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

Lec 13: October 6, 2021

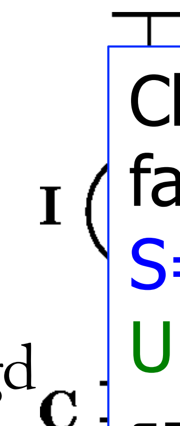
Inverter Performance

Scale V separately with Factor U

- $\tau_{gd}=Q/I=(CV)/I$
- $V' \rightarrow V/U$
  - $I'_d=(\mu C_{OX}S/2)((W/S)/(L/S)(V_{gs}/U-V_{TH}/U)^2$
- $I'_d \rightarrow S/U^2 \times I_d$
- $C' \rightarrow C/S$
- $\tau'_{gd} \rightarrow ((1/(SU)) / (S/U^2)) \times \tau_{gd}$
- $\tau'_{gd} \rightarrow (U/S^2) \times \tau_{gd}$
- $f' \rightarrow (S^2/U) \times f$

$f_{new}/f_{ideal}=10$

Ideal scale factors:
S=100
U=100
τ=1/100
f_{ideal}=100



Cheating factors:
S=100
U=10
τ=1/1000
f_{new}=1000

Power Density Impact

- $P = (1/2)CV^2 f$
- $P \rightarrow (1/S)\ (1/U^2)\ (S^2/U) = S/U^3$
- $P/A \rightarrow (S/U^3)\ /\ (1/S^2) = S^3/U^3$

- U=10 S=100
- $P/A \rightarrow 1000\ (P/A)$

Power Density Impact

- $P = (1/2)CV^2 f$
- $P \rightarrow (1/S)\ (1/U^2)\ (S^2/U) = S/U^3$
- $P/A \rightarrow (S/U^3)\ /\ (1/S^2) = S^3/U^3$

- U=10   S=100
- $P/A \rightarrow 1000\ (P/A)$

- Compare with ideal scaling:
- $P/A \rightarrow S^3 \times P$ (ideal scaling)
- $P/A \rightarrow 1{,}000{,}000\ (P/A)$ (ideal scaling)

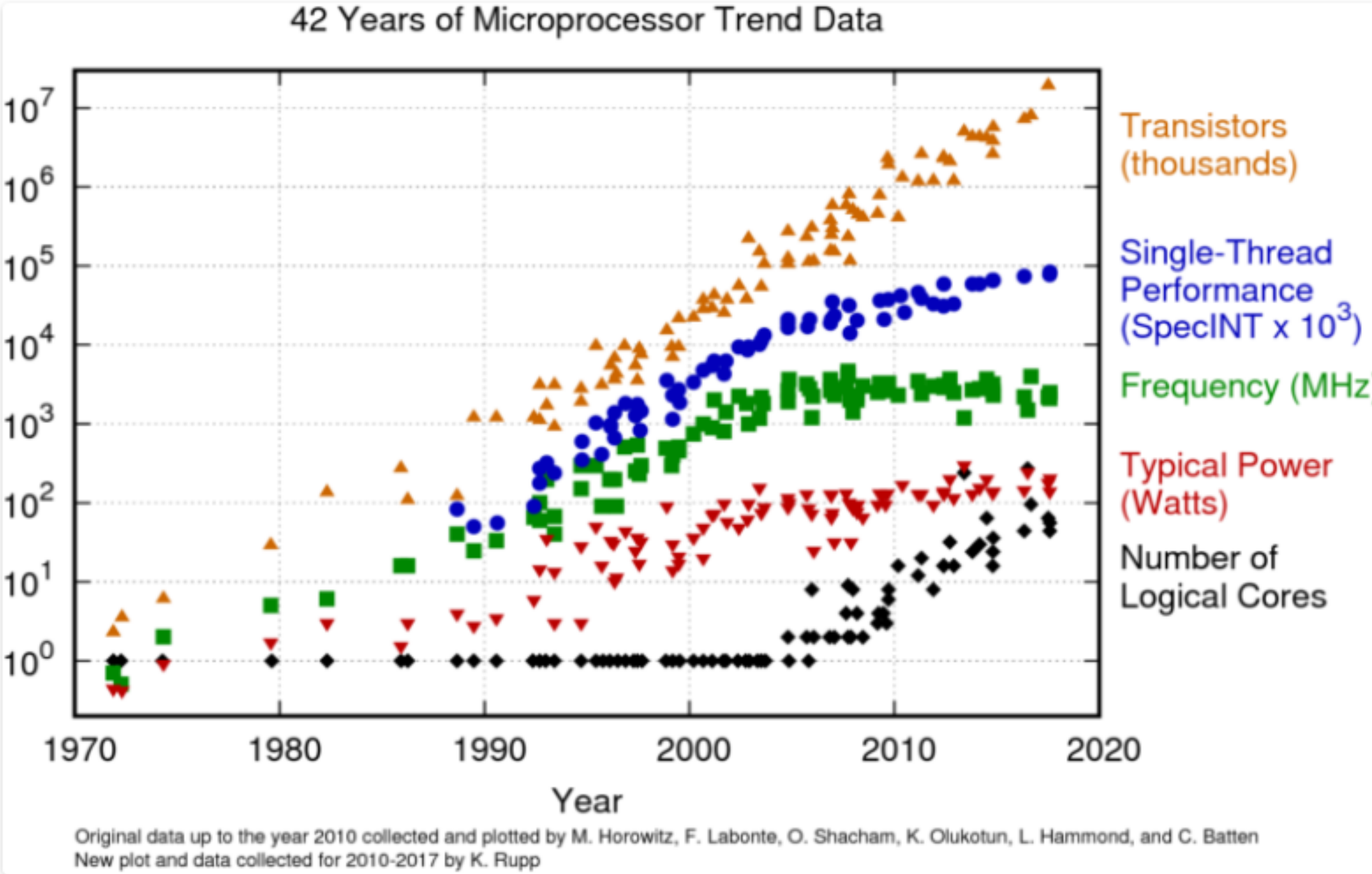
Scaling Methods

Table 3.8 Scaling scenarios for short-channel devices.

Parameter	**Relation**	**Constant Field**	**General Scaling**	**Constant Voltage**
W, L, t_{ox}		$1/S$	$1/S$	$1/S$
V_{DD}, V_T		$1/S$	$1/U$	1
N_{SUB}	$V/W_{depl}^{\;2}$	S	S^2/U	S^2
$Area$/Device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_{gate}	$C_{ox}WL$	$1/S$	$1/S$	$1/S$
k_n, k_p	$C_{ox}W/L$	S	S	S

