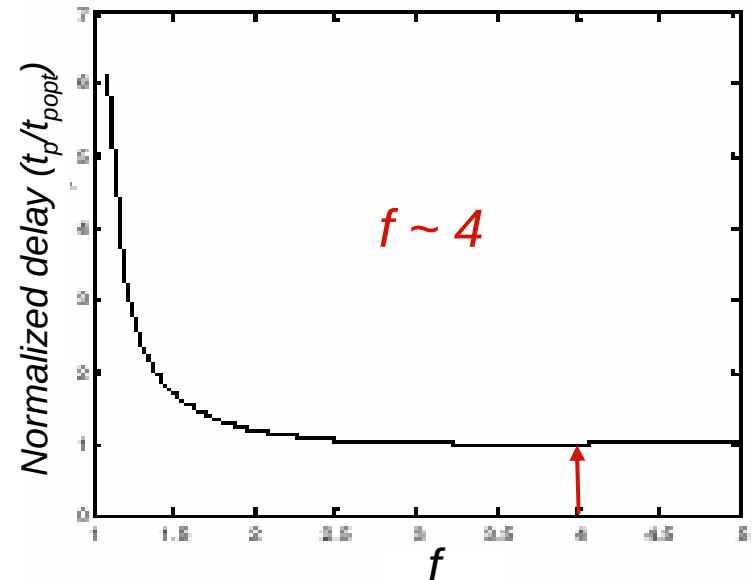
Impact of Self-Loading on t_p

No Self-Loading, $\gamma=0$



Optimal number of stages, $N = \ln(F)$

With Self-Loading $\gamma=1$



If $f < f_{opt}$ (too many stages) will result in delay to increase

Normalized delay function of F

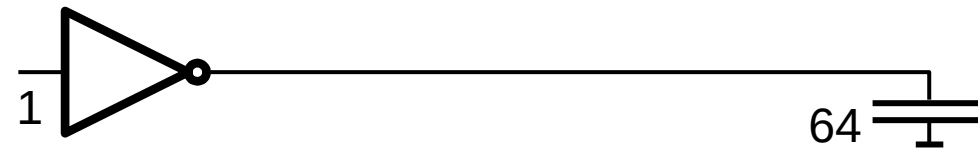
$$t_p = Nt_{p0} \left(1 + \sqrt[N]{F} / \gamma \right)$$

t_{popt}/t_{p0} for $\gamma=1$

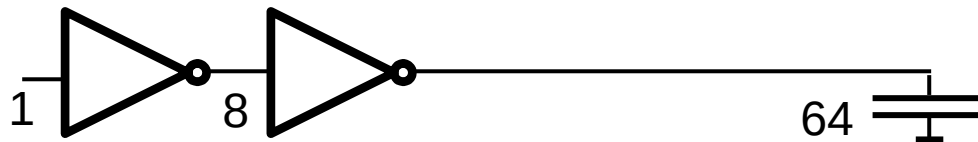
F	Unbuffered	Two Stage	Inverter Chain
10	11	8.3	8.3
100	101	22	16.5
1000	1001	65	24.8
10,000	10,001	202	33.1

