

High Speed Communication Circuits and Systems
Lecture 13
High Speed Digital Circuits

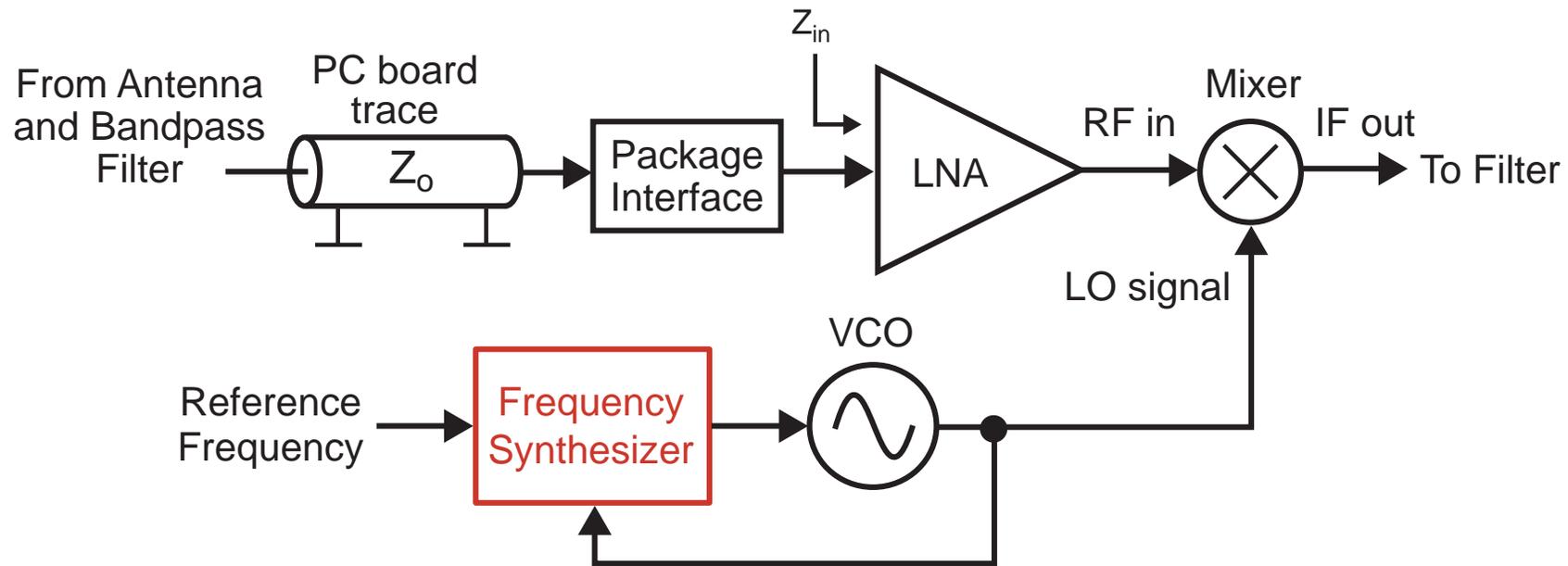
Michael H. Perrott

March 18, 2004

Copyright © 2004 by Michael H. Perrott

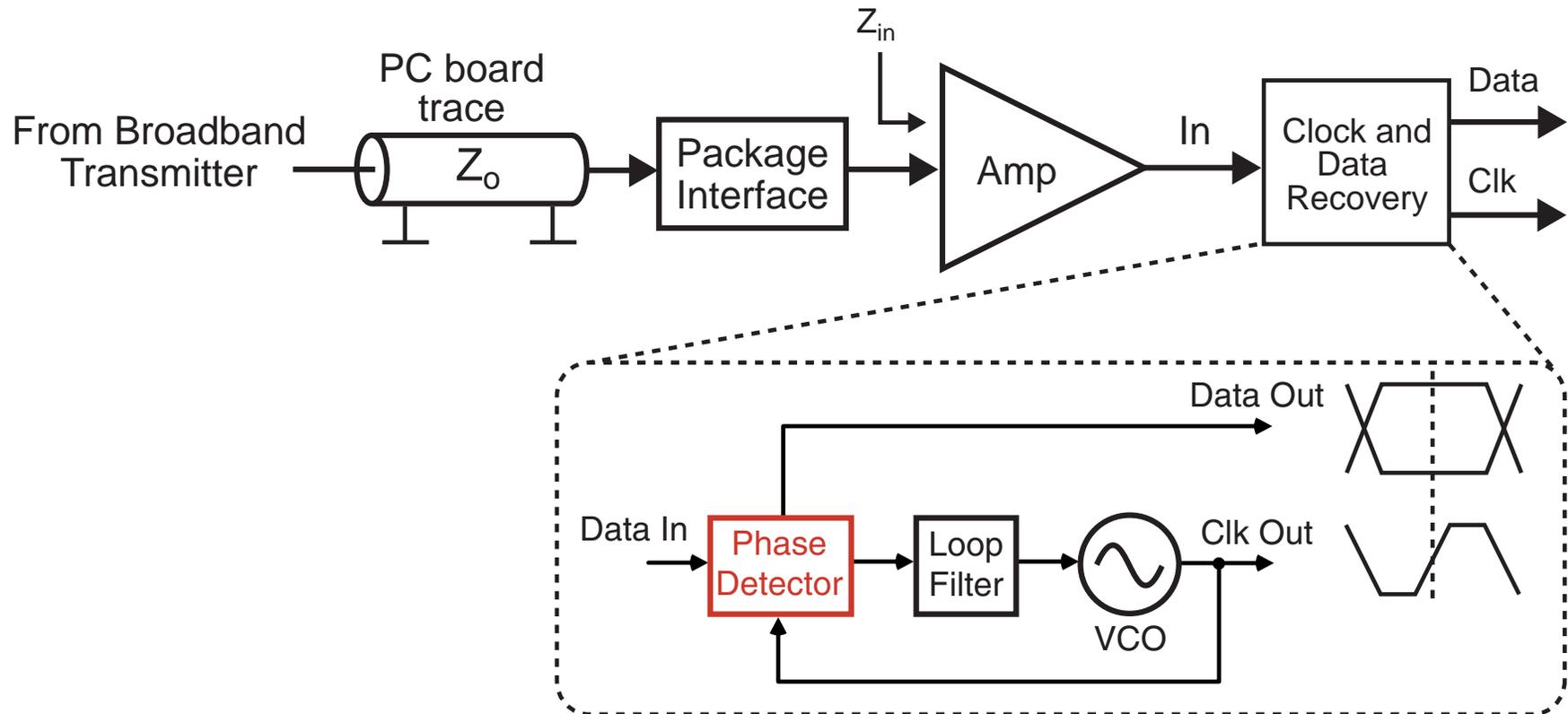
All rights reserved.

High Speed Digital Design in Wireless Systems



- **Primary application areas**
 - Divider within frequency synthesizer
 - High speed A/D's and D/A's in future wireless systems
- **Design Issues**
 - Speed – want it to be fast
 - Power – want low power dissipation
 - Noise – need to be careful of how it impacts analog circuits

High Speed Digital Design in High Speed Data Links

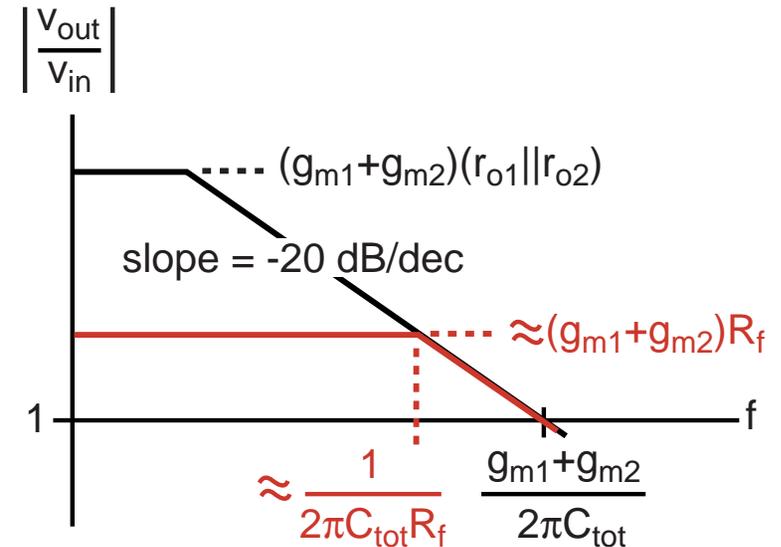
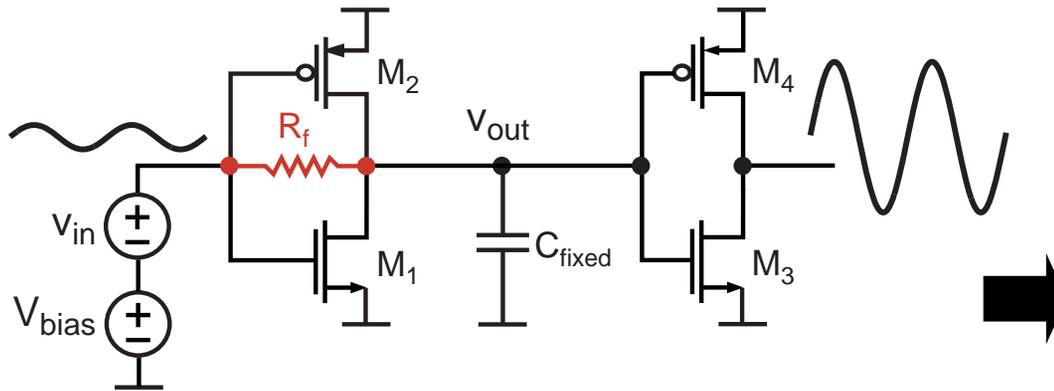


- **Primary application areas**
 - Phase detector within CDR
 - High speed A/D's and D/A's in future systems
- **Design Issues**
 - Same as wireless, but dealing with non-periodic signals

Note: much of the material to follow can be found in

J. Rabaey, “Digital Integrated Circuits: A Design Perspective”, Prentice Hall, 1996

The CMOS Inverter As An Amplifier (From Lecture 5)

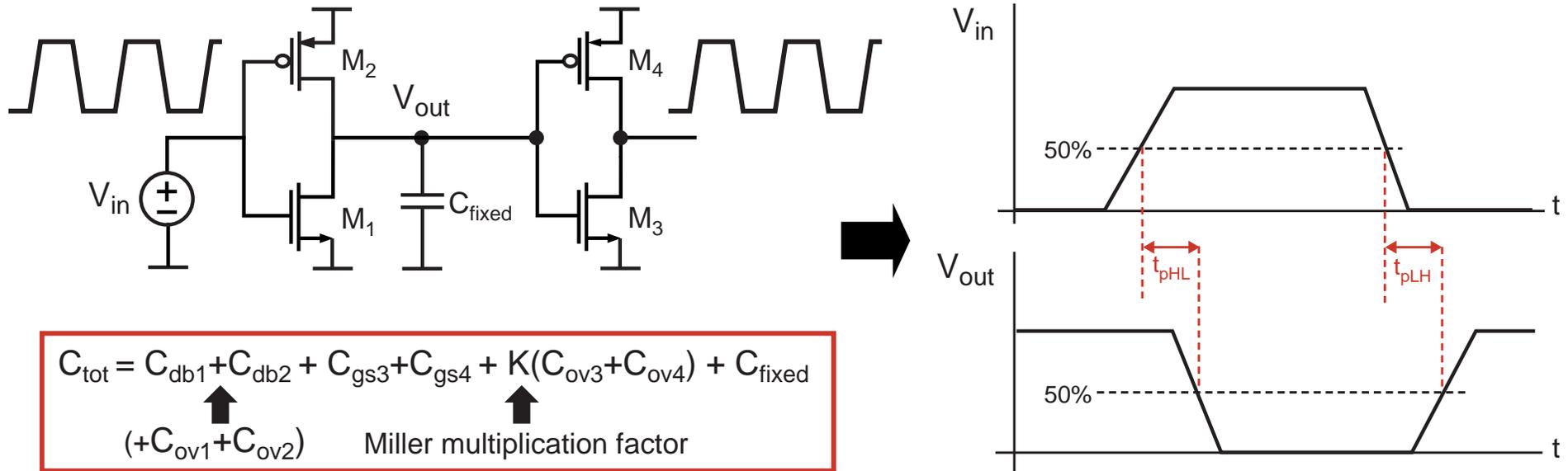


$$C_{\text{tot}} = C_{\text{db1}} + C_{\text{db2}} + C_{\text{gs3}} + C_{\text{gs4}} + K(C_{\text{ov3}} + C_{\text{ov4}}) + C_{\text{Rf}}/2 + C_{\text{fixed}}$$

\uparrow \uparrow
 $(+C_{\text{ov1}} + C_{\text{ov2}})$ Miller multiplication factor

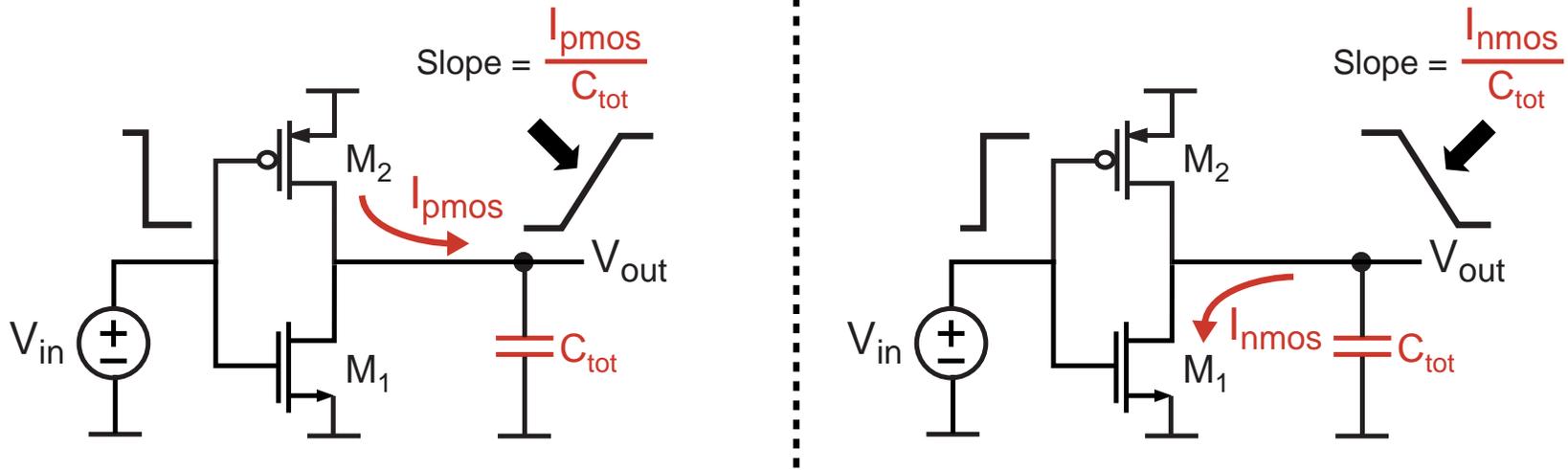
- Small signal assumption allows linearized modeling
- Key metric for speed: gain-bandwidth product ($= f_t$)
 - Strive for high transconductance to capacitance ratio ($= f_t$)
 - Increase speed by lowering gain (use low valued resistors)
 - Minimize capacitance for given level of transconductance
- How does digital design differ?

The CMOS Inverter as a Digital Circuit



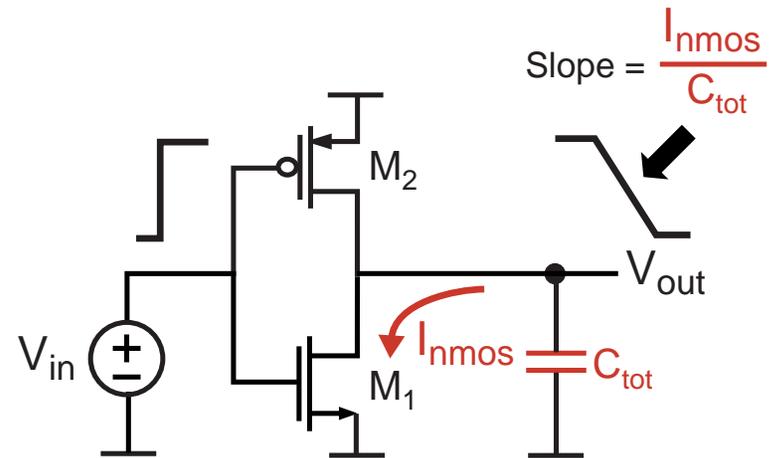
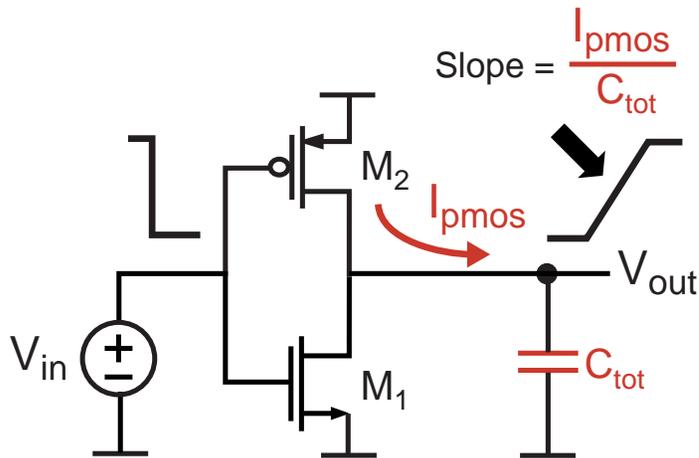
- Large signal variation prevents linearized modeling
 - We must examine nonlinear behavior of devices
- Key metric for speed: propagation delay
 - What device parameters influence this?
 - What are the tradeoffs?

