

ESE370: Circuit-Level Modeling, Design, and Optimization for Digital Systems

Lec 8: February 20, 2023

Performance Inverters and Gates





Previously

- ❑ Delay as RC-charging
- ❑ Transistor
 - Capacitance
 - Drive Current
 - Function of geometry (W/L)



Today

- τ -model
- Sizing
- Large Fanout
- Delay in Gates
- Data Dependent Delay
- Large Fanin



Transistor Sizing

- What happens to I_{ds} as a function of W ?

$$I_{DS} \approx v_{sat} C_{OX} W \left(V_{GS} - V_T - \frac{V_{DSAT}}{2} \right)$$

- What happens to C_g as a function of W ?

$$C_G \propto C_{ox} WL$$



Transistor Sizing

- What happens to I_{ds} as a function of W ?

$$I_{DS} \approx v_{sat} C_{OX} W \left(V_{GS} - V_T - \frac{V_{DSAT}}{2} \right)$$

- What happens to C_g as a function of W ?

$$C_G \propto C_{ox} WL$$

- **Conclude:** faster transistors present more load on their inputs



First Order Delay (preclass 1)

- $I_0 = I_{ds}$ of minimum size NMOS device
- $C_0 =$ gate capacitance of minimum size NMOS device

- $I_{drive} = WI_0$
- $C_g = WC_0$



First Order Delay (preclass 1)

- R_0 = Resistance of minimum size NMOS device
- C_0 = gate capacitance of minimum size NMOS device

- $R_{\text{drive}} = R_0/W$
- $C_g = WC_0$



τ model

- ❑ All delays are RC delays
- ❑ Always have an R_0C_0 term
- ❑ $\tau = R_0C_0$
- ❑ Express all delays in τ units
- ❑ Like λ units for measurement
 - Separate delay into
 - Technology dependent term $\tau = R_0C_0$
 - Technology independent coefficient

How to Size Transistors (preclass 2)

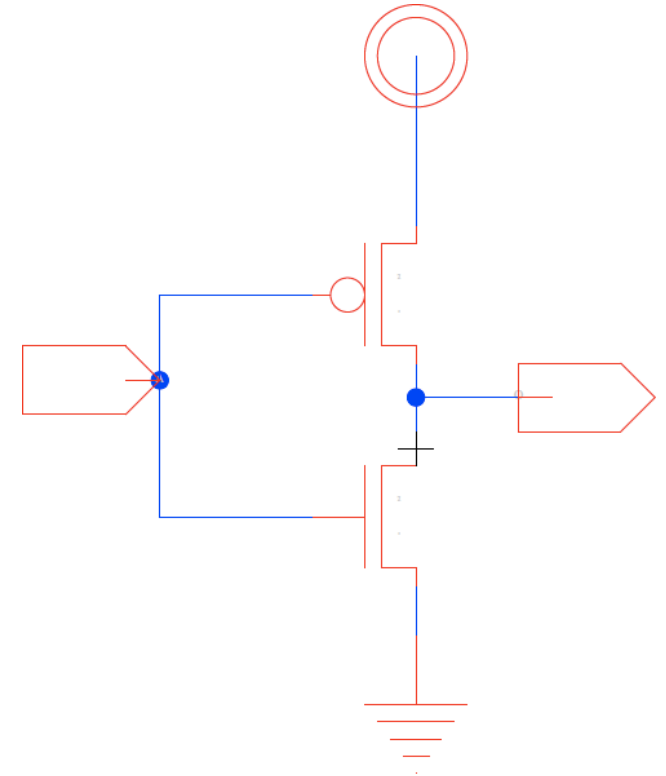
□ How should we size to equalize Rise and Fall?

■ Given:

■ $\mu_n = 500 \text{cm}^2/\text{Vs}$, $\mu_p = 200 \text{cm}^2/\text{Vs}$

■ $R_{\text{drive}} = R_0/2$ ($I_{\text{drive}} = 2I_0$)

$$I_{DS} = \mu C_{OX} \frac{W}{L} (V_{GS} - V_T)^2$$



How to Size Transistors (preclass 2)

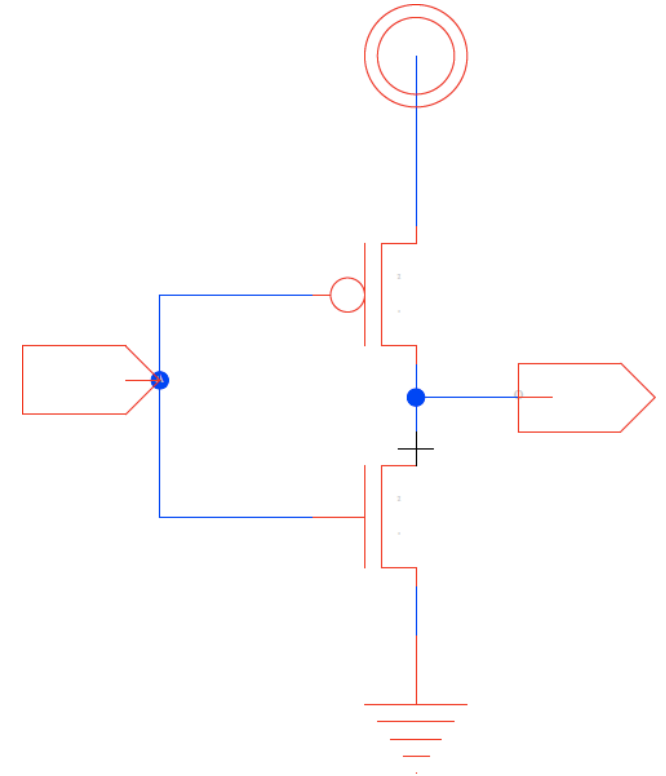
□ How should we size to equalize Rise and Fall?

■ Given:

■ $\mu_n = 500 \text{cm}^2/\text{Vs}$, $\mu_p = 200 \text{cm}^2/\text{Vs}$

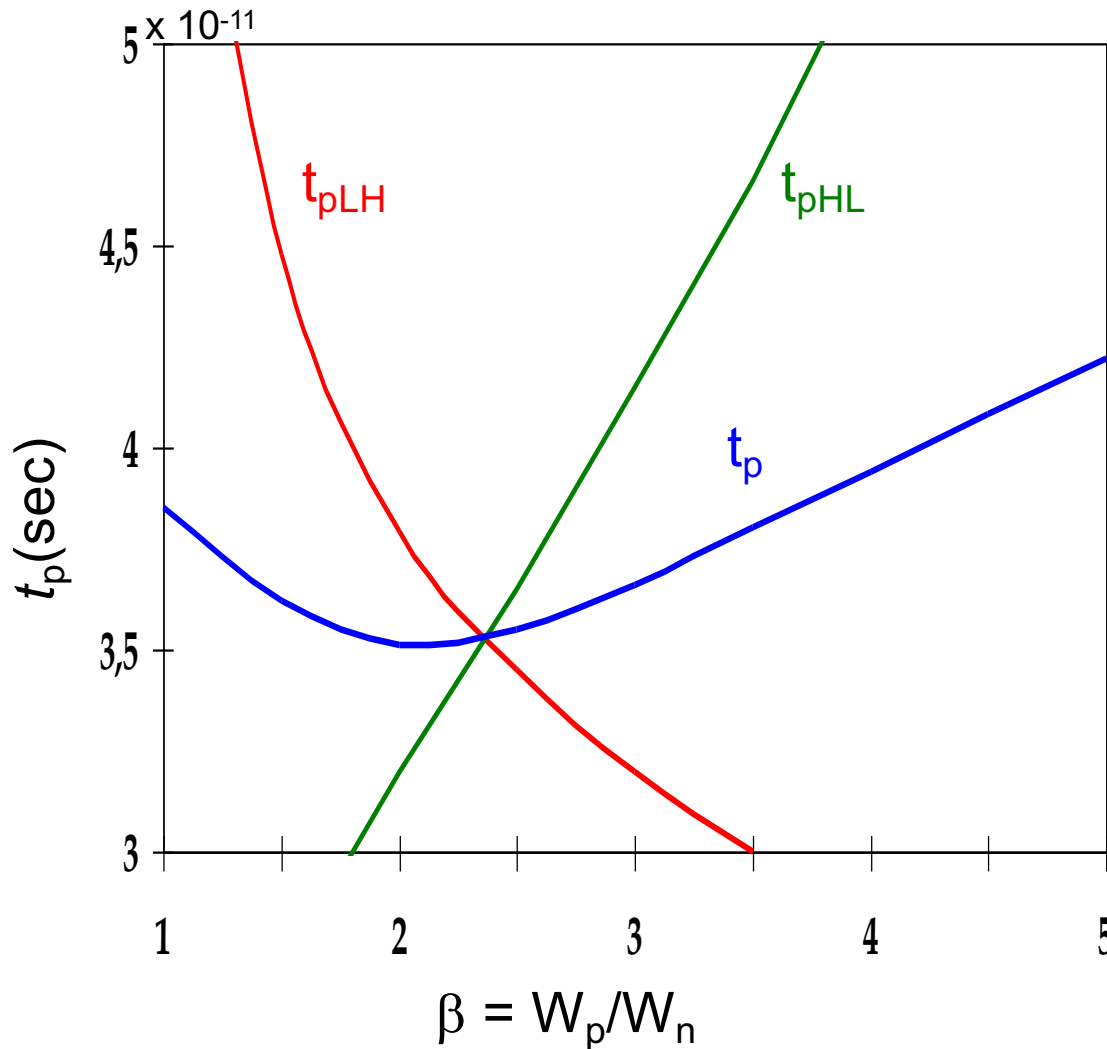
■ $R_{\text{drive}} = R_0/2$ ($I_{\text{drive}} = 2I_0$)

$$I_{DS} = \mu C_{OX} \frac{W}{L} (V_{GS} - V_T)^2$$



■ What is input capacitance for sized devices?