$U<S$

42 Years of uP Trend Data

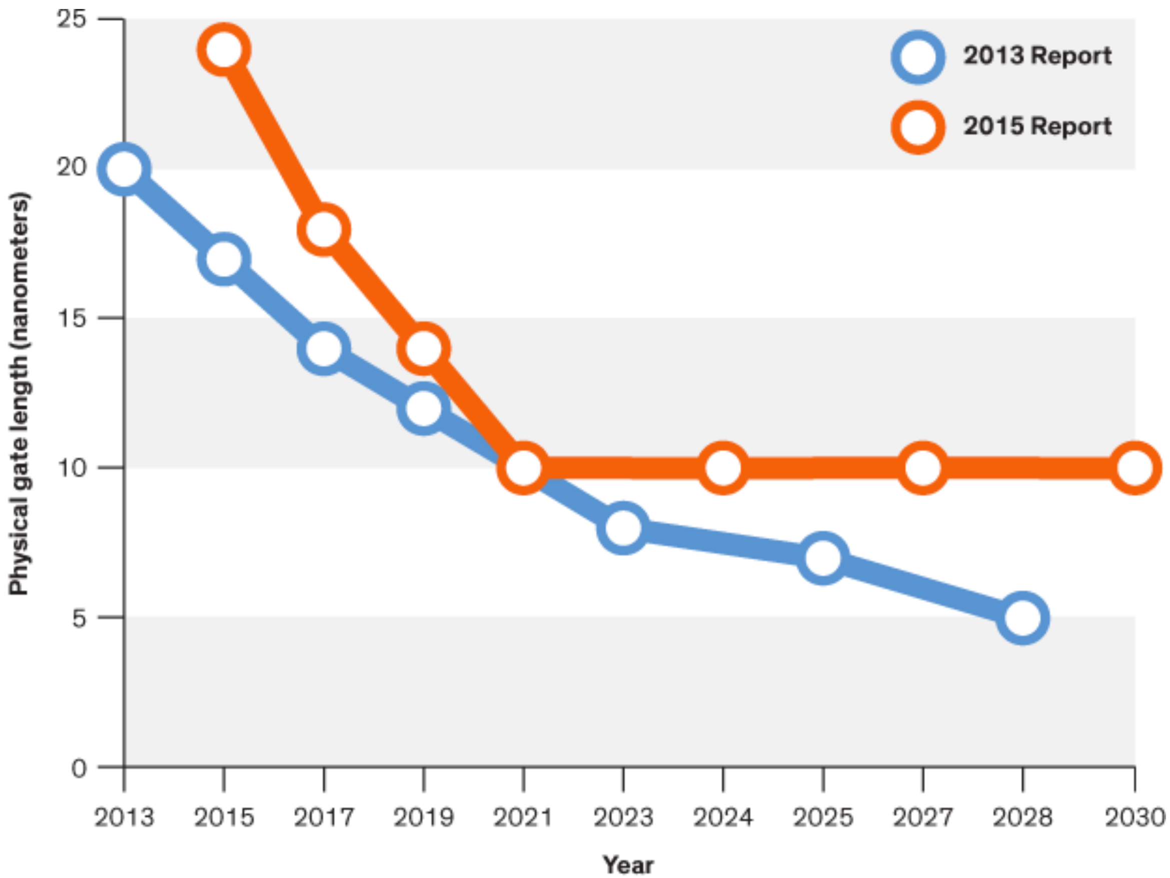


Conventional Scaling

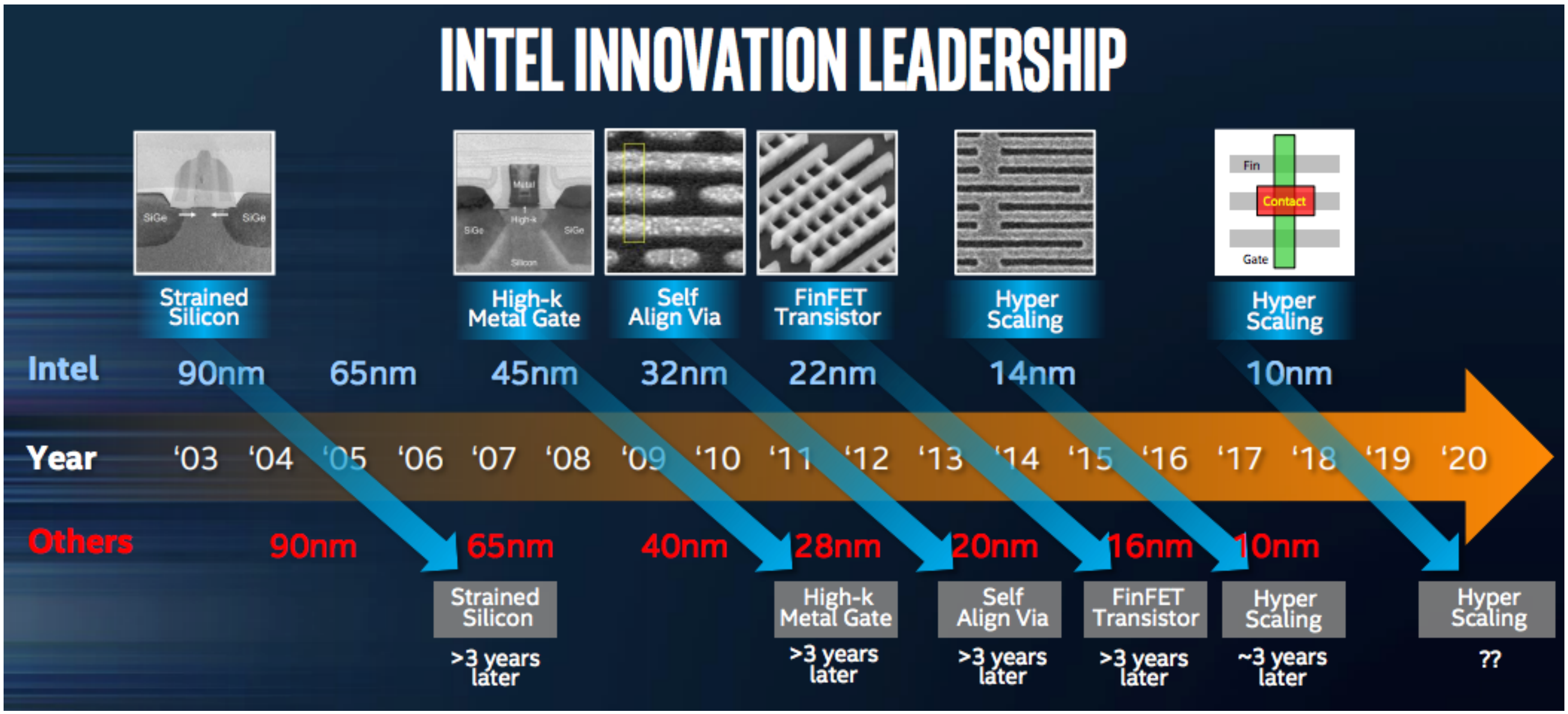
- Ends in your lifetime
- Perhaps already:
  - "Basically, this is the end of scaling."
    - May 2005, Bernard Meyerson, V.P. and chief technologist for IBM's systems and technology group

ITRS 2.0 Report 2015

- "After 2021, the report forecasts, it will no longer be economically desirable for companies to continue traditional transistor miniaturization in microprocessors."

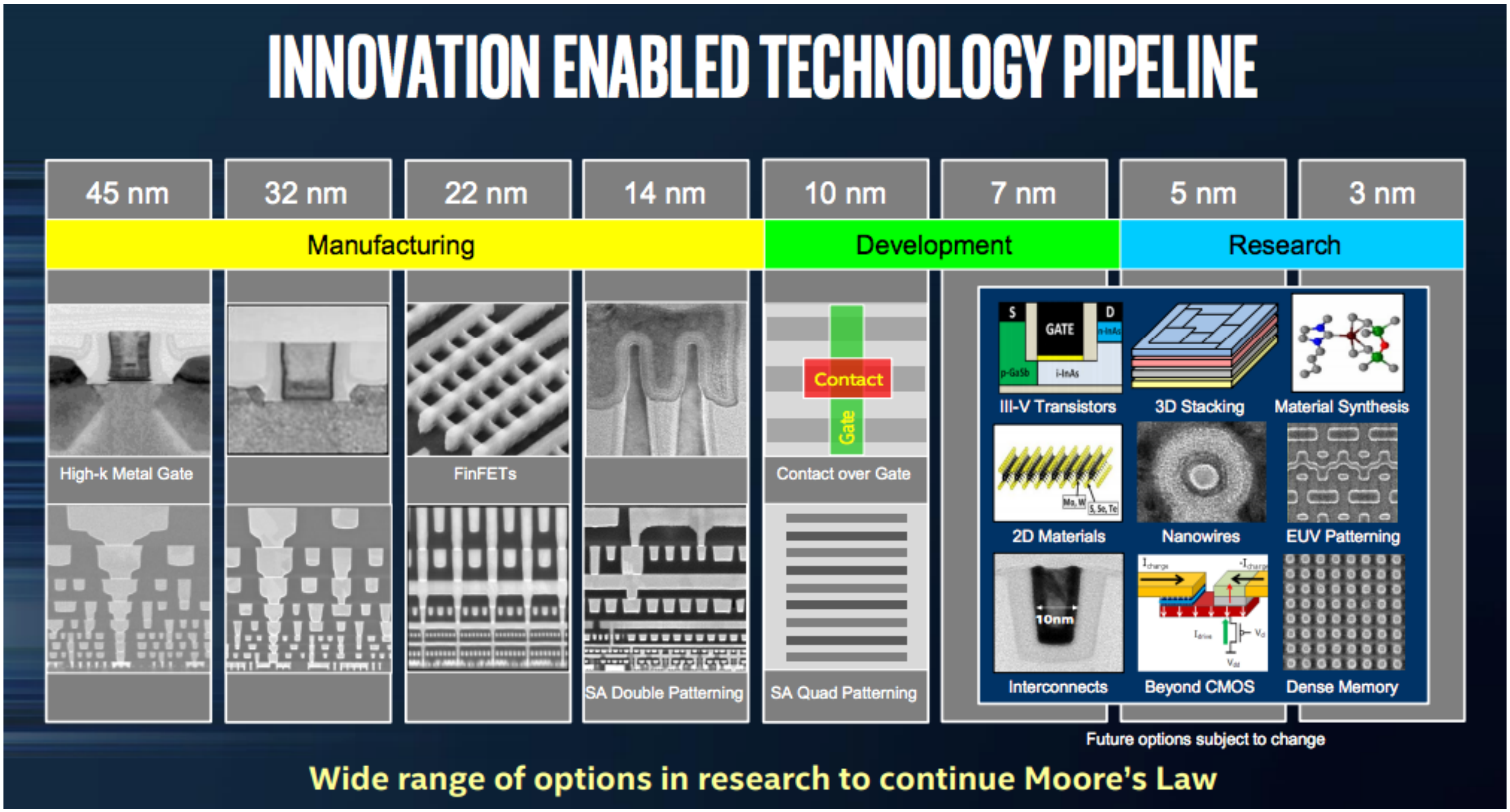


BUT…



Source:https://newsroom.intel.com/newsroom/wp-content/uploads/sites/11/2017/09/mark-bohr-on-continuing-moores-law.pdf

BUT…



Source:https://newsroom.intel.com/newsroom/wp-content/uploads/sites/11/2017/09/mark-bohr-on-continuing-moores-law.pdf

Previously

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

Today

- τ-model
- Sizing
- Large Fanout
- Capacitance Revisited
  - Miller Effect

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?

$$C_G \propto C_{ox} W L$$

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

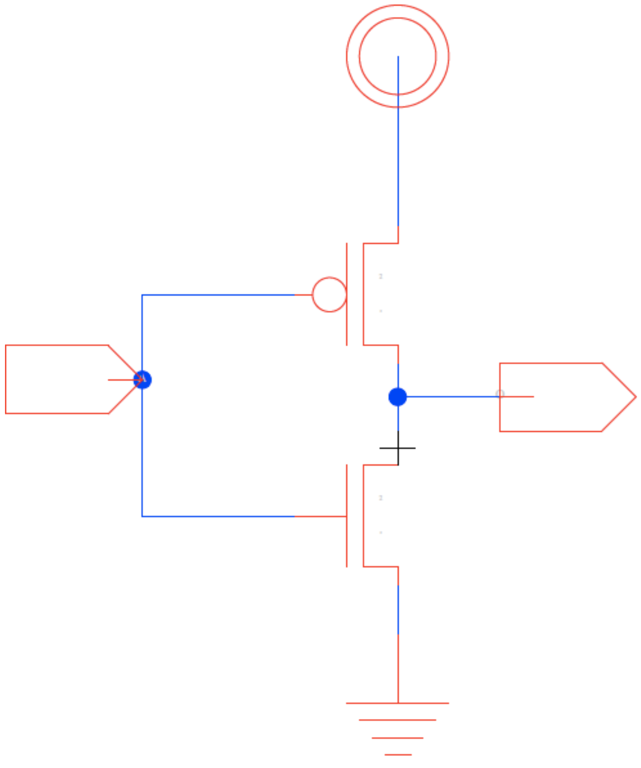
- $R_{drive} = R_0/W$
- $C_g = WC_0$

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)

