

EECS151/251A Introduction to Digital Design and ICs

Lecture 16: Inverter Chain Delay Sophia Shao



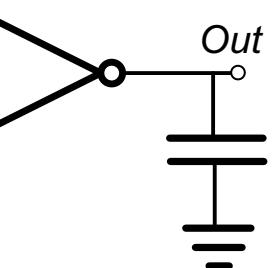
AMD Athlon: First 1GHz CPU in 2000!

The Athlon's arrival signaled the opening salvos in what was coined 'The Gigahertz War'. The Pentium III had a lead role in the 'Gigahertz War' against AMD's Athlon processors between 1999 and 2000. Ultimately it was AMD who crossed the finish line first, shipping the 1GHz Athlon days before Intel could launch theirs.

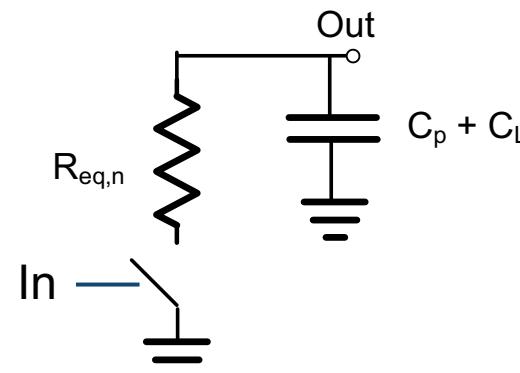


Review

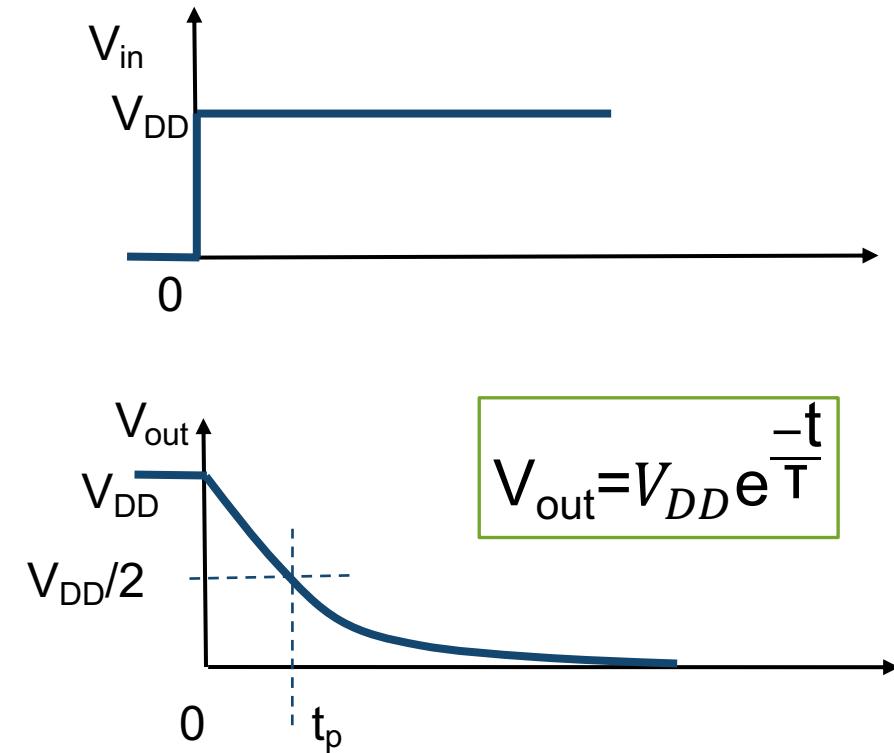
- CMOS Transistors and Gates
 - MOS transistor as a switch
 - Pull-up and Pull-down for CMOS design
 - CMOS Gates
- Inverter Delay
 - Delay affects achievable frequency.
 - Propagation delay from input to output
 - RC Delay model