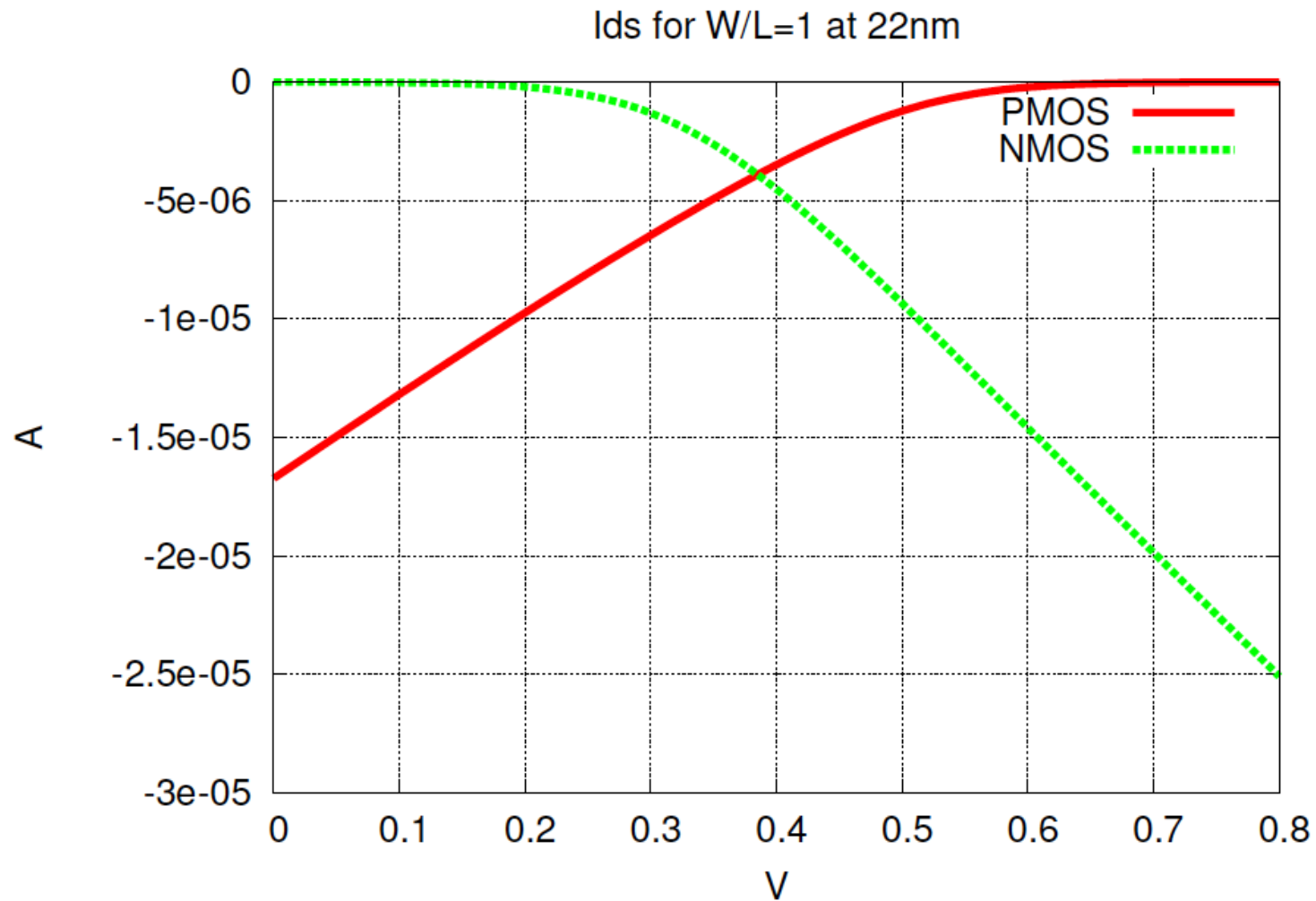
Size Transistors – Minimum Avg Delay



- β of 2.4 (= 31 k Ω /13 k Ω) gives symmetrical response
- β of 1.6 to 1.9 gives optimal performance



SPICE Simulation 22nm





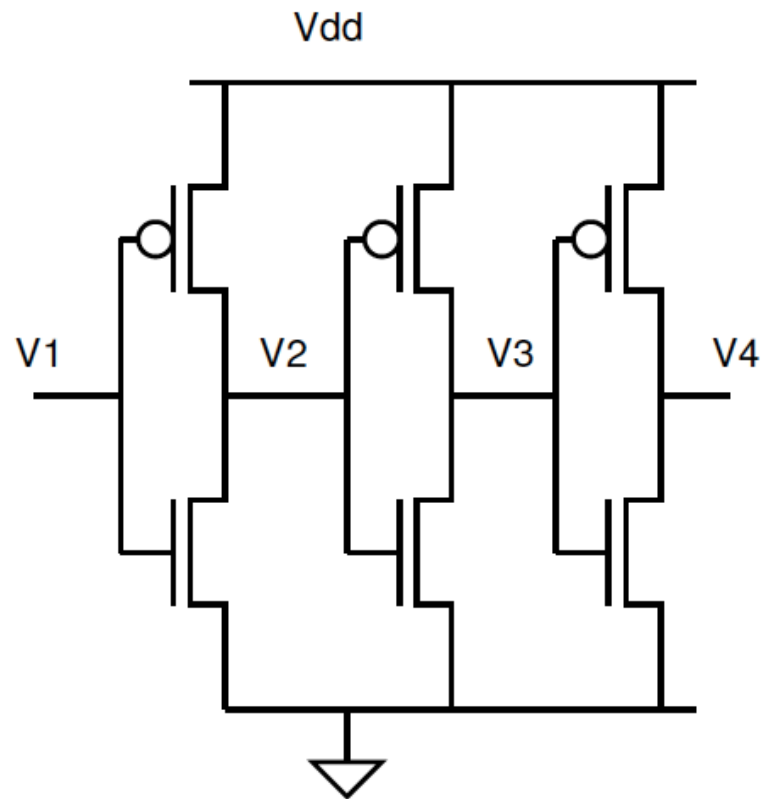
Equalizing Delay

- For simplicity, for today
 - Assume $W_p = W_n$ equalizes I_{ds}
 - I.e. $I_{0,n} = I_{0,p}$



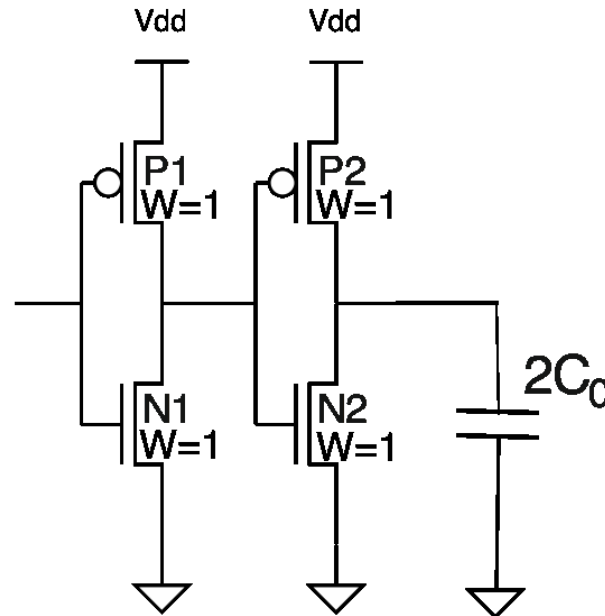
Inverter Sizing

- What is the impact of the delay if we double size of all the transistors?