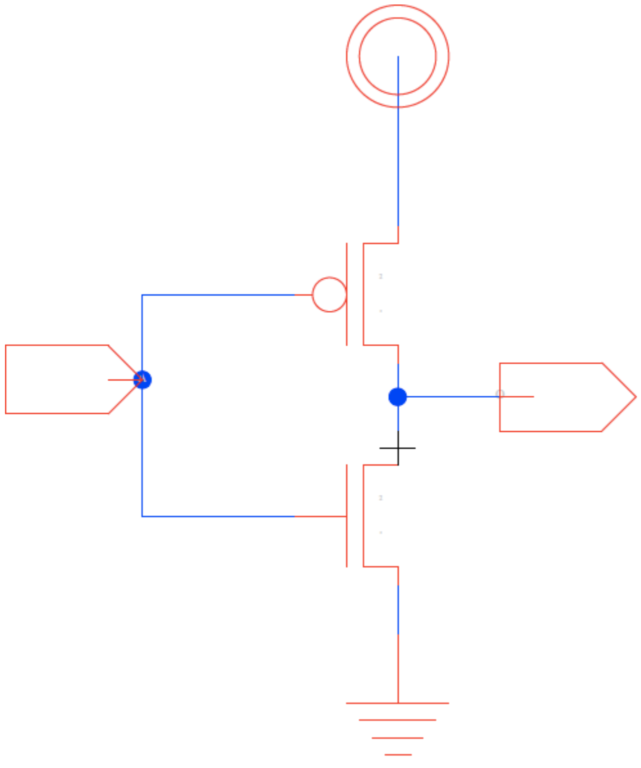
- How should we size to equalize Rise and Fall?
  - Given:
  - $\mu_n = 500 cm^2/Vs$, $\mu_p = 200 cm^2/Vs$
  - $R_{drive} = R_0/2$ ($I_{drive} = 2I_0$)



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

How to Size Transistors (preclass 2)

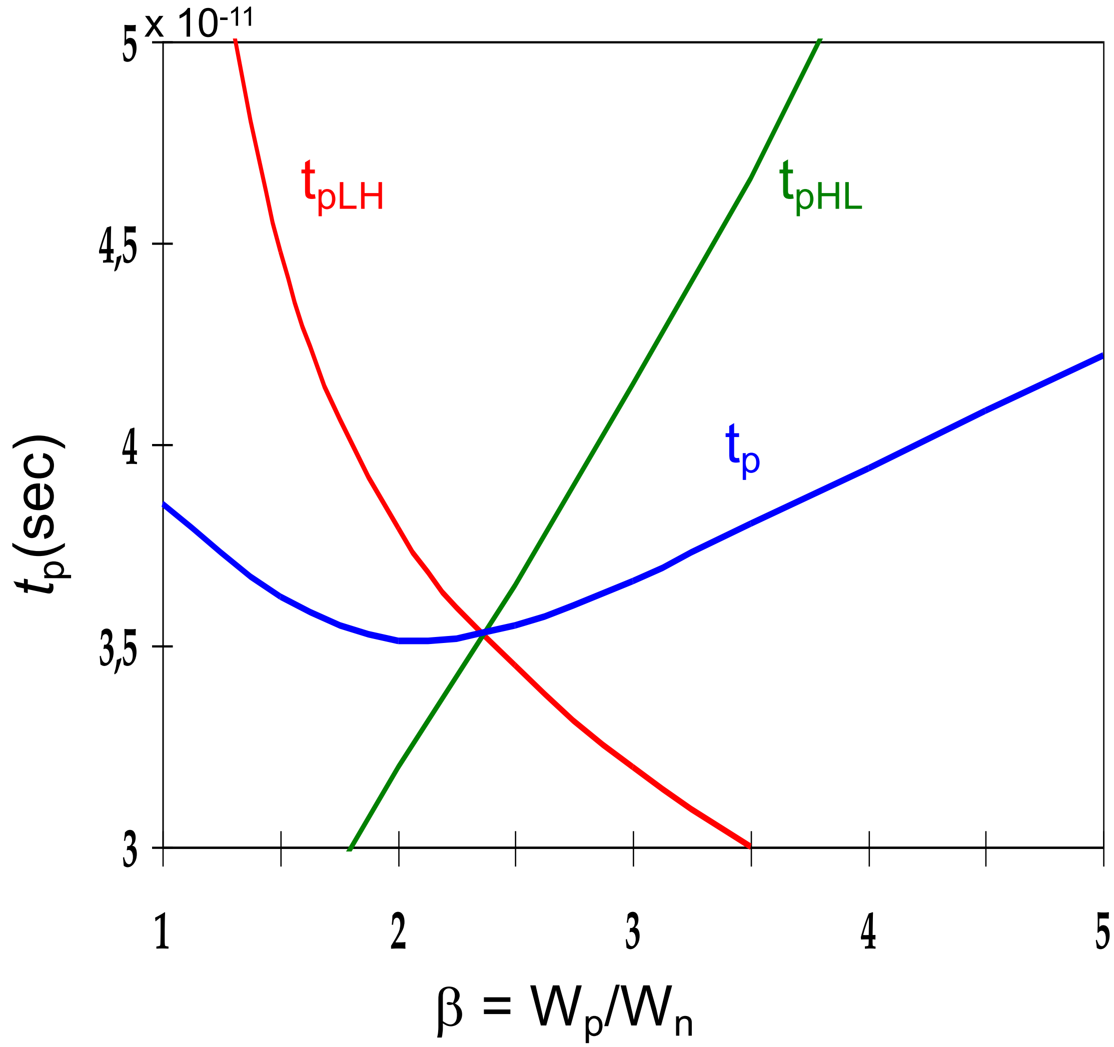
- How should we size to equalize Rise and Fall?
  - Given:
  - μ_n=500cm^2/Vs, μ_p=200cm^2/Vs
  - $R_{drive}=R_0/2$ ($I_{drive}=2I_0$)



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

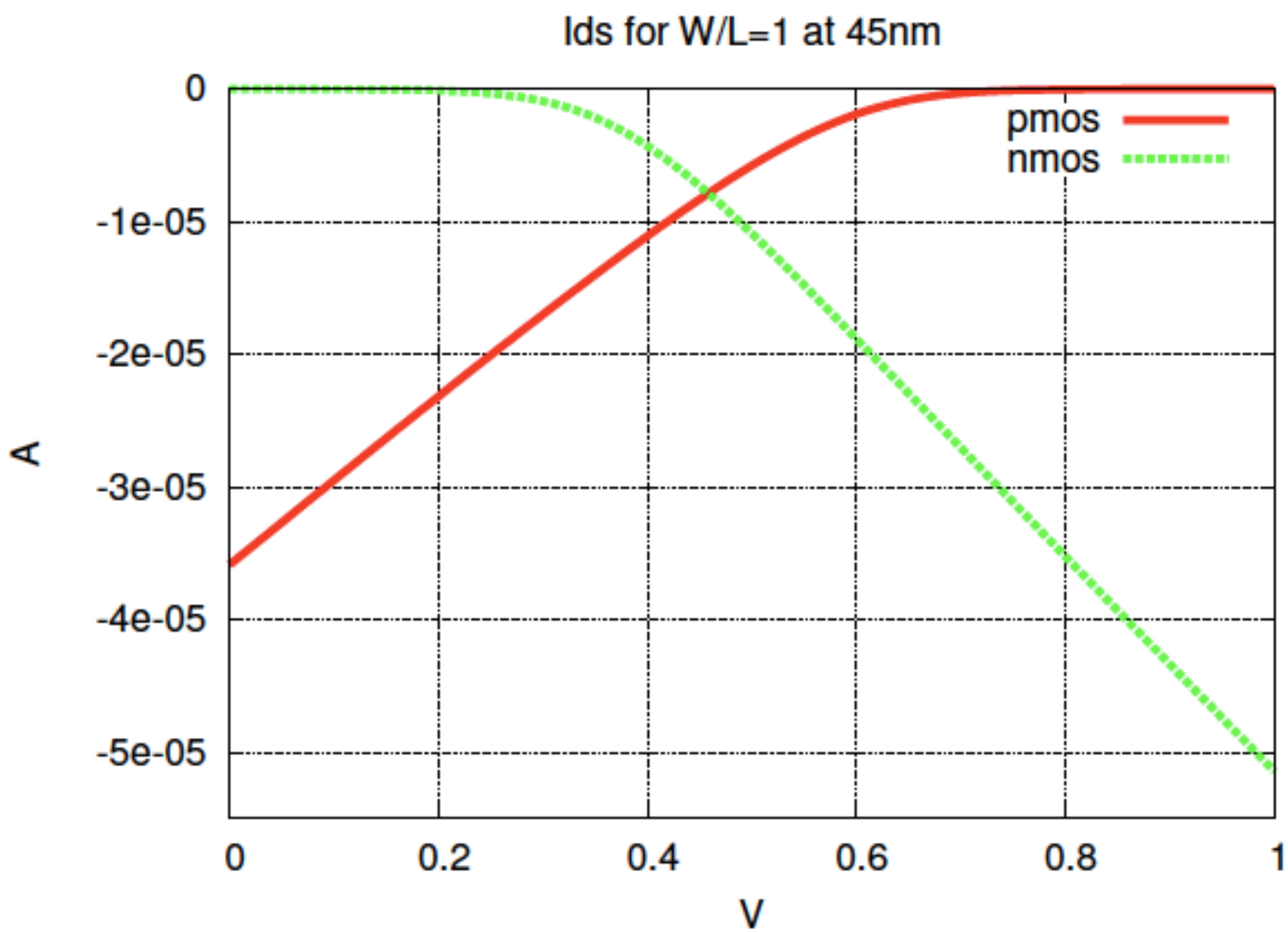
  - What is input capacitance for sized devices?