EECS151 L15 INVERTER CHAIN DELAY



Shao Fall 2022 © UCB



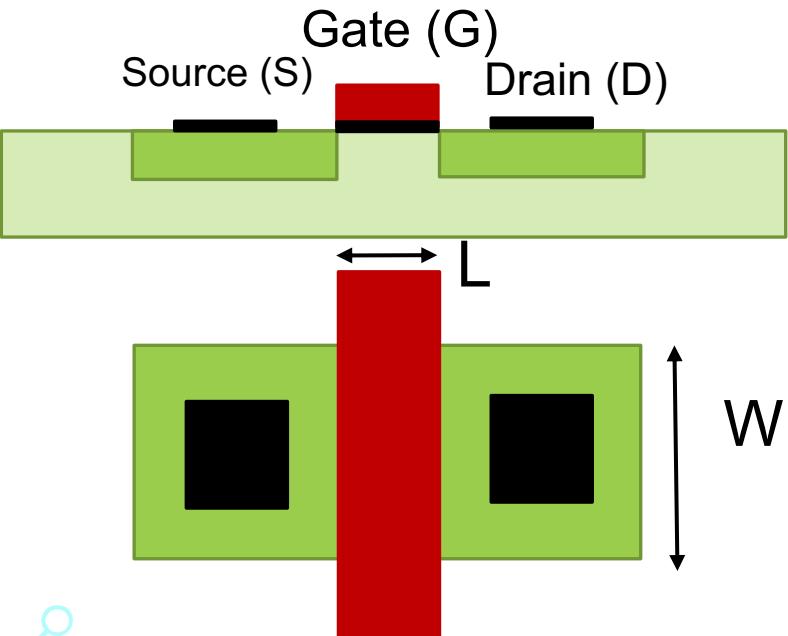
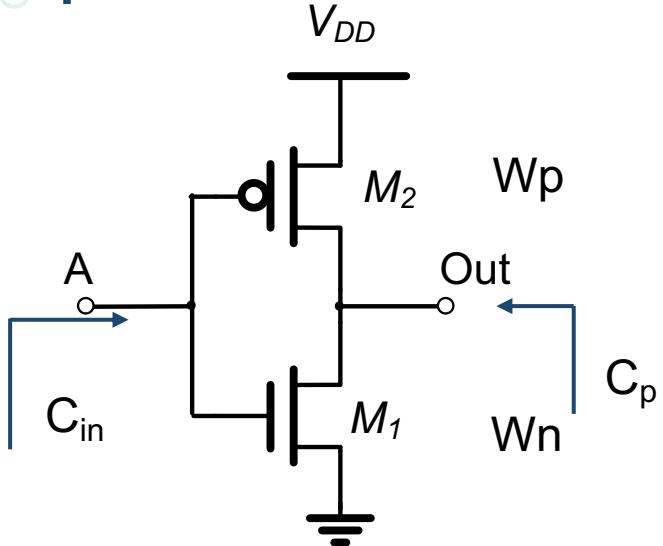
$$\tau = R_{eq,n}(C_p + C_L)$$

$$t_{p,HL} = (\ln 2) \tau = 0.7 R_{eq,n}(C_p + C_L)$$



- **Inverter Delay**
 - Inverter RC Components
 - Inverter Size
 - Inverter RC Delay Recap
- Inverter Chain
 - Path Delay
 - Minimize Inverter Chain Delay
- Generalize to Arbitrary Gates
 - Logical Effort

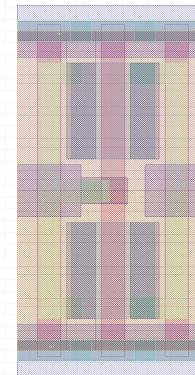
Capacitances



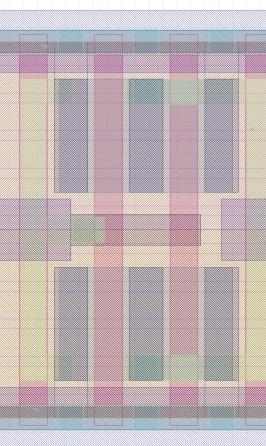
EECS151 L15 INVERTER CHAIN DELAY

4

INVX1



INVX2



- C_{in} (input capacitance)

