ESE3700: Circuit-Level Modeling, Design and Optimization for Digital Systems

Lec 9: February 26, 2024

Performance Inverters and Gates

Previously

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

Today

- τ-model
- Sizing
- Large Fanout
- Capacitance Revisited
  - Miller Effect
- Delay in Gates
- Data Dependent Delay
- Large Fanin

Transistor Sizing

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

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

- 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?

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

- 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

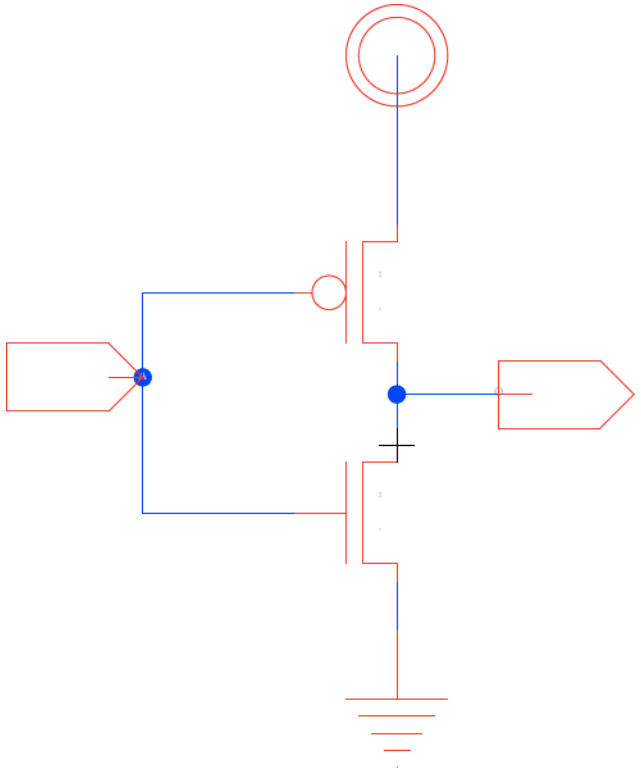
- $R_{drive} = R_0/W$
- $C_g = WC_0$

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)

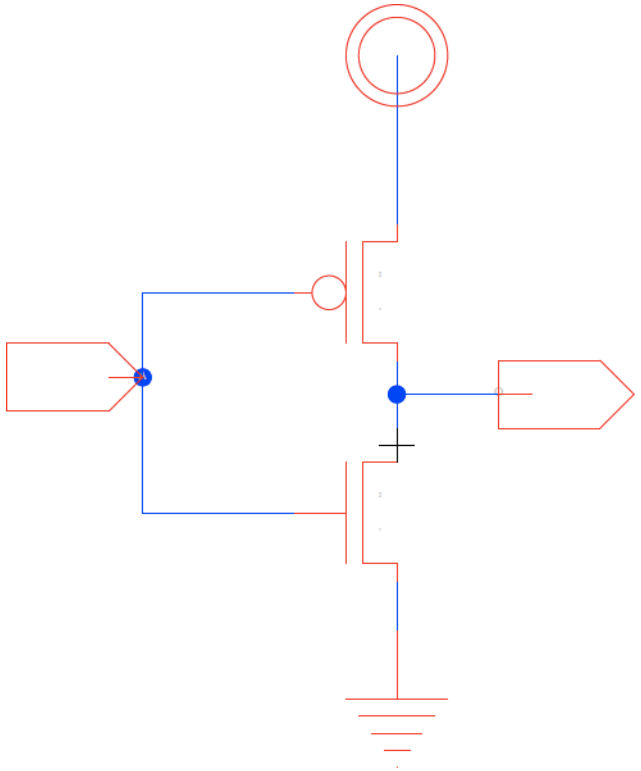
- How should we size to equalize Rise and Fall?
  - Given:
  - $\mu_n = 500 cm^2/Vs$, $\mu_p = 200 cm^2/Vs$
  - $R_{drive} = R_0/2$ ($I_{drive} = 2I_0$)



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

How to Size Transistors (preclass 2)

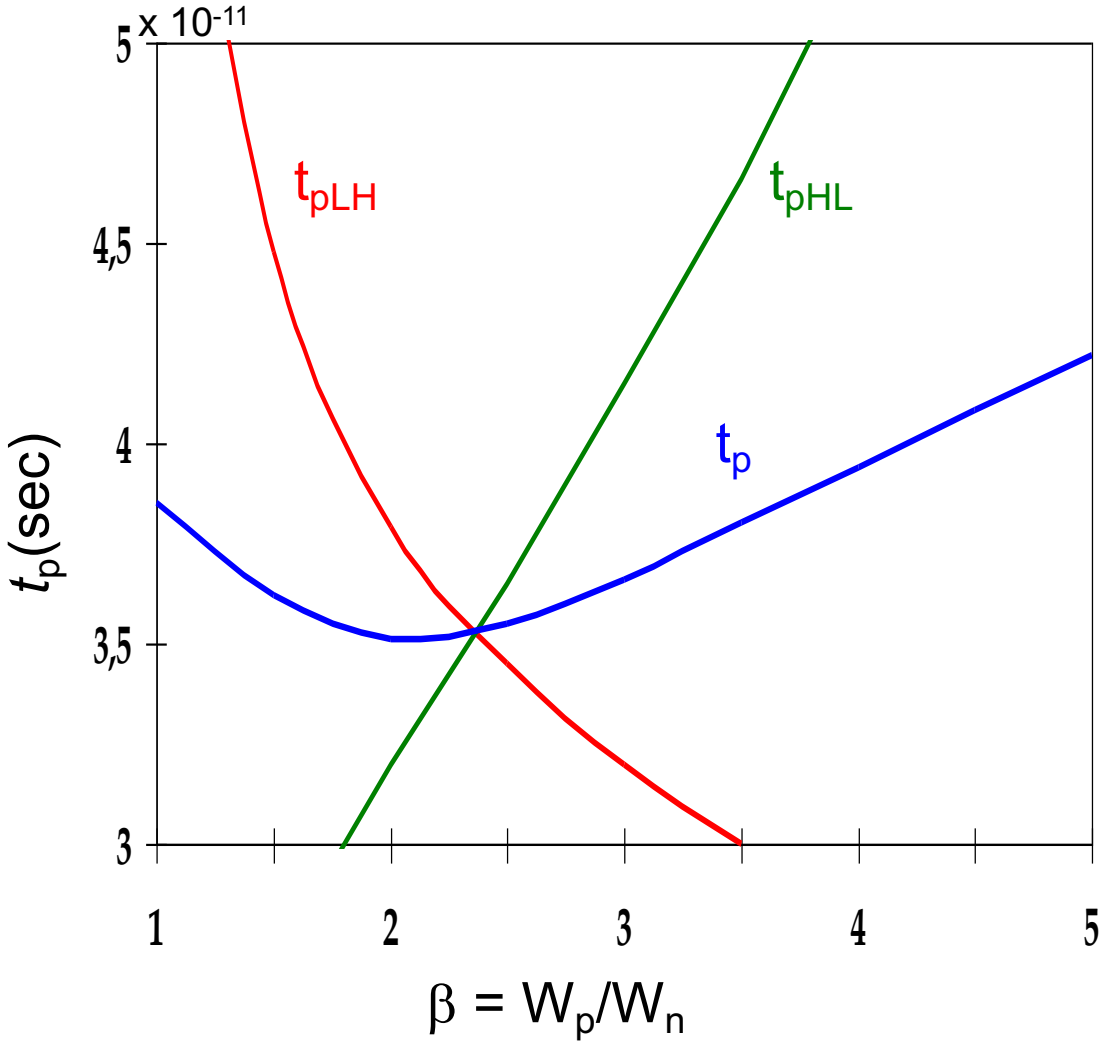
- How should we size to equalize Rise and Fall?
  - Given:
  - $\mu_n = 500 cm^2/Vs$, $\mu_p = 200 cm^2/Vs$
  - $R_{drive} = R_0/2$ ($I_{drive} = 2I_0$)



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

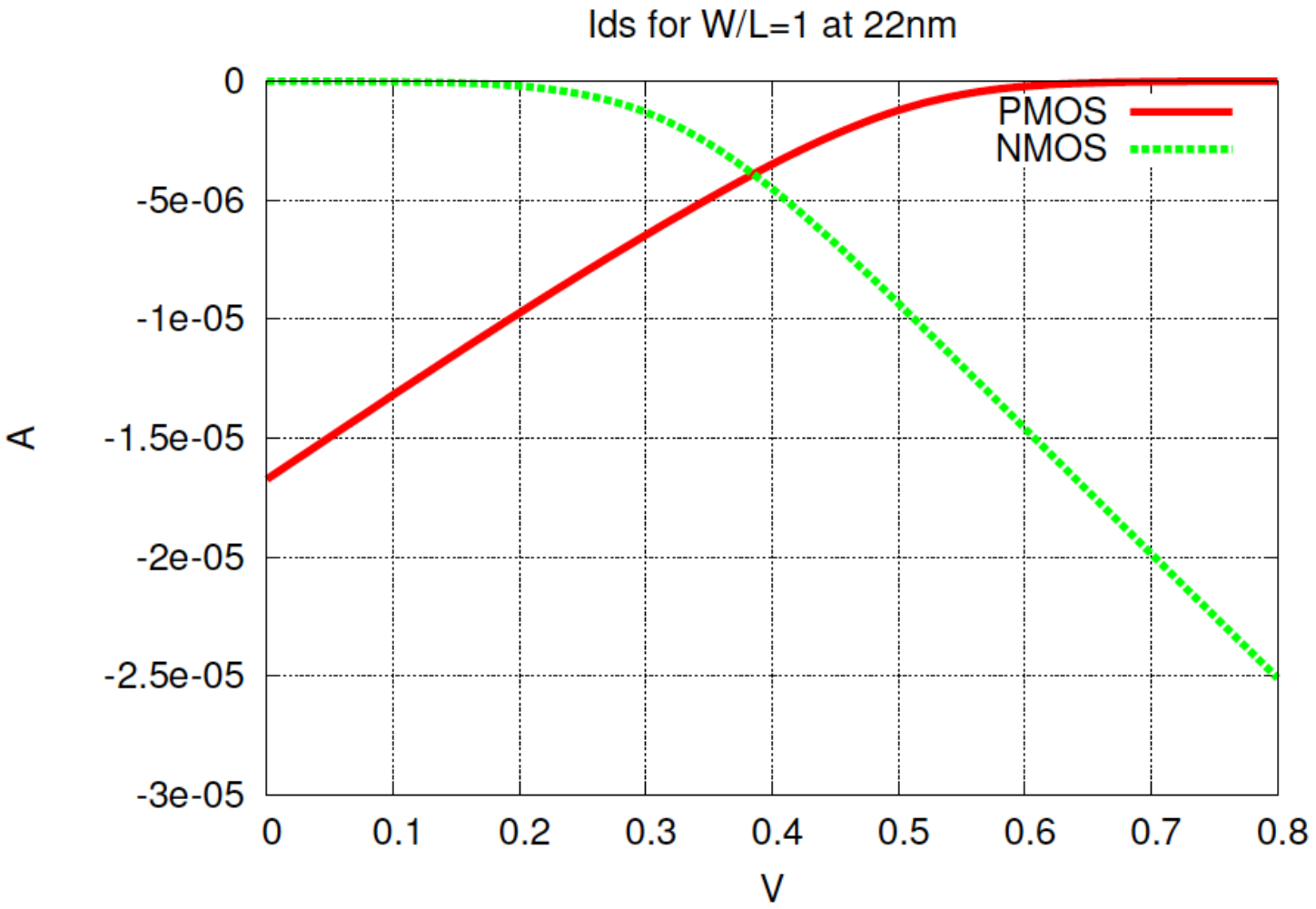
  - What is input capacitance for sized devices?