Key Issue for High Speed – Fast Rise and Fall Times



- **For digital circuit, propagation delays primarily set by rise and fall times**
 - **Rise and fall times set by slew rate**
 - Slew rate: ratio of driving current to load capacitance
 - **Faster speed obtained with higher slew rates**
 - **Key performance metric: current drive/capacitance**
 - Compare with analog: transconductance/capacitance

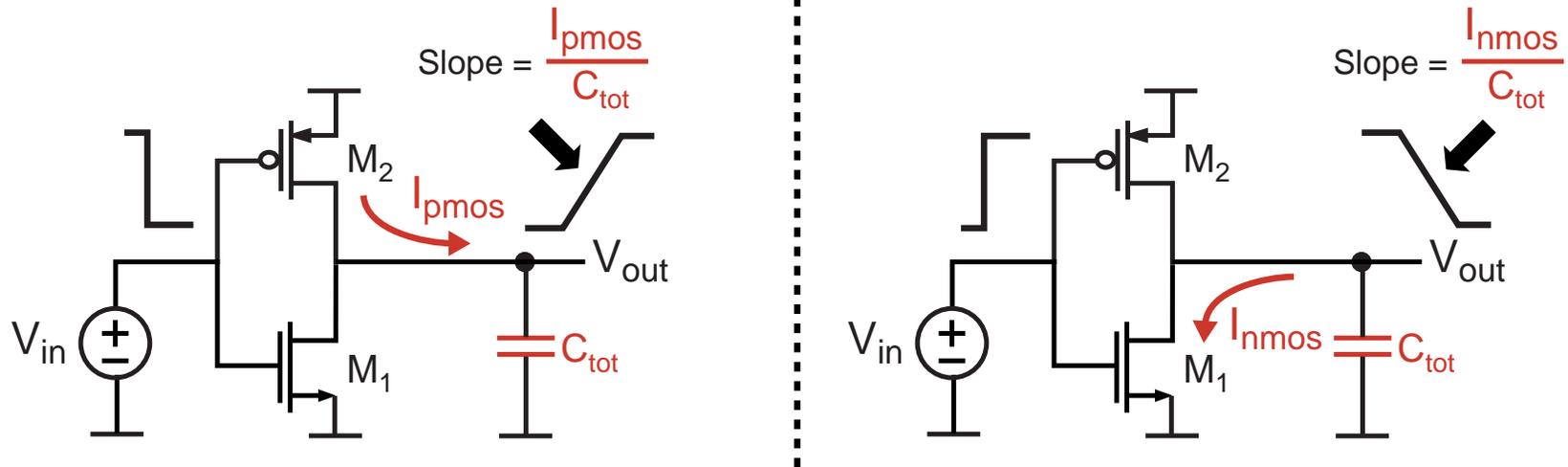
Designing for High Speed



■ Design parameters

- Voltage supply (and voltage swing)
 - Scaling of NMOS and PMOS devices
 - Relative to each other
 - In an absolute sense
 - Circuit architecture (impacts drive current/capacitance ratio)
- ## ■ Key focus point: how is drive current and capacitance influenced by these parameters?
- Focus on voltage and sizing issues first

Impact of Voltage and Sizing on Drive Current



- **Rigorous analysis is difficult**
 - Transistor goes through different regions of operation as load capacitance is charged (i.e., cutoff, triode, saturation)
 - Transistor physics is changing over time
 - Velocity saturation is becoming an important issue
- **We need a simple approach for intuition**
 - Assume device is in saturation the entire time load capacitor is being charged

Note: redo velocity saturation stuff to follow (make more rigorous)
Examine Device Current in Saturation (from Lec 5)

- We classically assume that MOS current is calculated as

$$I_D = \frac{\mu_n C_{ox} W}{2 L} (V_{gs} - V_T)^2$$

- Which is really

$$I_D = \frac{\mu_n C_{ox} W}{2 L} (V_{gs} - V_T) V_{dsat,l}$$

- $V_{dsat,l}$ corresponds to the saturation voltage at a given length, which we often refer to as ΔV
- It may be shown that

$$V_{dsat,l} \approx \frac{(V_{gs} - V_T)(LE_{sat})}{(V_{gs} - V_T) + (LE_{sat})} = (V_{gs} - V_T) \parallel (LE_{sat})$$

- If $V_{gs} - V_T$ approaches LE_{sat} in value, then the top equation is no longer valid
 - We say that the device is in velocity saturation

Analytical Device Modeling in Velocity Saturation (Lec 5)

- If L small (as in modern devices), then velocity saturation will impact us for even moderate values of $V_{gs} - V_T$

$$I_D = \frac{\mu_n C_{ox} W}{2 L} (V_{gs} - V_T) [(V_{gs} - V_T) || (L E_{sat})]$$

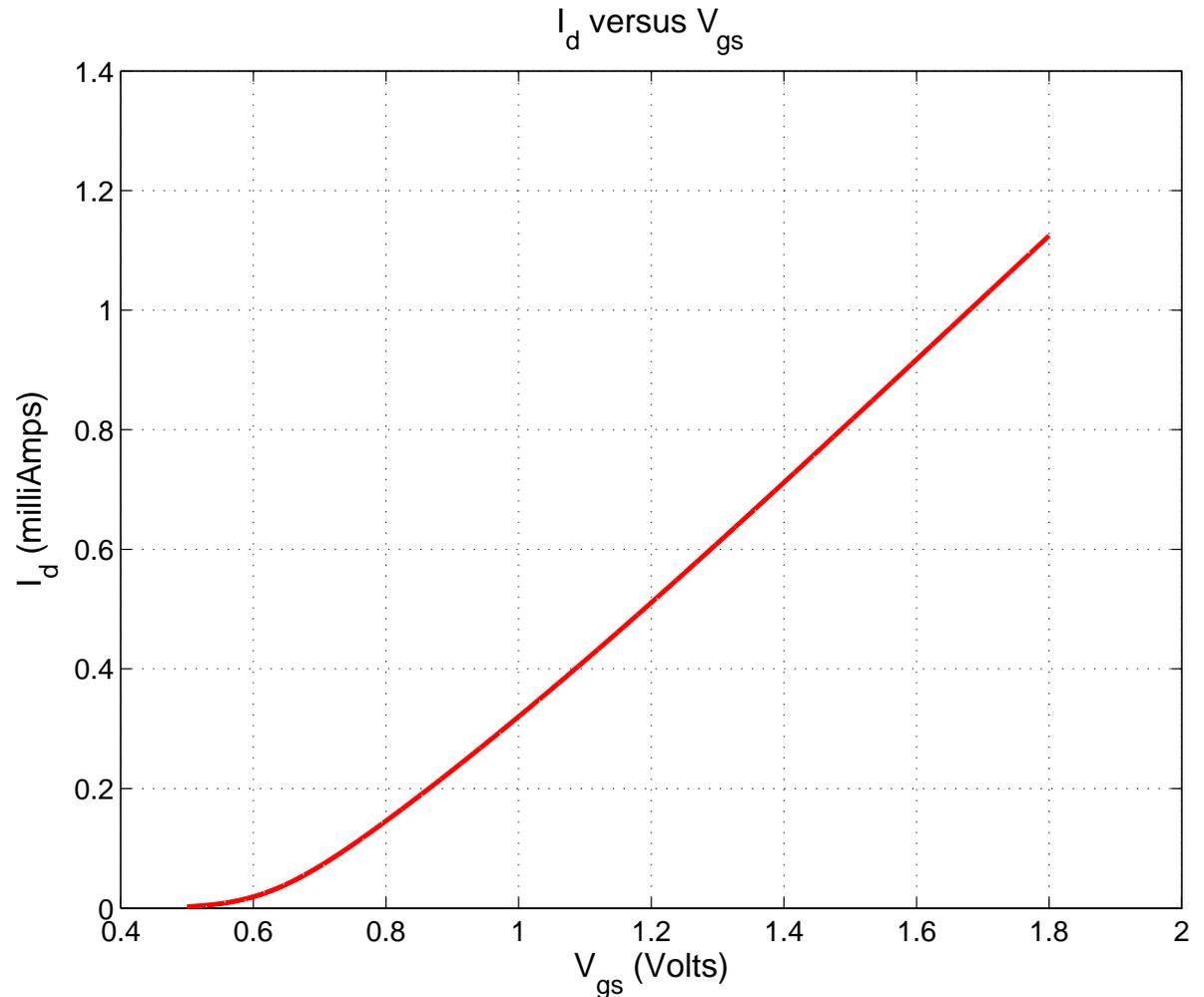
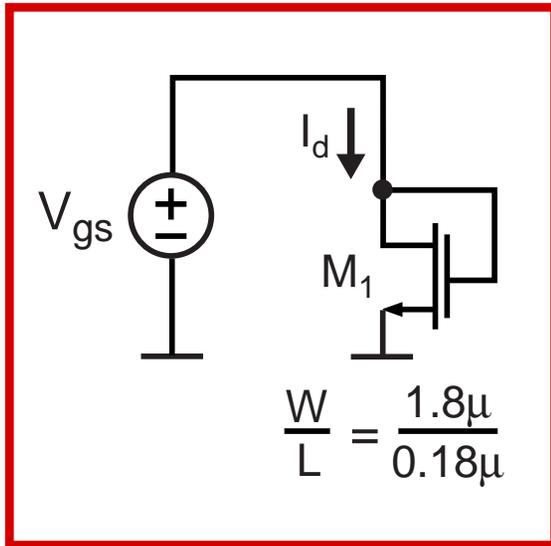
$$\Rightarrow I_D \approx \frac{\mu_n C_{ox}}{2} W (V_{gs} - V_T) E_{sat}$$

- Current increases linearly with $V_{gs} - V_T$
 - Current no longer depends on L !
-
- Note: above is extreme case of velocity saturation!
 - In practice, modern devices operate somewhere between square law and extreme velocity saturation

Useful References for Velocity Saturation

- **For a physics approach**
 - See Lundstrom et.al., “Essential Physics of Carrier Transport in Nanoscale MOSFETS”, IEEE Transactions on Electron Devices, Jan 2002
- **For an engineering model**
 - See Toh et. al., “An Engineering Model for Short-Channel MOS Devices”, JSSC, Aug 1988, pp 950-958
- **In this class**
 - We will simply do a quick experimental hack job at assessing its impact

Investigate Velocity Saturation Issue for 0.18 μ Device



- Linear curve for I_d versus V_{gs}
 - Velocity saturation is indeed an issue
 - How does this impact digital design?

Impact of Voltage and Sizing On Drive Current

Square Law Device

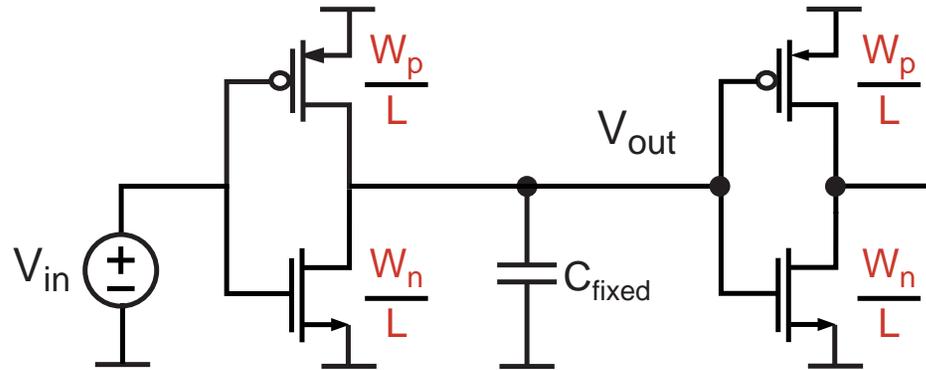
$$I_D = \frac{\mu_n C_{ox} W}{2L} (V_{gs} - V_T)^2$$

Velocity Saturated Device

$$I_D \approx \frac{\mu_n C_{ox}}{2} W (V_{gs} - V_T) E_{sat}$$

- **Voltage supply**
 - Drive current increases with higher drive voltage
- **Width**
 - Current scales proportionally
- **Length**
 - Current scales inversely proportional for square-law device
 - No dependence for purely velocity saturated device

Impact of Voltage and Sizing on Capacitance



$$C_{\text{tot}} = C_{\text{db1}} + C_{\text{db2}} + C_{\text{gs3}} + C_{\text{gs4}} + K(C_{\text{ov3}} + C_{\text{ov4}}) + C_{\text{fixed}}$$

- **Voltage supply (and voltage swing)**
 - Has no impact on capacitance (to first order)
- **Sizing of NMOS and PMOS devices**
 - Input capacitance proportional to product of width and length of transistor

$$C_{gs} \propto WL$$

- Junction and overlap capacitance proportional to W

$$C_{db} \propto W, \quad C_{ov} \propto W$$

Designing For High Speed

- Want the highest ratio of drive current to load capacitance
- Increased supply voltage

Drive current 

Capacitance 

- Increased transistor width

Drive current 

Capacitance 

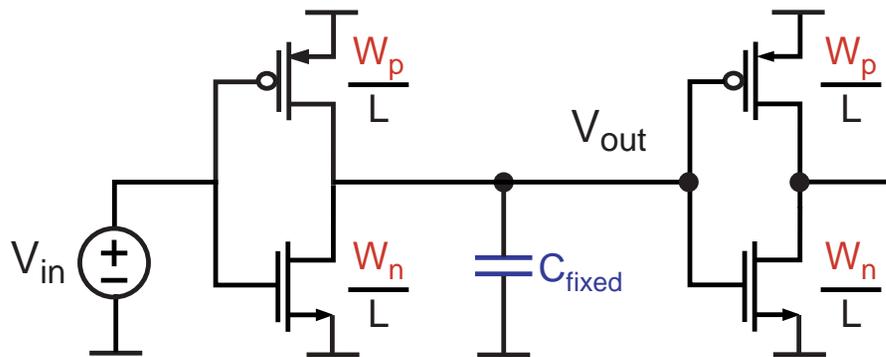
- Increased transistor length

Drive current  or 

Capacitance 

Want high voltage supply and small length to achieve high speed

Setting of Transistor Width for High Speed

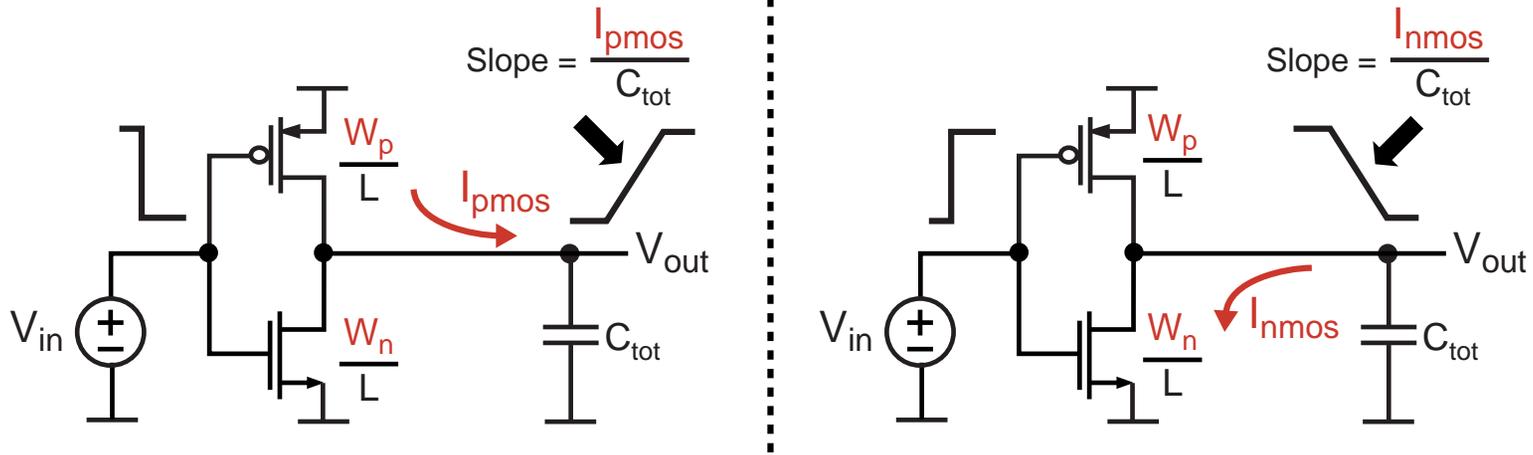


$$C_{\text{tot}} = \frac{C_{\text{db1}} + C_{\text{db2}} + C_{\text{gs3}} + C_{\text{gs4}} + K(C_{\text{ov3}} + C_{\text{ov4}})}{1} + C_{\text{fixed}}$$

$$C_{\text{tot}} = W_n C_{\text{ndevices}} + W_p C_{\text{pdevices}} + C_{\text{fixed}}$$

- **Intrinsic performance of device not influenced by W**
 - Current/capacitance ratio (considering only device capacitance) is constant with changing W (to first order)
- **Within circuit, speed is improved by increasing W when C_{fixed} is significant with respect to device capacitance**
 - W should be chosen such that device capacitance equals or exceeds fixed wiring capacitance

Relative Sizing of NMOS and PMOS Devices



■ Comparison of NMOS and PMOS current drive

Square Law Device

NMOS
$$I_D = \frac{\mu_n C_{ox} W}{2 L} (V_{gs} - V_{Tn})^2$$

PMOS
$$I_D = \frac{\mu_p C_{ox} W}{2 L} (|V_{gs}| - V_{Tp})^2$$

Velocity Saturated Device

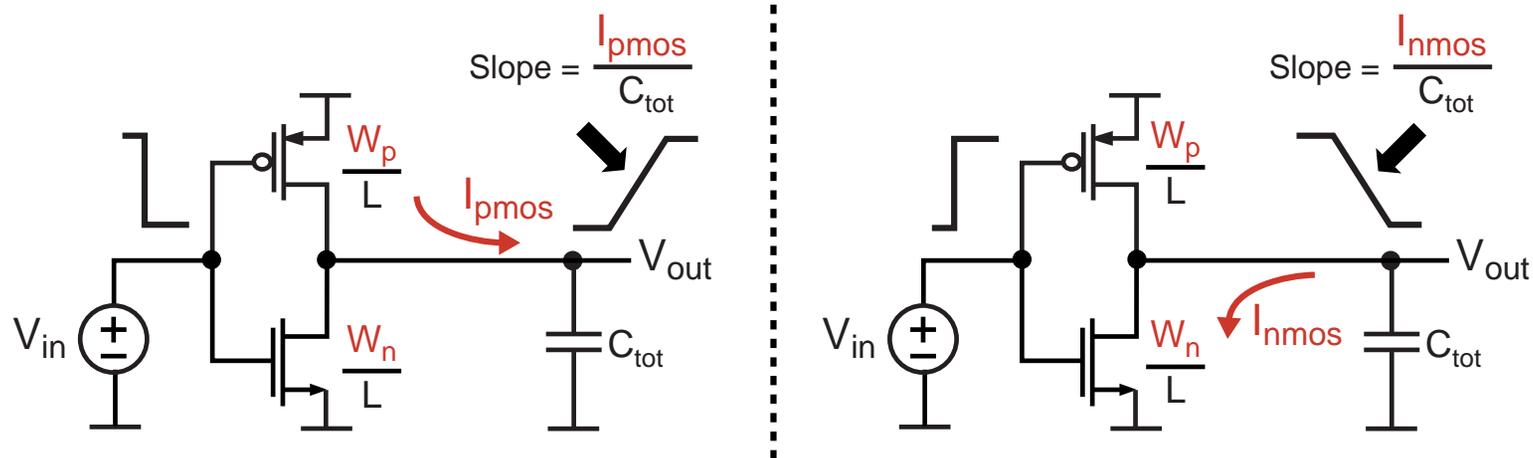
$$I_D \approx \frac{\mu_n C_{ox}}{2} W (V_{gs} - V_{Tn}) E_{sat}$$

$$I_D \approx \frac{\mu_p C_{ox}}{2} W (|V_{gs}| - V_{Tp}) E_{sat}$$

Primary difference – mobility values (μ_n versus μ_p)