- Largely set by the gate cap $C_g \sim WL$
- Inverter: $C_{in} = C_{g,pmos} + C_{g,nmos}$
- ($W \rightarrow 2W$) $\Rightarrow 2x C_{in}$

- C_p (parasitic capacitance)

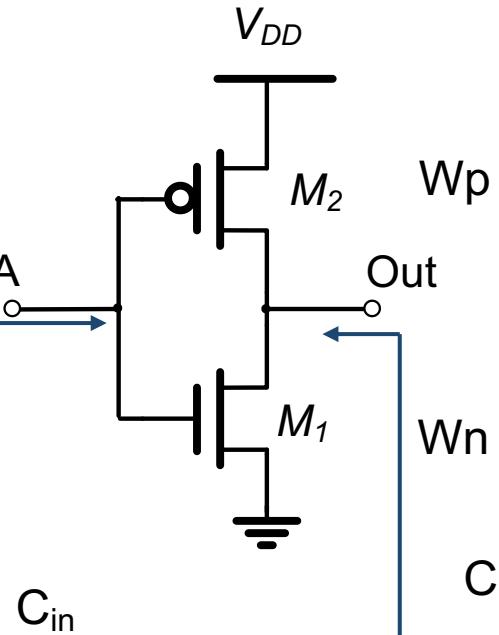
- Largely set by the drain cap $C_d \sim W$ (drain area/perimeter)
- Inverter: $C_p = C_{d,pmos} + C_{d,nmos}$
- ($W \rightarrow 2W$) $\Rightarrow 2x C_p$

- $C_d = \gamma C_g (\gamma = 1)$

- Inverter: $C_p = \gamma C_{in}$

Inverter Sizing

- A size “2” inverter:
 - $W_{p,inv_x2} = 2 * W_{p,inv_x1}$
 - $W_{n,inv_x2} = 2 * W_{n,inv_x1}$
- Doubling the gate size (by doubling W_s):