Size Transistors – Minimum Avg Delay



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

SPICE Simulation 22nm

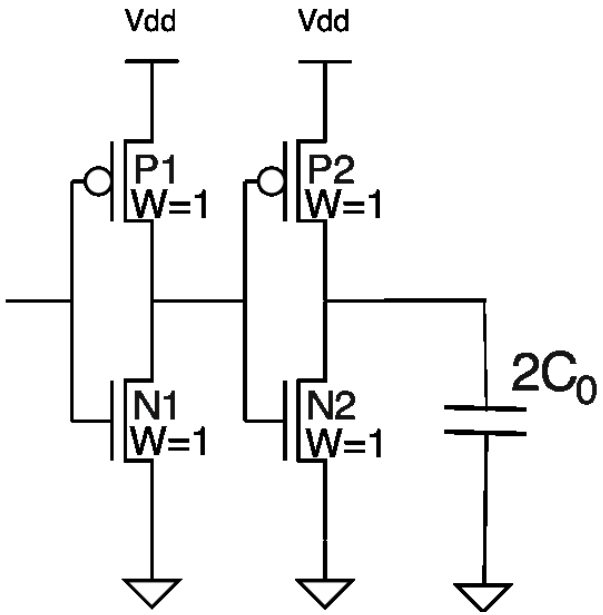


Equalizing Delay

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

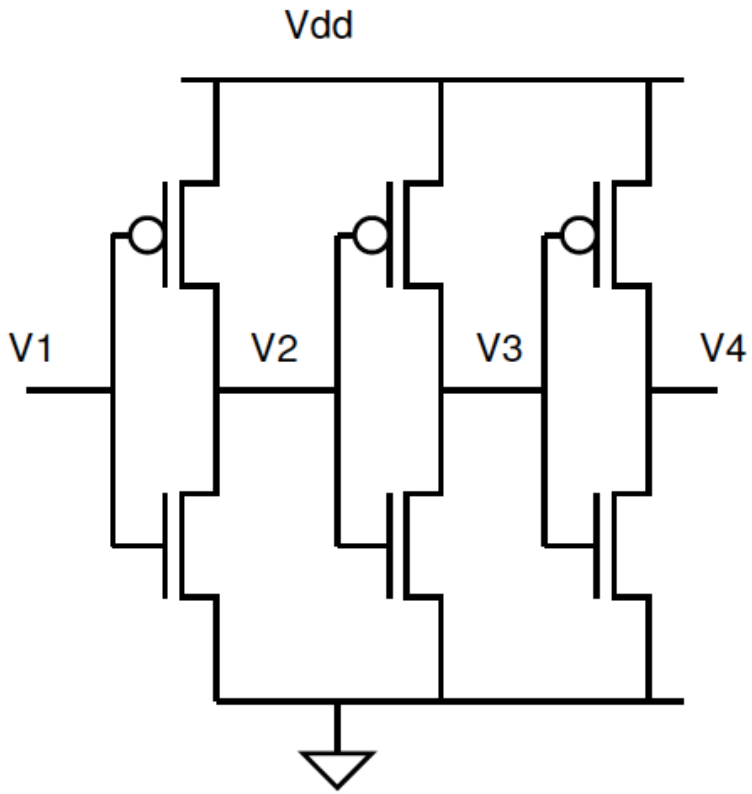
Multistage Delay

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



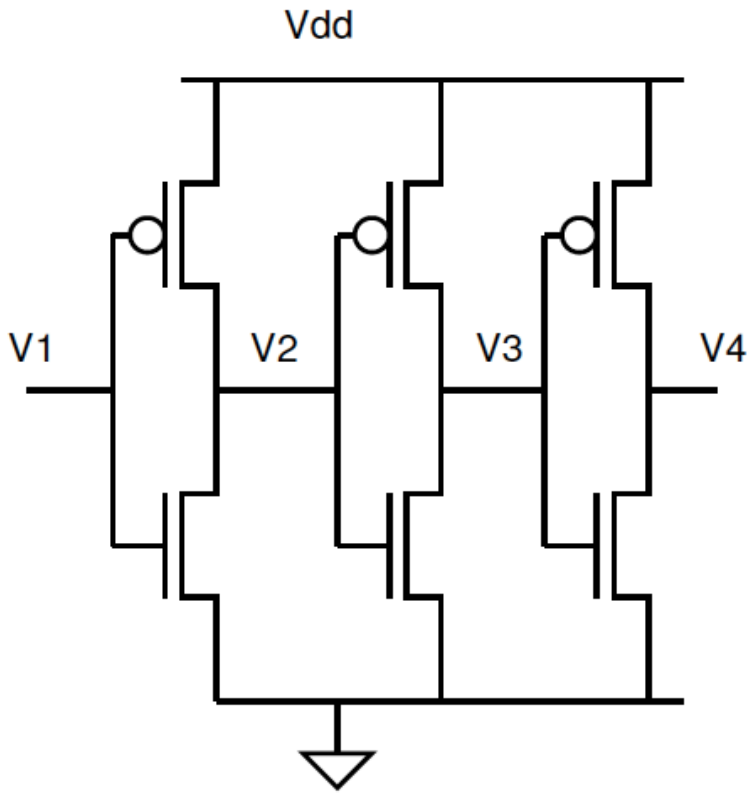
Inverter Sizing

- Delay from V1 to V4 of all min size transistors?



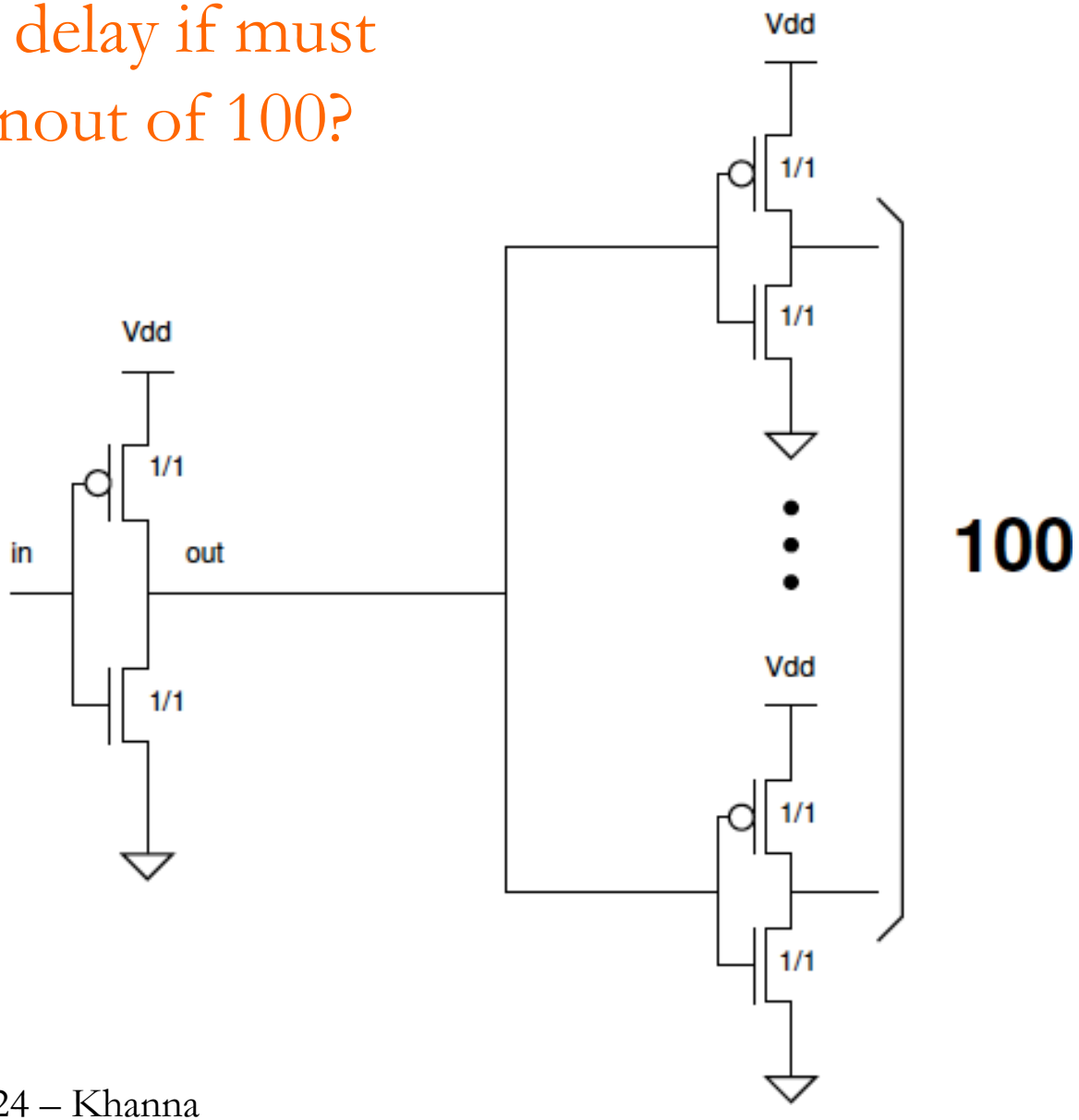
Inverter Sizing

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



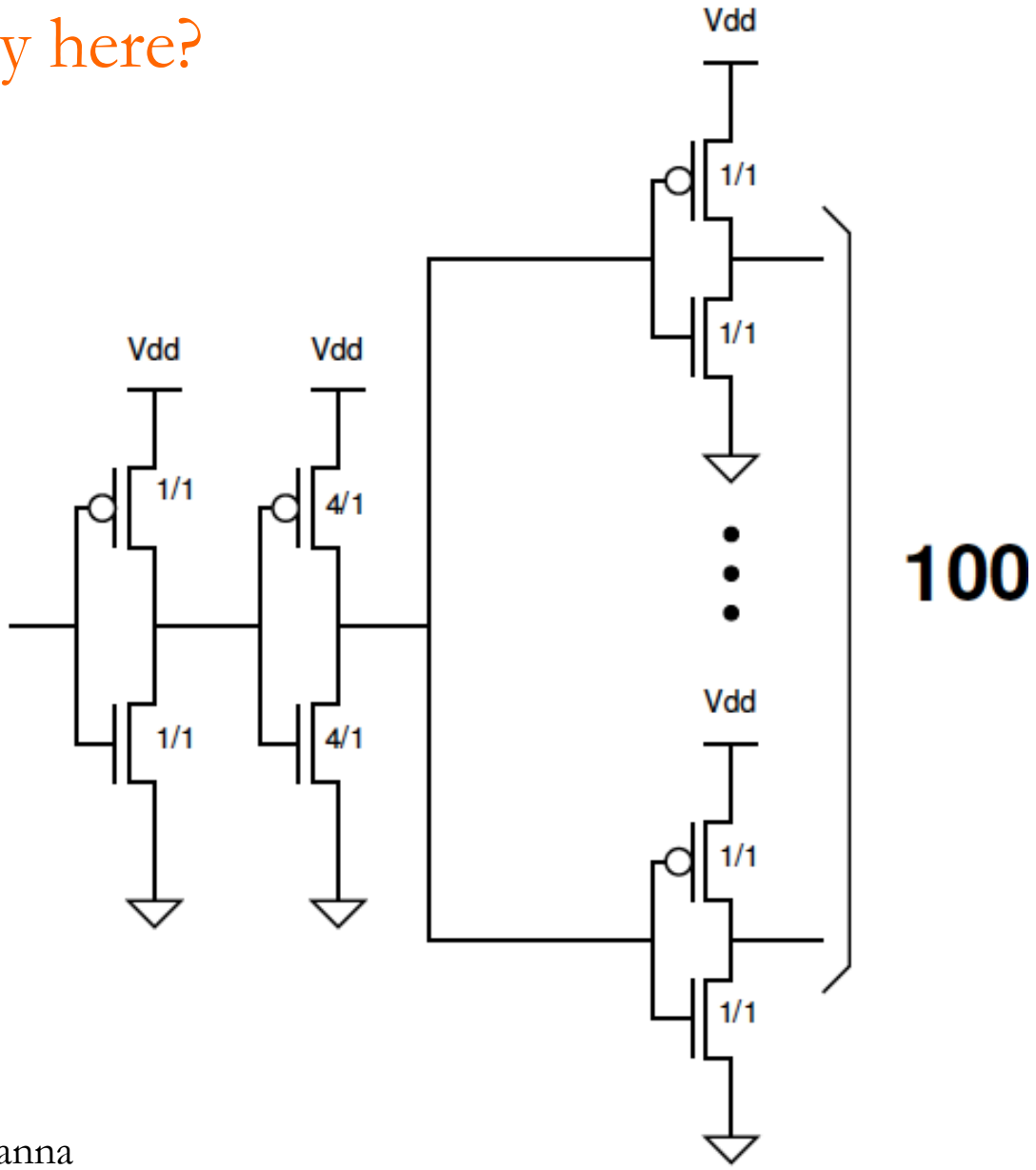
Large Fanout Delay (preclass 3)

- What is delay if must drive fanout of 100?