■ Capacitance relationships the same for NMOS and PMOS

Relative Sizing to Match Propagation Delays

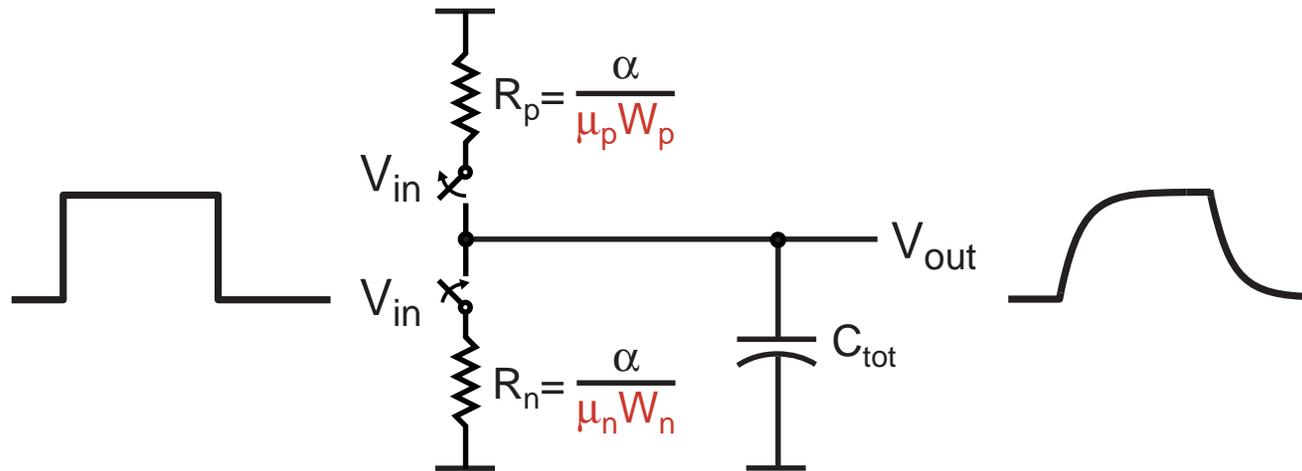


- **Equate drive currents to get same slope when charging and discharging load capacitance**
 - Assume minimum L for NMOS and PMOS for high speed
 - Choose W values to accommodate difference between NMOS and PMOS mobility values

$$\Rightarrow \frac{W_p}{W_n} = \frac{\mu_n}{\mu_p} \approx 2.5 \quad (\text{for } 0.18\mu \text{ CMOS})$$

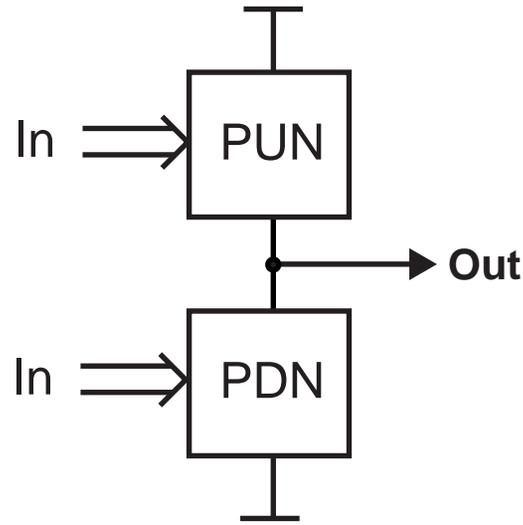
- **Size PMOS devices 2.5 times larger than NMOS!**

Modeling Propagation Delays with Resistance



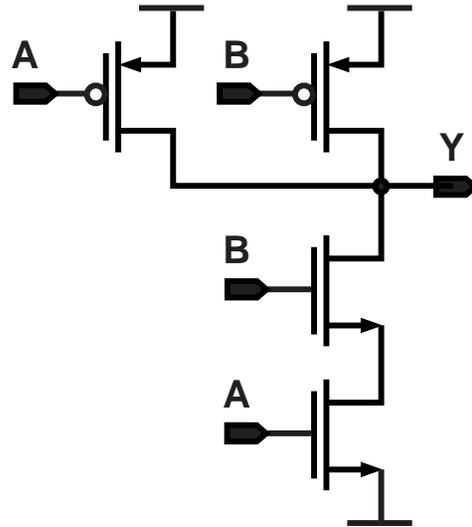
- We can visualize impact of relative transistor sizing between NMOS and PMOS by using switched resistances to represent their current drive
 - Choose α parameter to match propagation times of actual circuit (assume α has same value for NMOS and PMOS)
 - We see that increasing mobility or width reduces resistance
 - Intuitively illustrates impact of these parameters on drive current
- To match propagation delays, set $R_p = R_n \Rightarrow \frac{W_p}{W_n} = \frac{\mu_n}{\mu_p}$

Complementary CMOS Logic



- **Composed of pull-up and pull-down networks that are duals of each other**
 - Each network composed of NAND (series connection) and/or NOR (parallel connection) functions
- **Advantage**
 - No static power (except leakage)

Example: NAND Gate



- **Boolean function**

$$Y = \overline{A \cdot B}$$

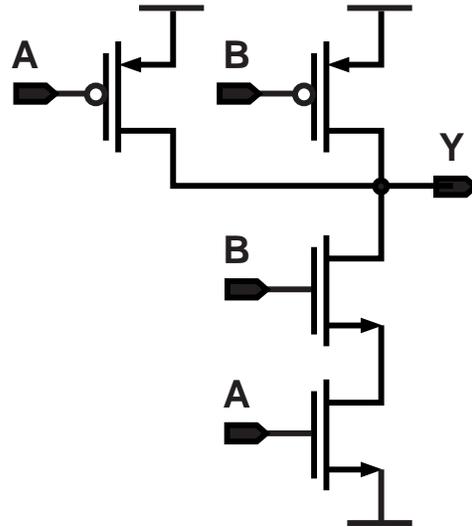
- **PDN performs NAND operation**

$$\text{PDN} = \overline{A \cdot B} \Rightarrow \text{series NMOS}$$

- **PUN is dual of PDN**

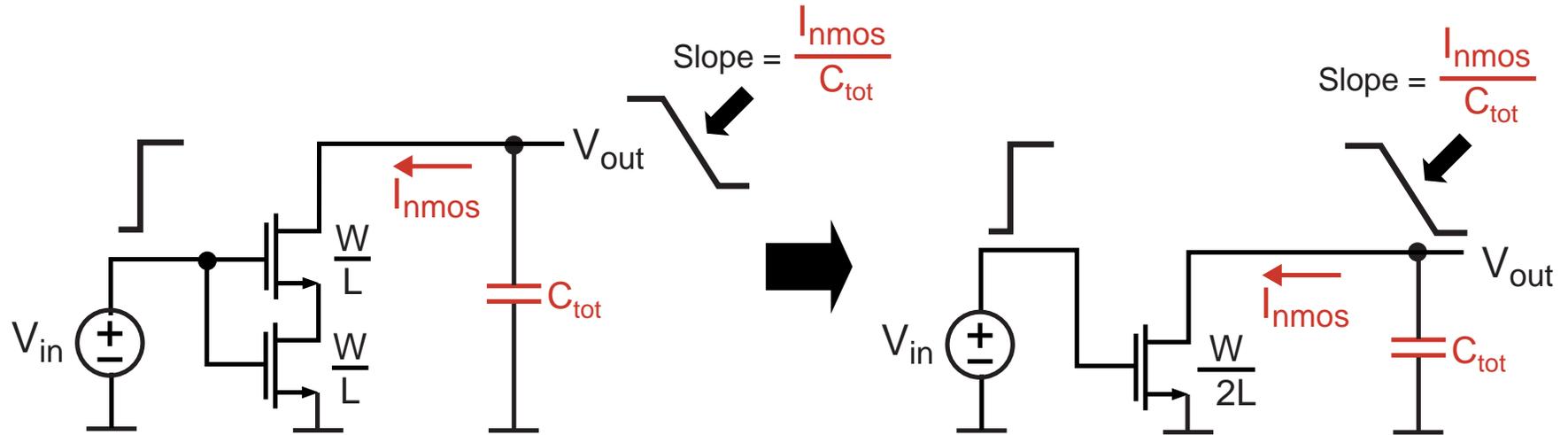
$$\text{PUN} = \overline{\overline{\overline{A \cdot B}}} = \overline{\overline{A + B}} \Rightarrow \text{parallel PMOS}$$

Modeling Dynamic Performance of NAND Gate



- Assume NMOS devices are same size and PMOS devices are same size
- Modeling of parallel devices (in PUN above) is straightforward
 - Simply represent with parallel switched resistors
- Modeling of series devices (in PDN above) is not immediately obvious
 - We need to do further investigation

Equivalent Transistor Model of Stacked Transistors



- **Drive current is created only when both devices are on**
 - We can hook gates together without loss of generality
 - Resulting configuration is equivalent (at least to first order) to a single device with twice the length
- **Issue – if device velocity saturated, what’s the impact?**

Square Law Device

$$I_D = \frac{\mu_n C_{ox} W}{2L} (V_{gs} - V_T)^2$$

Velocity Saturated Device

$$I_D \approx \frac{\mu_n C_{ox}}{2} W (V_{gs} - V_T) E_{sat}$$

Let's Do A Test

- In Hspice, simulate the output current of an NMOS transistor with a given V_{gs} bias
 - Vary the length of the transistor
 - Scale the current by the length

- For square law device

$$L \cdot I_D = L \cdot \frac{\mu_n C_{ox} W}{2 L} (V_{gs} - V_T)^2$$

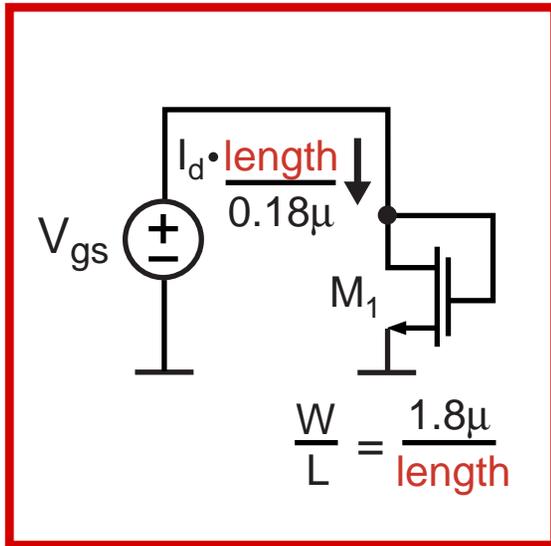
- Product independent of length

- For velocity saturated device

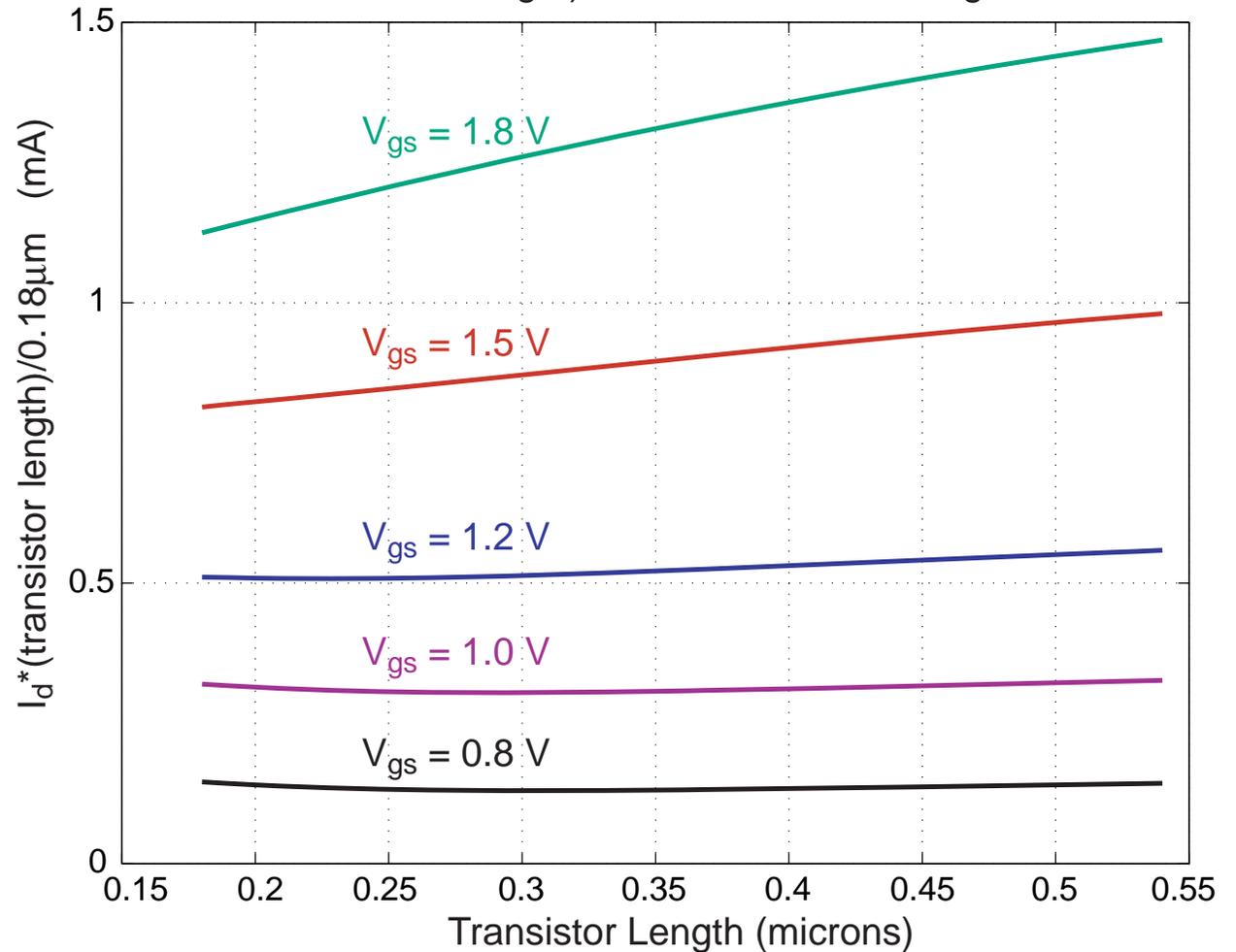
$$L \cdot I_D \approx L \cdot \frac{\mu_n C_{ox}}{2} W (V_{gs} - V_T) E_{sat}$$

- Product increases with length

Length Normalized Drain Current – 0.18 μ NMOS Device

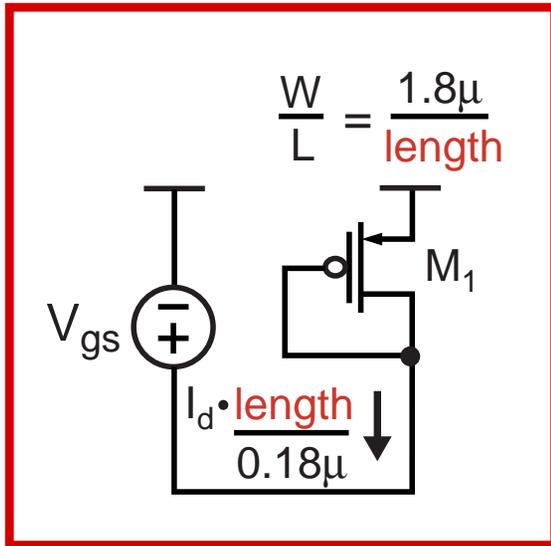


NMOS Drain Current (Multiplied by Normalized Transistor Length) Versus Transistor Length