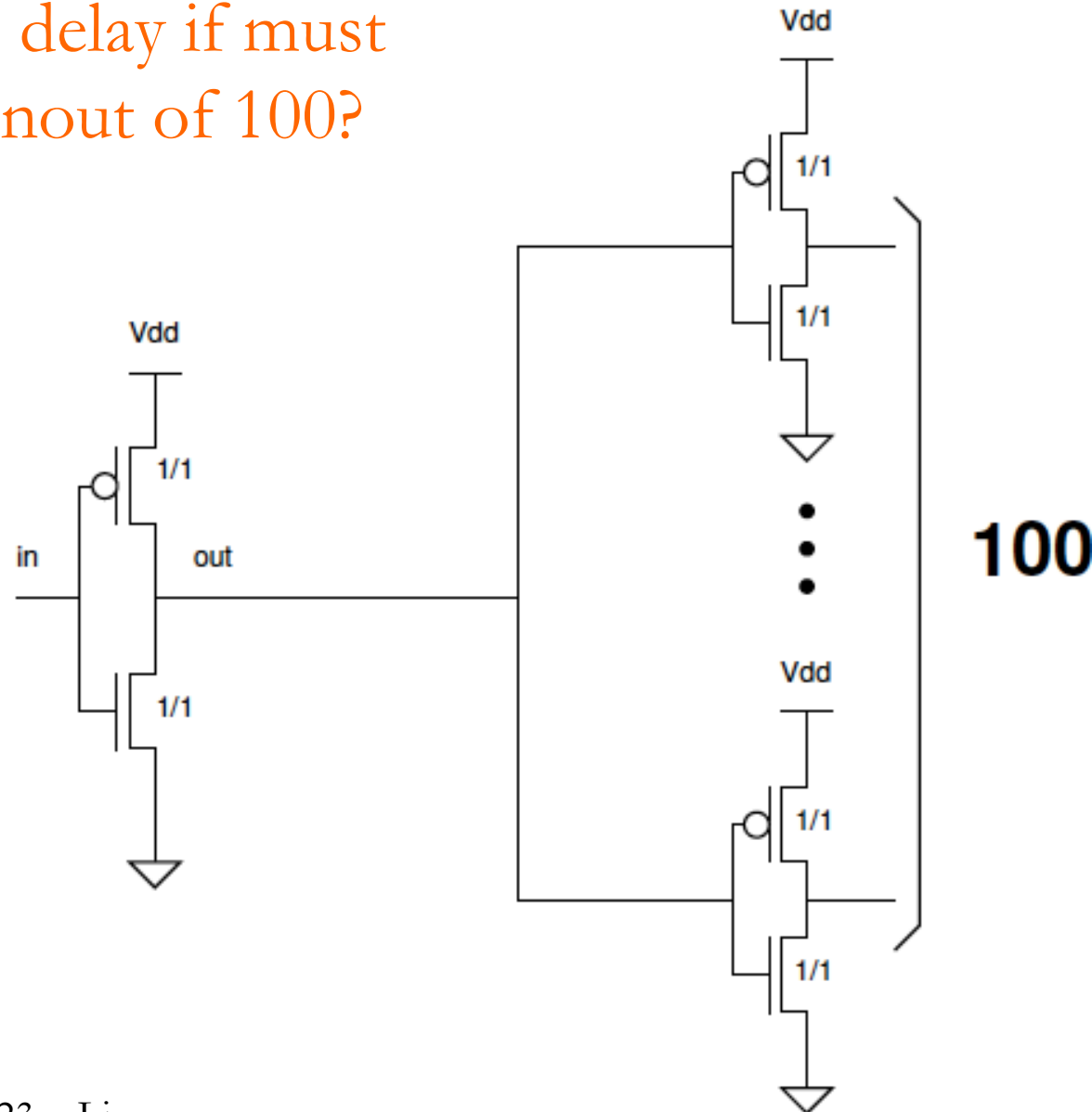
Multistage Delay

- Total delay = sum of stage delays
- What is delay here?
 - From (P1,N1) to final capacitive load



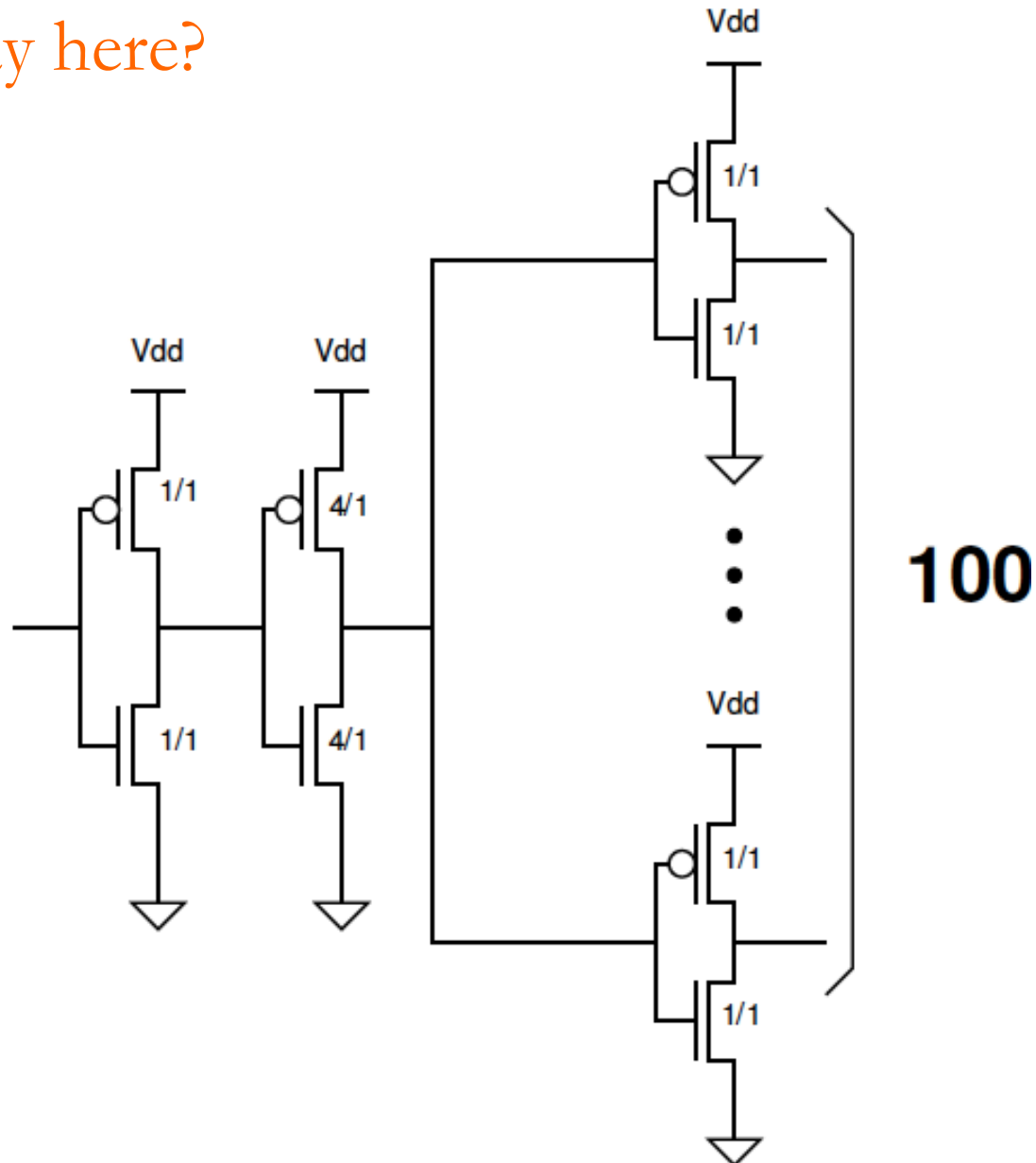
Large Fanout Delay (preclass 3)

- What is delay if must drive fanout of 100?



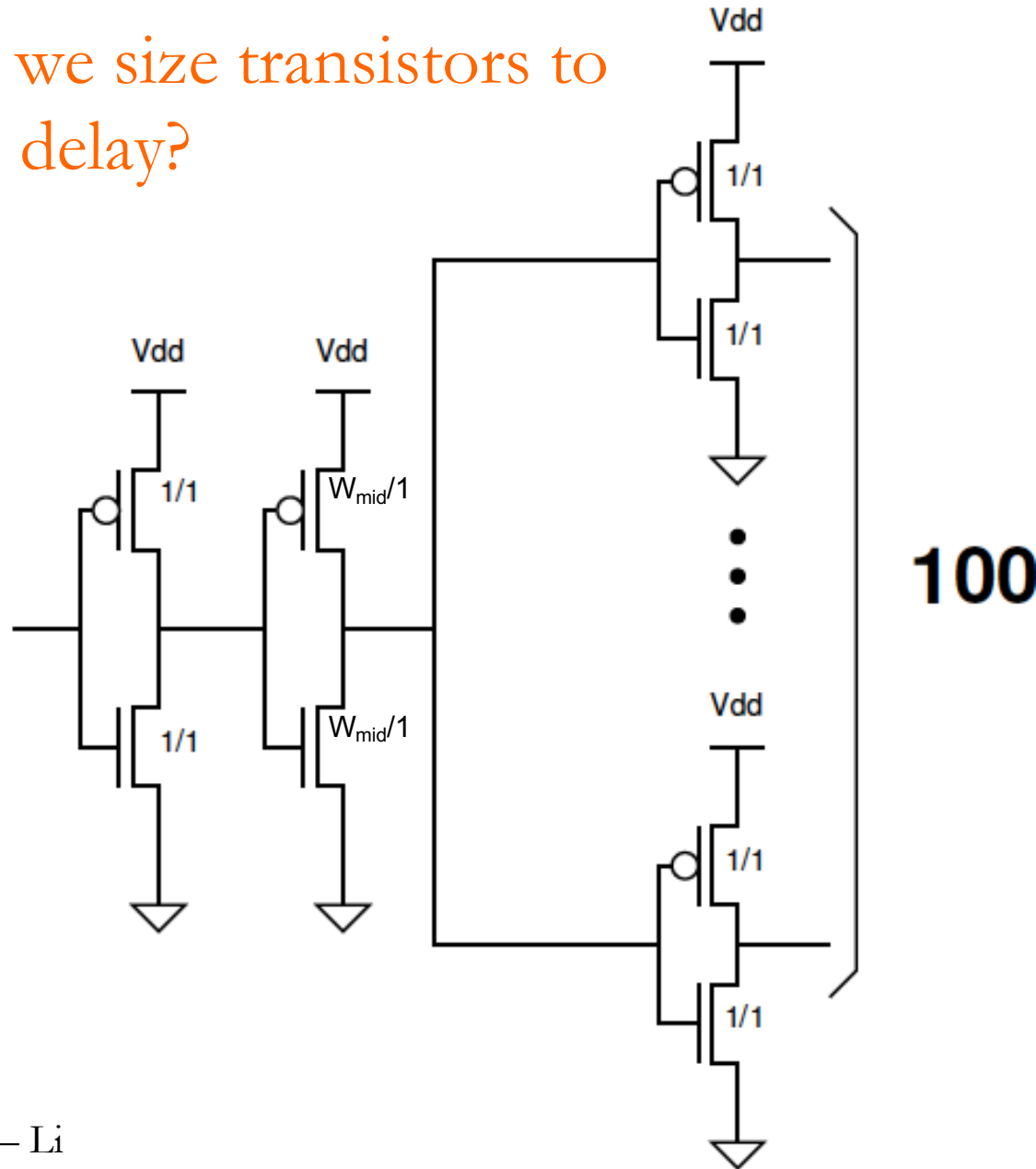
Graduated Fanout Delay (preclass 3)

- What is delay here?