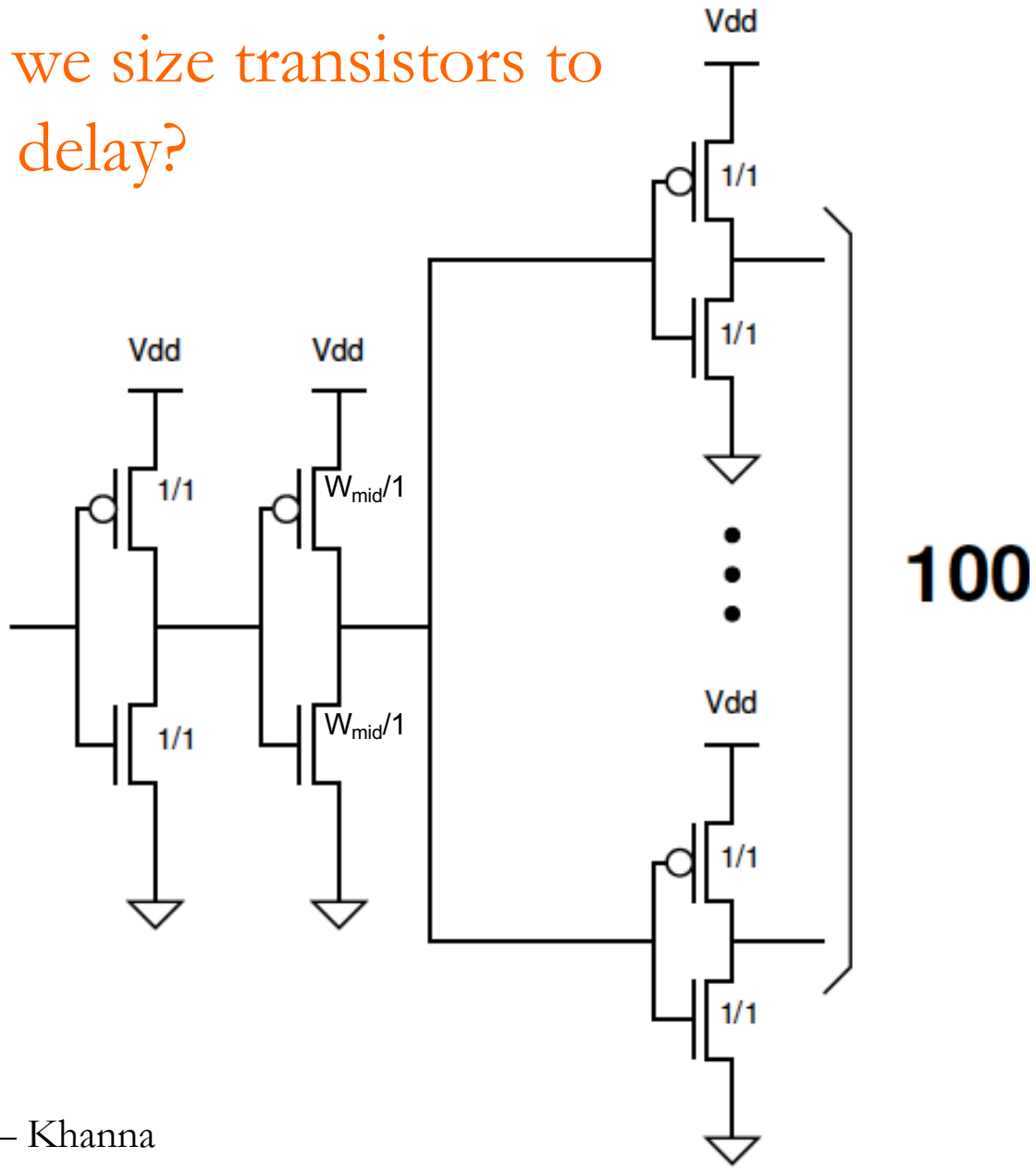
Graduated Fanout Delay (preclass 3)

- What is delay here?



Optimize Fanout Delay (preclass 4)

- How can we size transistors to minimize delay?



Optimizing (preclass 4)

- Derivate to minimize

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

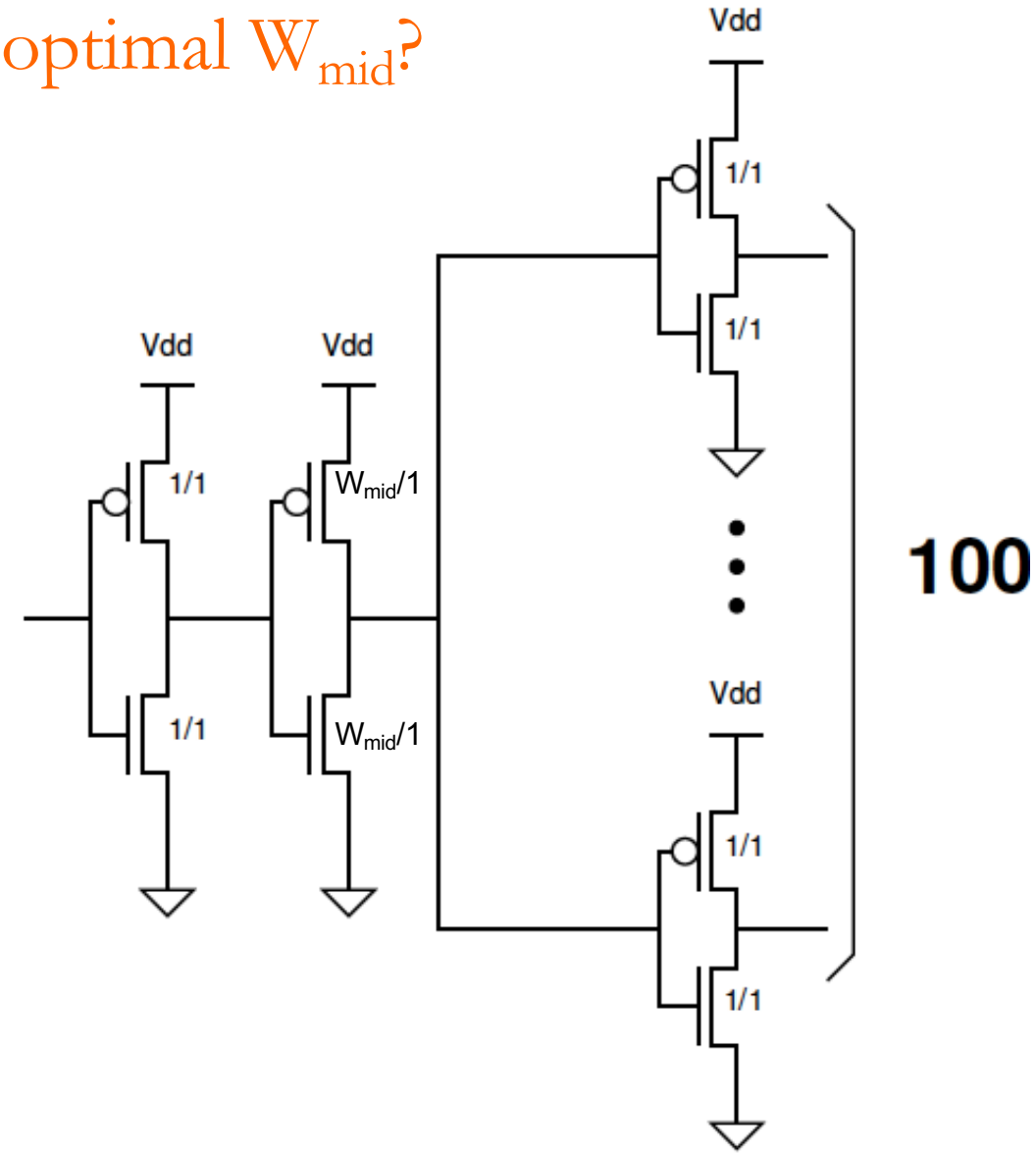
$$\frac{\partial \tau_{est}}{\partial W_{mid}} = 0$$

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

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

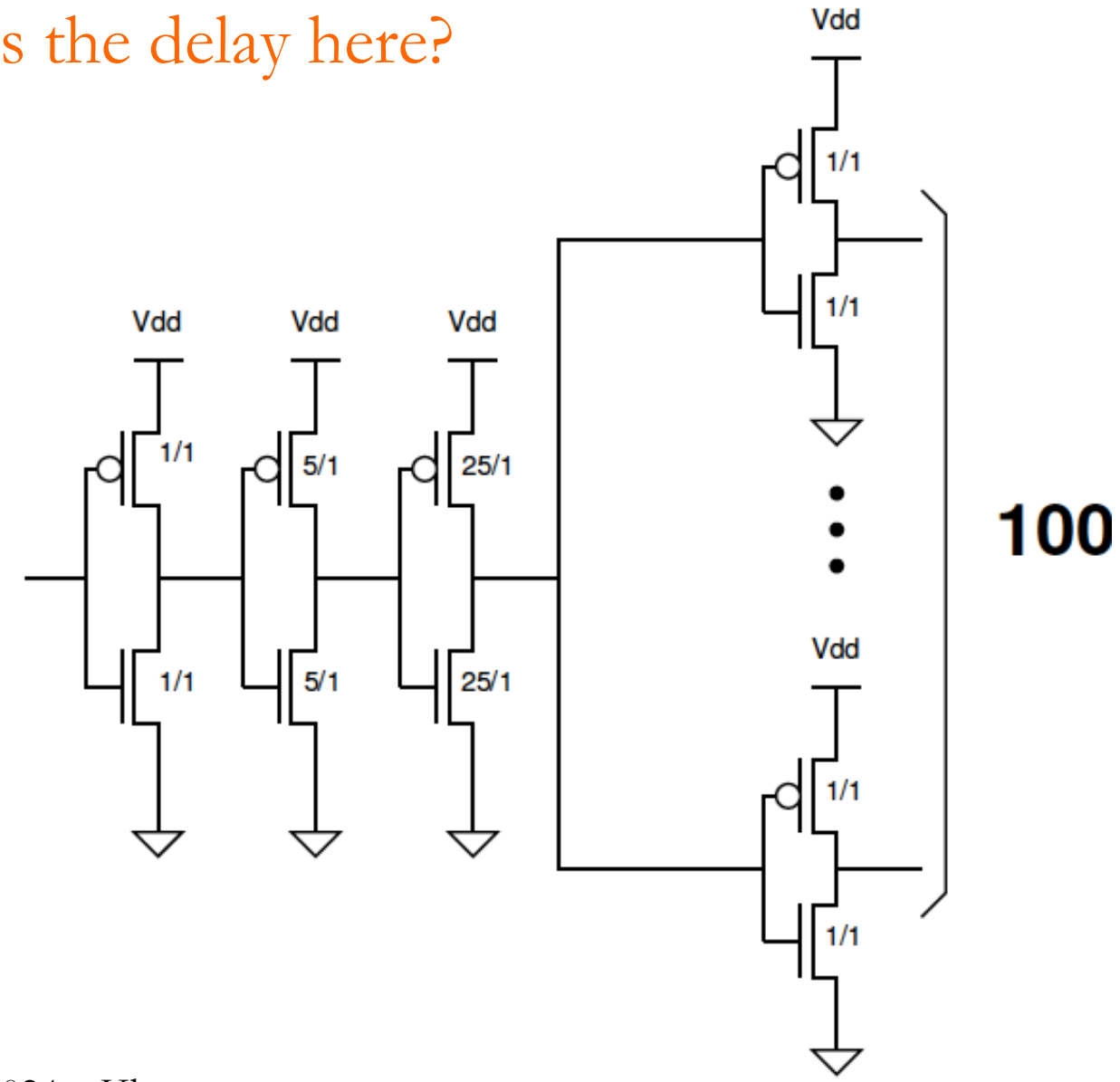
Delay? (preclass 4)

- Delay at optimal W_{mid}?