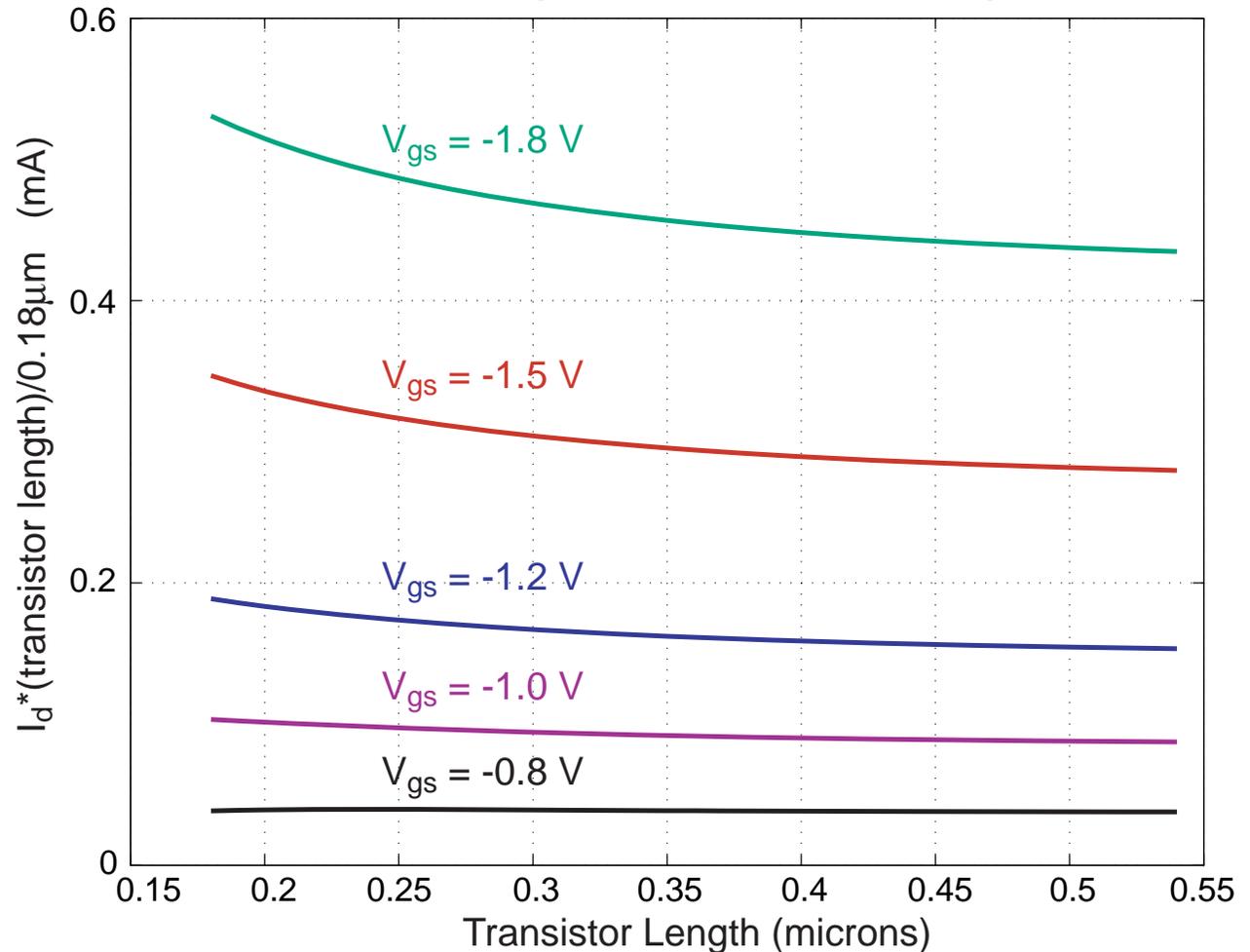


- Product is relatively constant – square law behavior for L

Length Normalized Drain Current – 0.18 μ PMOS Device

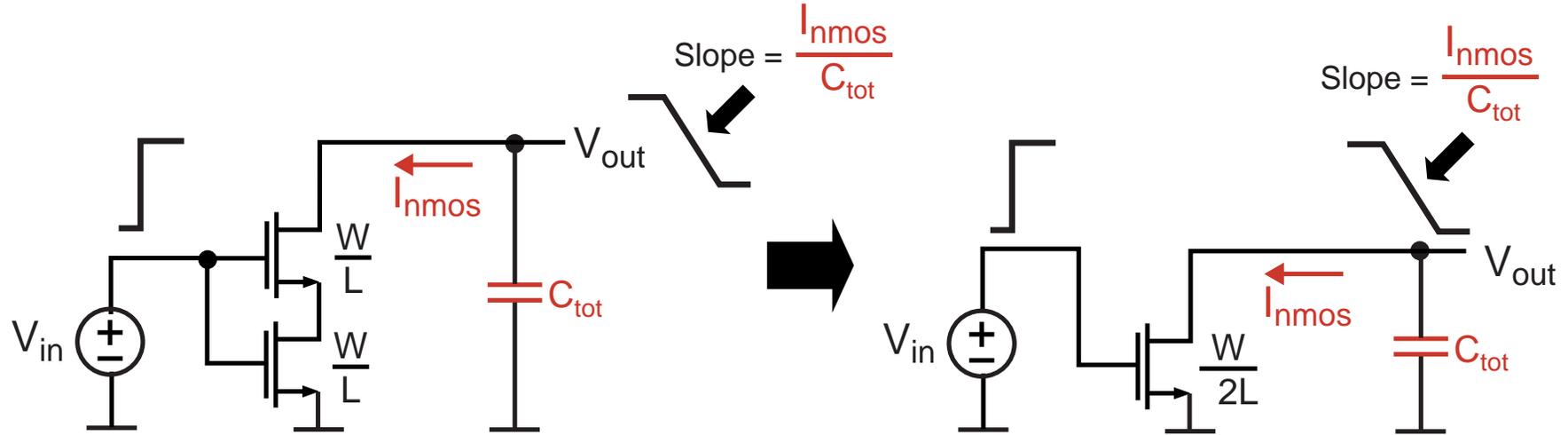


PMOS Drain Current (Multiplied by Normalized Transistor Length) Versus Transistor Length



- Product is relatively constant – square law behavior for L

Back to Dynamic Modeling of Stacked Transistors

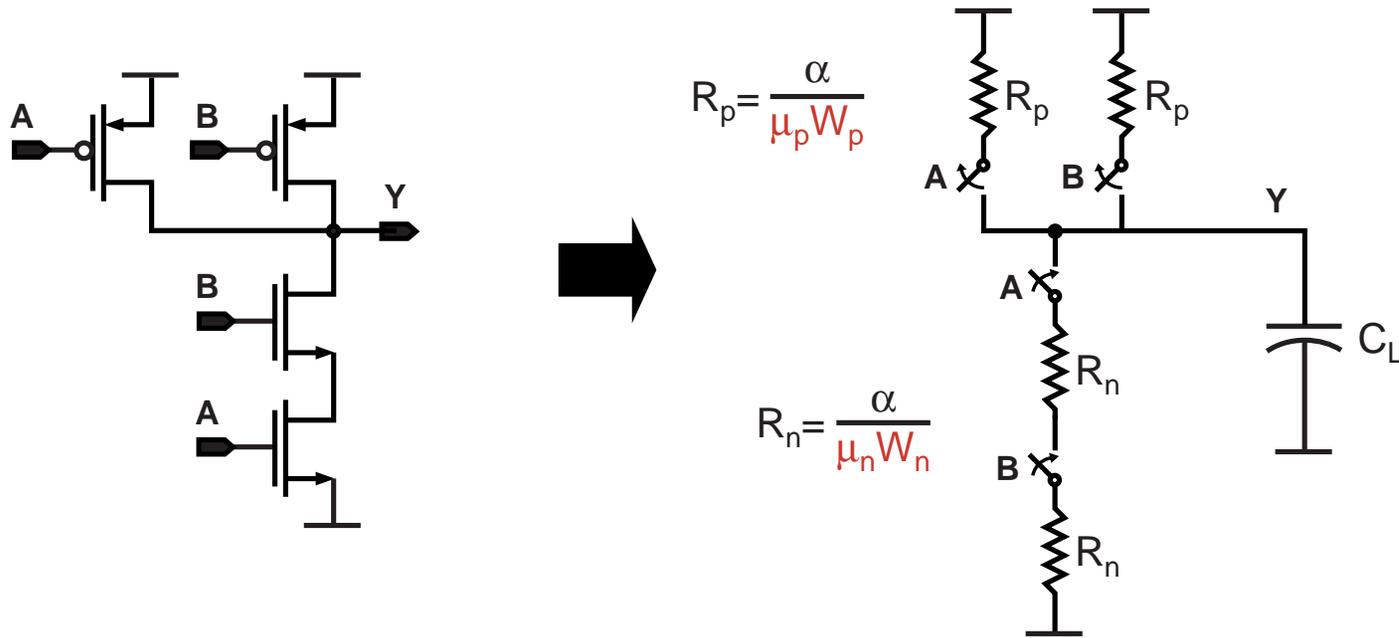


- Since we can assume square law behavior with respect to impact of L

$$\Rightarrow I_D \propto \frac{1}{L}$$

- Model with two switched resistors in series
 - Represents the fact that we have half the drive current

Dynamic Model of NAND Gate

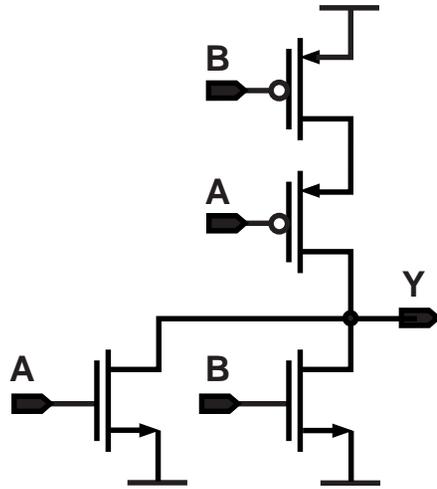


- To match worst case propagation delays

$$\Rightarrow 2R_n = R_p$$

$$\Rightarrow W_p = \frac{1}{2} \left(\frac{\mu_n}{\mu_p} \right) W_n = 1.25W_n \text{ (for } 0.18\mu \text{ CMOS)}$$

Another Example: NOR Gate



- **Boolean function**

$$Y = \overline{A + B}$$

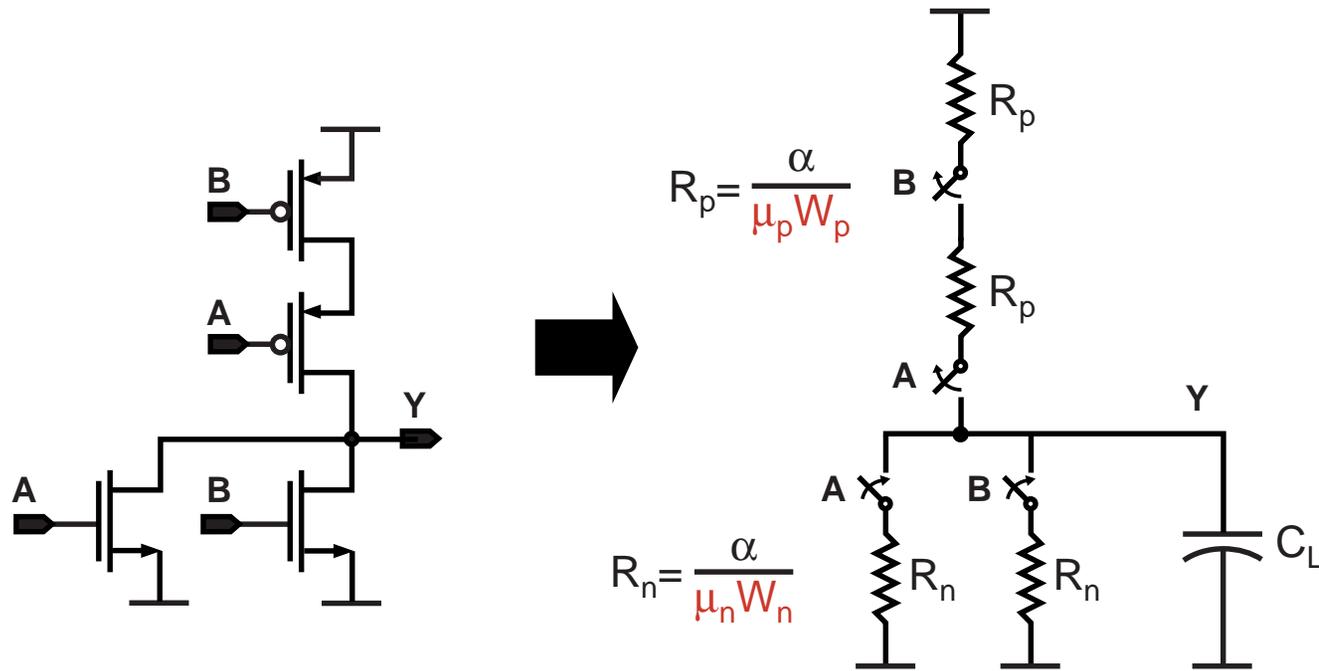
- **PDN performs NAND operation**

$$\text{PDN} = \overline{A + B} \Rightarrow \text{parallel NMOS}$$

- **PUN is dual of PDN**

$$\text{PUN} = \overline{\text{PDN}} = \overline{\overline{A + B}} = \overline{\overline{A} \cdot \overline{B}} \Rightarrow \text{series PMOS}$$

Dynamic Model of NOR Gate

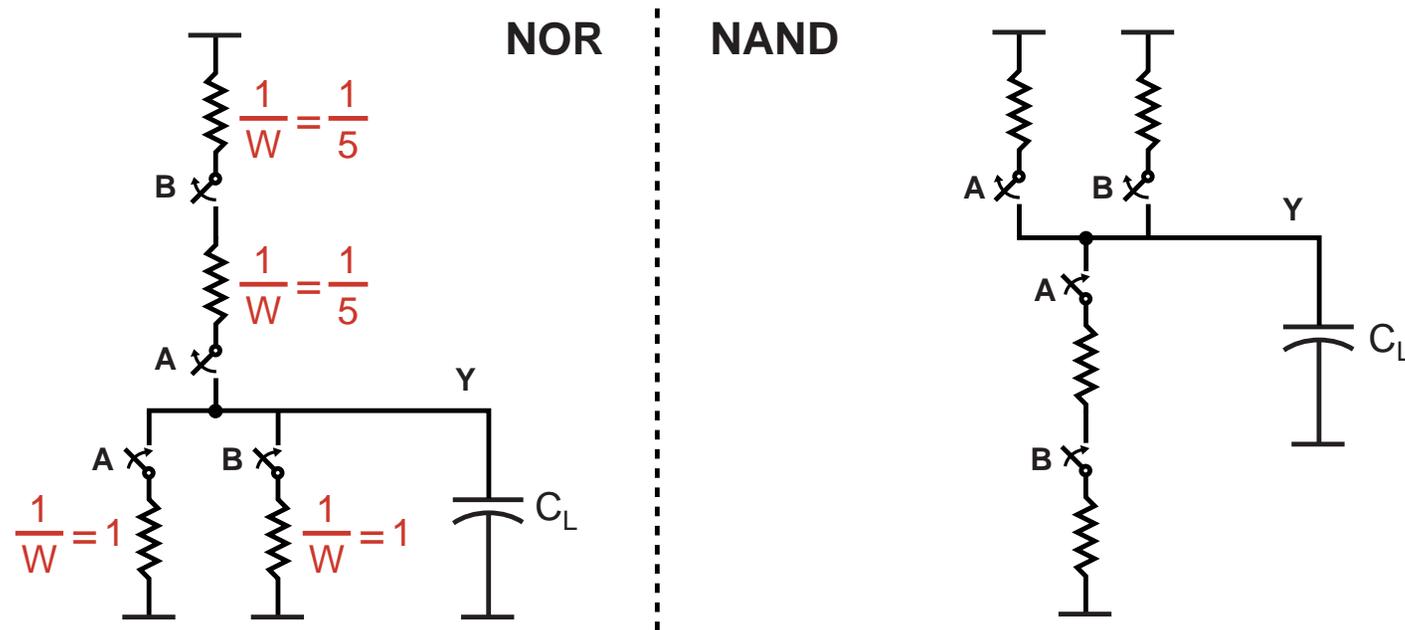


- To match worst case delays

$$\Rightarrow 2R_p = R_n$$

$$\Rightarrow W_p = 2 \left(\frac{\mu_n}{\mu_p} \right) W_n = 5W_n \text{ (for } 0.18\mu \text{ CMOS)}$$

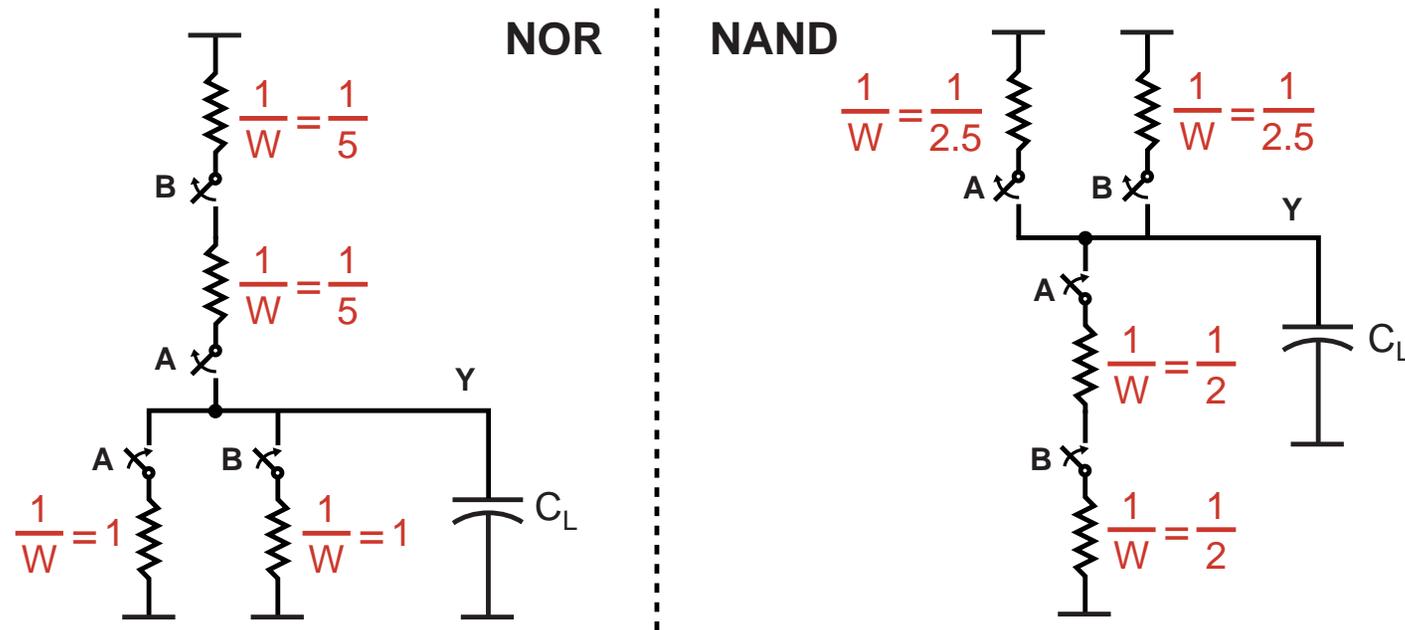
Comparing the Dynamic Performance of Gates (Step 1)