Size Transistors – Minimum Avg Delay

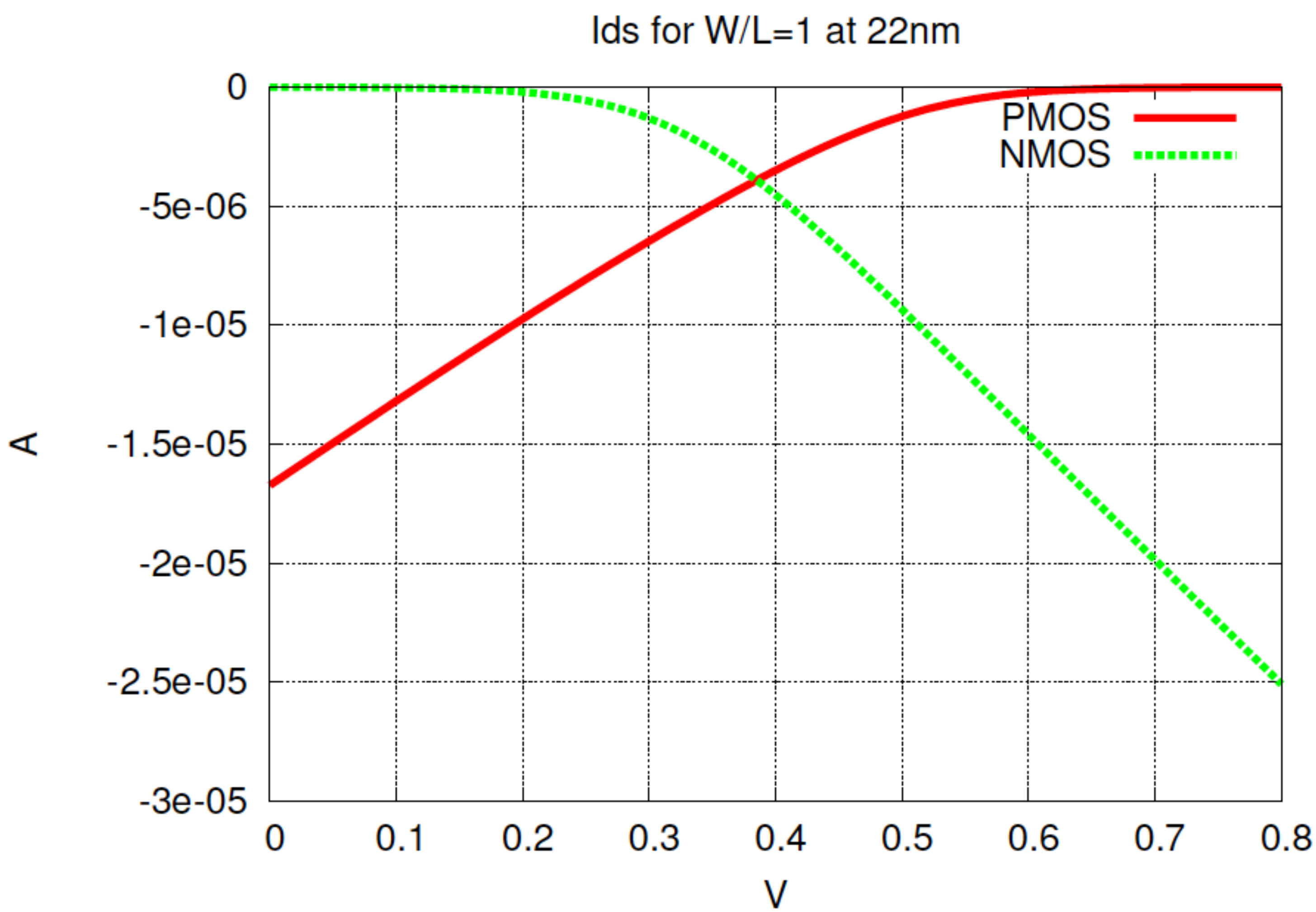


- β of 2.4 (= 31 kΩ/13 kΩ) gives symmetrical response
- β of 1.6 to 1.9 gives optimal performance