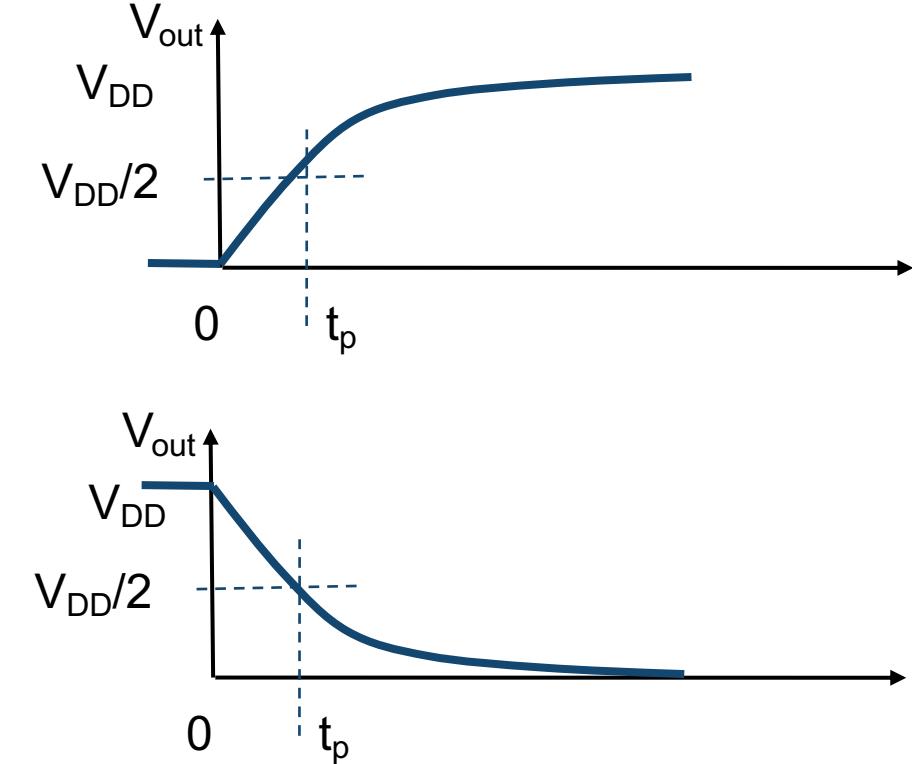
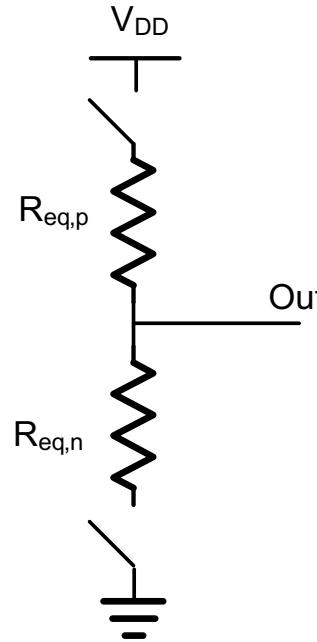
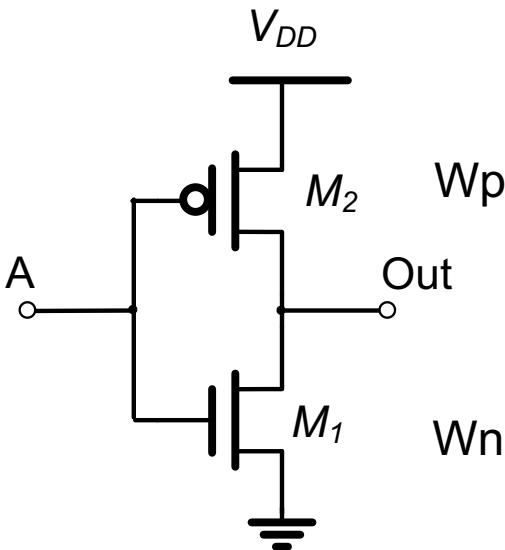


- Doubles C_{in}
- Doubles C_p
- Halves equivalent gate resistance
 - Delivers 2x current to flow

$$R = \rho \frac{L}{TW}$$

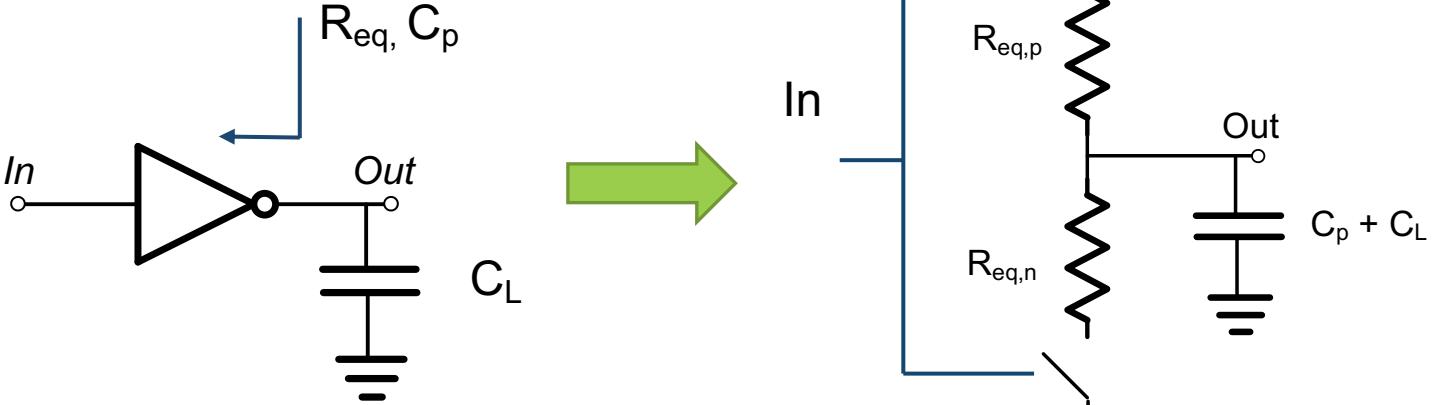
Inverter W_p vs W_n

- Optimal W_p/W_n to have the same low-to-high and high-to-low delays.