■ NOR

- Normalize performance by setting NMOS widths to 1
- PMOS widths set to 5 to match propagation delay

Comparing the Dynamic Performance of Gates (Step 2)



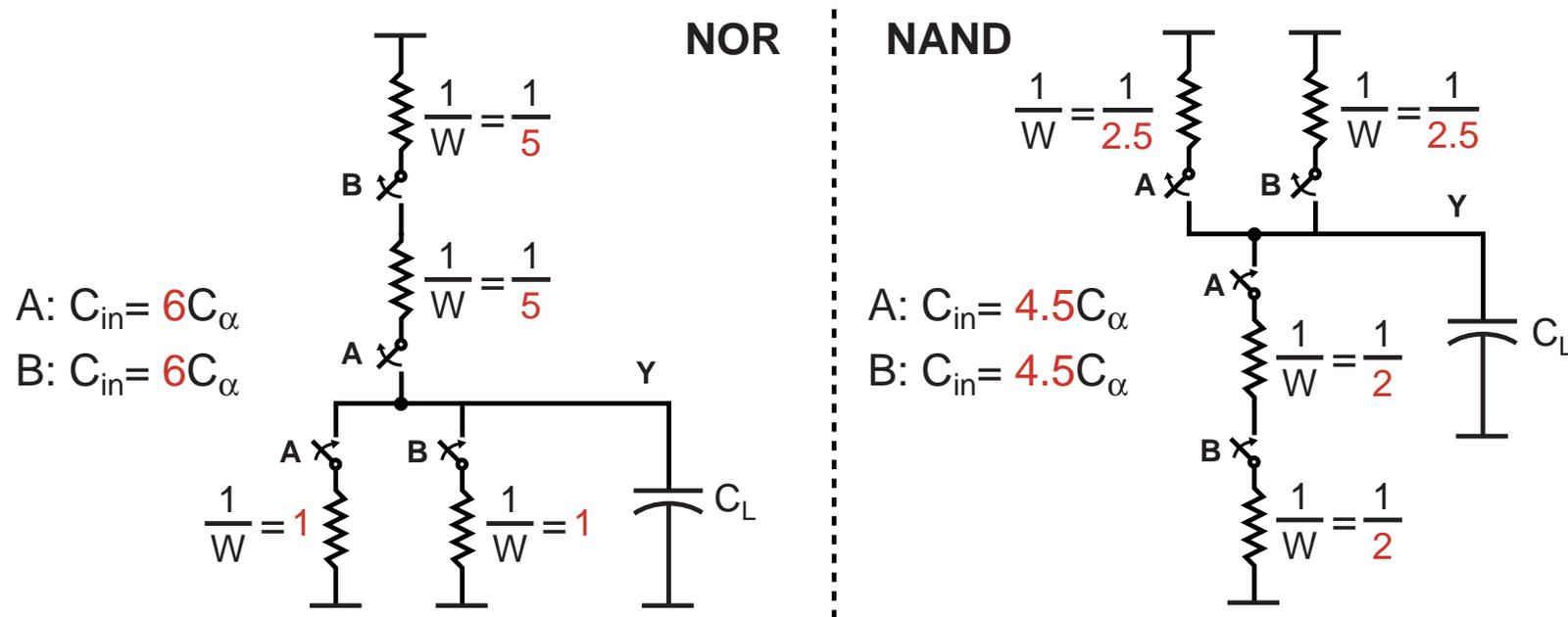
■ NOR

- Normalize performance by setting NMOS widths to 1
- PMOS widths set to 5 to match NMOS propagation delay

■ NAND

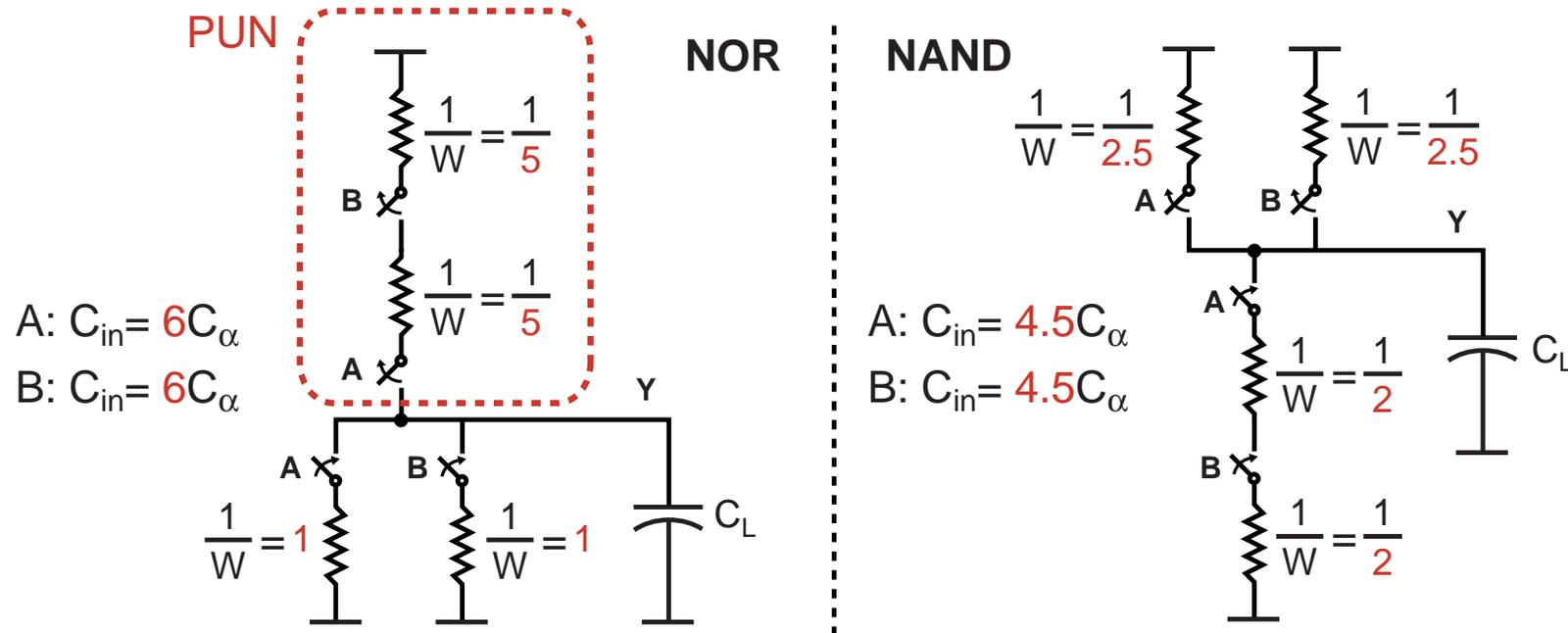
- Match NOR by setting NMOS widths to 2
- PMOS widths set to 2.5 to match NMOS propagation delay

Comparing the Dynamic Performance of Gates (Step 3)



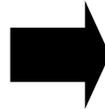
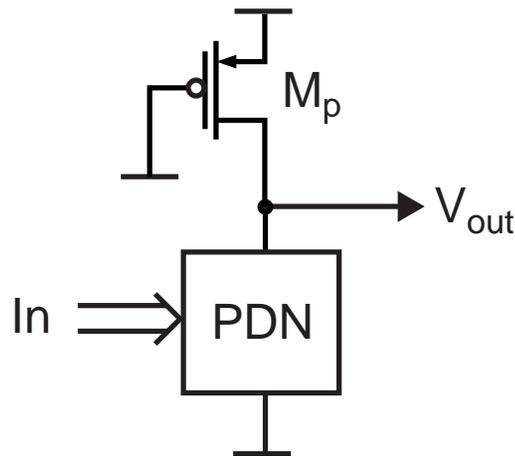
- Compare the input device capacitance of each gate
 - Proportional to width of devices connected to a given input
 - Define C_α as a capacitance scaling factor
 - Includes impact of C_{ox} , L, etc.
- We see that the NAND gate is faster than the NOR gate
 - Ratio of current drive to capacitance is higher

Issue – Stacked PMOS Transistors Lower Performance

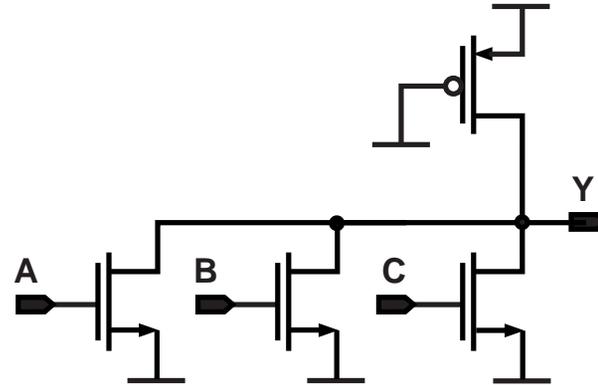


- Why is NOR performance worse than the NAND?
 - PMOS create dominant portion of capacitive load
 - Stacked PMOS require even larger devices
- Can we eliminate the impact of the PMOS devices on input capacitance (i.e. eliminate the PUN)?
 - Could achieve higher speed!

Technique 1 to Eliminate PUN: Pseudo-NMOS



Example: 3 input NOR gate



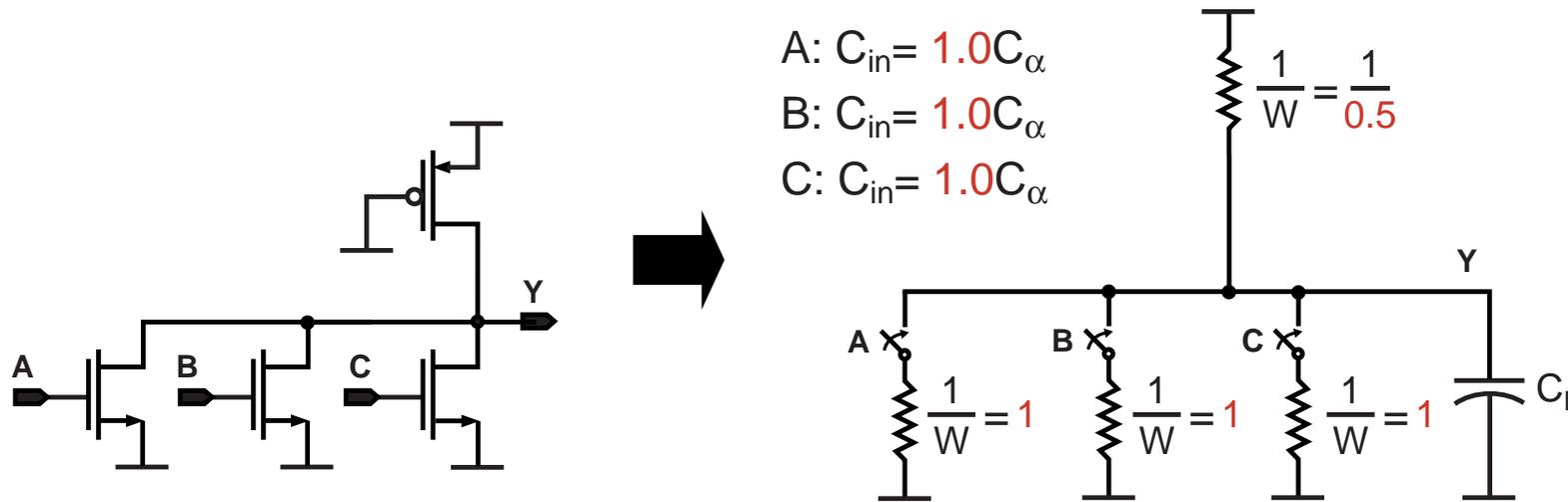
■ Benefit

- Substantial reduction in input capacitance – faster speed!

■ Negatives

- Static power consumption
- Asymmetric propagation delays (falling edge faster)
- Output logic levels set by ratio of NMOS to PMOS width
 - Rule of thumb: Set R_p/R_n to 4 (or more)
 - Alternate rule of thumb: Set $W_p = W_n/2$

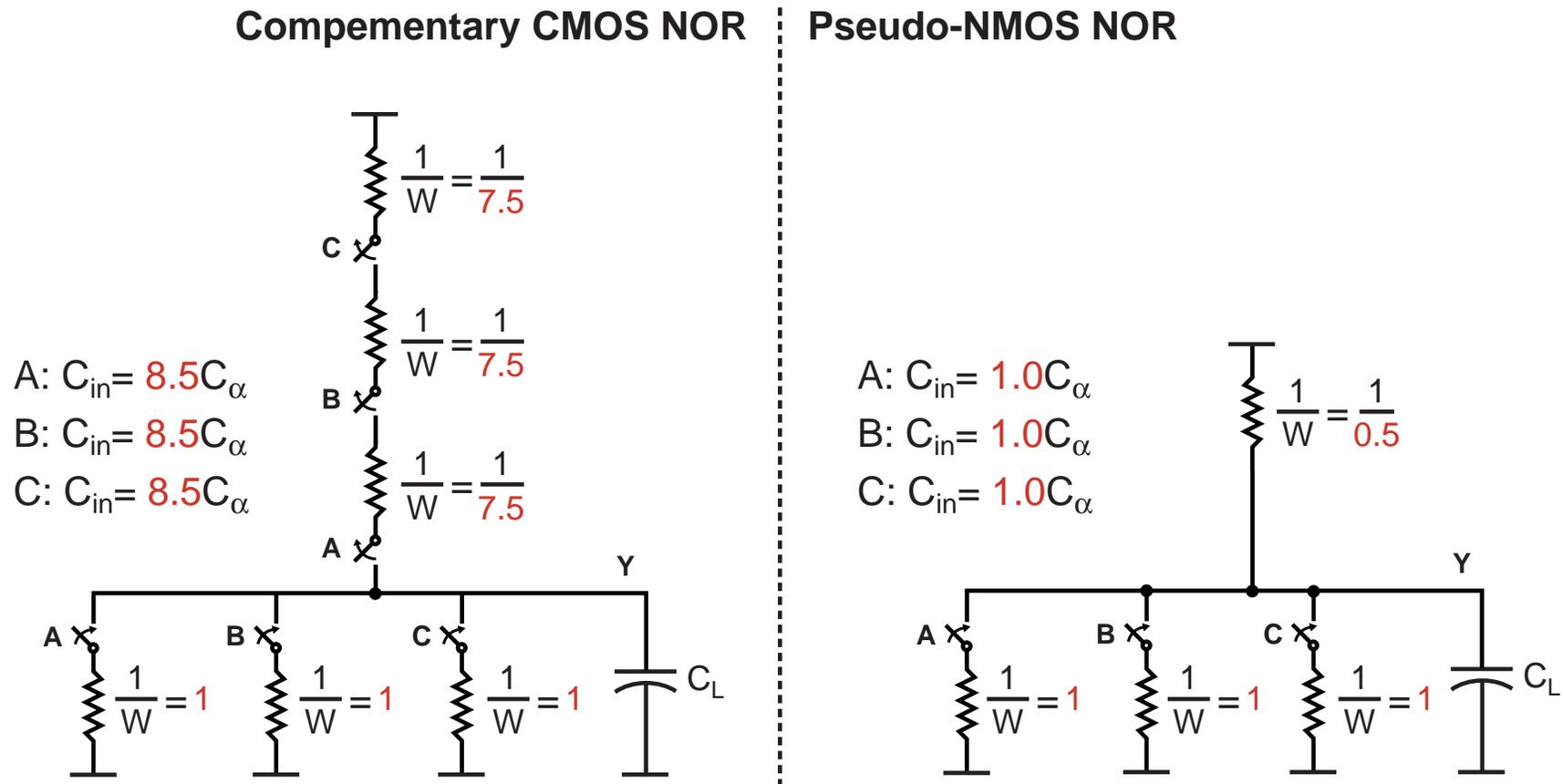
Dynamic Model for Pseudo-NMOS



- Arbitrarily choose NMOS width to be 1
 - Set PMOS width to be 1/2 according to rule of thumb on previous slide
- Note that negative edge transition at output is 5 times faster than the positive edge transition at output

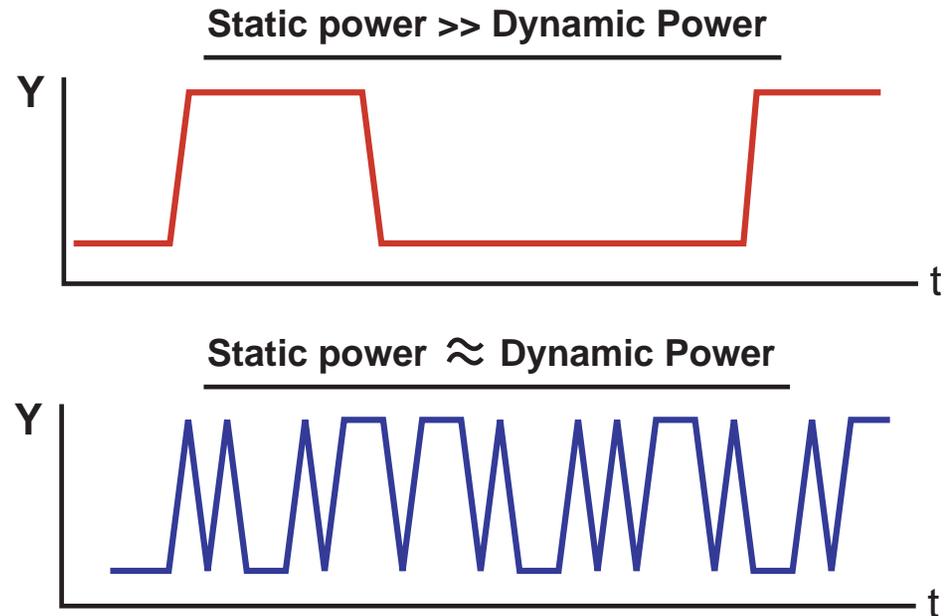
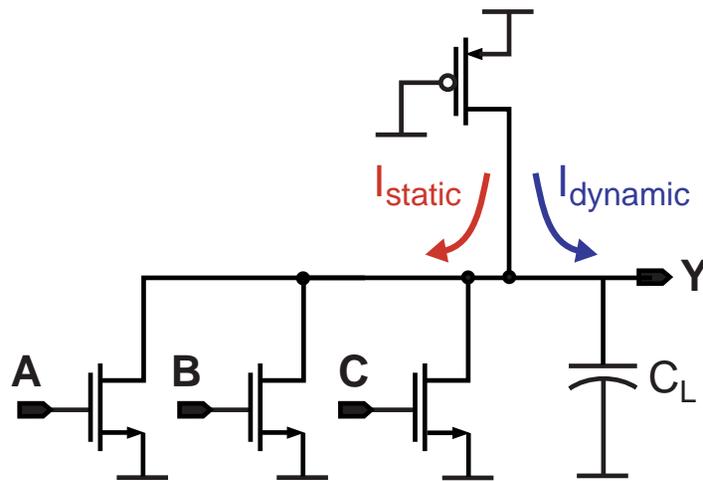
$$\frac{R_p}{R_n} = \frac{\mu_n W_n}{\mu_p W_p} = 2.5 \frac{1}{0.5} = 5 \quad (\text{for } 0.18\mu \text{ CMOS})$$