SPICE Simulation 45nm



SPICE Simulation 22nm

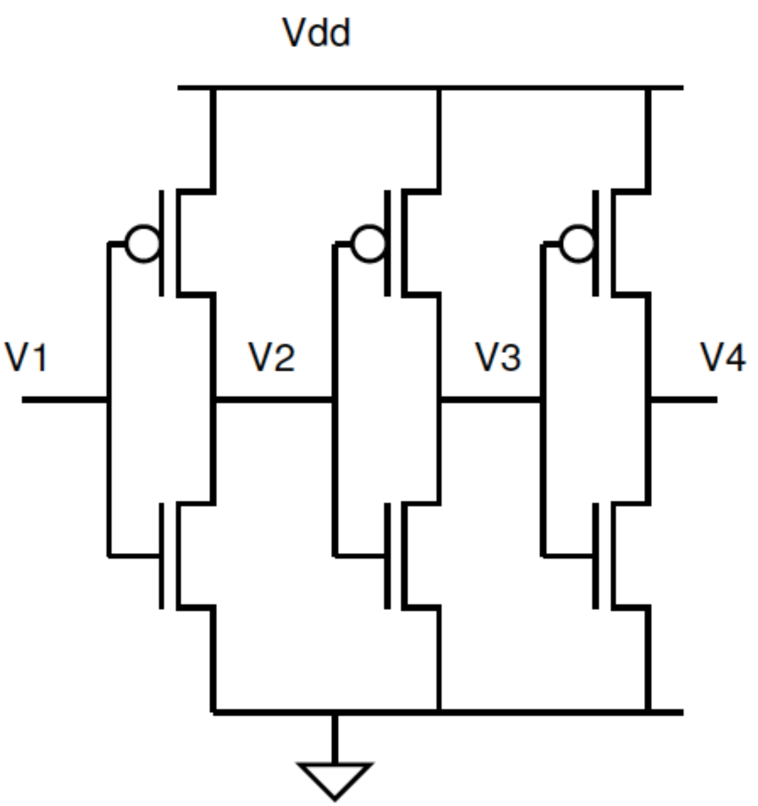


Equalizing Delay

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

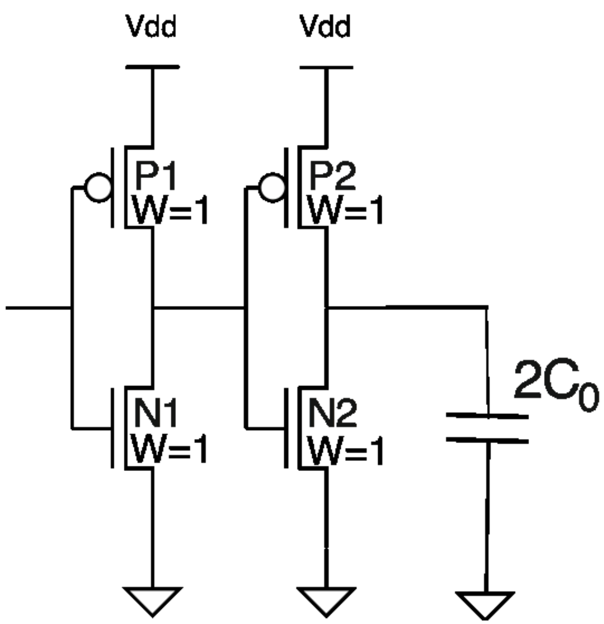
Inverter Sizing

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



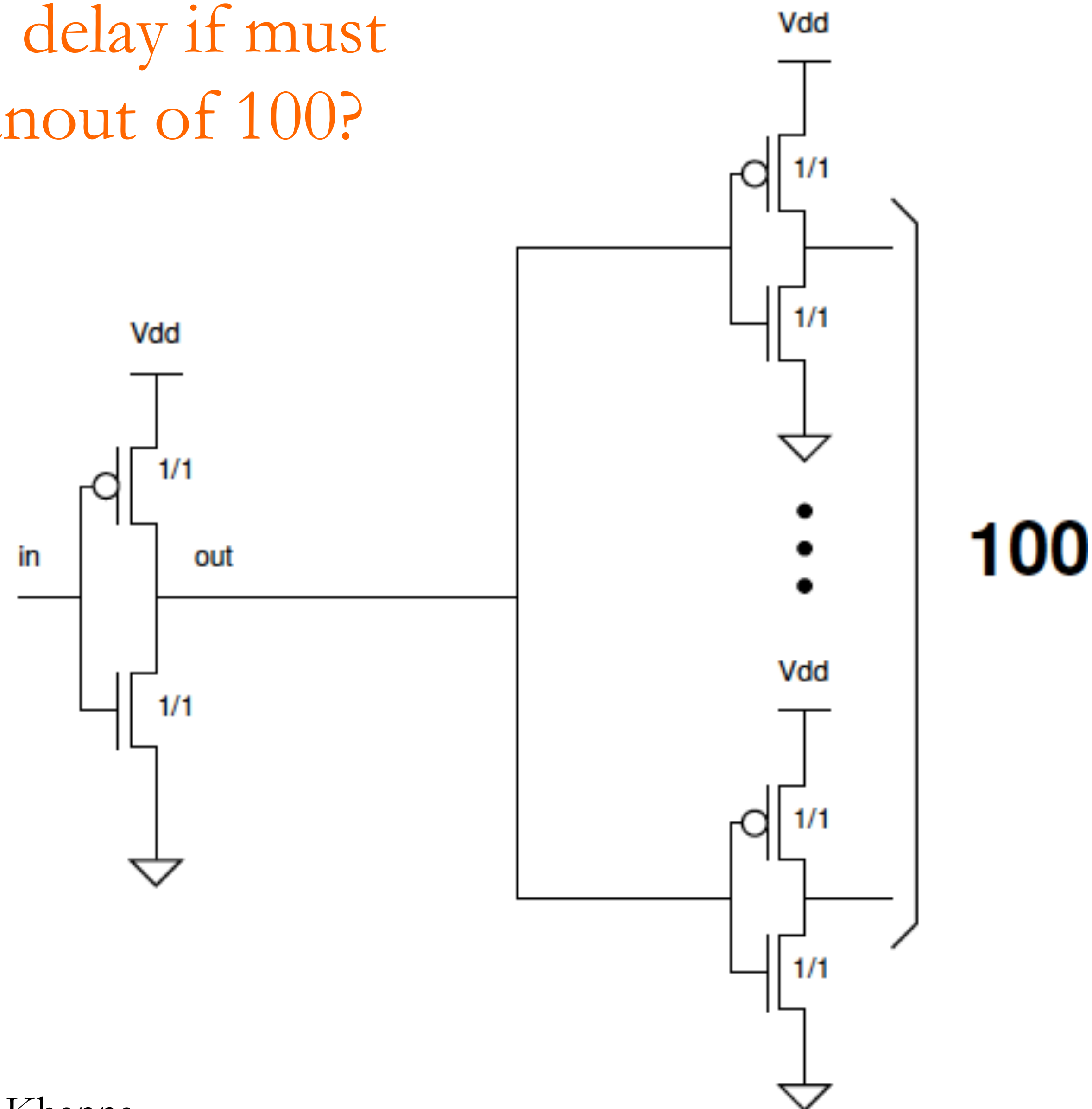
Multistage Delay

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



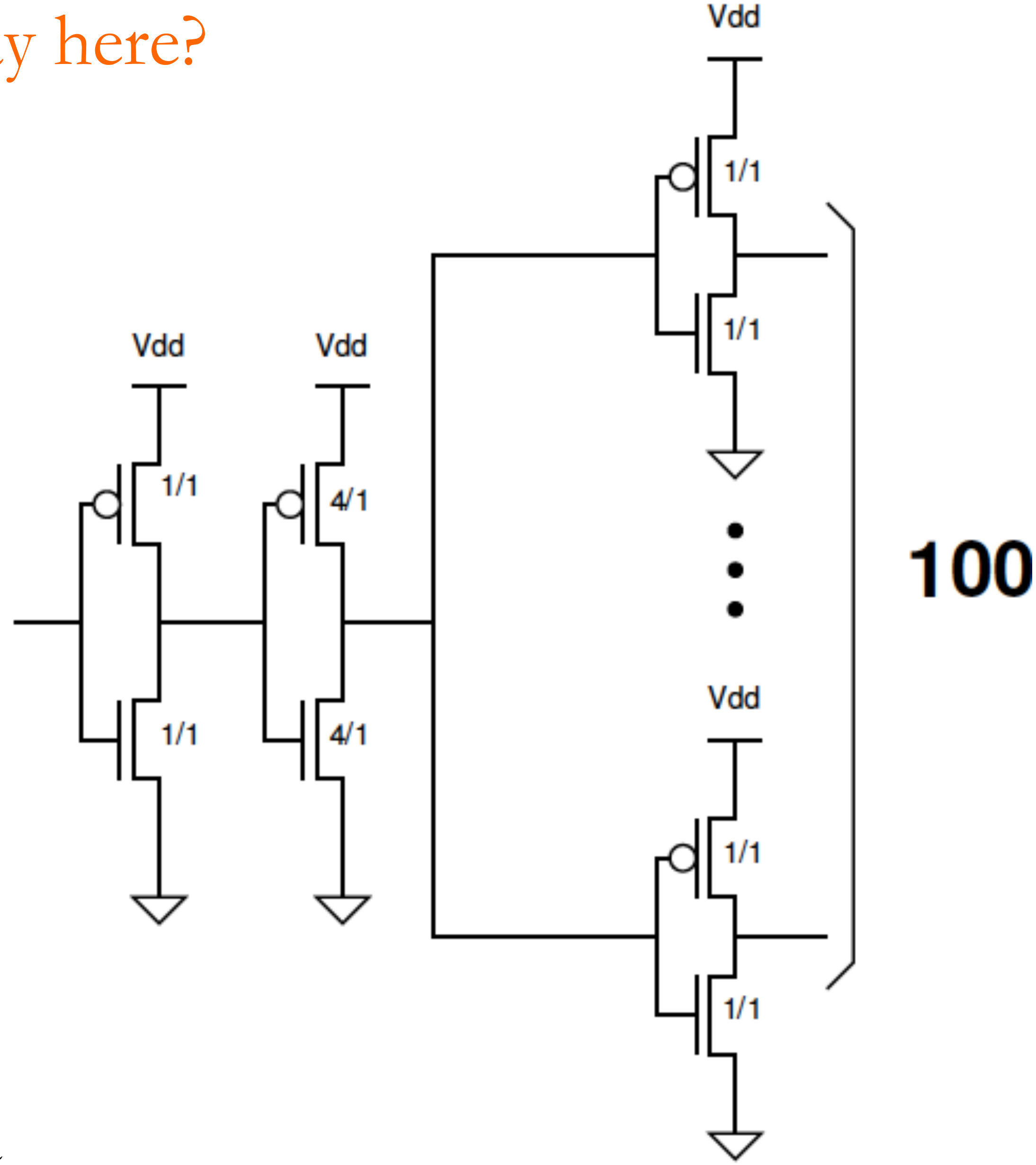
Large Fanout Delay (preclass 3)

- What is delay if must drive fanout of 100?



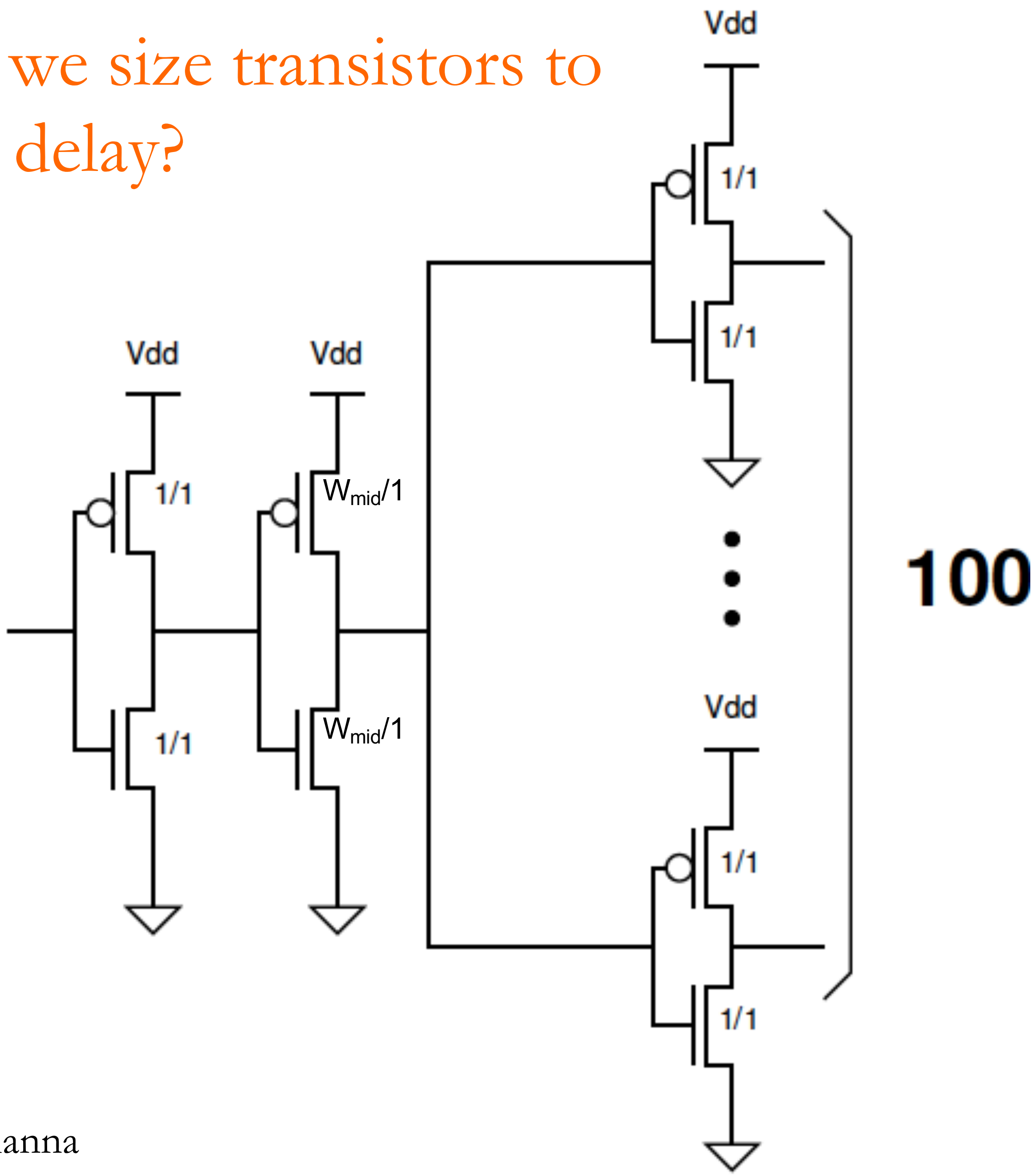
Graduated Fanout Delay (preclass 3)

- What is delay here?



Optimize Fanout Delay (preclass 4)

- How can we size transistors to minimize delay?



Optimizing (preclass 4)

- Derivate to minimize

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

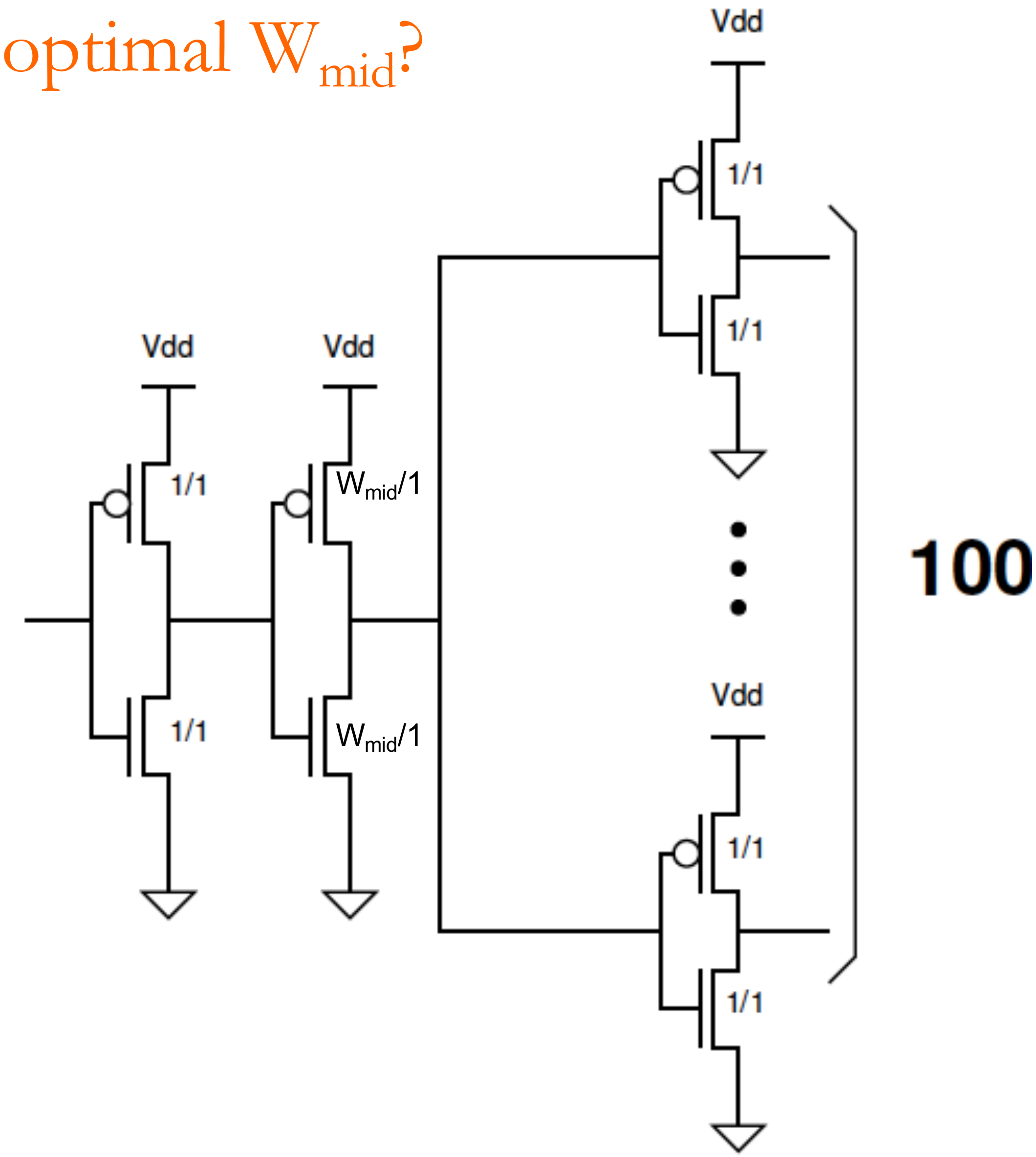
$$\frac{\partial \tau_{est}}{\partial W_{mid}} = 0$$