- In the past, to have equal resistance, $\rho_p > \rho_n$, $W_p > W_n$ $R = \rho \frac{L}{TW}$
- In modern processes (FinFET), $\rho_p = \rho_n$, $W_p = W_n$

Inverter RC Delay



- $t_p = \ln 2 * R_{eq}(C_p + C_L)$
 - Parasitic/Intrinsic delay: $\ln 2 * R_{eq}C_p$
- $t_p = \ln 2 * R_{eq} (\gamma C_{in} + C_L) = \ln 2 * R_{eq}C_{in}(\gamma + C_L/C_{in})$
 - $\gamma = 1$
- $t_p = \tau_{INV}(1+f)$
 - $\tau_{INV} = \ln 2 * R_{eq}C_{in}$
 - **τ_{INV} is independent of transistor sizes.**
 - **Fanout = $f = C_L/C_{in}$**

Normalized Delay to τ_{INV}
 $D(\text{inv}) = t_p / \tau_{INV} = 1 + f$

Parasitic delay

Effort delay

Quiz: Inverter Delay

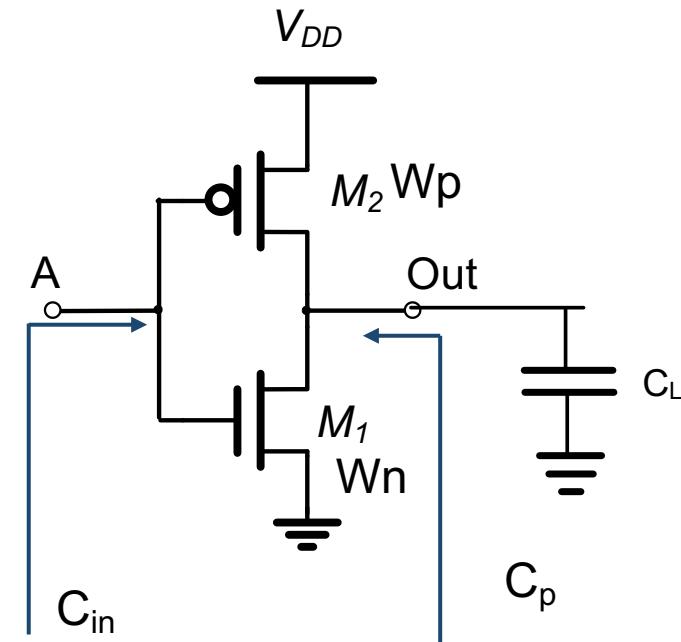
- How does the inverter delay change if the widths of both PMOS and NMOS of the inverter are doubled?

- $t_p = \ln 2 * R_{eq} C_{in} (1 + C_L / C_{in}) = \tau_{INV}(1+f)$
- $f = \text{fanout} = C_L / C_{in}$

A: $t_{p,inv_x2} > t_{p,inv_x1}$

B: $t_{p,inv_x2} < t_{p,inv_x1}$

C: $t_{p,inv_x2} = t_{p,inv_x1}$



Administrivia

- Lab 6 starts this week.
- Homework 5 out today.
- Project release next week.