Buffer Design

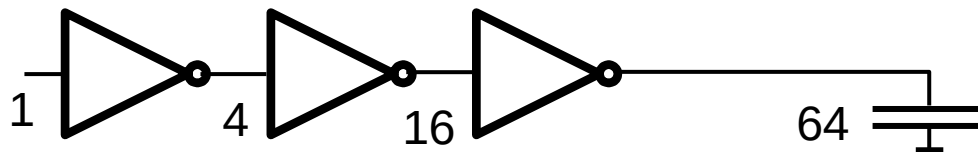
$$f = F^{1/N}$$



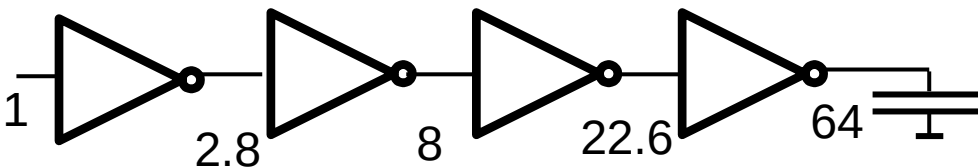
N	f	t _p
1	64	65



2	8	18
---	---	----



3	4	15
---	---	----

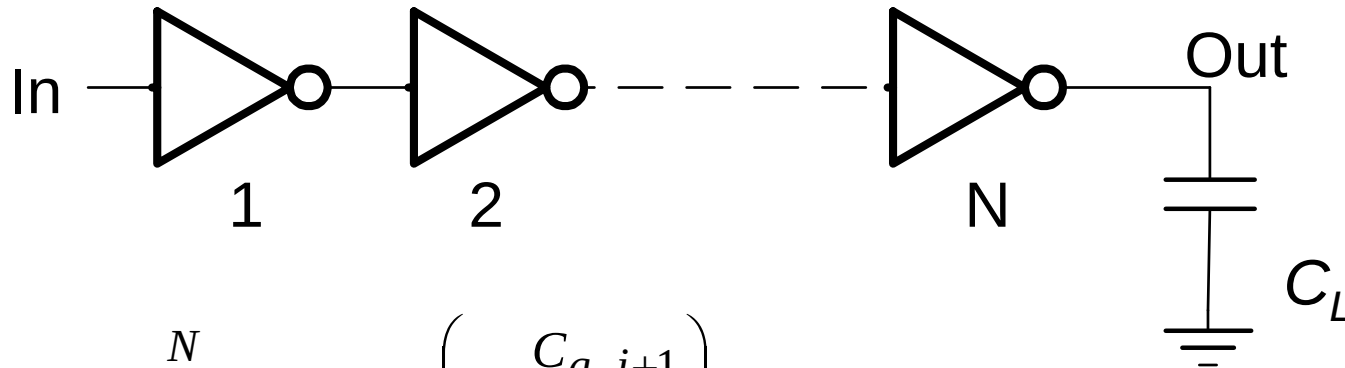


4	2.8	15.3
---	-----	------

Sizing Logic Paths for Speed

- ❑ Frequently, input capacitance of a logic path is constrained
- ❑ Logic also has to drive some capacitance
- ❑ Example: ALU load in an Intel's microprocessor is 0.5pF
- ❑ How do we size the ALU datapath to achieve maximum speed?
- ❑ We have already solved this for the inverter chain – can we generalize it for any type of logic?

Buffer Example



$$\text{Chain Delay} = \sum_{j=1}^N t_{p,j} = t_{p0} \left(1 + \frac{C_{g,j+1}}{\gamma C_{g,j}} \right), \quad \text{with } C_{g,N+1} = C_L$$

(in units of τ_{inv})

For given N : $C_{g,j+1}/C_{g,j} = C_{g,j}/C_{g,j-1}$

Optimal fanout (f): $C_{g,j+1}/C_{g,j} \sim 4$

How to generalize this to any logic path?

Designing Combinational Logic Circuits

Static Vs Dynamic Circuits

Static Circuits

At every point in time (except during the switching transients) each **gate output is connected to either V_{DD} or V_{SS}** via a low-resistive path.

The outputs of the gates **assume at all times the value of the Boolean function**, implemented by the circuit (ignoring, once again, the transient effects during switching periods).

This is in contrast to the **dynamic** circuit class, which relies on temporary storage of signal values on the capacitance of high impedance circuit nodes.

Resulting gate is simpler and faster, but increased sensitivity to noise....

Static and Dynamic CMOS Circuit Families

❑ Static:

- **Complementary CMOS**
 - (robustness, low power, large fan-in expensive in terms of area and performance)
- **Ratioed Logic (pseudo-NMOS, DCVSL)**
 - (simple and fast at the expense of reduced NM and static power)
- **Pass-Transistor Logic (Transmission Gate)**
- -attractive for specific circuits: MUX, XOR-dominated logic such as Adders)

❑ Dynamic:

- good for fast and complex gates, design process is harder due to parasitic effects, leakage puts an upper limit on the operating frequency of the circuit
 - **Domino Logic**
 - **np-CMOS**

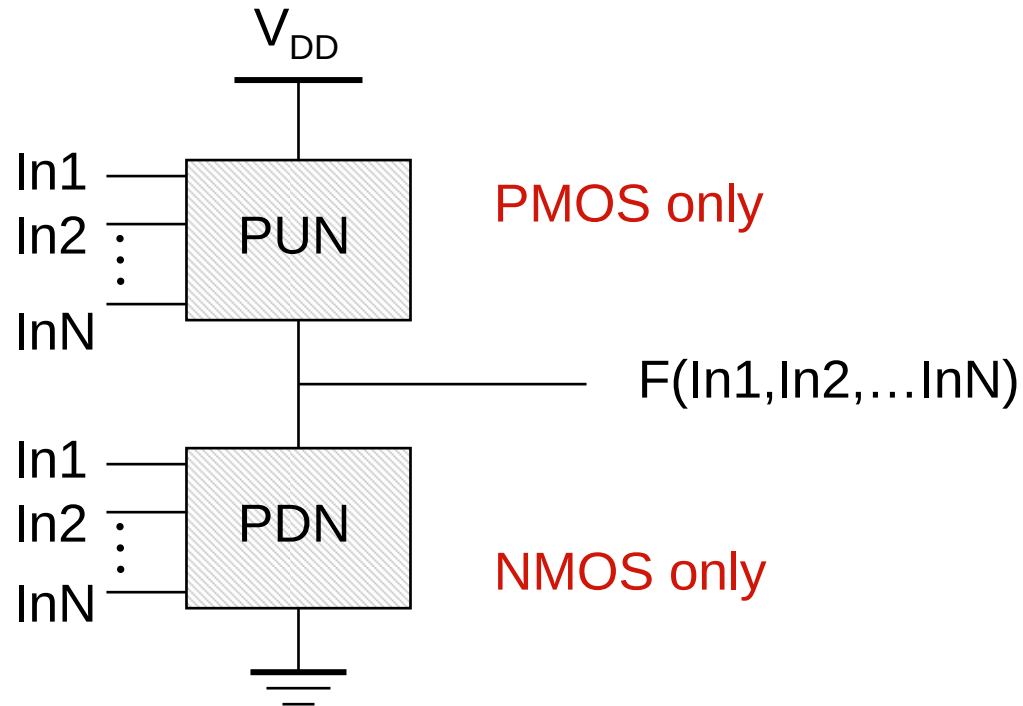
Which style is best?