Optimize Fanout Delay (preclass 4)

- How can we size transistors to minimize delay?





Optimizing (preclass 4)

□ Derivate to minimize

$$\tau_{est} = R_0 \times 2W_{mid}C_0 + \frac{R_0}{W_{mid}} \times 200C_0$$

$$\frac{\partial \tau_{est}}{\partial W_{mid}} = 0$$

$$2R_0C_0 - \frac{200}{W_{mid}^2} R_0C_0 = 0$$

$$W_{mid} = \sqrt{100} = 10$$



Lesson

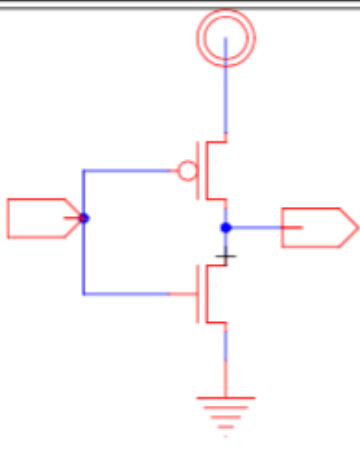
- ❑ Don't drive large fanout with a single stage
- ❑ Must scale up over a number of stages
- ❑ ...but not too many
- ❑ Exact number will be technology dependent

Gates



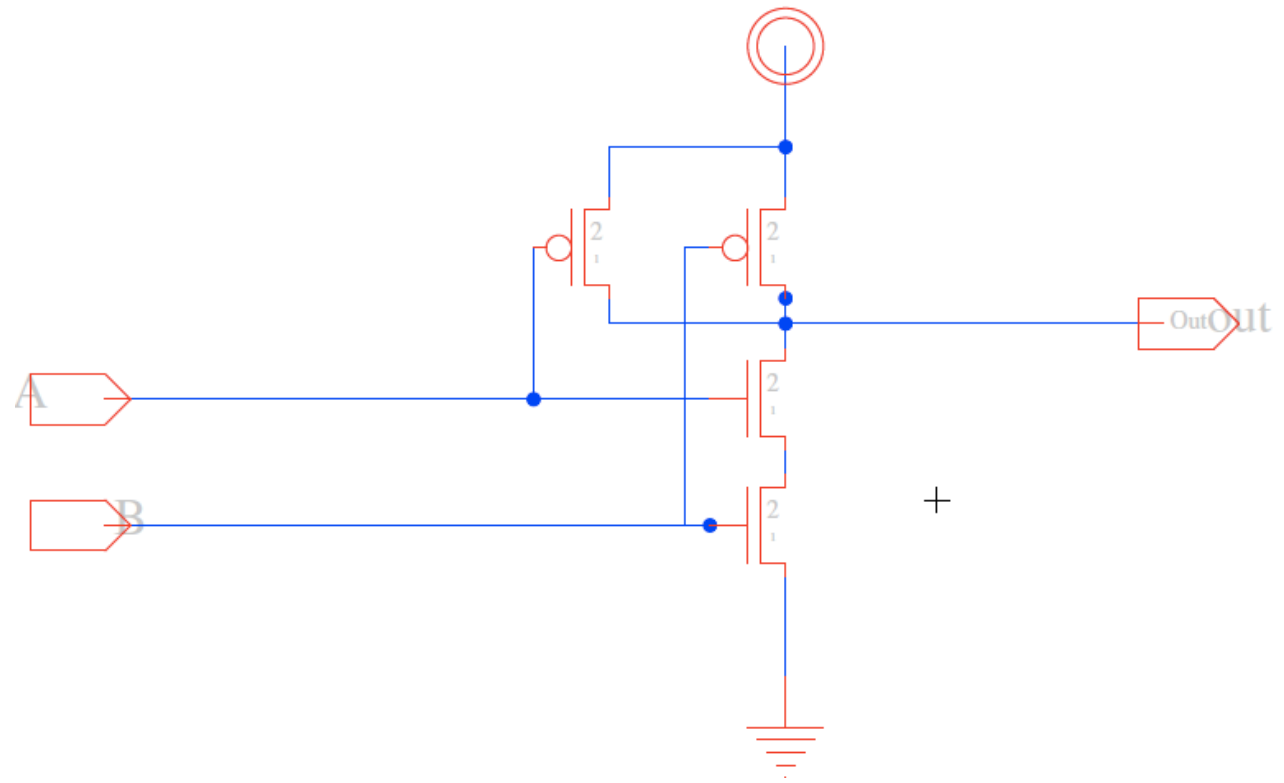
Inverter Performance (preclass 5, row 1)

- Sized for $R_0/2$ drive resistance, ($R_0=R_{n0}$)

	$R_{p0} = R_{n0}$			$R_{p0} = 2R_{n0}$		
	W_p	W_n	C_a	W_p	W_n	C_a
	2	2	$4C_0$	4	2	$6C_0$

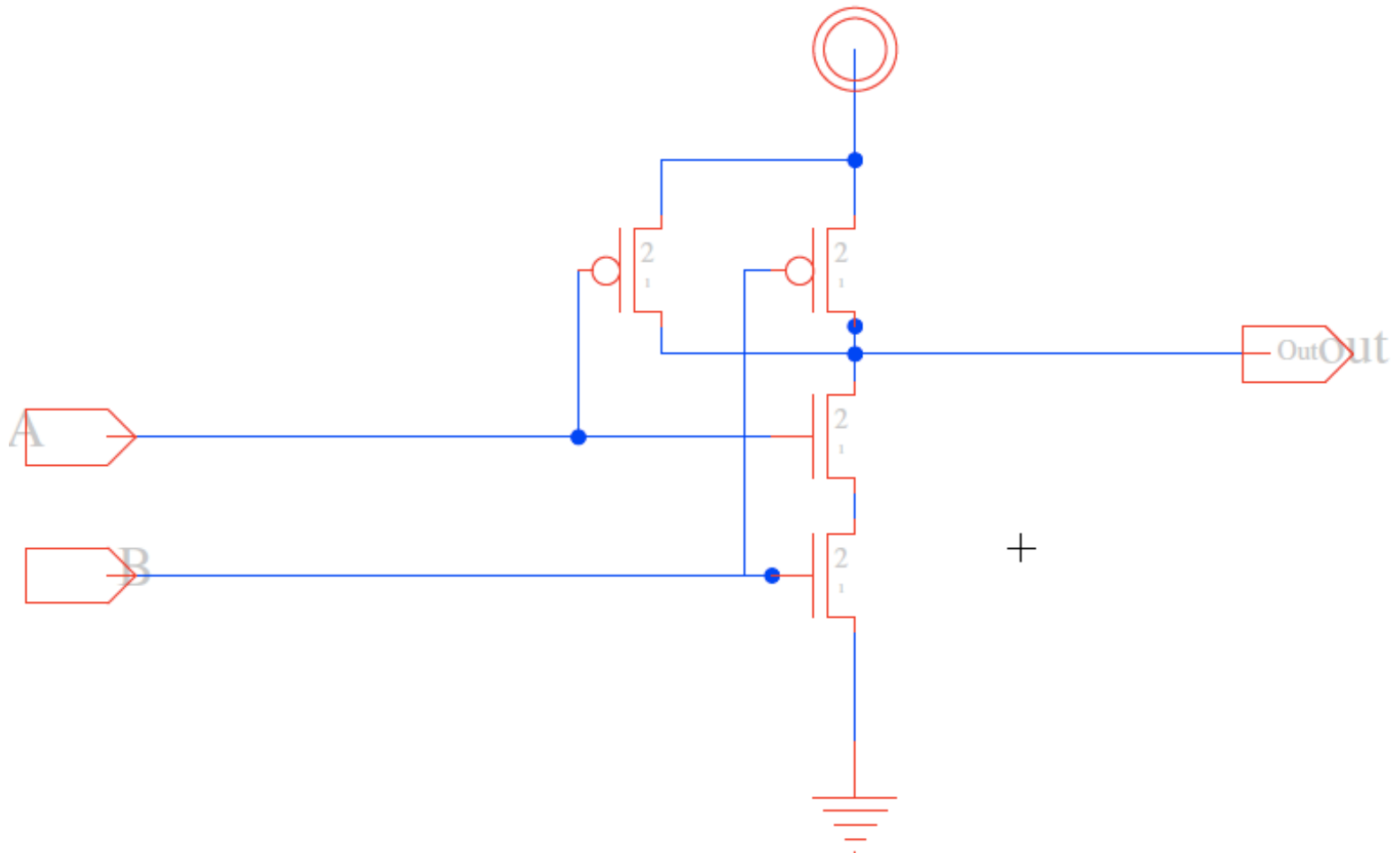
Data Dependent Delay

- Drive resistance depends on input values
 - Delay depends on input data
 - Analyze using worst case delay