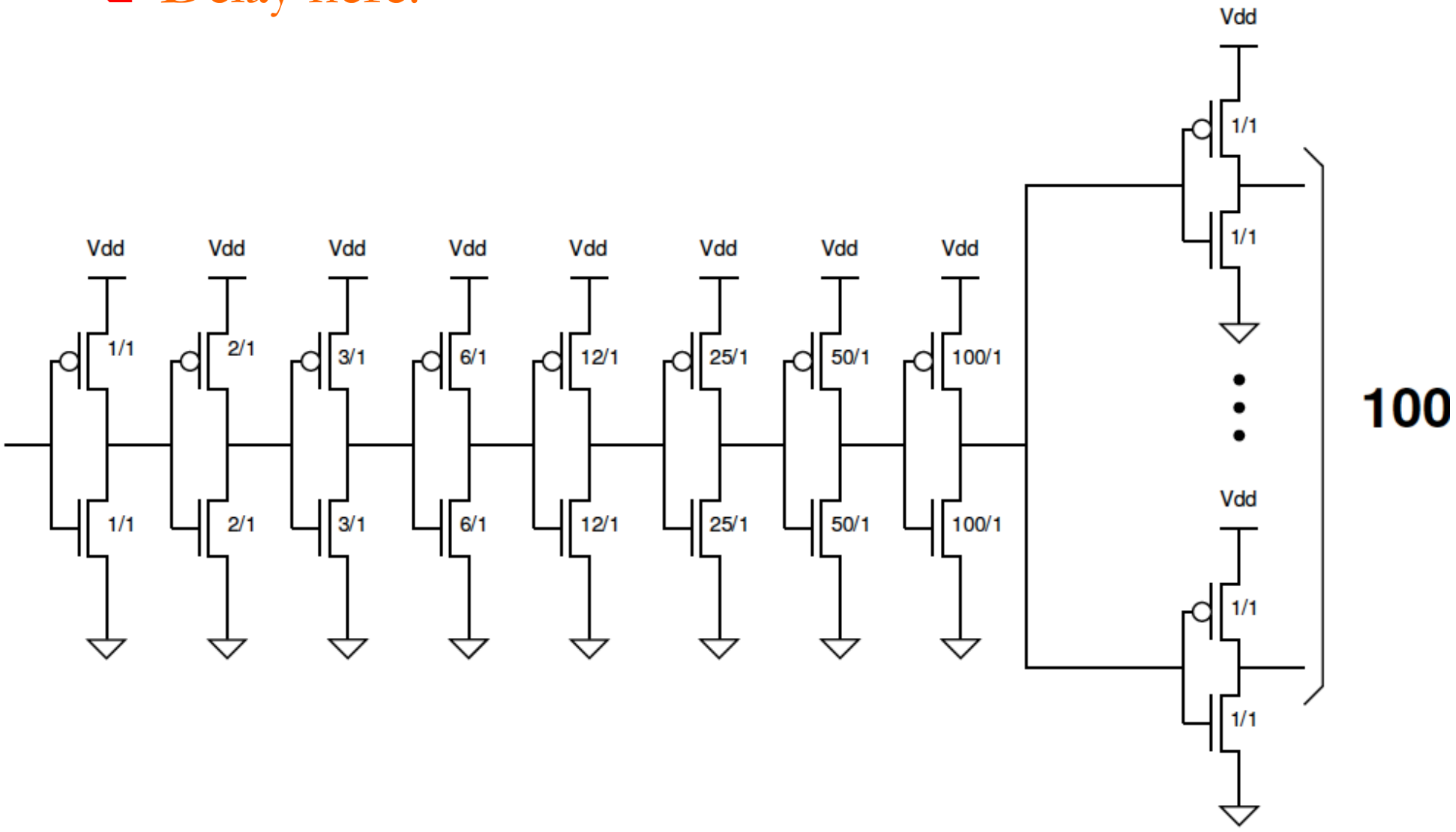
Try Again with More Stages (preclass 6)

- What is the delay here?

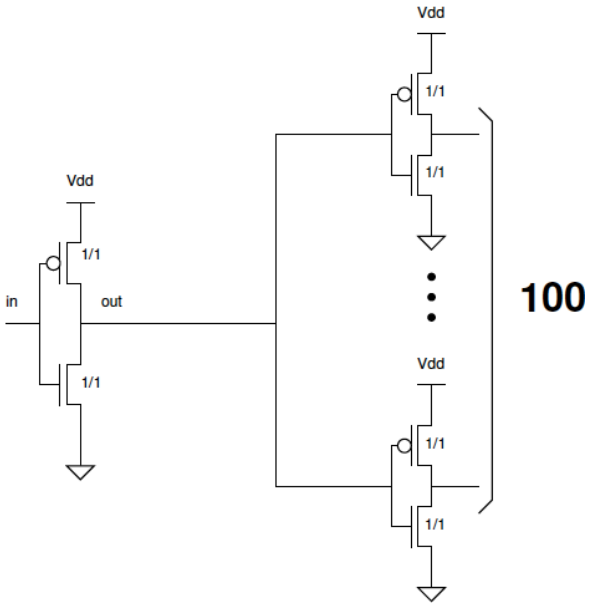


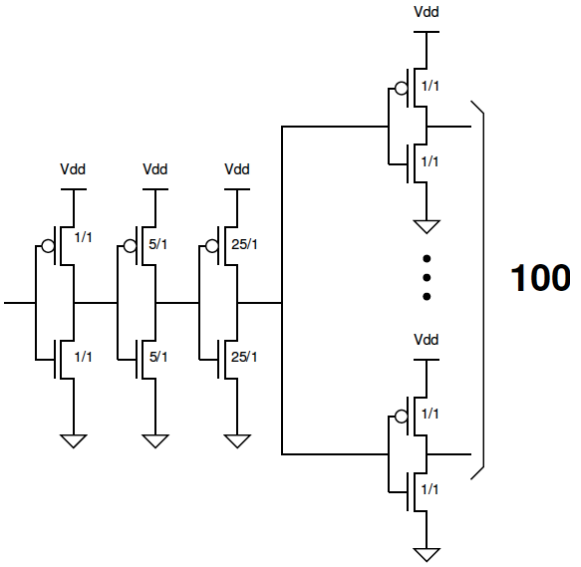
…and Again (preclass 5)

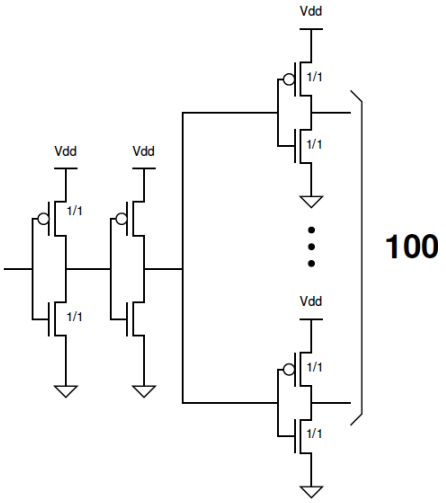
- Delay here?

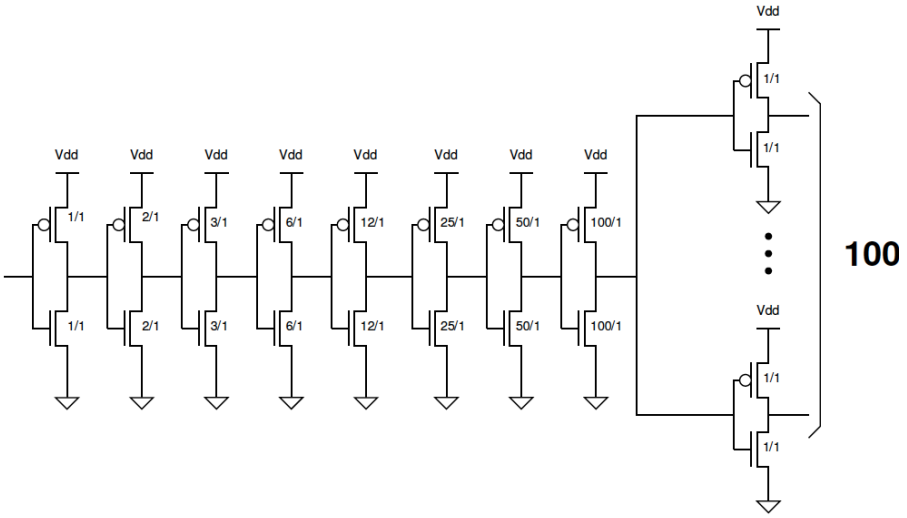


Delay Summary









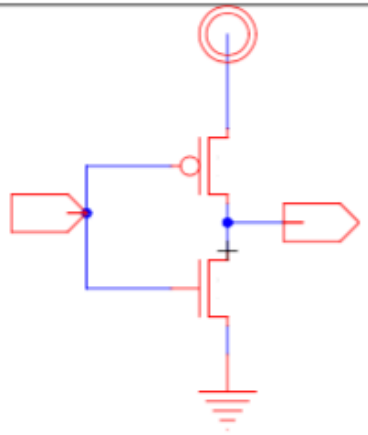
Lesson

- 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

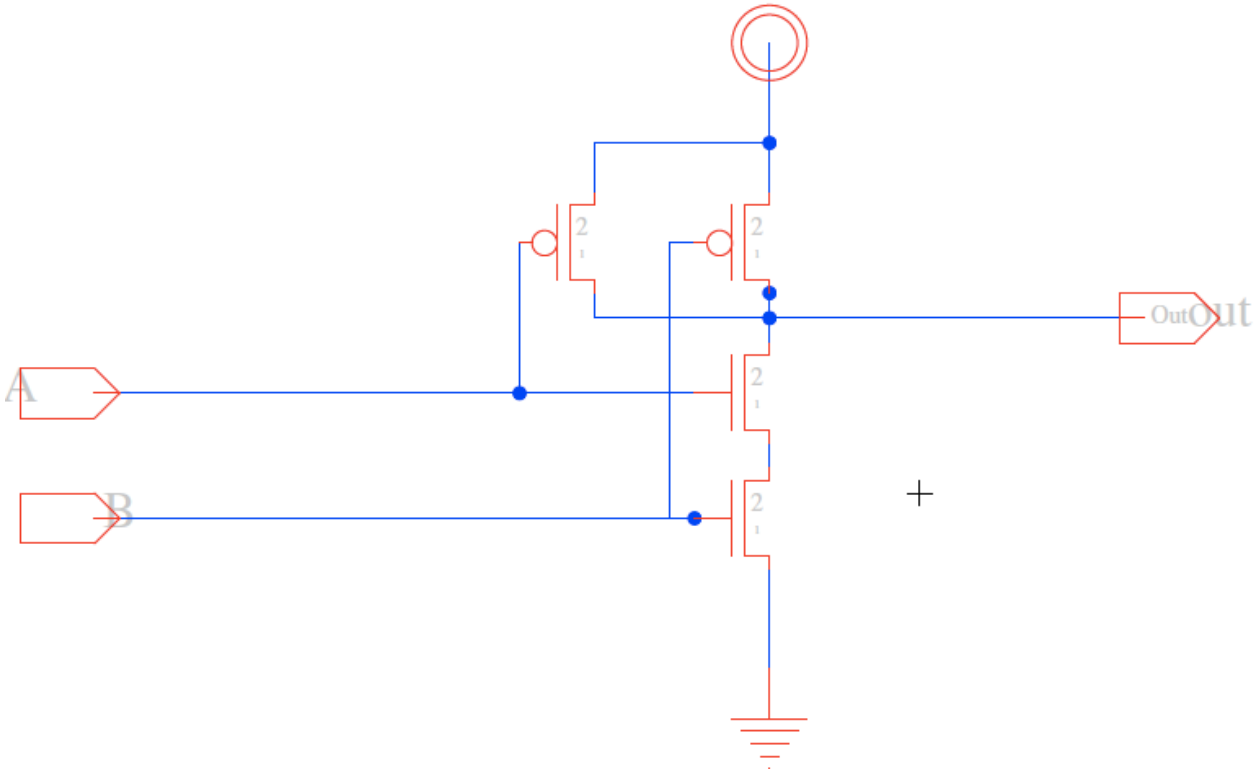
Inverter Performance (preclass 5, row 1)