Comparison of Complementary CMOS vs Pseudo-NMOS



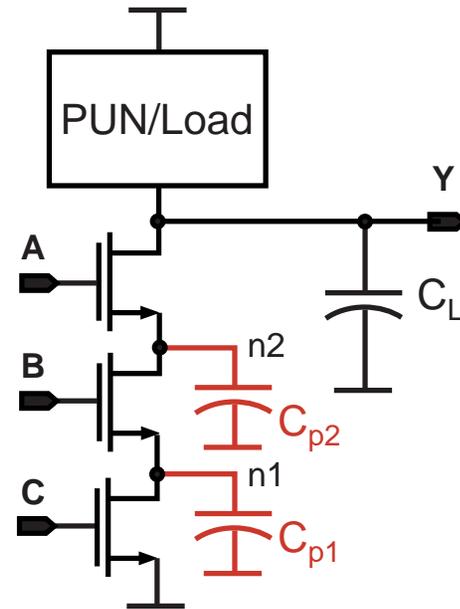
- For same negative transition propagation delay
 - Pseudo-NMOS has nearly 1/10 the input capacitance
- In practice, may want to scale up the pseudo-NMOS sizes to get faster positive transition propagation delay

The Issue of Static Power Dissipation



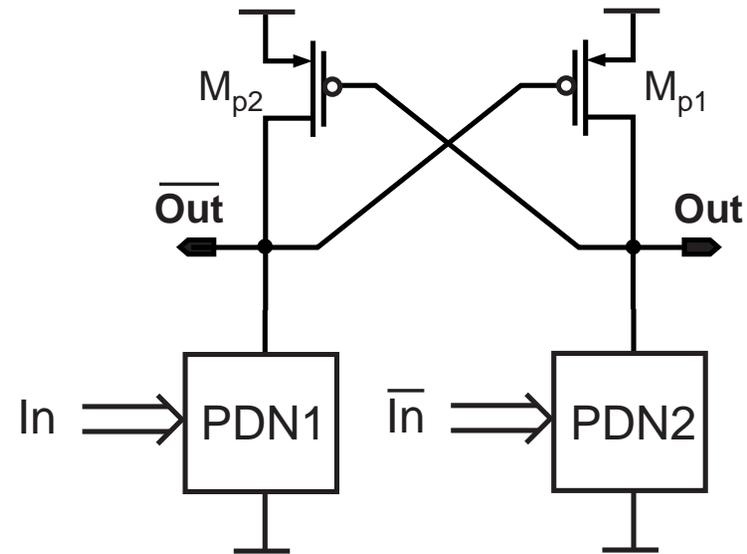
- **Ratio of dynamic power to static power depends on transition activity of output**
 - **For low transition activity, static power is dominant**
 - Could potentially turn off PMOS during quite times?
 - **For high transition activity, static and dynamic power may be similar in value**
 - Pseudo-NMOS can save power due to reduced capacitive loading

Sizing PDN Transistors for High Speed



- **Diffusion capacitance exists on intermediate nodes**
 - Different effective cap load for each PDN transistor
 - Example: transistor C must discharge C_L , C_{p2} , C_{p1}
 - Transistor drive compromised by the floating nodes
 - Example: transistor A has reduced drive for $V_{n2} > 0$
- **Design tips for highest speed**
 - Increase the width of devices farthest from output (trans. C)
 - Place signals that transition last closest to output (trans. A)

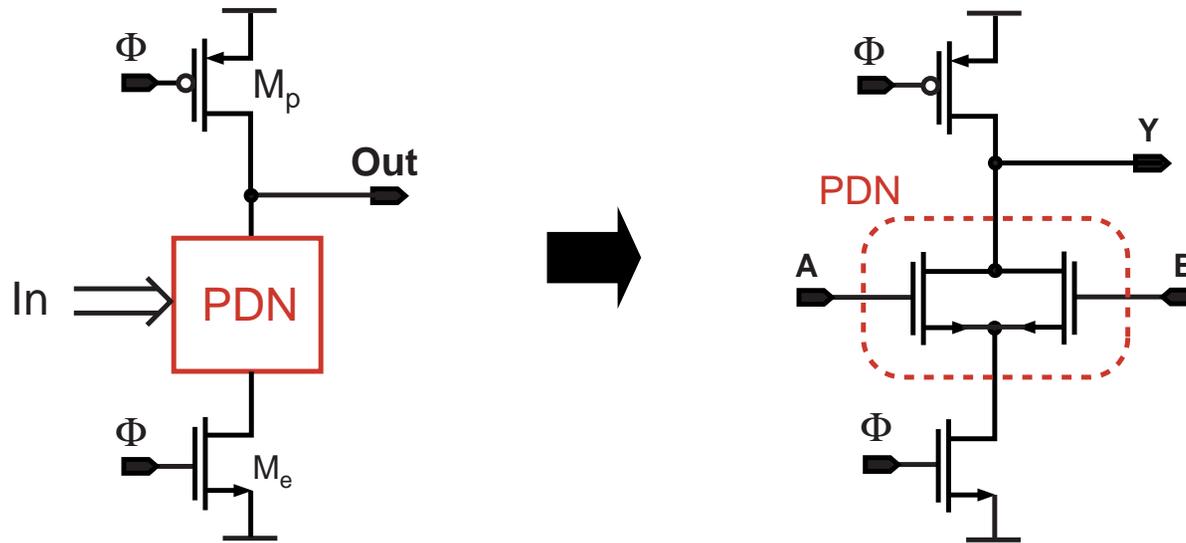
Technique 2 to Eliminate PUN - DCVSL



- **Differential Cascade Voltage Swing Logic (DCVSL)**
 - Employs differential logic structure
 - Faster speed than complementary CMOS
 - No static power dissipation
 - Great for interface between power supply domains
- **Issues**
 - Slower than Pseudo-NMOS (PMOS gates load output)
 - More power than complementary CMOS

Technique 3 to Eliminate PUN (or PDN): Dynamic Logic

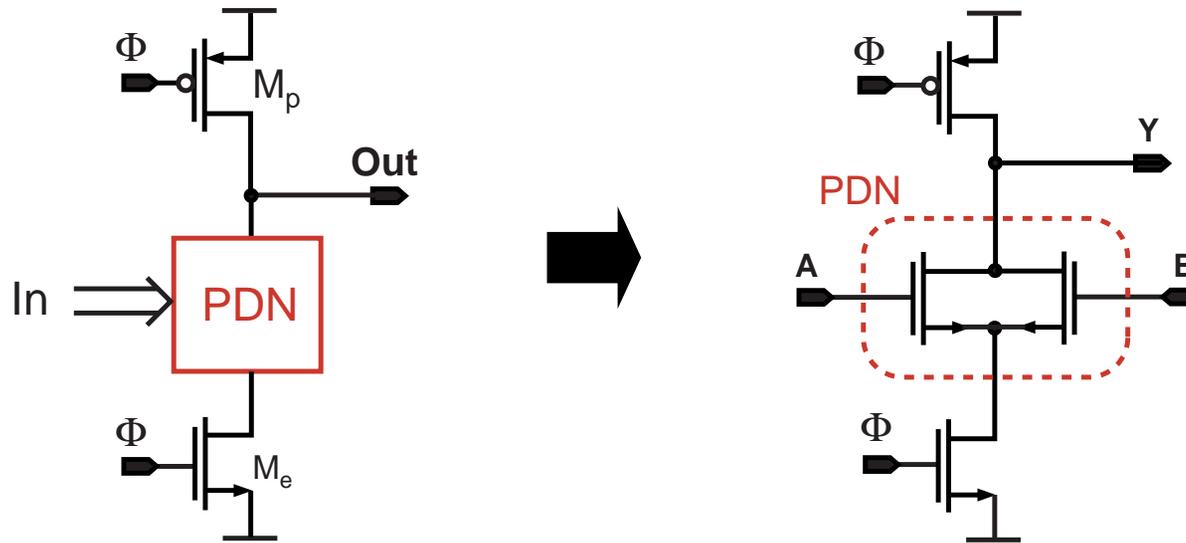
Example: 2 input NOR gate



- Use a clock, Φ , to gate the load and PDN network
 - $\Phi = 0$
 - Precharge the output node
 - Shut off current to PDN
 - $\Phi = 1$
 - Turn off the precharge device
 - Send current to PDN so that it “evaluates” inputs

The Pros and Cons of Dynamic Logic

Example: 2 input NOR gate



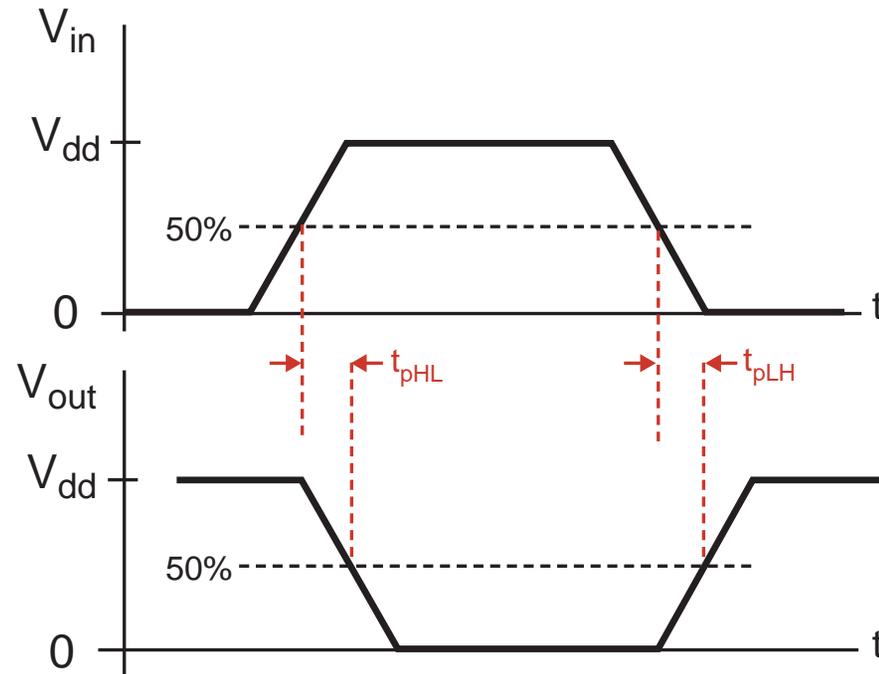
■ Benefits

- High speed (but lower speed than Pseudo-NMOS due to precharge time requirement)
- No static power, non-ratioed, and low number of transistors

■ Issues

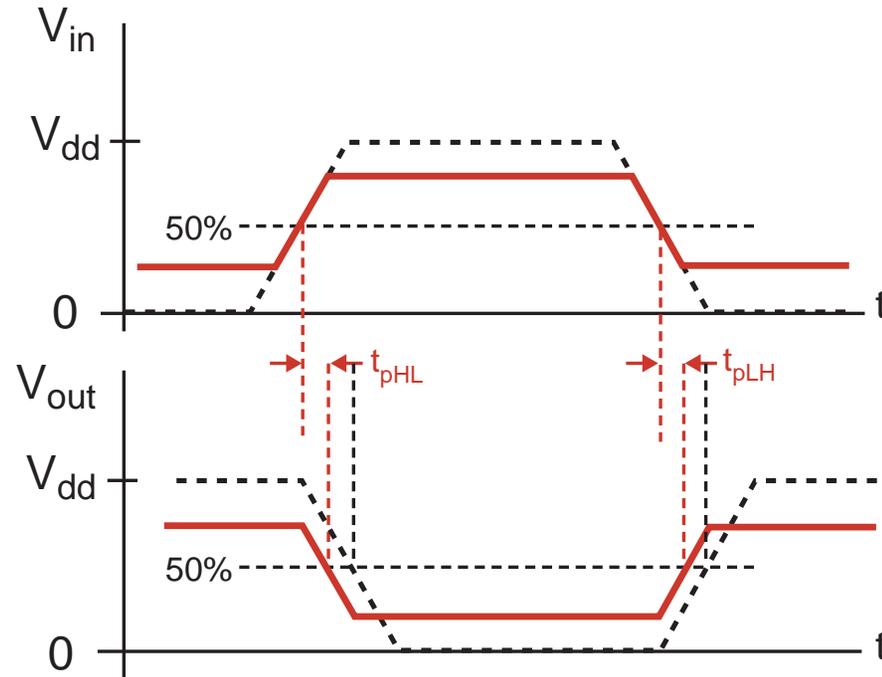
- High design complexity – cascading requires care
- Large clock load, minimum clock speed due to leakage

Increasing Speed By Reducing Voltage Swing



- The propagation delay is defined as time between input and output crossing at 50% amplitude
- We found that increased voltage is beneficial for speed
 - Increased V_{gs} leads to increased drive current to capacitance ratio
- What if we could keep high drive current to capacitance ratio AND reduce the swing?

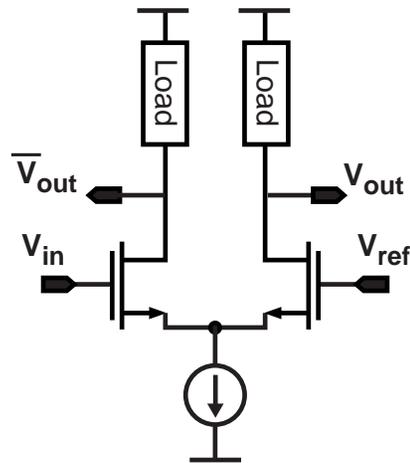
Impact of Reduced Swing with Same Drive Current



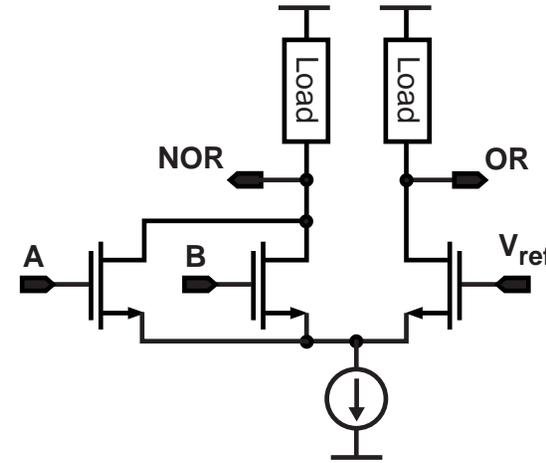
- Propagation time reduced!
- How do we reduce the swing AND achieve high drive current to capacitance ratio?