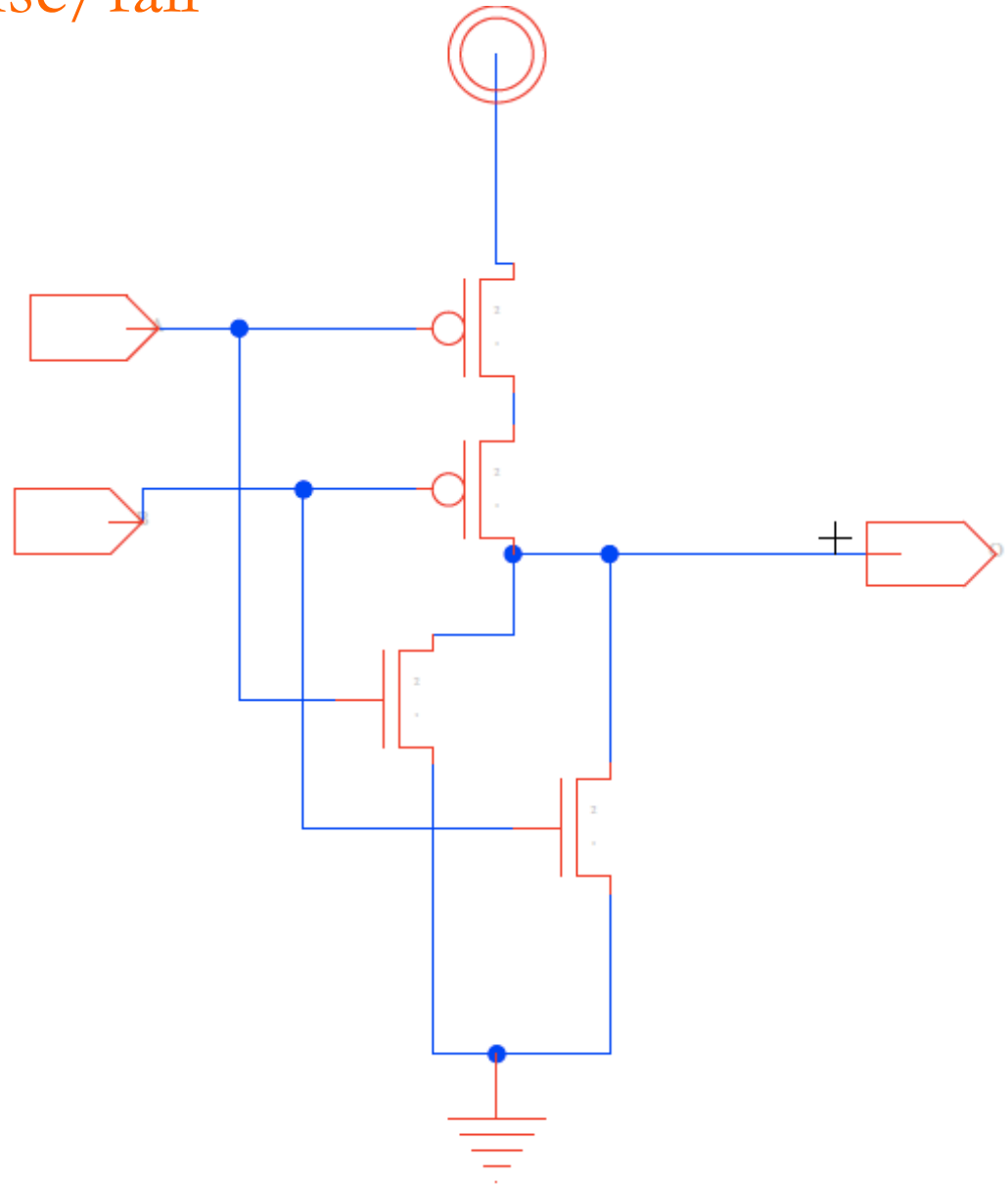
Transistor Sizing (preclass 5, row 2)

- How should we size to equalize worst-case rise/fall times for $R_{\text{drive}} = R_0/2$?



Transistor Sizing (preclass 5, row 3)

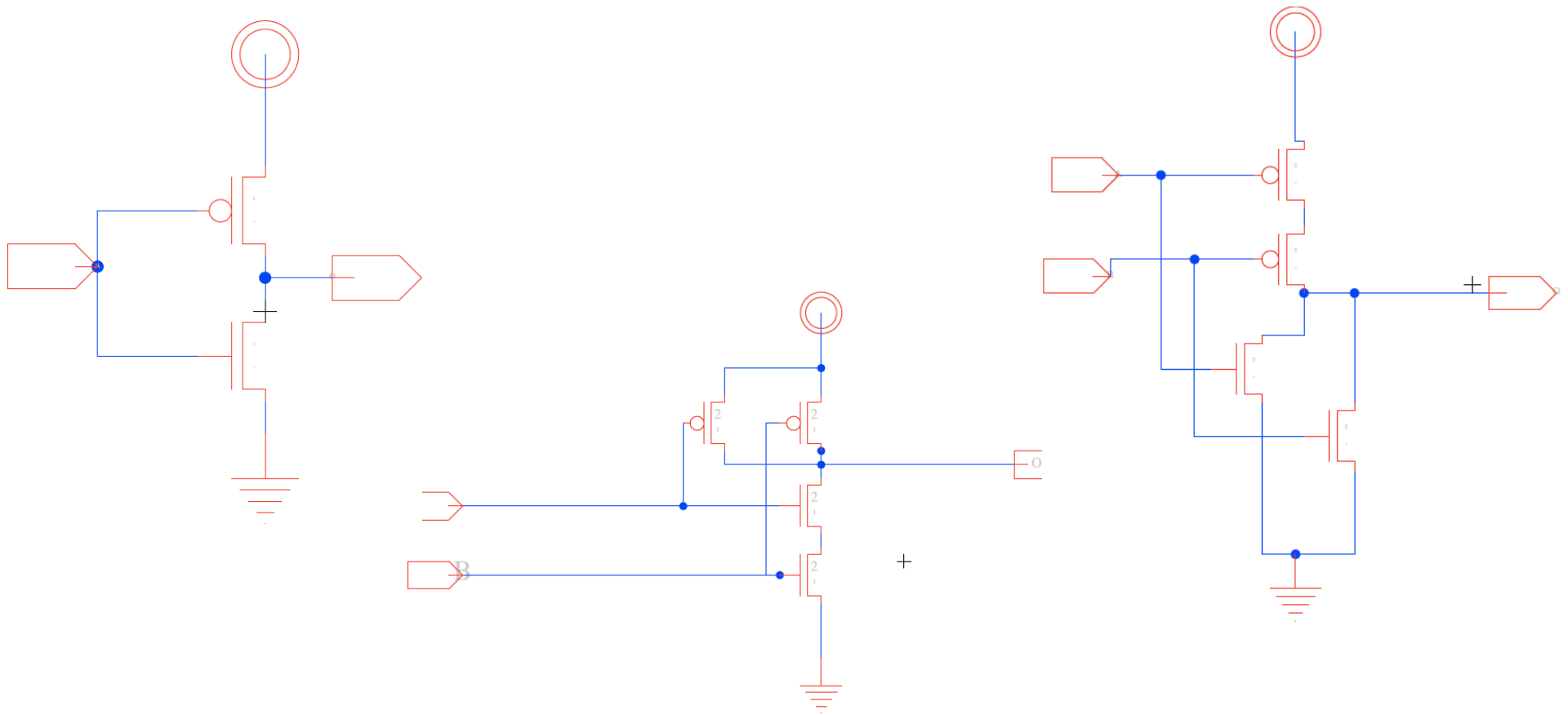
- How size for equal rise/fall for $R_{\text{drive}} = R_0/2$?





Input Load

- Input capacitance per input in each case?



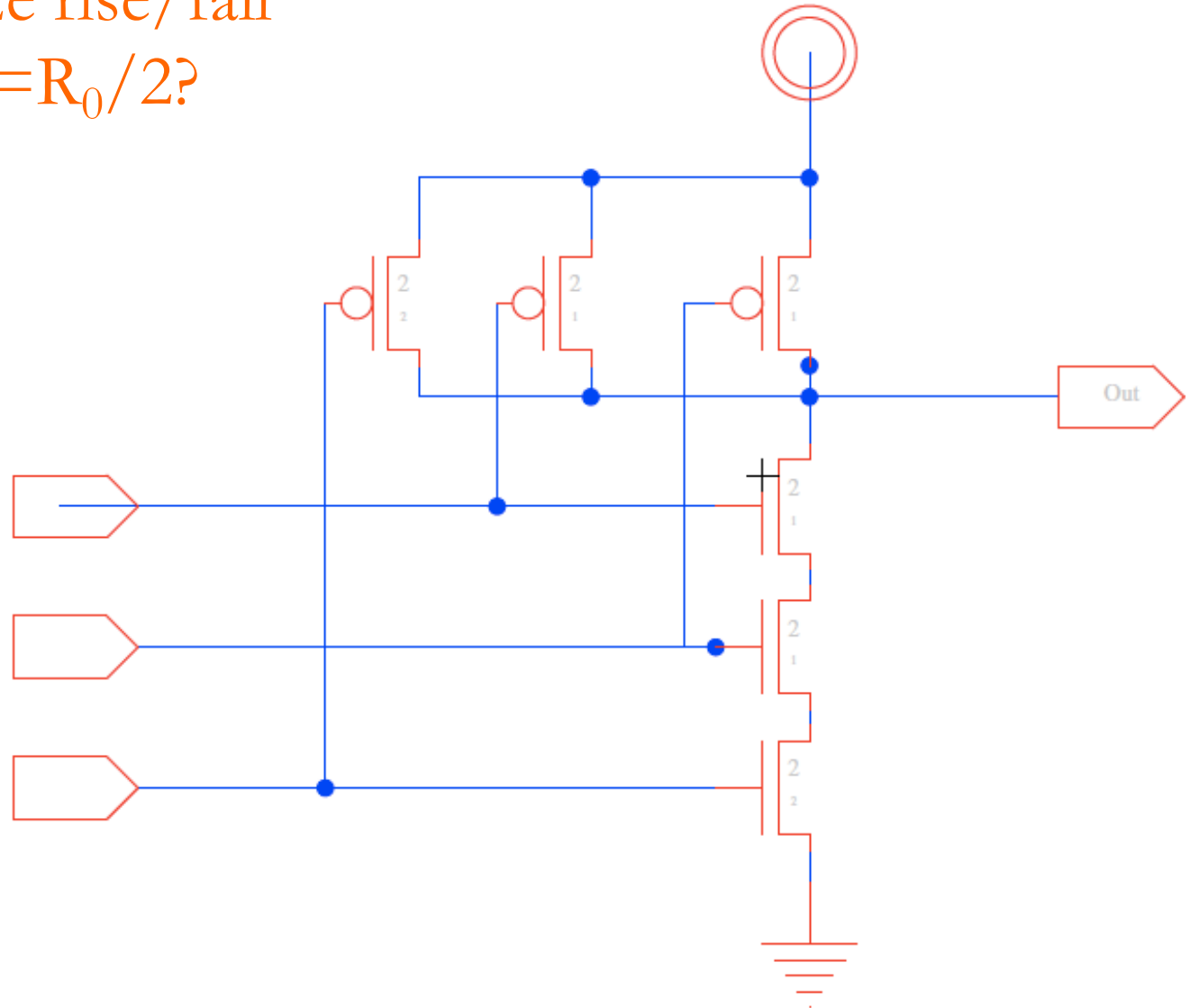


Observe

- Ratio of Input Load Capacitance to Output Drive Strength: $C_{\text{InLoad}}/I_{\text{ds}}$
 - Differs with gate function
 - Gate efficiency