- 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	$6C_0$

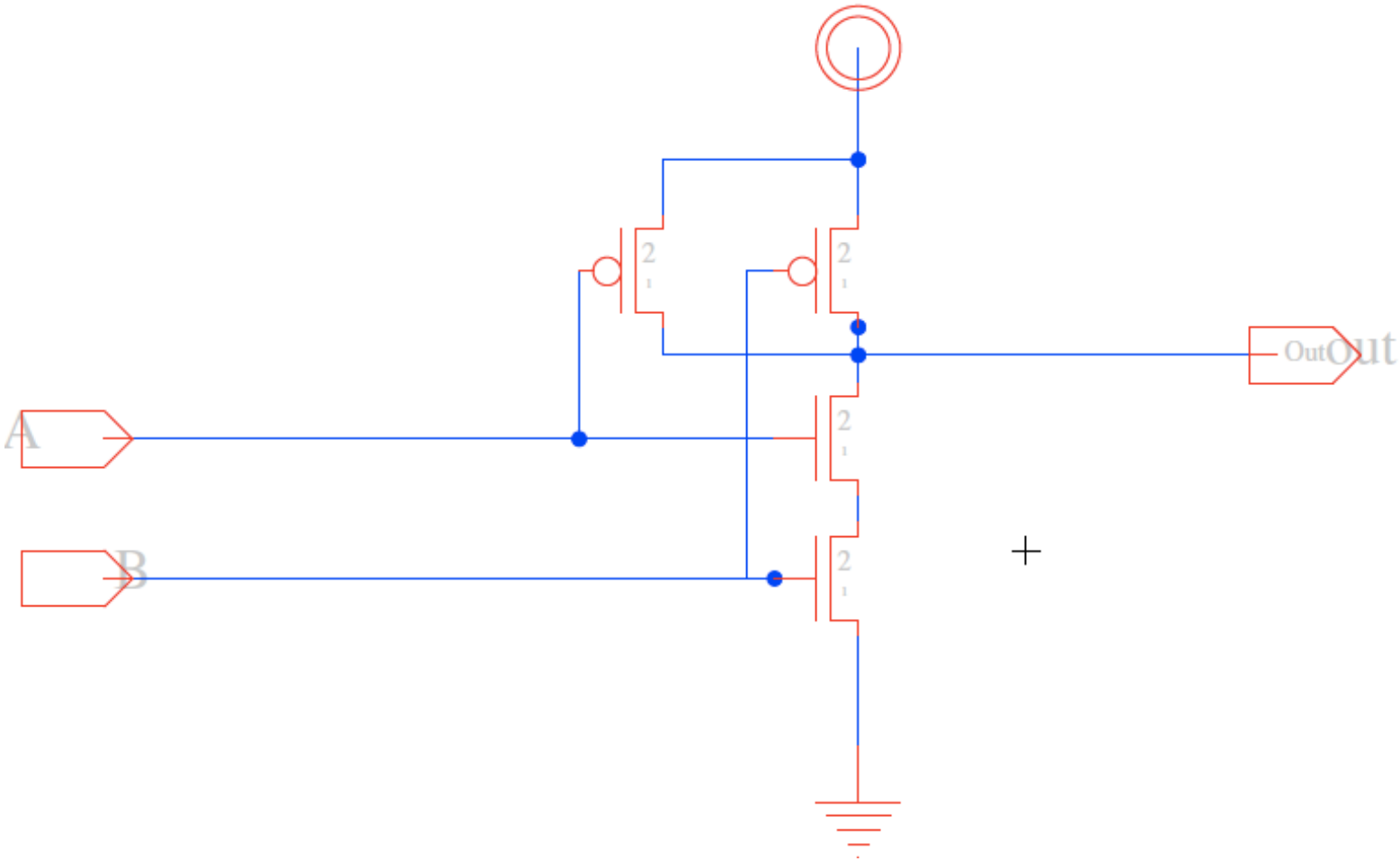
Data Dependent Delay

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



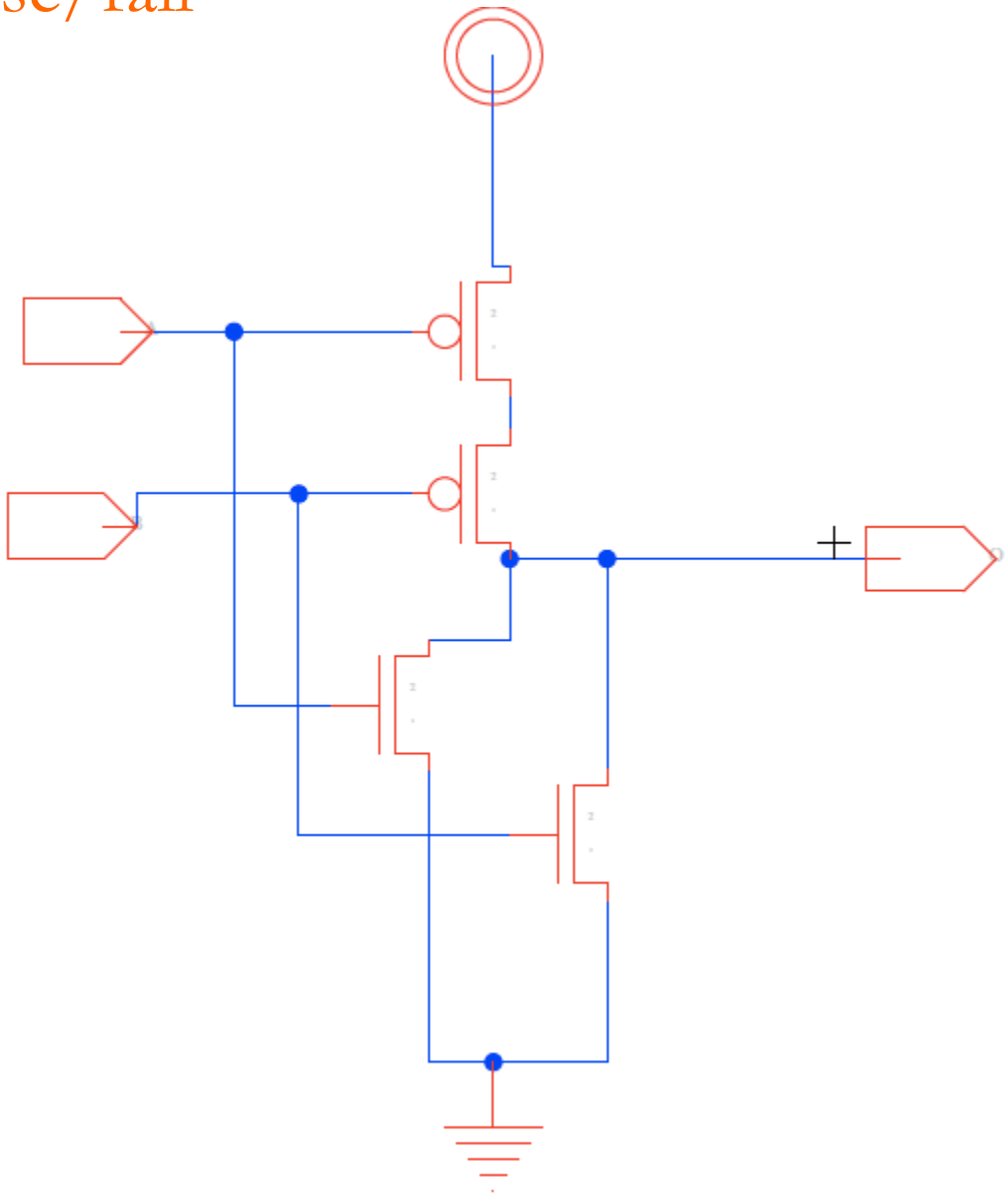
Transistor Sizing (preclass 6, row 2)

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



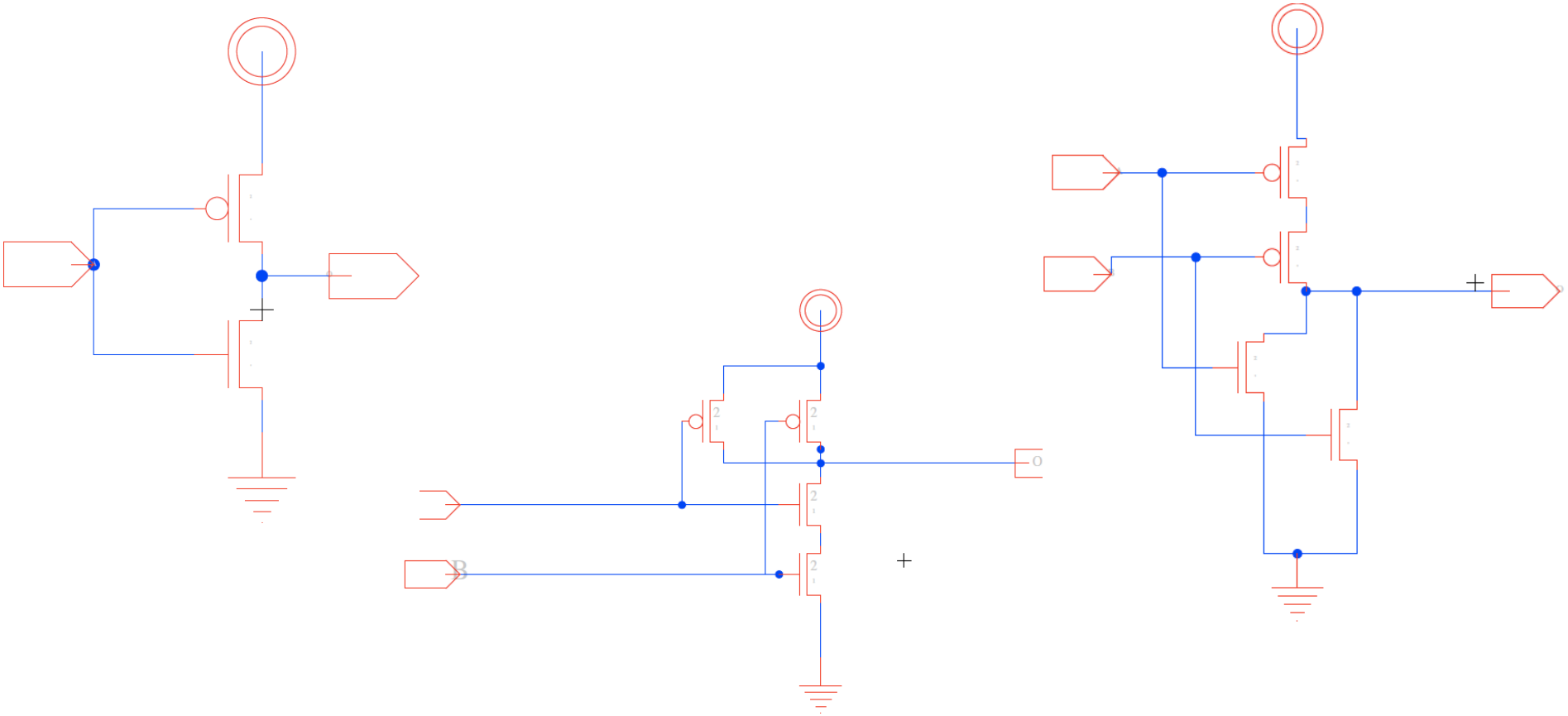
Transistor Sizing (preclass 6, row 3)

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



Input Load

- Input capacitance per input in each case?