Technique 4 to Eliminate PUN: Source-Coupled Logic

Basic building block: differential pair

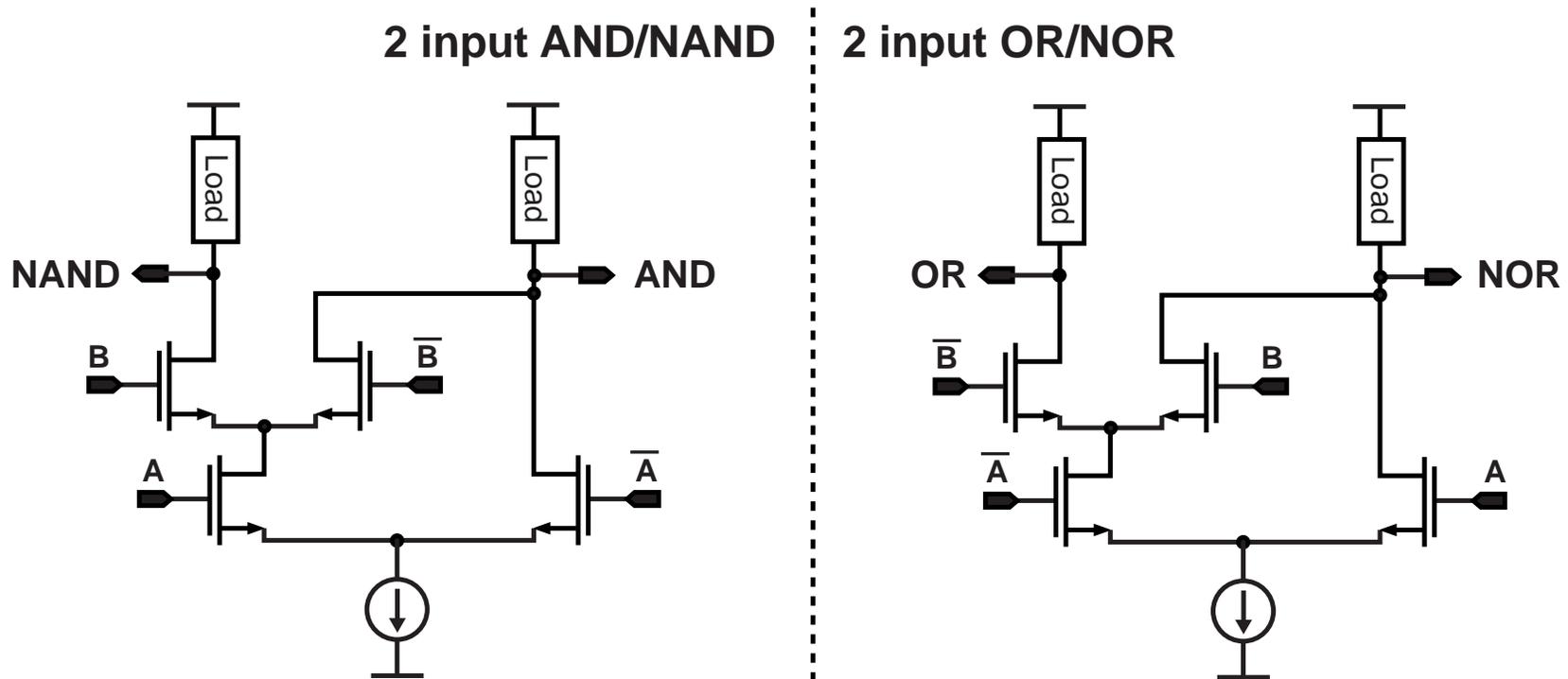


Example: 2 input OR/NOR gate



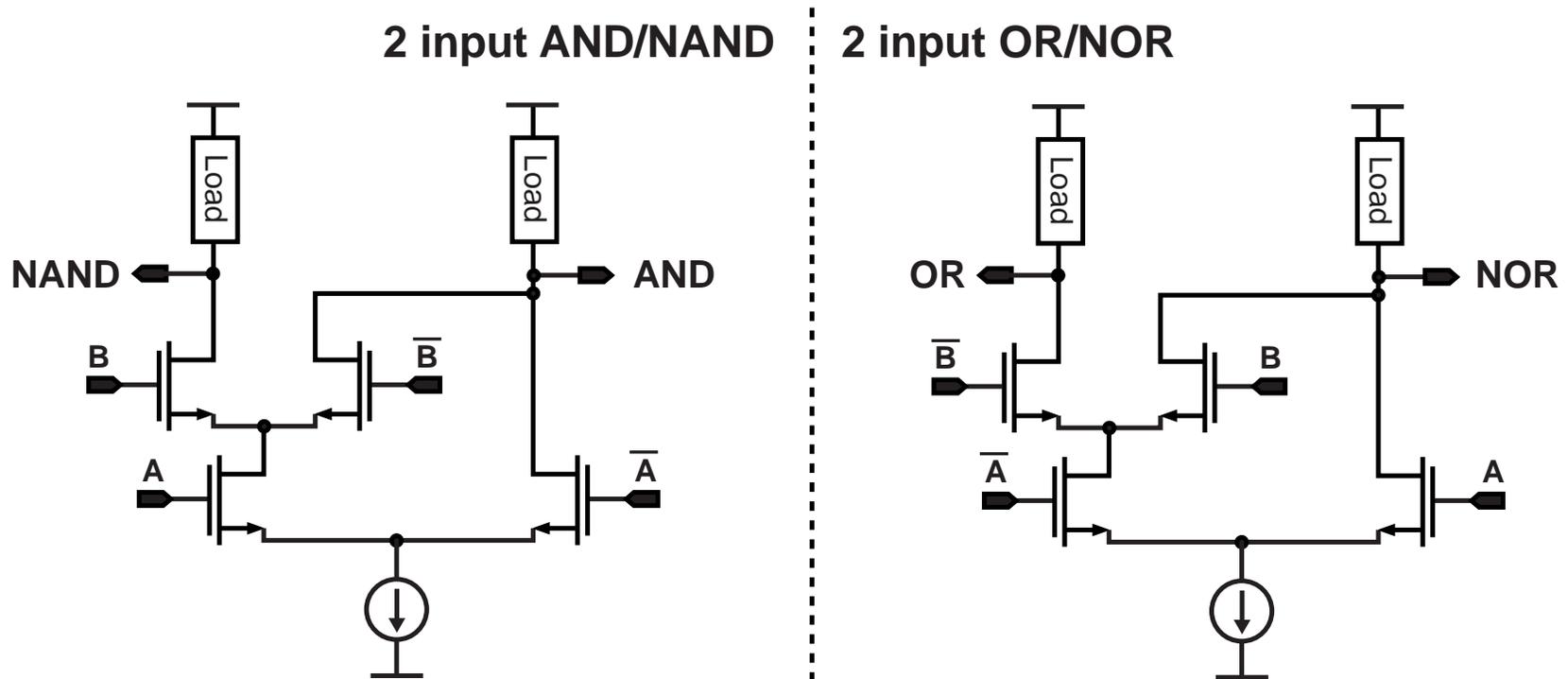
- **Single-ended version – V_{ref} set by bias network**
- **High speed achieved through**
 - Small signal swings
 - Leveraging of a fast amplifier structure
- **Load can be implemented in a variety of ways**
 - Resistor: highest speed, but large area
 - Diode connected PMOS (or NMOS): slower, but small area
 - PMOS in triode region: high speed, but complicated biasing

Logic Realization Using Differential SCL



- Employs differential signaling (no V_{ref})
 - More robust and higher noise margin than singled-ended version
- Ordering of signals yields AND/NAND versus OR/NOR

Comparison of Differential SCL to Full Swing Logic



■ Advantages

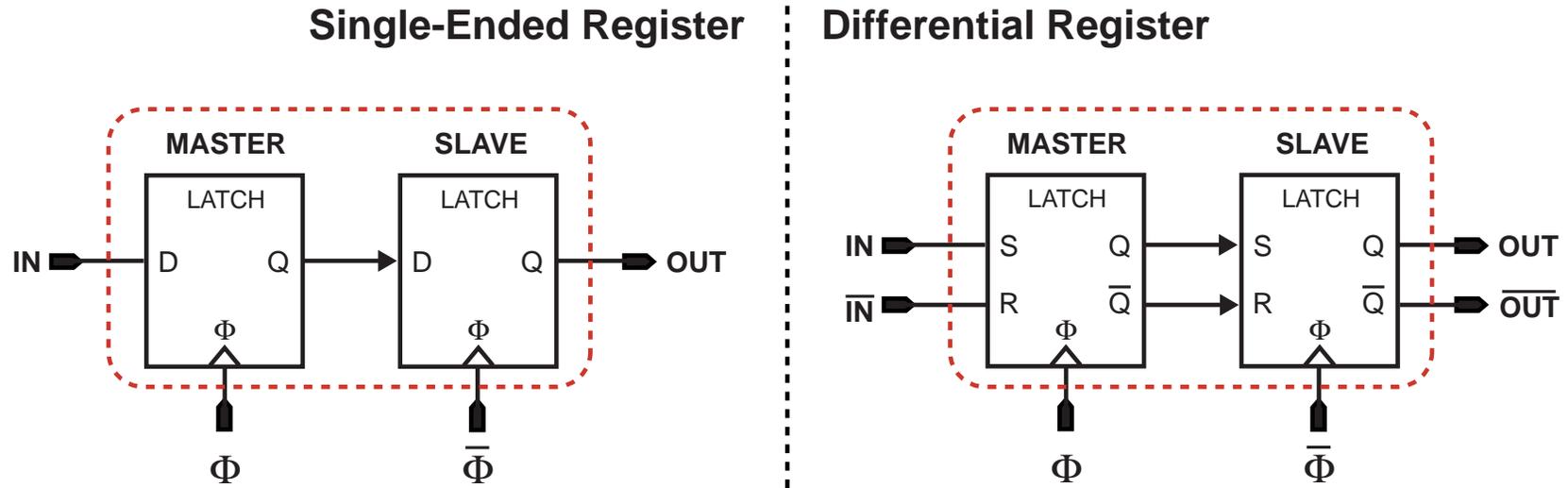
- Much faster speed ($> 2X$ with resistor loads)
- Quieter on supplies (good when analog parts nearby)

■ Disadvantages

- Static current, need for biasing networks
- Logic implementation more clumsy

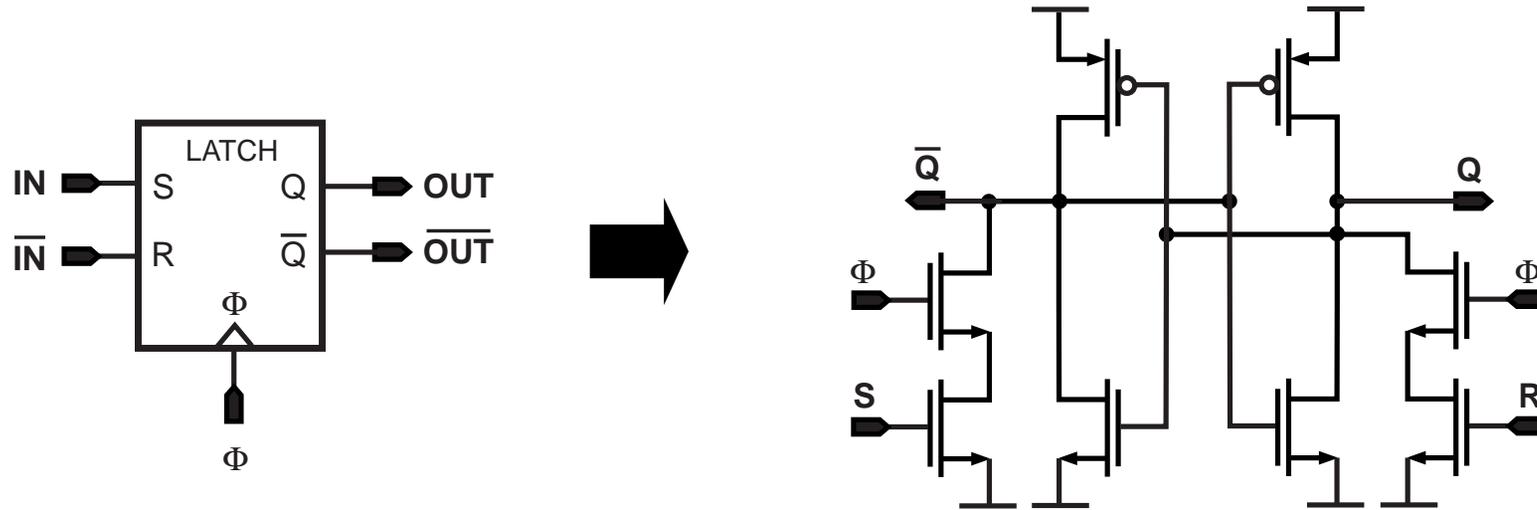
Registers

Edge-triggered Registers



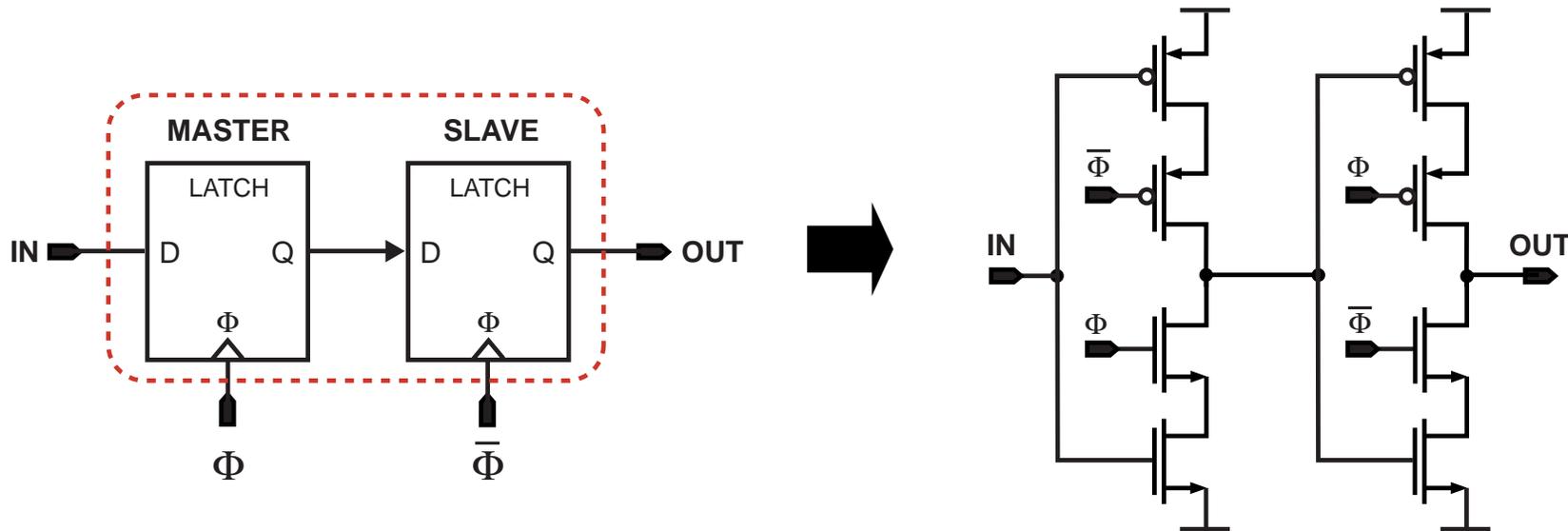
- Achieved by cascading two latches that are transparent out of phase from one another
- Two general classes of latches
 - Static – employ positive feedback
 - Robust
 - Dynamic – store charge on parasitic capacitance
 - Smaller, lower power in most cases
 - Negative: must be refreshed (due to leakage currents)

Static Latches



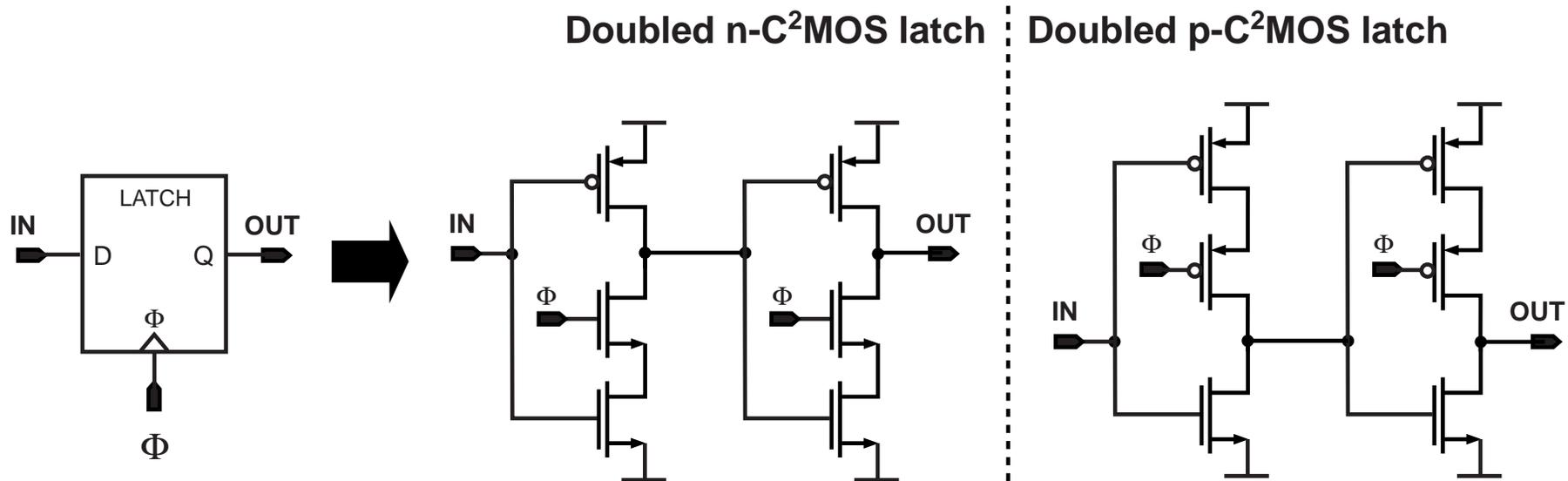
- **Classical case employs cross-coupled NAND/NOR gates to achieve positive feedback**
- **Above example uses cross-coupled inverters for positive feedback**
 - Set, reset, and clock transistors designed to have enough drive to overpower cross-coupled inverters
 - Relatively small number of transistors
 - Robust

Dynamic Latches



- **Leverage CMOS technology**
 - High quality switches with small leakage available
 - Can switch in and store charge on parasitic capacitances quite reliably
- **Achieves faster speed than full swing logic with fewer transistors**
- **Issues: higher sensitivity to noise, minimum refresh rate required due to charge leakage**

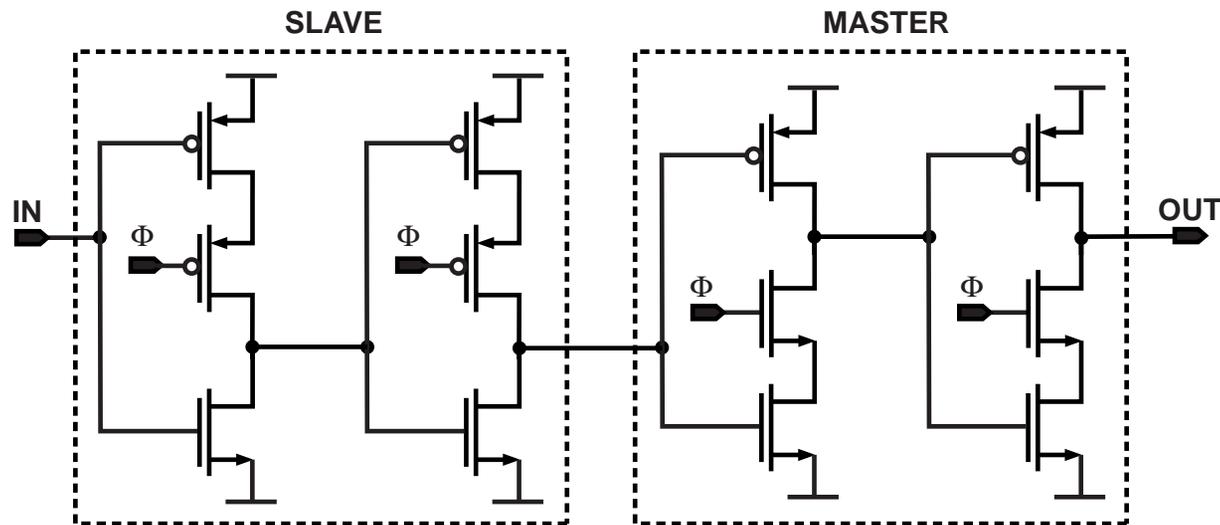
True Single Phase Clocked (TSPC) Latches