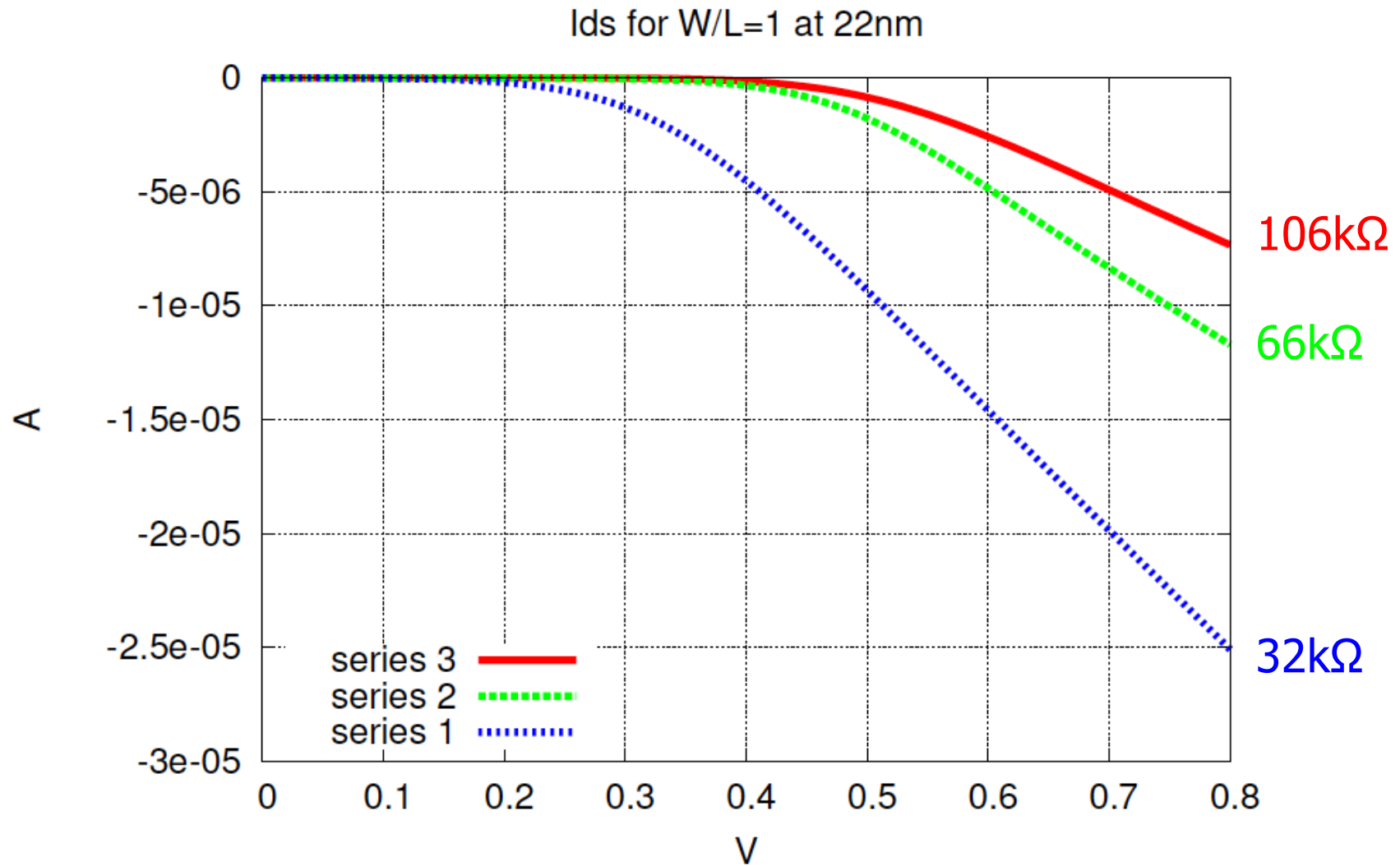
Transistor Sizing (preclass 5, row 4)

- Size equalize rise/fall times $R_{drive} = R_0/2?$





Series Transistors





Increasing Fanin (preclass 6)

- What happens to input capacitance as fanin (k) increases
 - Keeping output drive the same
 - E.g. $R_{\text{drive}} = R_0/2$
- k -input nand gate has what input capacitance?



Fanin

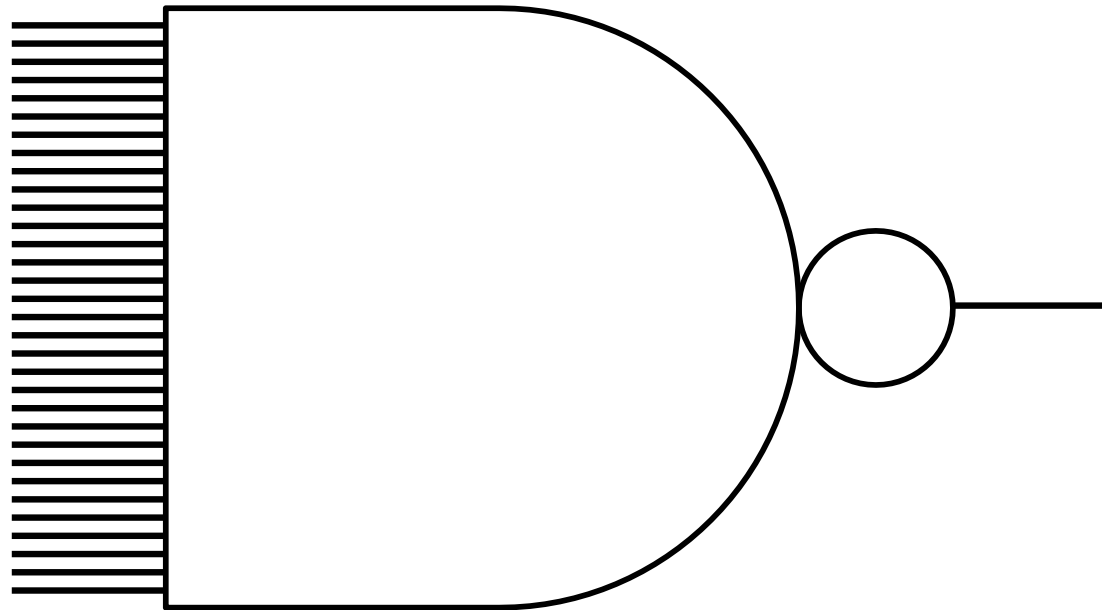
- ❑ **Conclude:** gates slow down with fanin
 - Less drive per input capacitance
 - $C_{\text{InLoad}}/I_{\text{ds}}$ increases

nand32 (preclass 7, row 1)

□ single-stage nand32

$R_{n0} = R_{p0}$ case only

- Delay with $R_0/2$ input drive and $4C_0$ load?

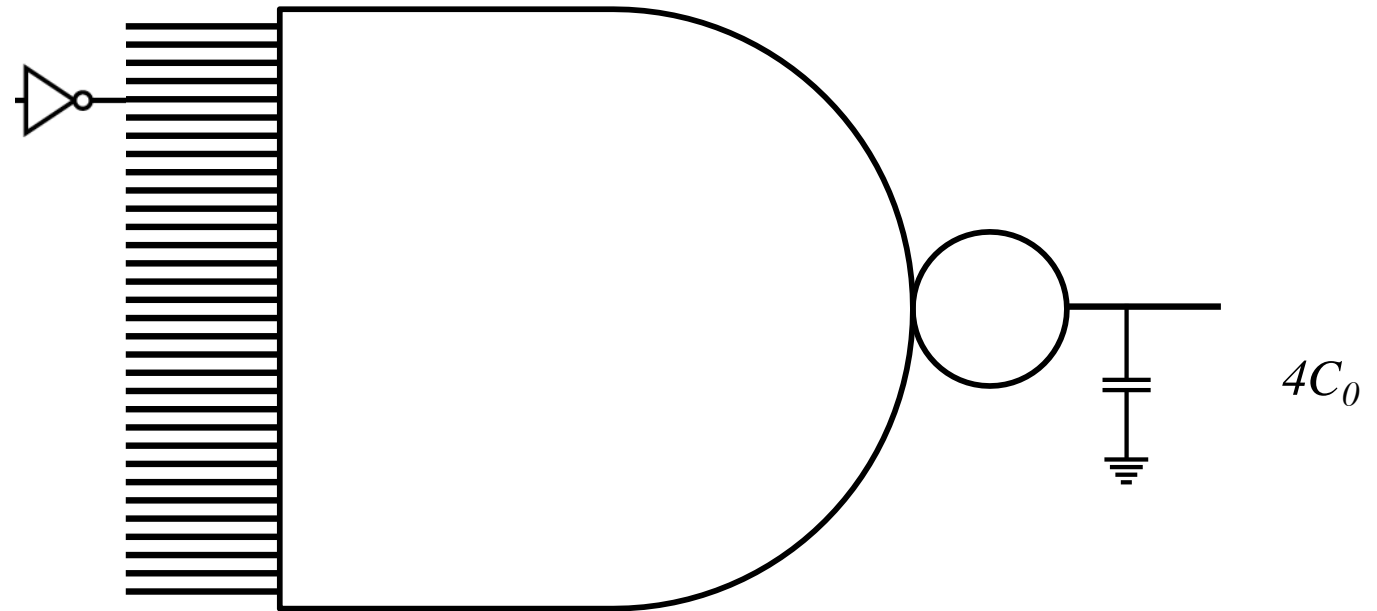


nand32 (preclass 7, row 1)