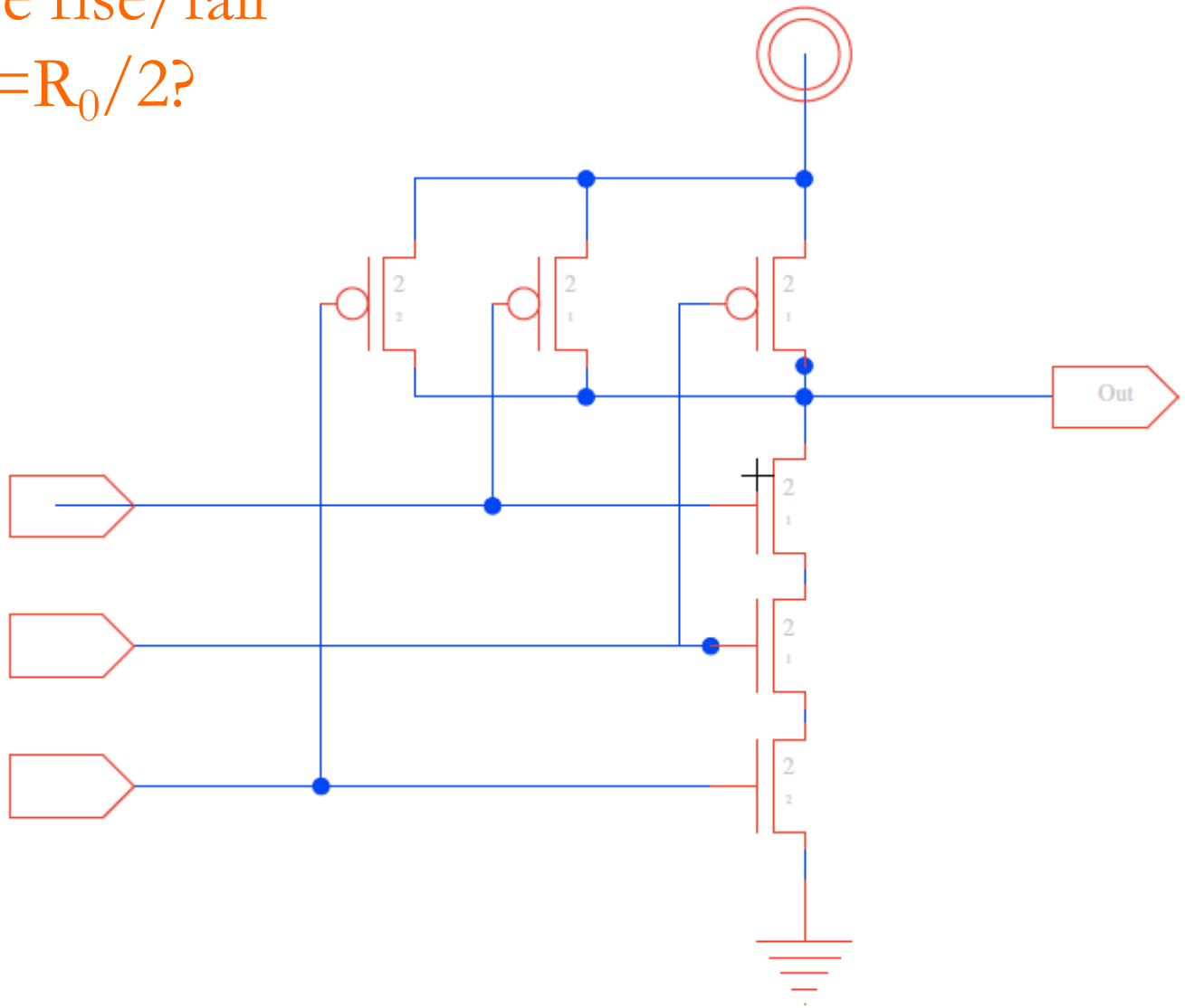


Observe

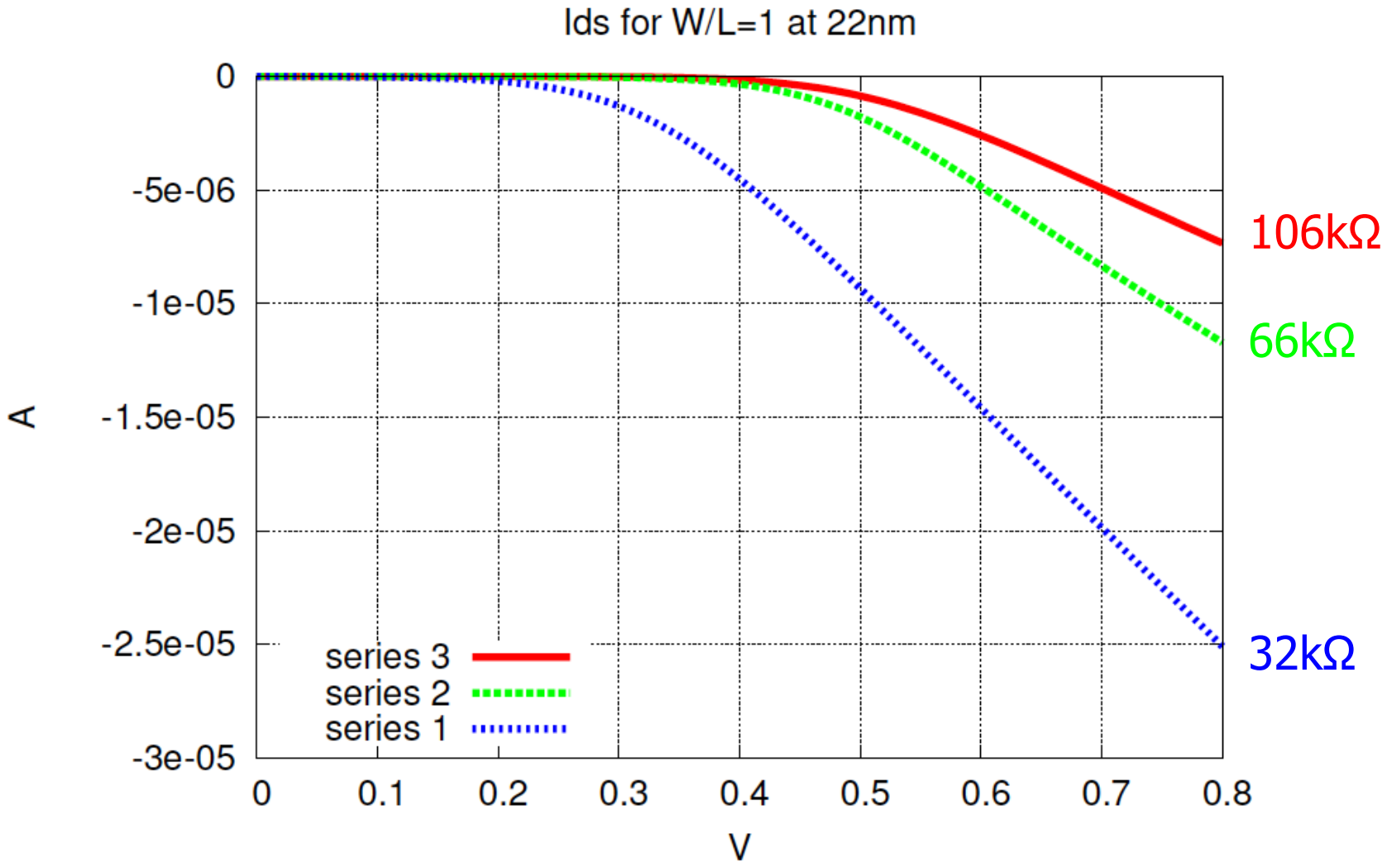
- Ratio of Input Load Capacitance to Output Drive Strength: C_{InLoad}/I_{ds}
  - Differs with gate function
  - Gate efficiency

Transistor Sizing (preclass 6, row 4)

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



Series Transistors



Increasing Fanin (preclass 7)

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

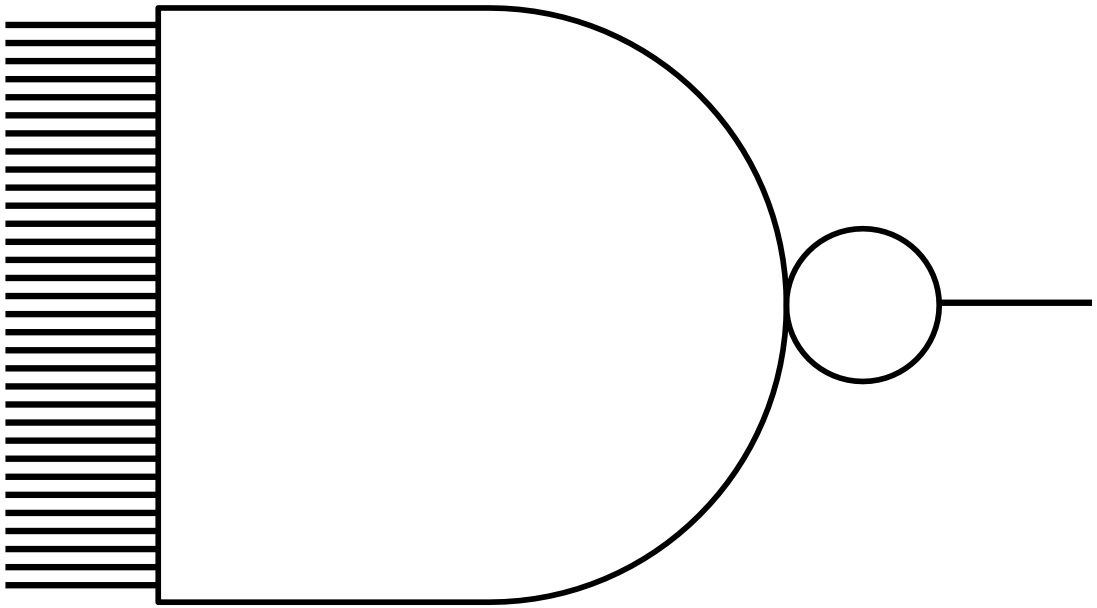
Fanin

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

nand32 (preclass 8, row 1)

- single-stage nand32
  - Delay with $R_0/2$ input drive and $4C_0$ load?

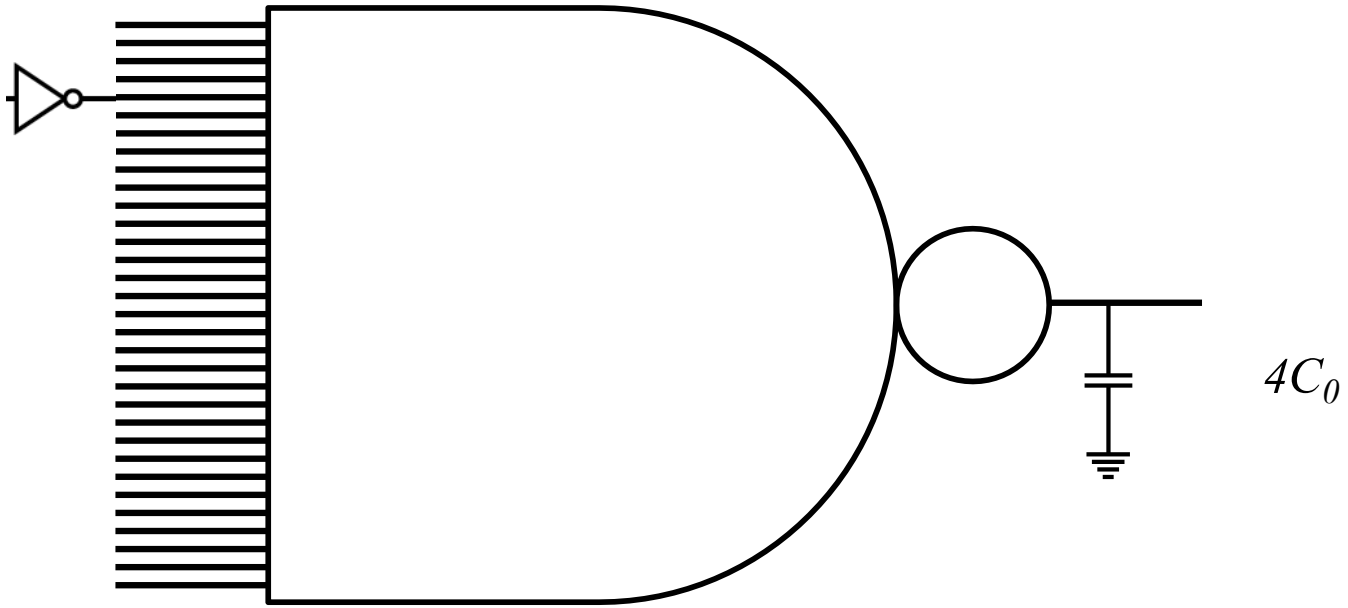
$R_{n0}=R_{p0}$ case only



nand32 (preclass 8, row 1)

- single-stage nand32
  - Delay with $R_0/2$ input drive and $4C_0$ load?

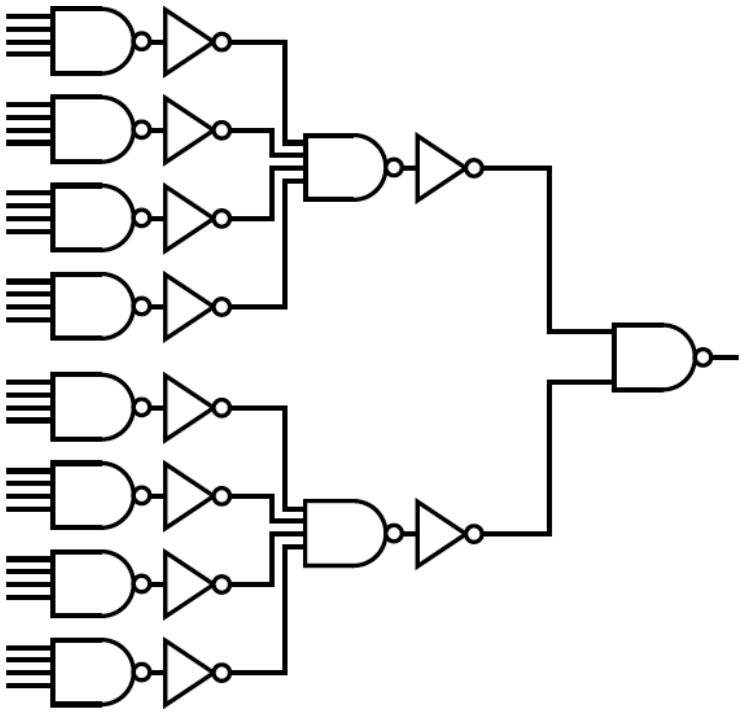
$R_{n0}=R_{p0}$ case only



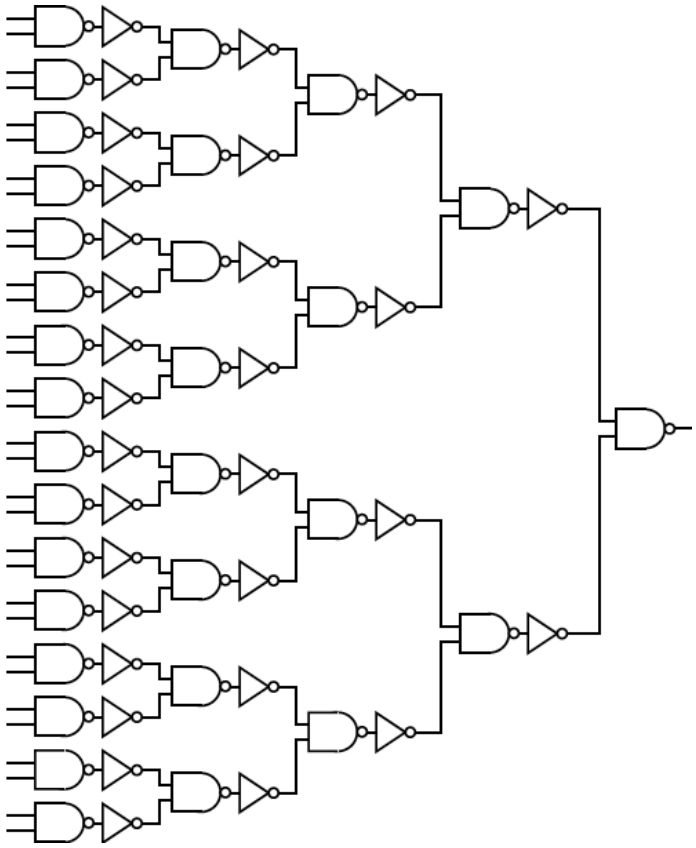
Which is Faster? (preclass 8, rows 2&3)