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

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

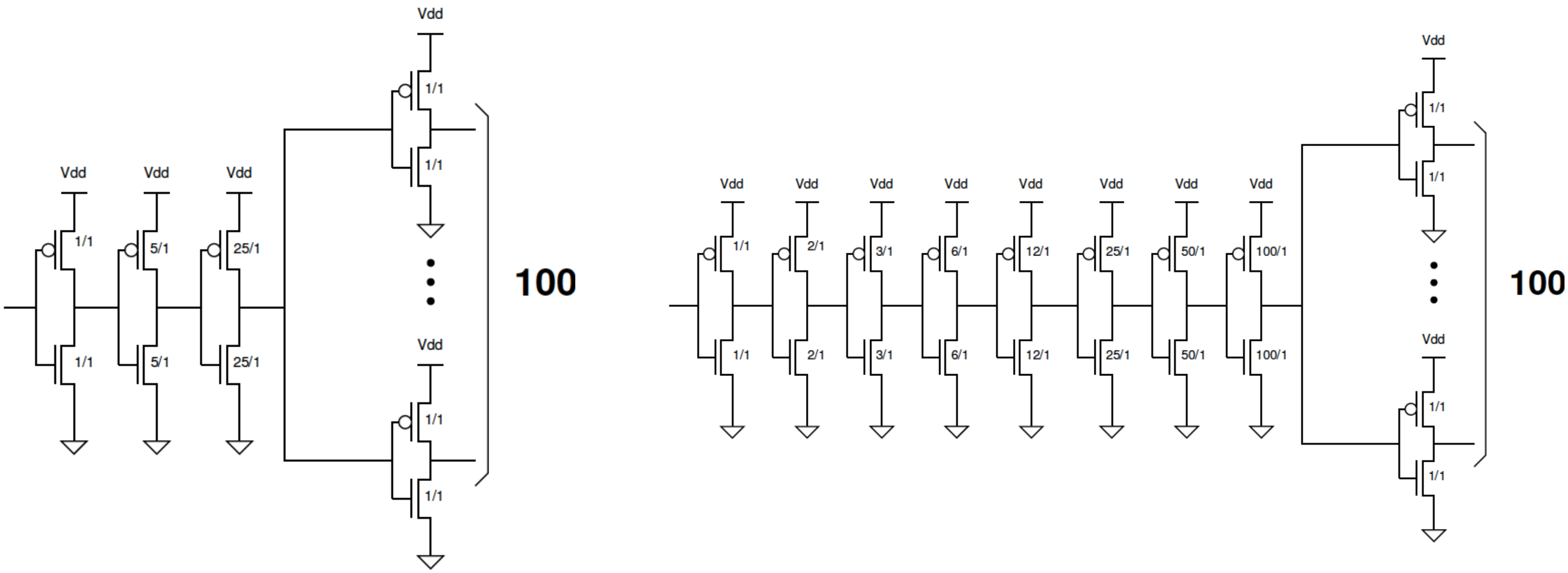
Delay? (preclass 4)

- Delay at optimal W_{mid}?



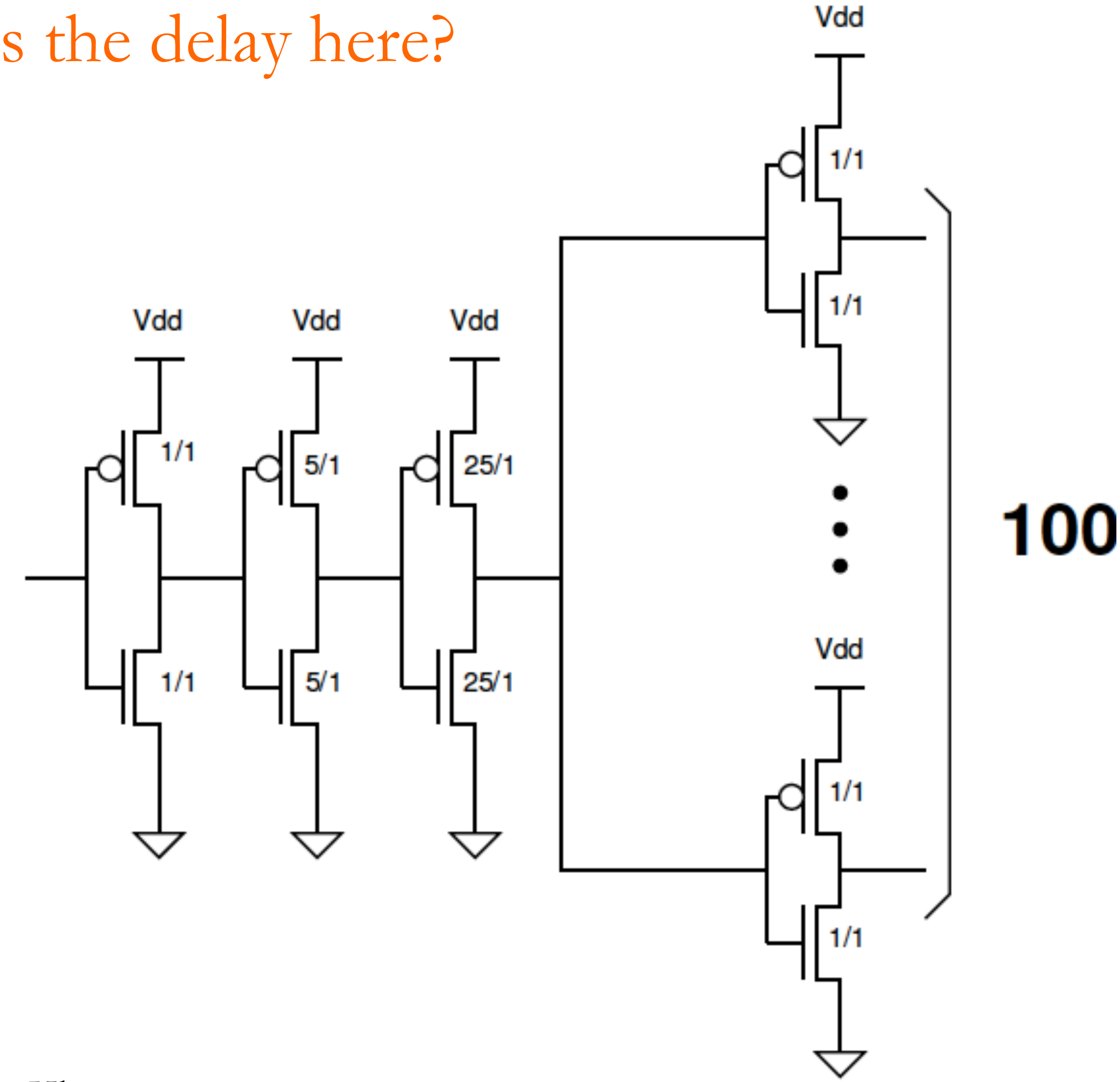
Try Again with More Stages (preclass 5)

- What is the delay here?