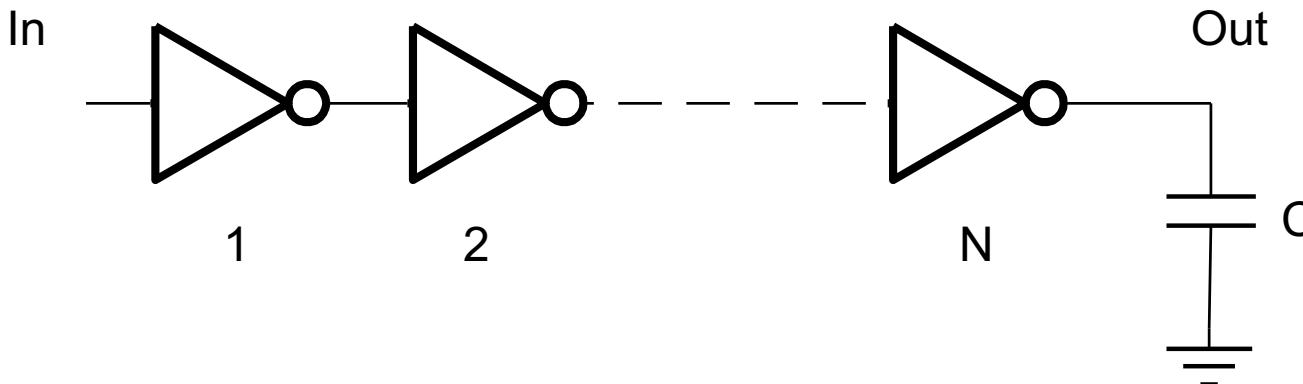


EECS151 L15 INVERTER CHAIN DELAY

10

- Inverter Delay
 - Inverter RC Components
 - Inverter Size
 - Inverter RC Delay Recap
- Inverter Chain
 - Path Delay
 - **Minimize Inverter Chain Delay**
- Generalize to Arbitrary Gates
 - Logical Effort

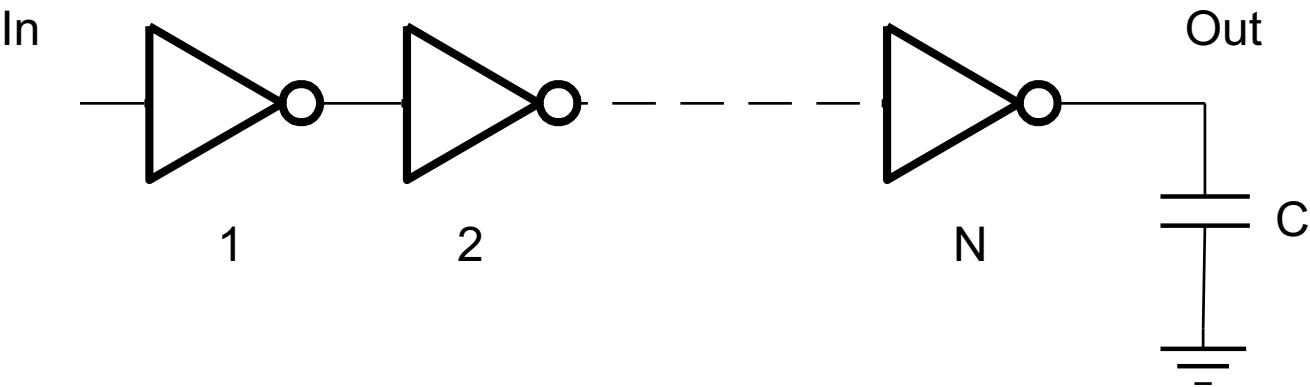
Inverter Chain



$$D = 1 + f$$

- $C_{L,i} = C_{in,i+1}$ (the load of the i-th stage is the input capacitance of the next stage.)
- N: # of inverters in the chains
- $C_{in,1} = 1$
- Given N and C_L , how to size each inverter in the chain to achieve minimum delay?