- nand32

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



nand4-inv-nand4-inv-nand2

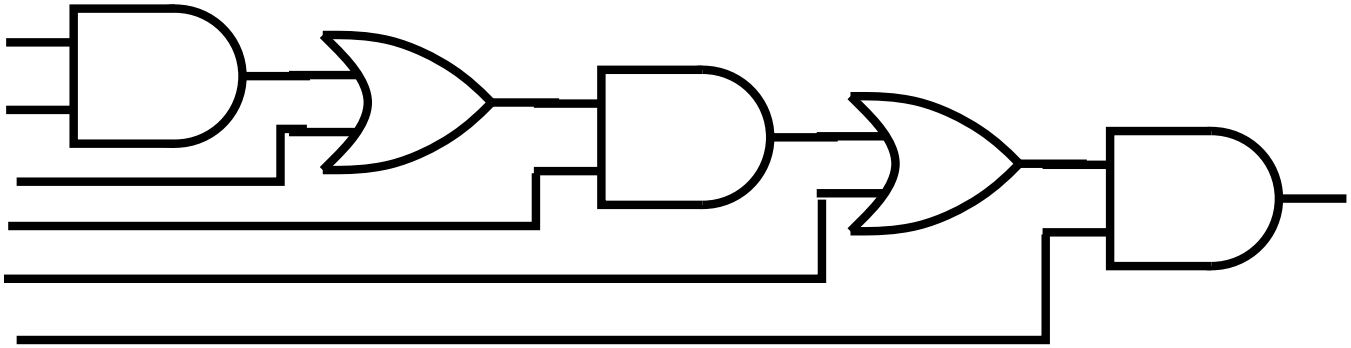


$(\text{nand2-inv})^4\text{-nand2}$

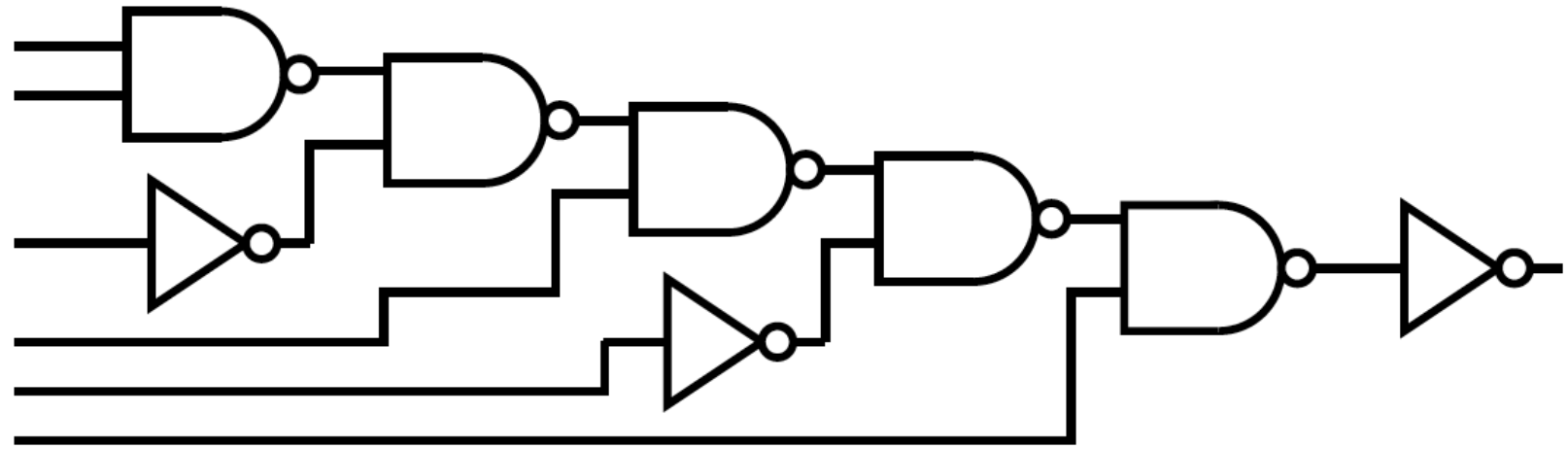
Lesson

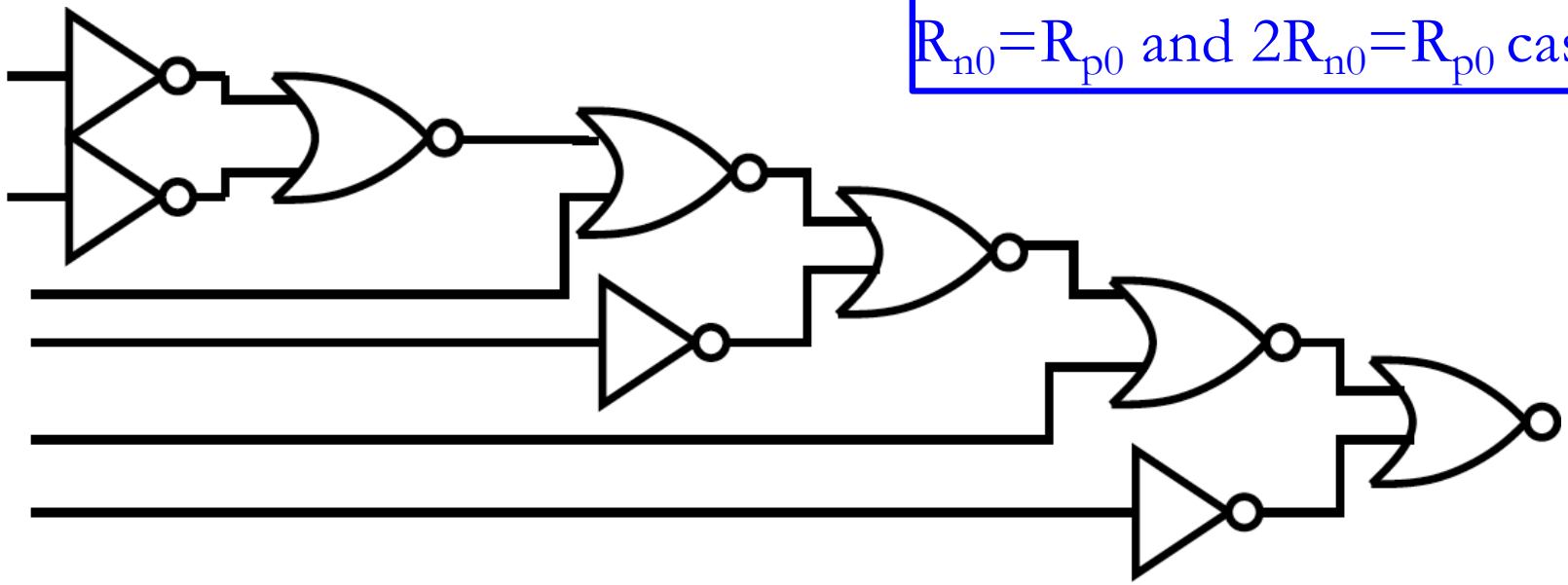
- Large gates are slow / inefficient
  - High capacitive load / drive current
- Small gates can be inefficient
  - Need many stages
- Staging over moderate size gates minimizes delay
- Exact size will be technology dependent

And-Or Chain



Delay of each implementation? (preclass 9)

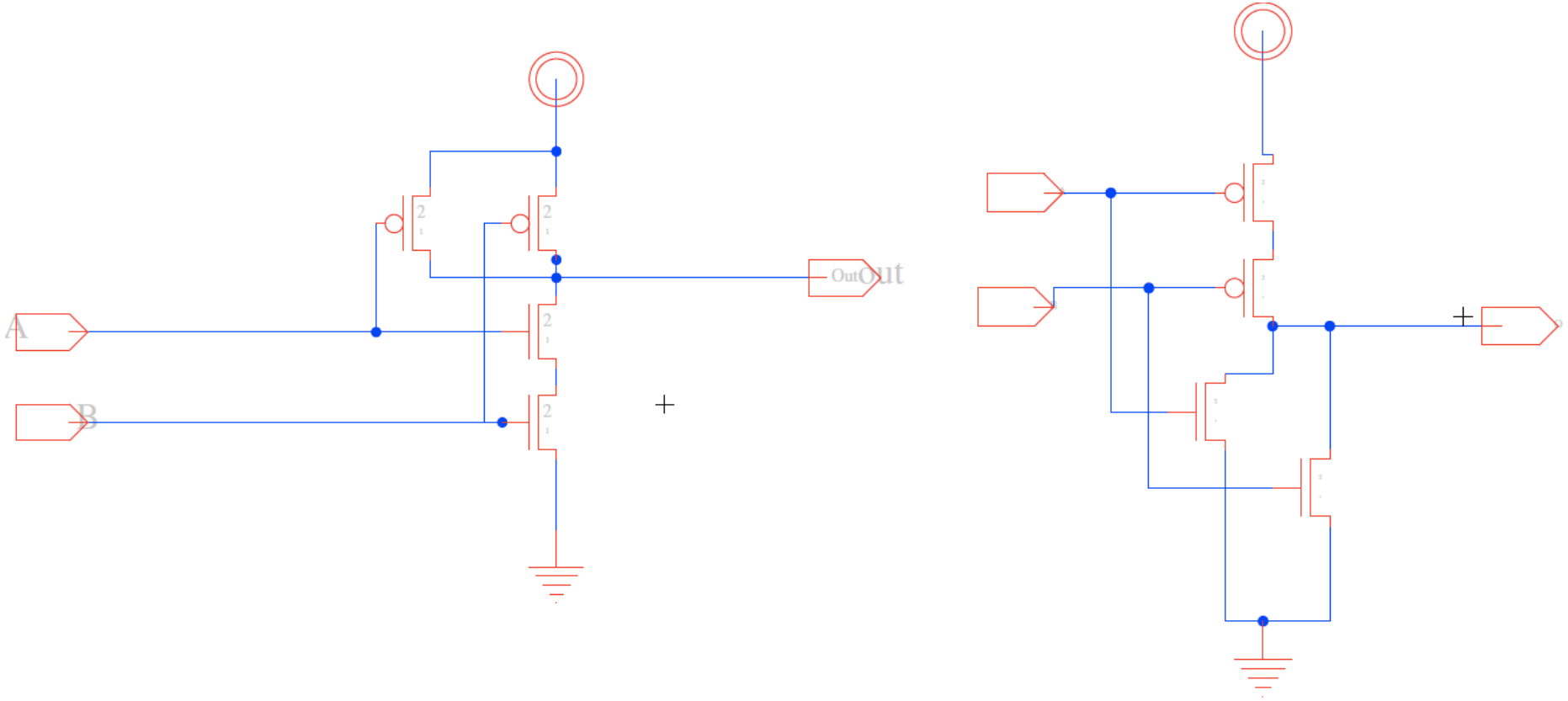




$R_{n0}=R_{p0}$ and $2R_{n0}=R_{p0}$ cases

Take Away?