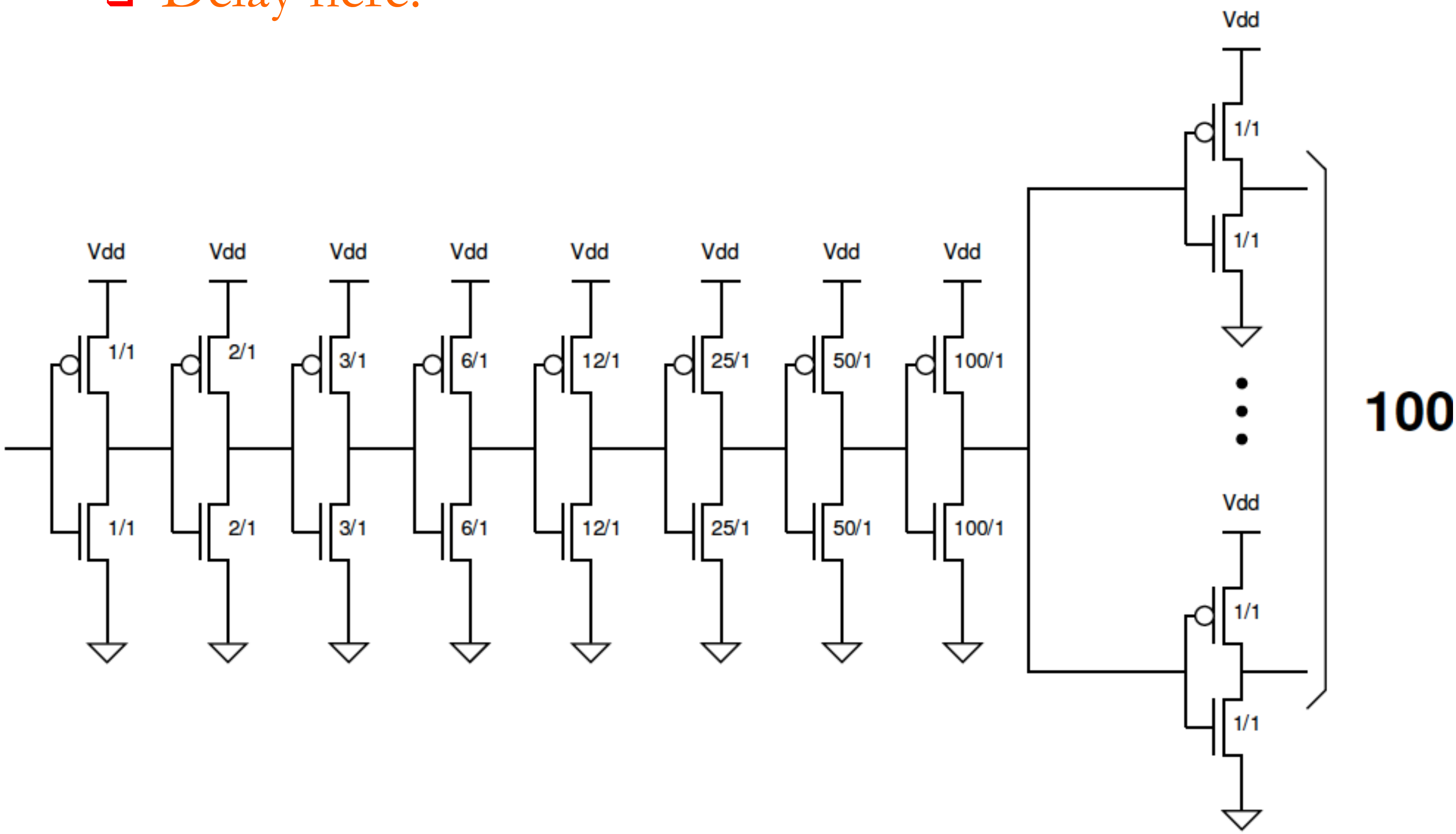
Try Again with More Stages (preclass 5)

- What is the delay here?

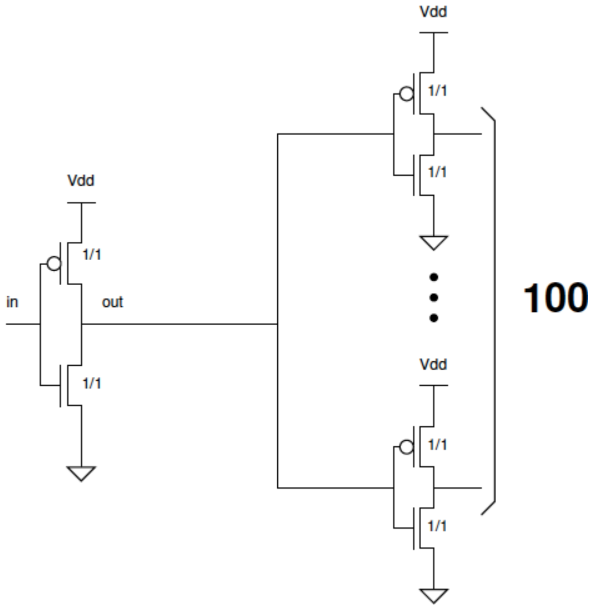


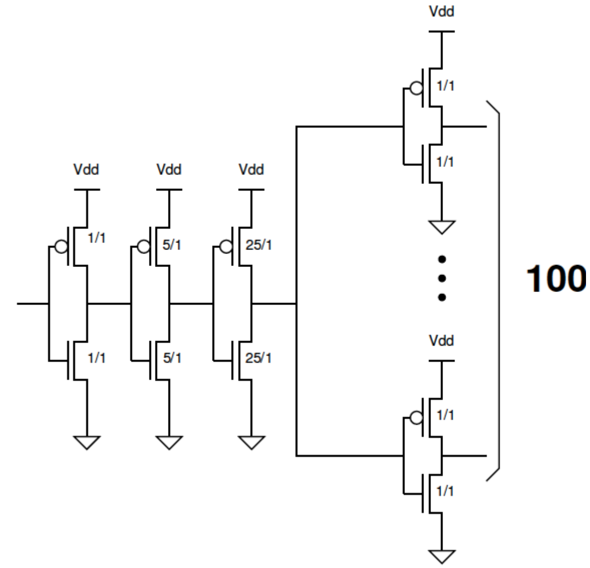
…and Again (preclass 5)

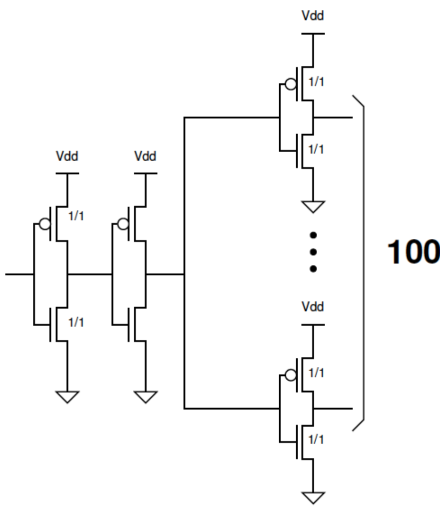
- Delay here?