□ single-stage nand32

$R_{n0} = R_{p0}$ case only

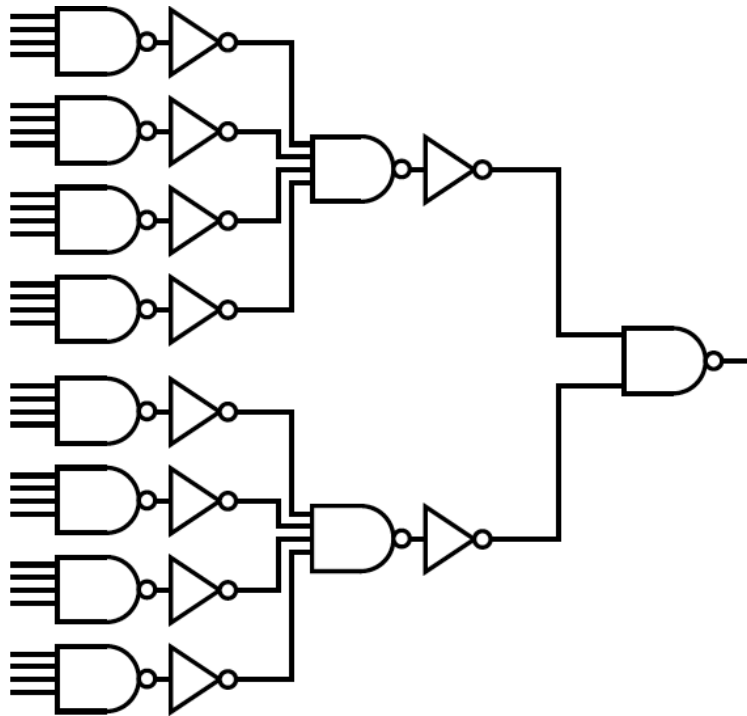
- Delay with $R_0/2$ input drive and $4C_0$ load?



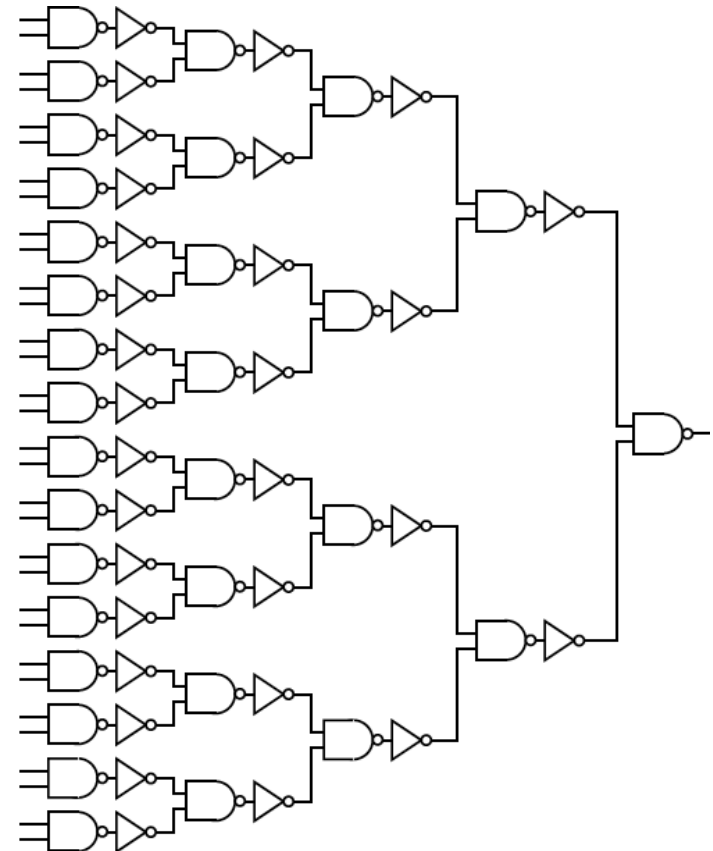
Which is Faster? (preclass 7, rows 2&3)

□ nand32

$R_{n0} = R_{p0}$ case only



nand4-inv-nand4-inv-nand2



$(\text{nand2-inv})^4\text{-nand2}$



Lesson

- ❑ Large gates are slow / inefficient
 - High capacitive load / drive current
- ❑ Small gates can be inefficient
 - Need many stages
- ❑ Staging over moderate size gates minimizes delay
- ❑ Exact size will be technology dependent



Ideas

- ❑ First order reason in $\tau = R_0C_0$ units
- ❑ Scaling everything up doesn't help
- ❑ Gates have different efficiencies
 - Drive strength per unit input capacitance
- ❑ Without velocity saturation
 - Reason to prefer nand over nor
- ❑ With velocity saturation
 - nands and nors are similar efficiency
- ❑ Large fanin and fanout slow gates
 - Decompose into stages
 - ...but not too many



Admin

- ❑ Drop date today 2/20
- ❑ HW 5 out now
 - A lot of SPICE
 - Start early
 - Create your schematics, icons and test schematics with care to minimize the time spent



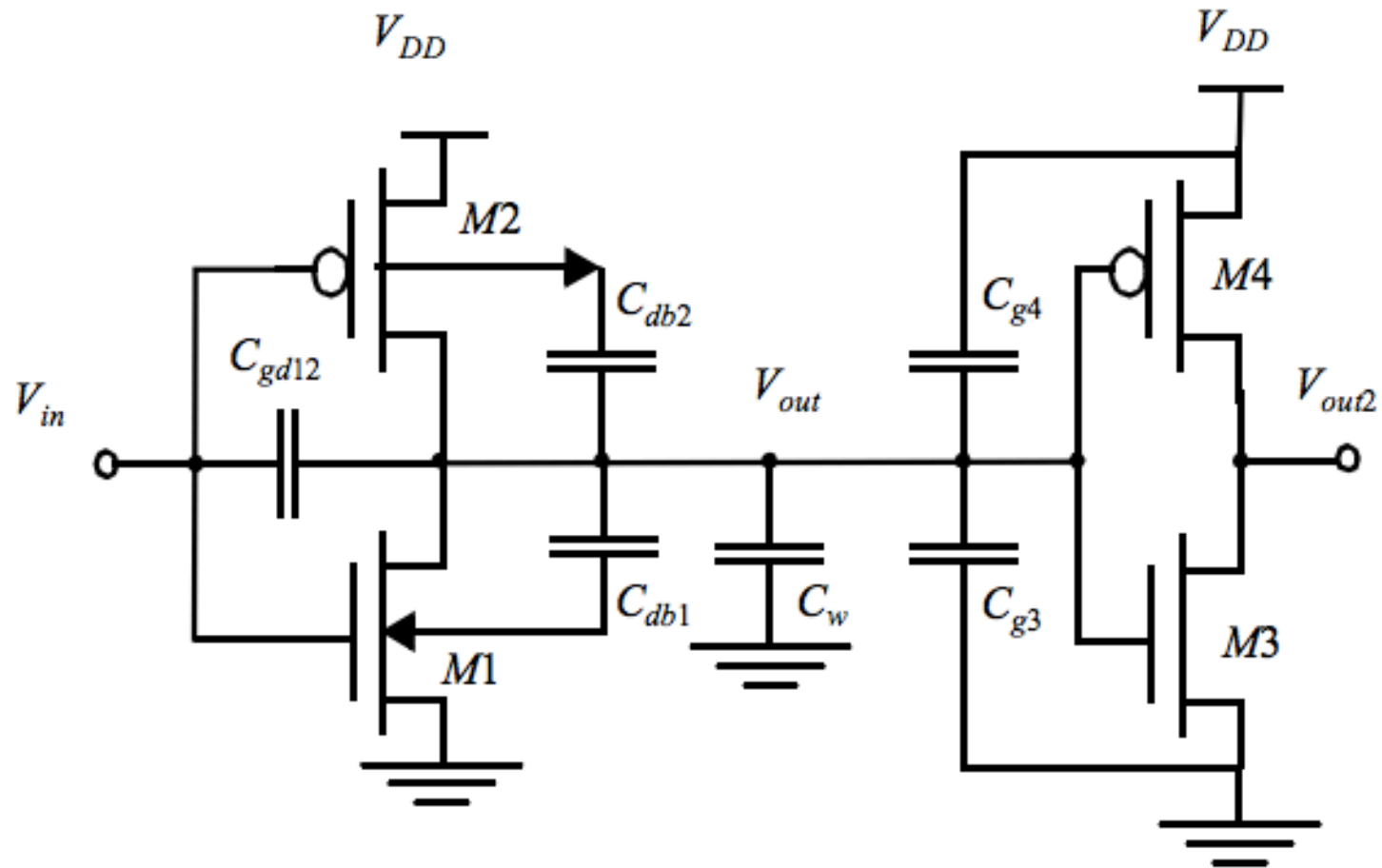
Acknowledgement

- Prof. André DeHon (University of Pennsylvania)
- Prof. Tania Khanna (University of Pennsylvania)