- nor vs. nand



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

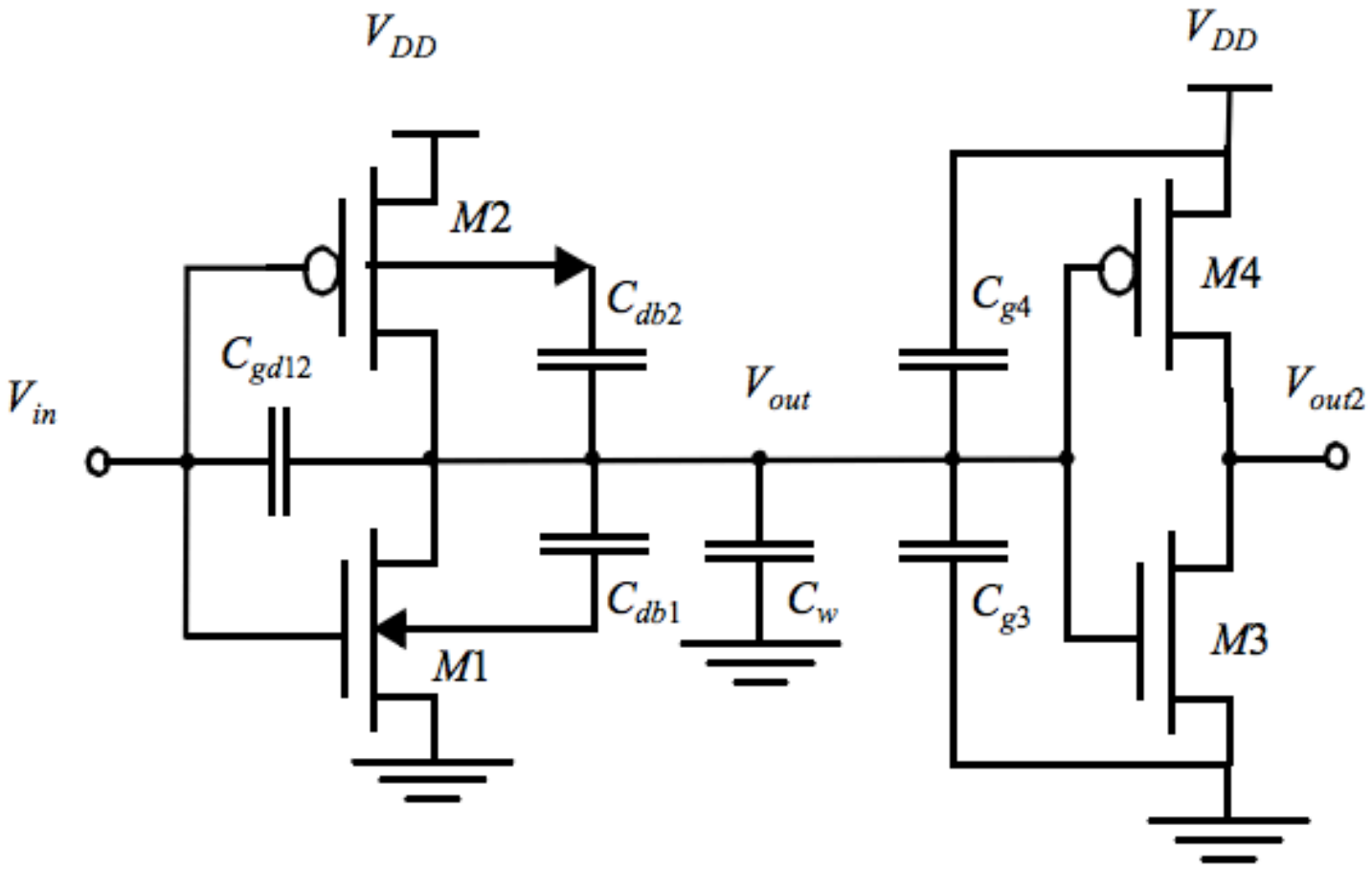
- HW4 due Wednesday 2/28
- Drop date is tomorrow 2/27

Acknowledgement

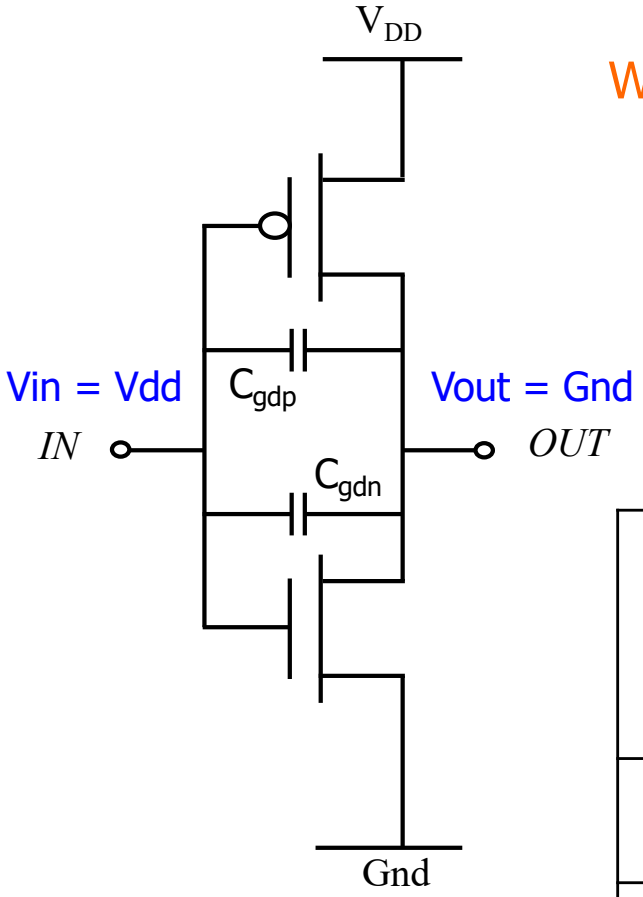
- Prof. André DeHon (University of Pennsylvania)
- Prof. Jing Li (University of Pennsylvania)

Miller Effect (Optional)

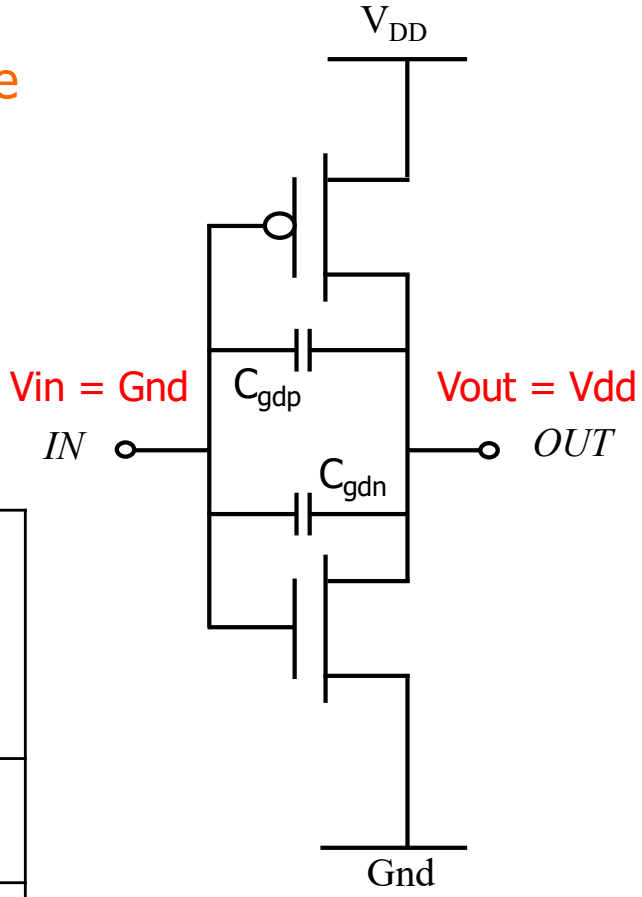
Capacitance Reminder



Charge on Capacitors (preclass 6)



What is charge, Q, on each of the capacitors when Vin=Vdd and Vin=Gnd?



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

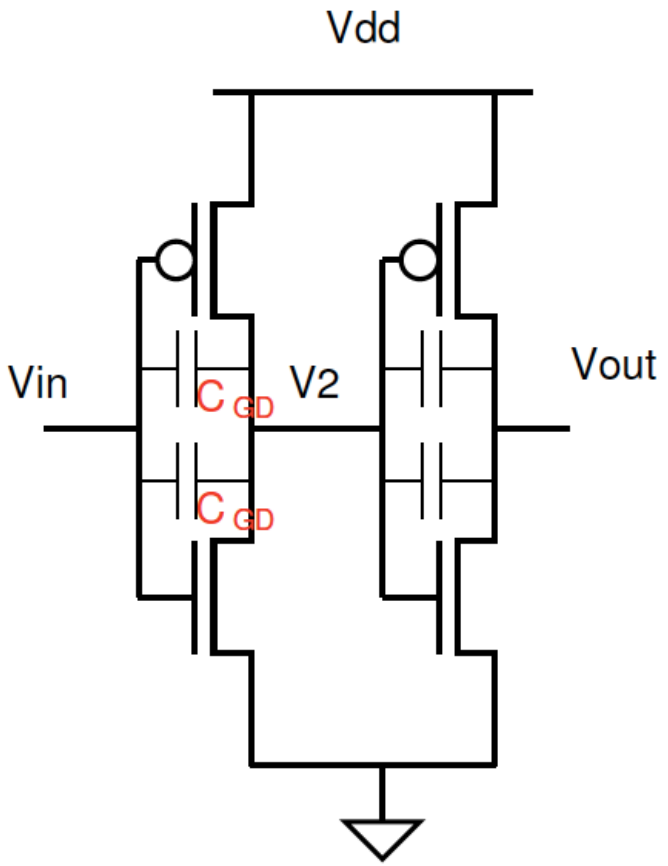
Questions

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

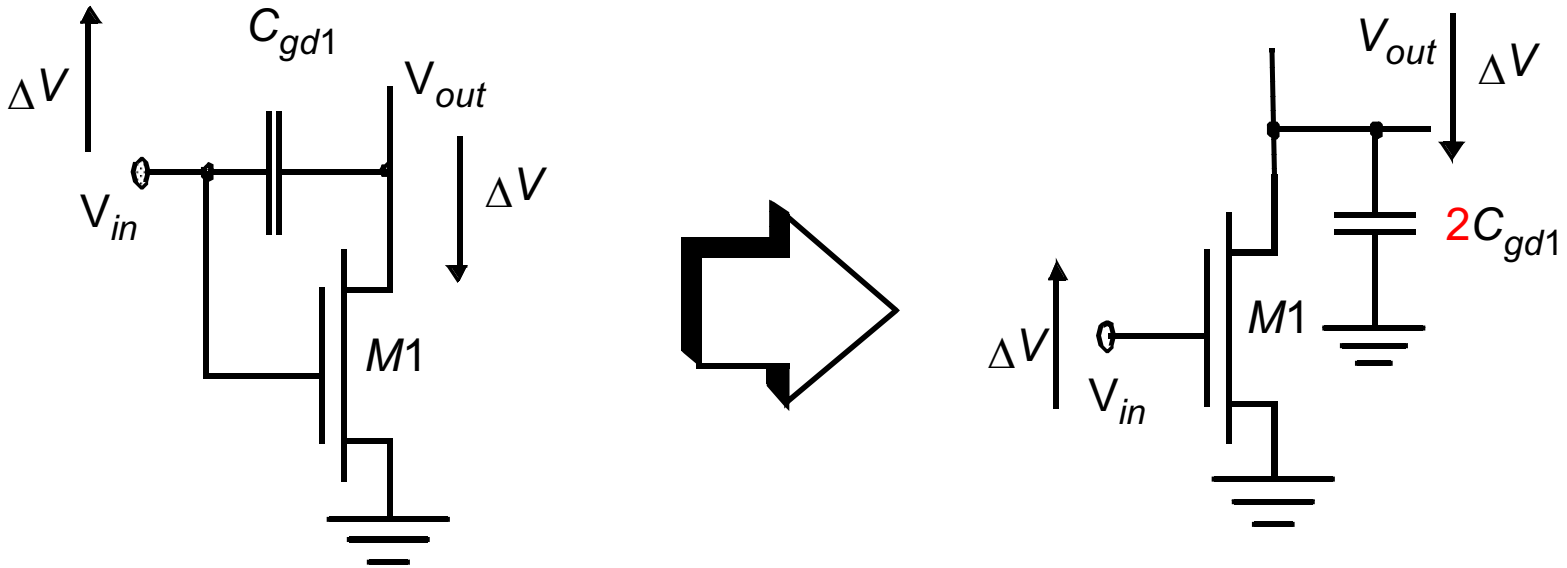
Vin	**PMOS** C_{gdp}	**NMOS** C_{gdn}
Vdd		
Gnd		

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."