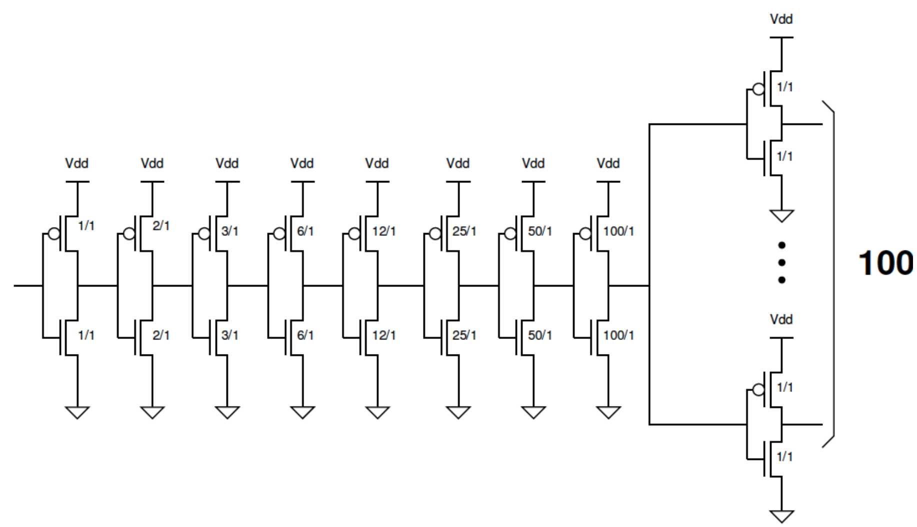


Delay Summary





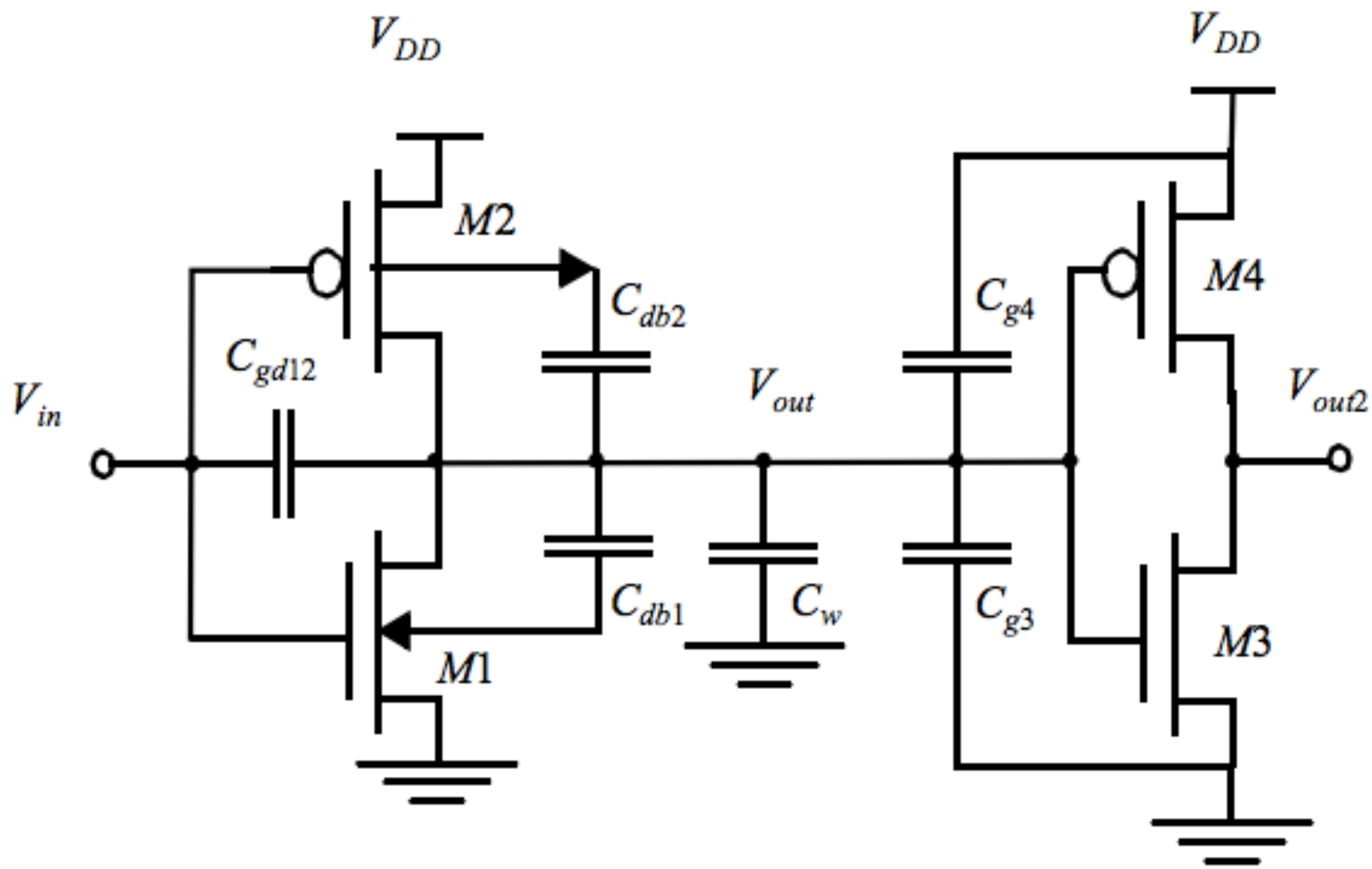




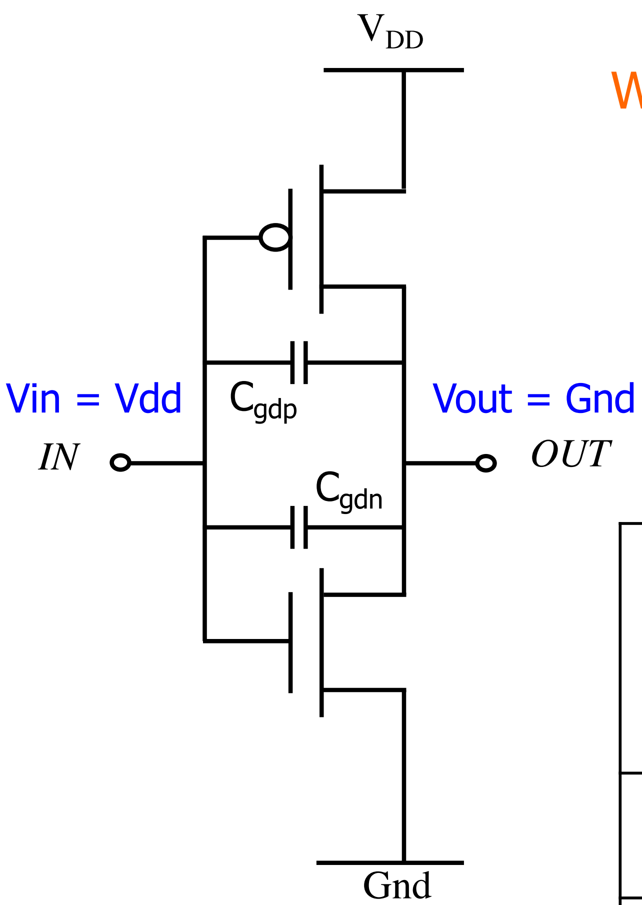
Lesson

- 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

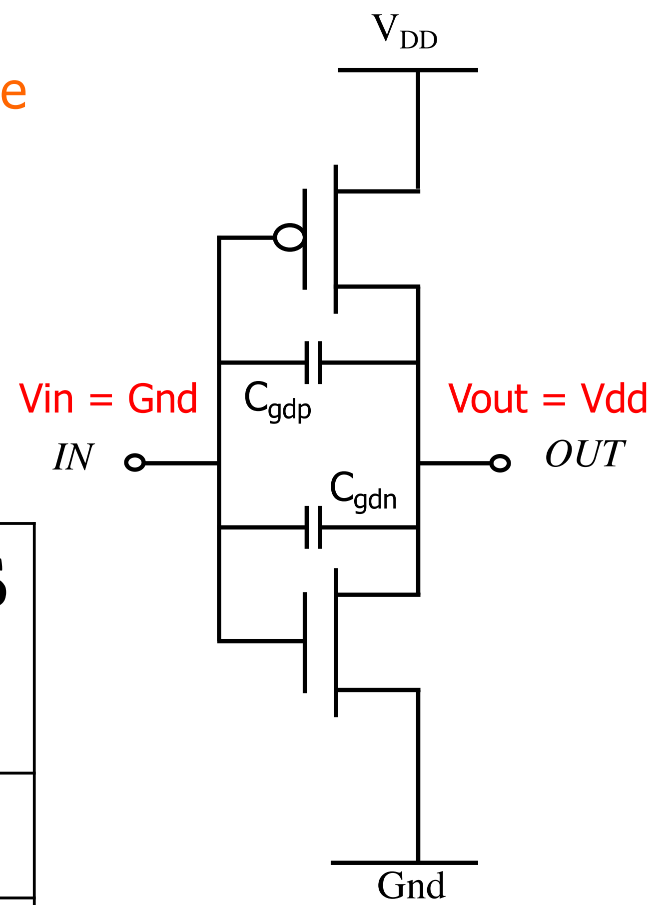
Capacitance Reminder



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		

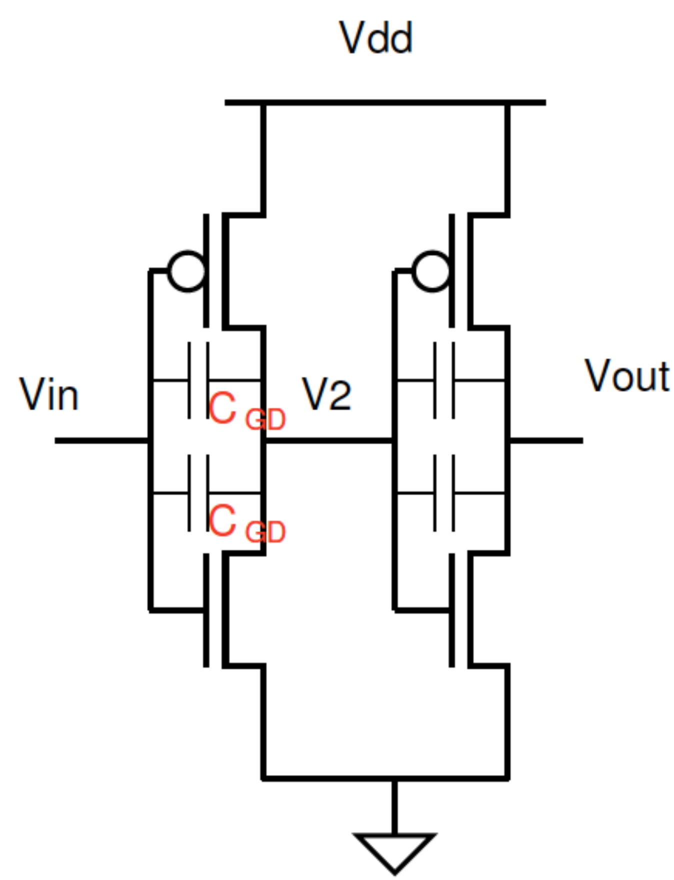
Questions

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

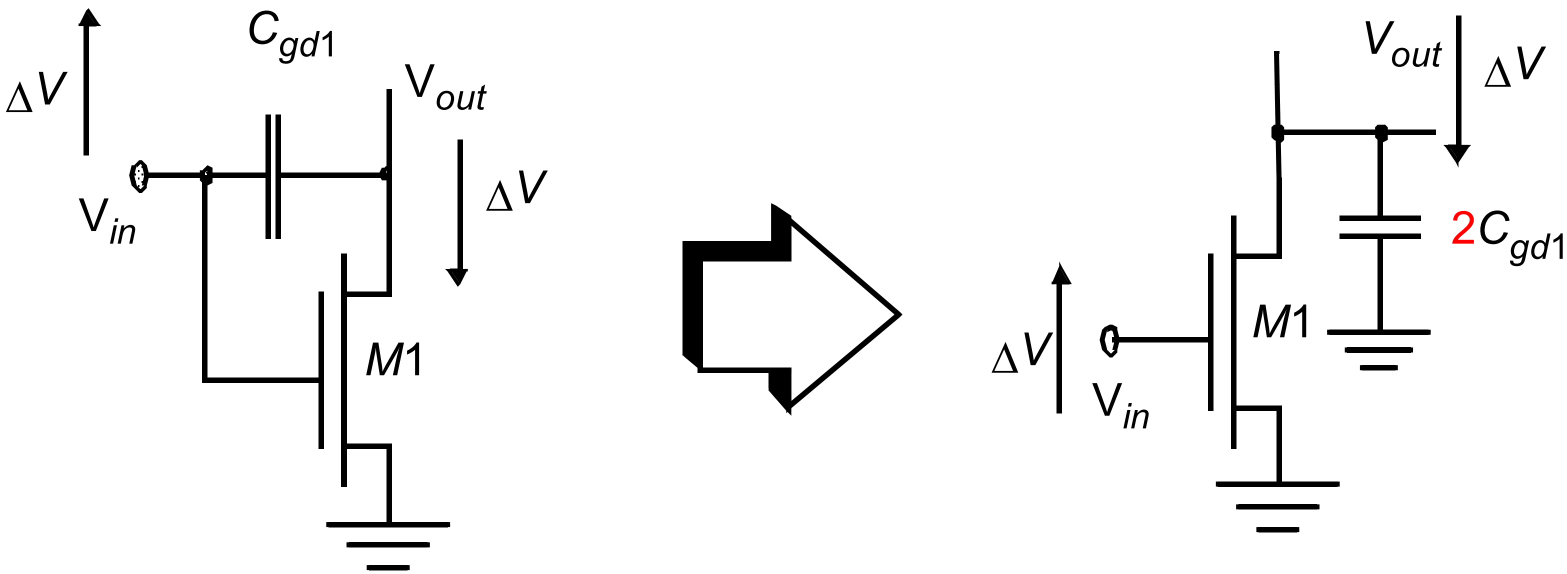
Vin	**PMOS** C_{gdp}	**NMOS** C_{gdn}
Vdd		
Gnd		

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."

Ideas

- First order delay reason in $\tau = R_0C_0$ units
- Scaling everything up doesn't help
- Drive large capacitive loads in stages
  - But not too many

Admin

- HW4 due Friday

- Drop date is on Monday
  - Email me if you want to talk about your standing in the class