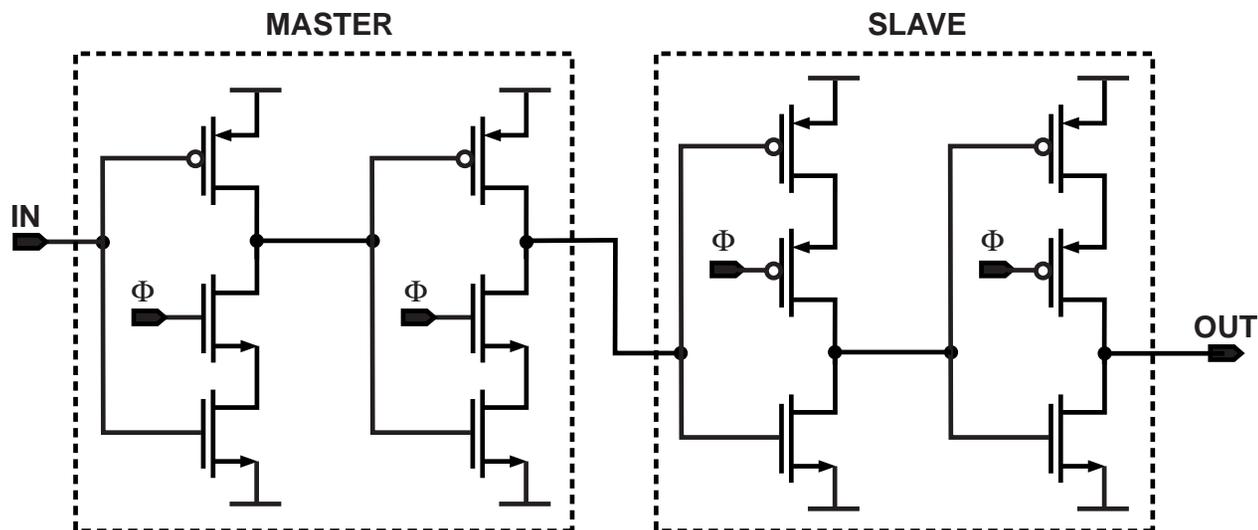
- **Allow register implementations with only one clock!**
 - Latches made transparent at different portions of clock cycle by using appropriate latch “flavor” – n or p
 - n latches are transparent only when Φ is 1
 - p latches are transparent only when Φ is 0
- **Benefits: simplified clock distribution, high speed**

Example TSPC Registers

- Positive edge-triggered version

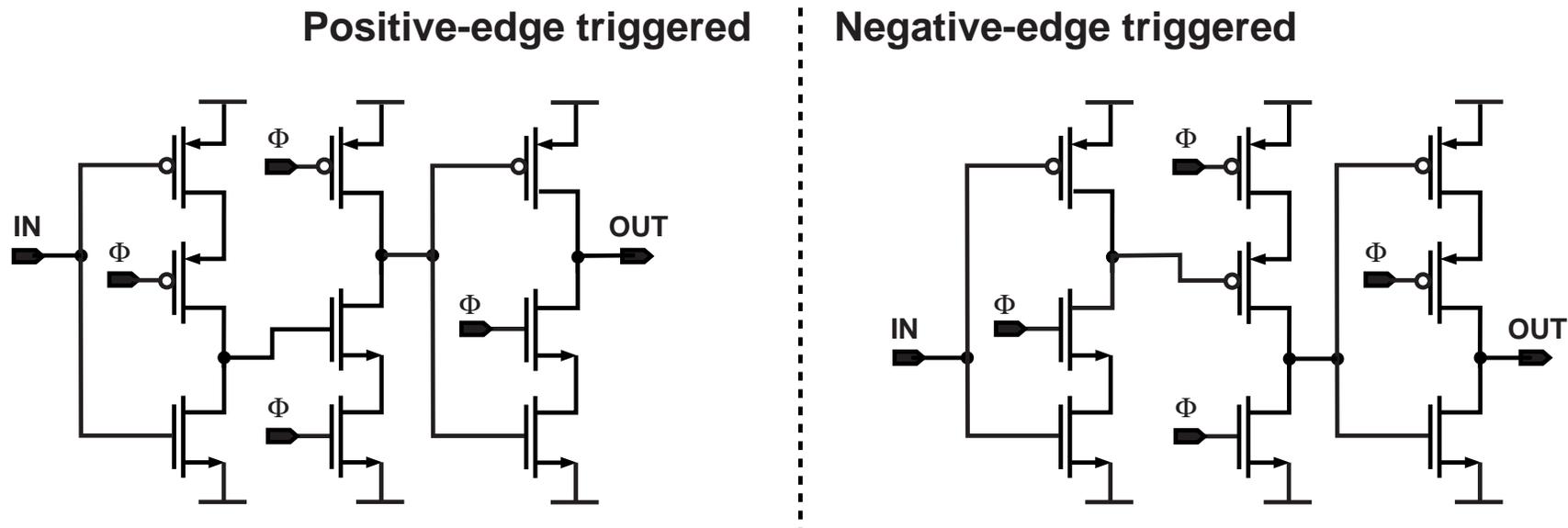


- Negative edge-triggered version



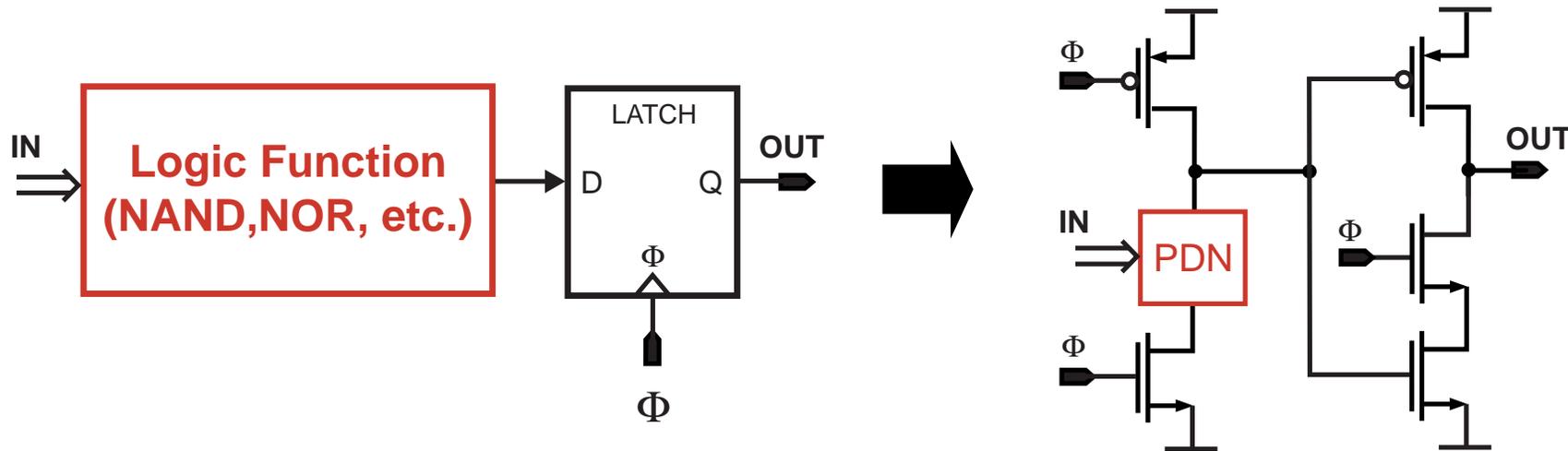
A Simplified Approach to TSPC Registers

- **Clever implementation of TSPC approach can be achieved with reduced transistor count**



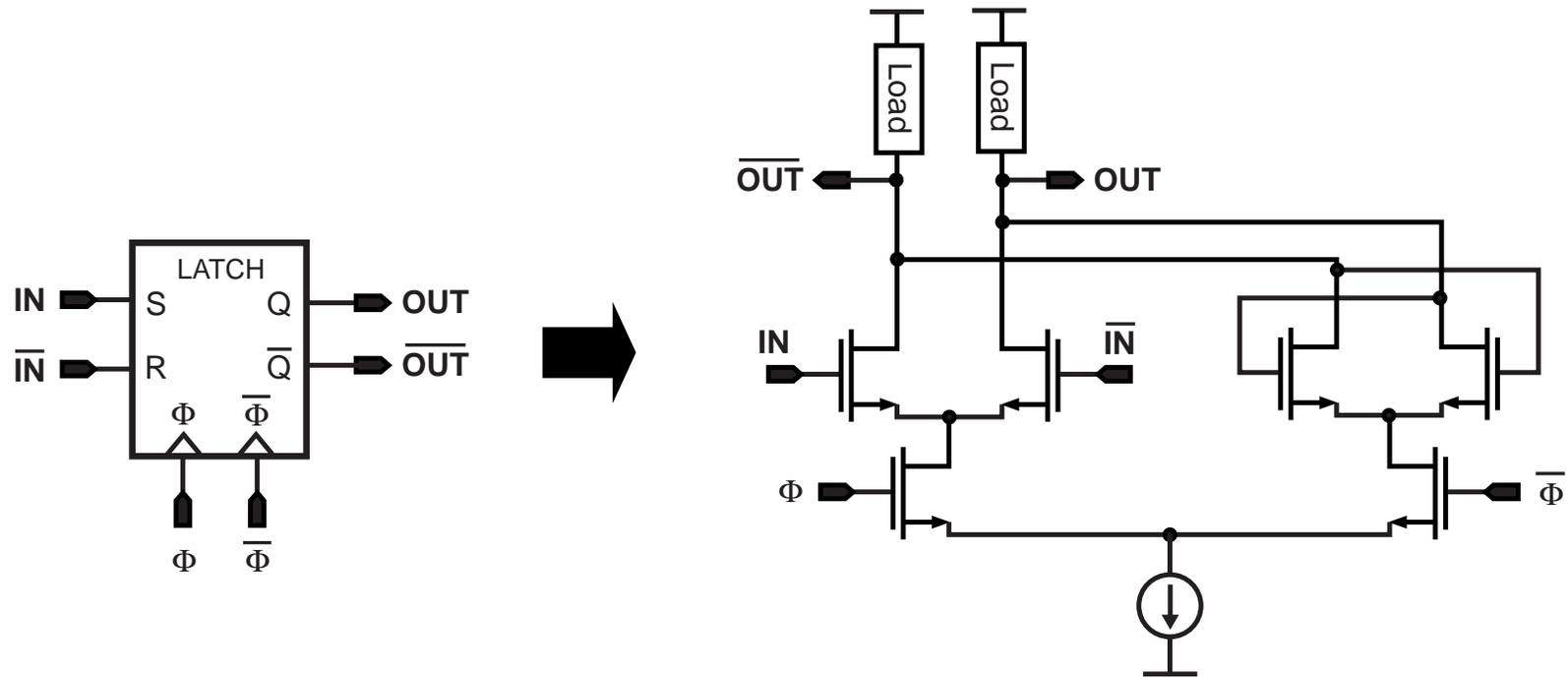
- **For more info on TSPC approach, see**
 - J. Yuan and C. Svensson, "New Single-Clock CMOS Latches and Flipflops with Improved Speed and Power Savings", JSSC, Jan 1997, pp 62-69

Embedding of Logic within Latches



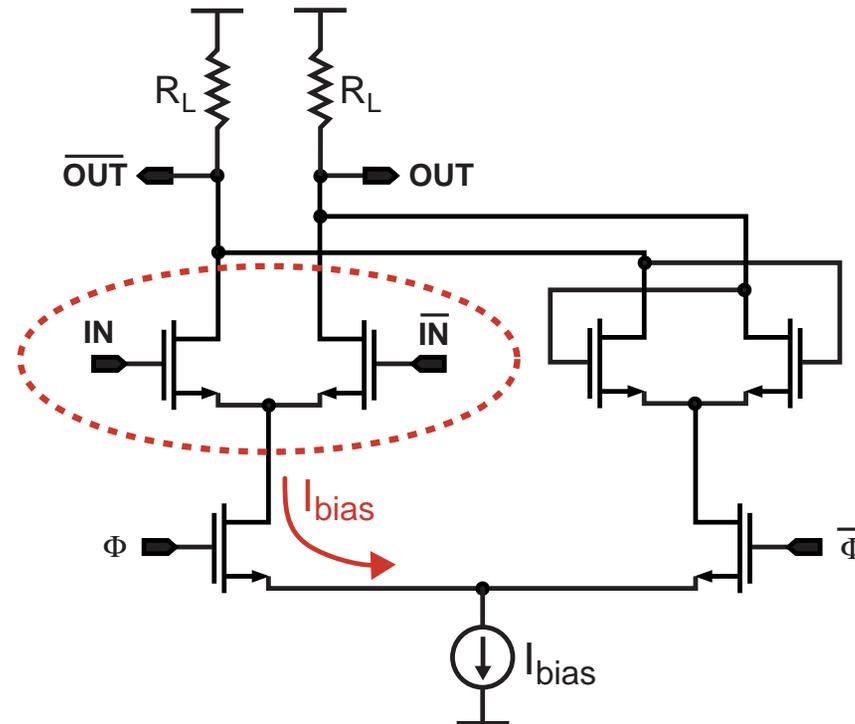
- We can often increase the speed of a logic function fed into a latch through embedding
 - Latch slowed down by extra transistors, but logic/latch combination is faster than direct cascade of the functions
- Method can be applied to both static and dynamic approaches
 - Dynamic approach shown above

Highest Speed Achieved with Differential SCL Latch



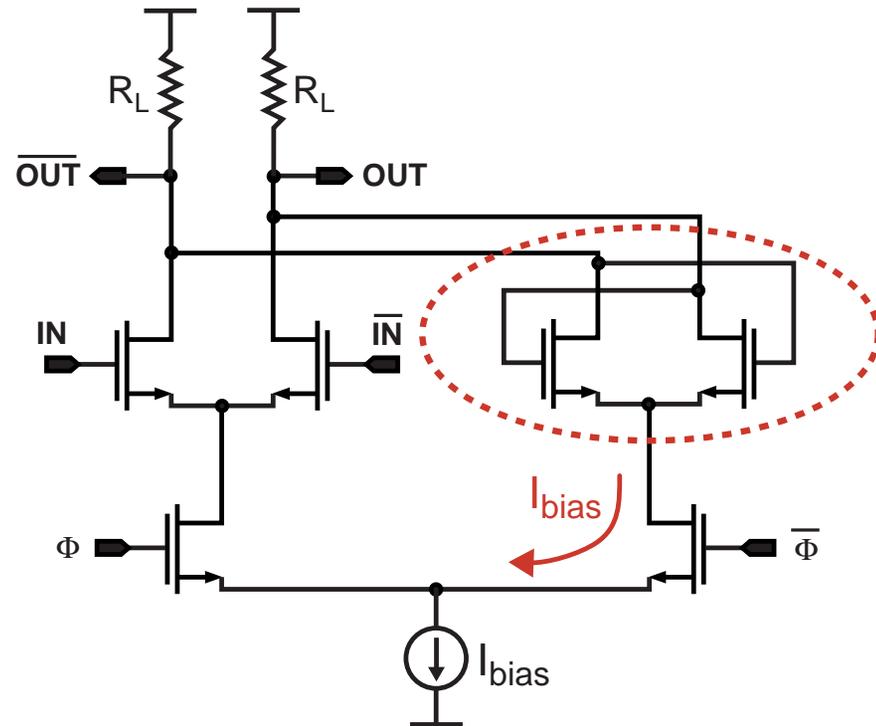
- **Employs positive feedback for memory**
 - Realized with cross-coupled NMOS differential pair
- **Method of operation**
 - Follow mode: current directed through differential amplifier that passes input signal
 - Hold mode: current shifted to cross-coupled pair

Design of Differential SCL Latch with Resistor Loads



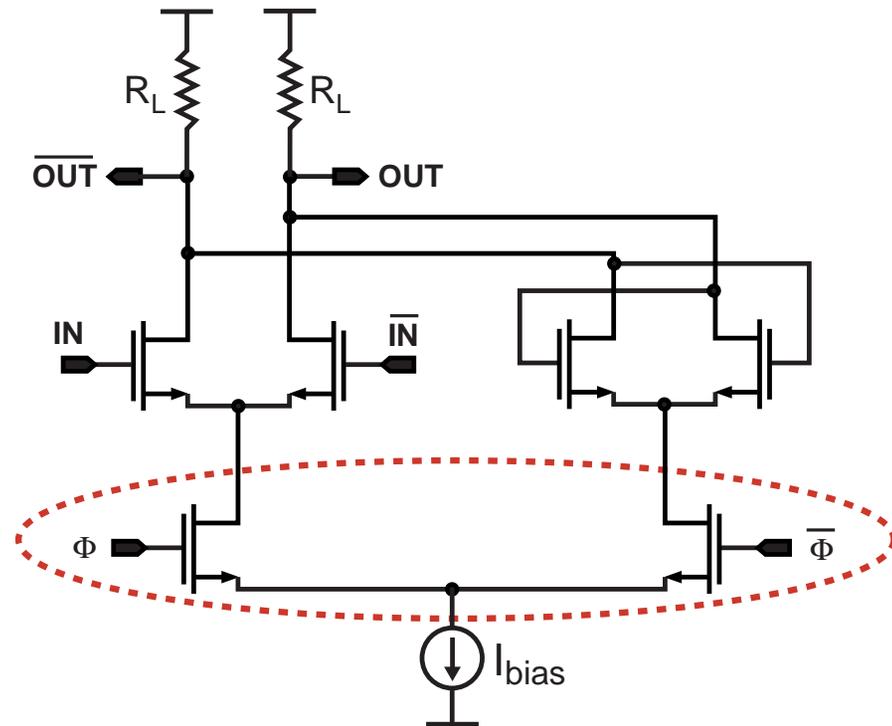
- Step 1: Design follower amplifier to have gain of 1.75 to 2 using simulated g_m technique from Lecture 5

Design of Differential SCL Latch with Resistor Loads



- Step 1: Design follower amplifier to have gain of 1.75 to 2 using simulated g_m technique from Lecture 5
- Step 2: For simplicity, size cross-coupled devices the same as computed above (or make them slightly smaller)

Design of Differential SCL Latch with Resistor Loads



- Step 1: Design follower amplifier to have gain of 1.75 to 2 using simulated g_m technique from Lecture 5
- Step 2: For simplicity, size cross-coupled devices the same as computed above (or make them slightly smaller)
- Step 3: Choose clock transistors roughly 20% larger in width (they will be in triode, and have lower drive)