.....depends on ease of design, performance, power, area and robustness.

Complementary CMOS Logic

- ❑ Full rail-to-rail swing; **high noise margins**
- ❑ Logic levels not dependent upon the relative device sizes; **ratioless**
- ❑ Always a path to Vdd or Gnd in steady state; **low output impedance**
- ❑ Extremely **high input resistance**; nearly zero steady-state input current
- ❑ No direct path steady state between power and ground; **no static power dissipation**
- ❑ Propagation delay function of load capacitance and resistance of transistors

Static Complementary CMOS

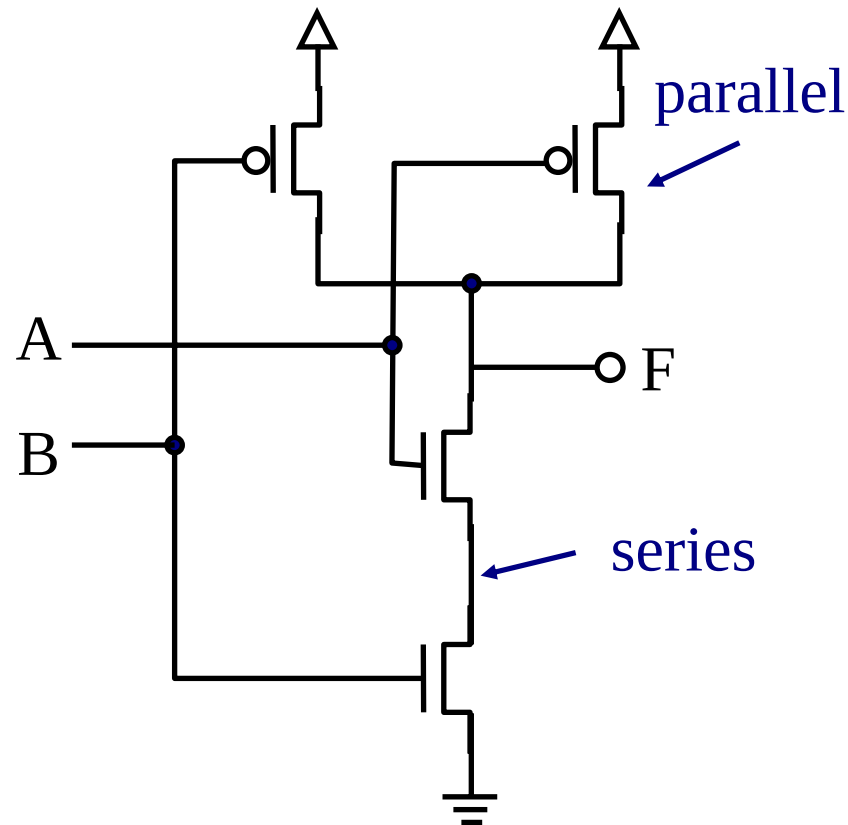


PUN and PDN are **dual** logic networks

Dual Networks

- Dual networks: parallel connection in PUN = series connection in PDN, vice-versa
- If CMOS gate implements logic function F :
 - PUN implements function $\overline{F}$
 - PDN implements function $G = F$

Example: NAND gate



Key Properties of Complementary CMOS Gates: Snapshot

High noise margins

V_{OH} and V_{OL} are at V_{DD} and GND , respectively.

No static power consumption

There never exists a direct path between V_{DD} and V_{SS} (GND) in steady-state mode.

Comparable rise and fall times:

(under appropriate sizing conditions)