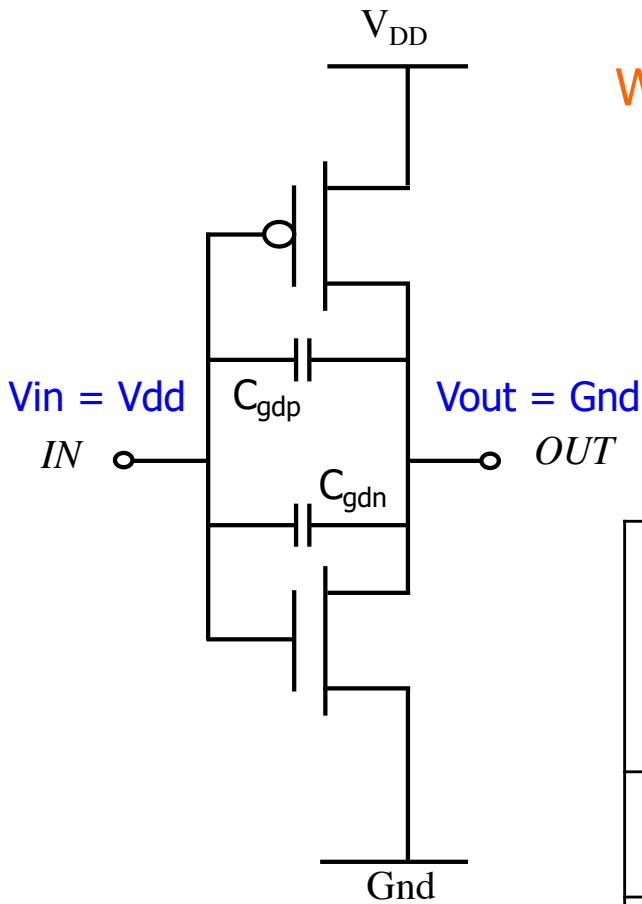
Miller Effect (Optional, not required)



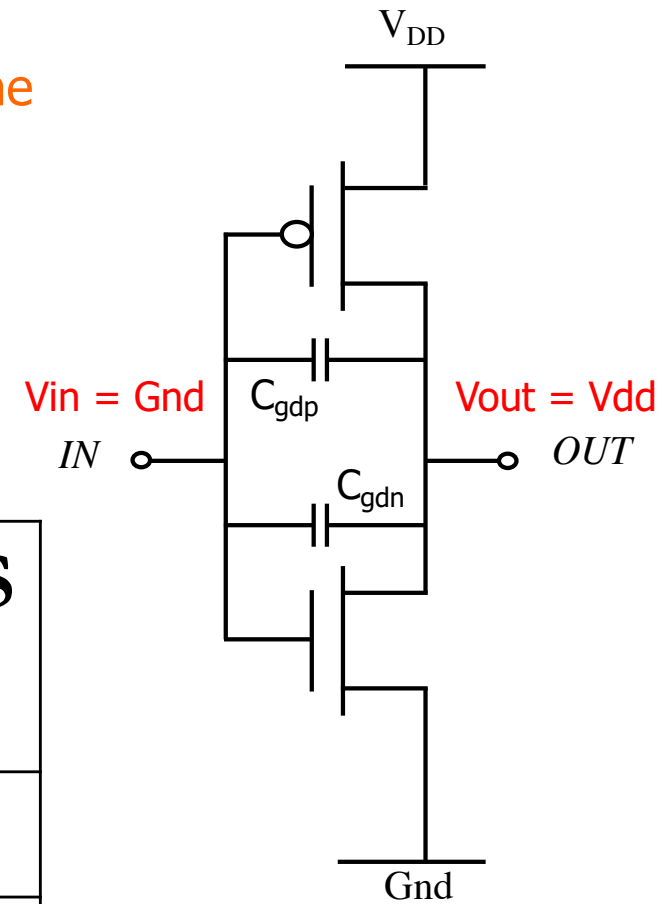
Capacitance Reminder



Charge on Capacitors



What is charge, Q , on each of the capacitors when $V_{in} = V_{DD}$ and $V_{in} = Gnd$?



	PMOS	NMOS
V_{in}	C_{gdp}	C_{gdn}
V_{DD}		
Gnd		



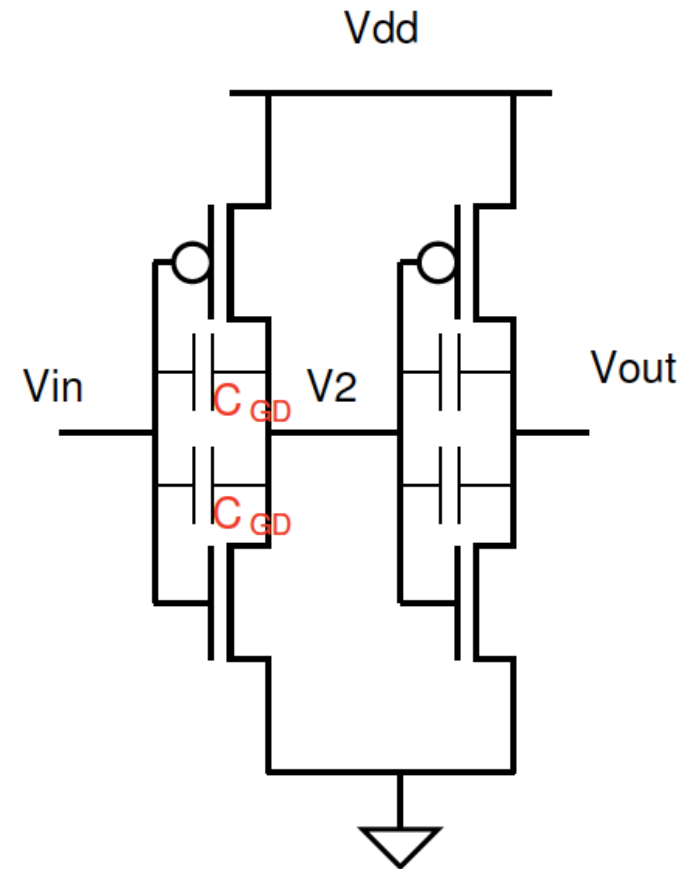
Questions

- ❑ What is ΔQ on each C_{gd} when input switched?
- ❑ Assuming $\Delta V = V_{dd}$, what is equivalent capacitance?

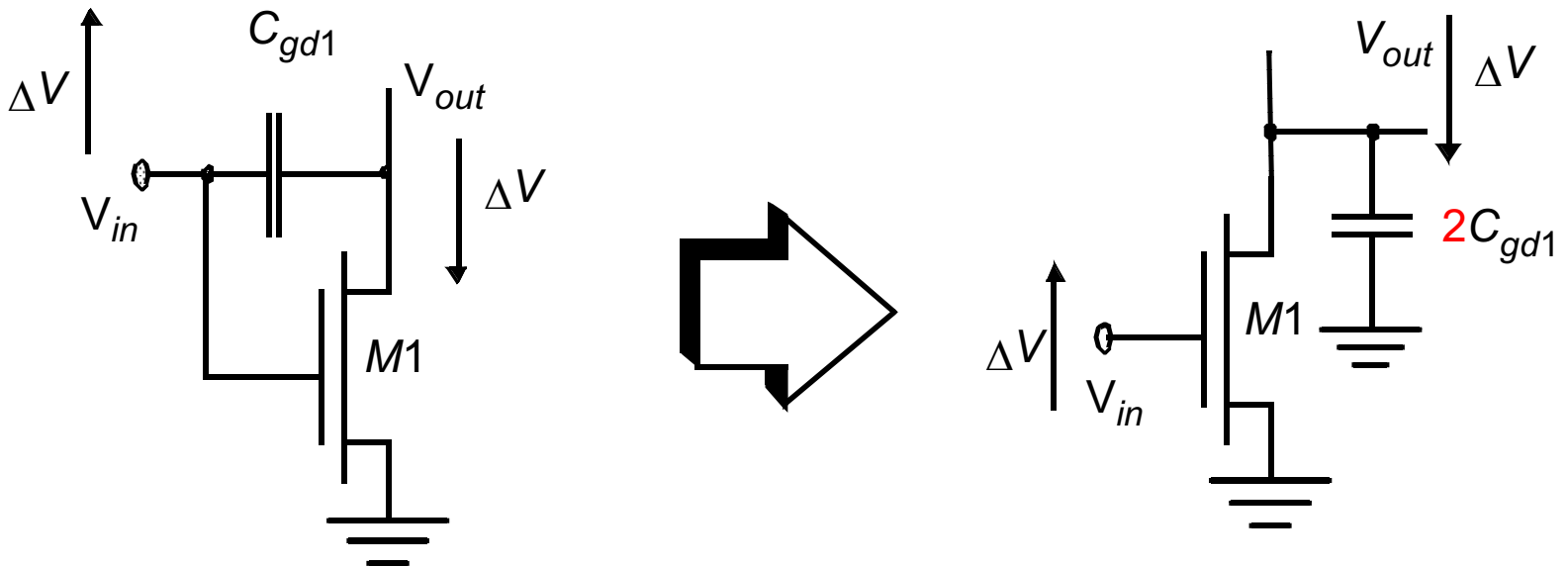
	PMOS	NMOS
Vin	C_{gdp}	C_{gdn}
Vdd		
Gnd		

Miller Effect For an Inverter

- ❑ Feedback capacitance (C_{gd}) between input and output must swing $2 V_{dd}$
- ❑ Or...behaves same as a double-sized capacitor on the output



Miller Effect For an Inverter



“A capacitor experiencing identical but opposite voltage swings at both its terminals can be replaced by a capacitor to ground, whose value is two times the original value.”