Inverter Chain



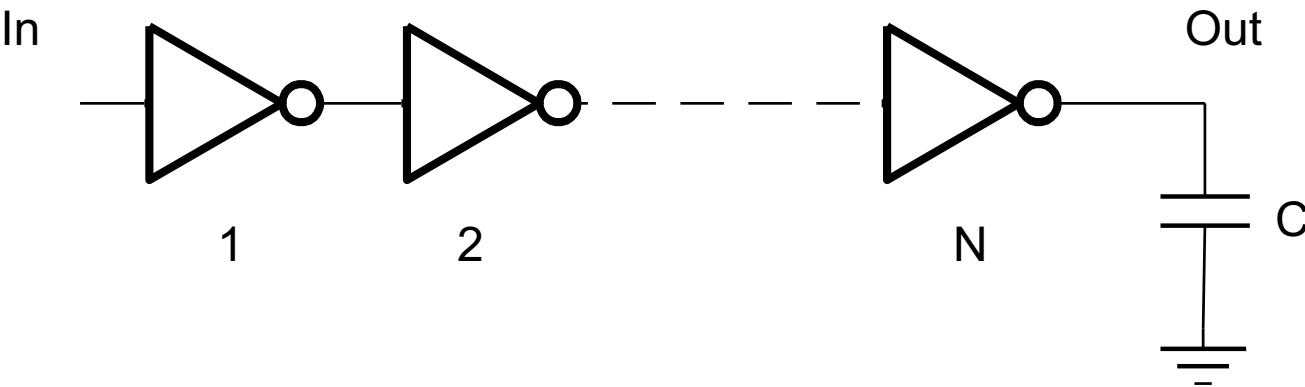
$$D = 1 + f$$

- Path Delay = $t_{p1} + t_{p2} + \dots + t_{pN}$
= $(1+f_1) + (1+f_2) + \dots + (1+f_N)$

- $f_1 = C_{in,2}/C_{in,1}$, $f_2 = C_{in,3}/C_{in,2}$, ...

- Path Fanout = $C_L/C_{in,1} = f_1 f_2 \dots f_N$

Minimizing the delay of an inverter chain



$$D = 1 + f$$

- Delay is minimized when each stage has the same delay.

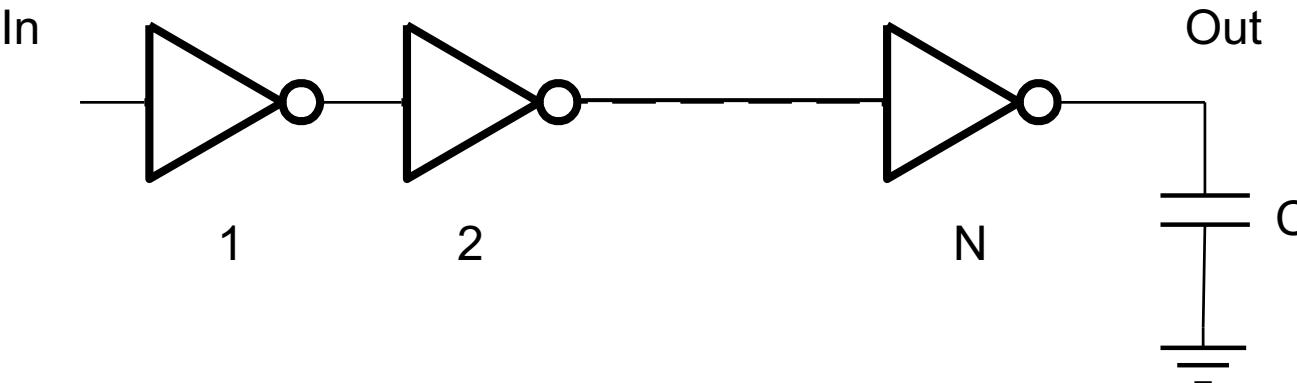
$$\bullet f = \sqrt[N]{F} = \sqrt[N]{C_L/C_{in,1}}$$

- **Minimum path delay:**

$$\begin{aligned}\bullet D &= (1+f_1) + (1+f_2) + \dots + (1+f_N) \\ &= Nf + N = N \sqrt[N]{F} + N\end{aligned}$$

- Delay = $t_{p1} + t_{p2} + \dots + t_{pN}$
= $(1+f_1) + (1+f_2) + \dots + (1+f_N)$
- $f_1 = C_{in,2}/C_{in,1}, f_2 = C_{in,3}/C_{in,2}, \dots$
- Path Fanout $F = C_L/C_{in,1} = f_1 f_2 \dots f_N$

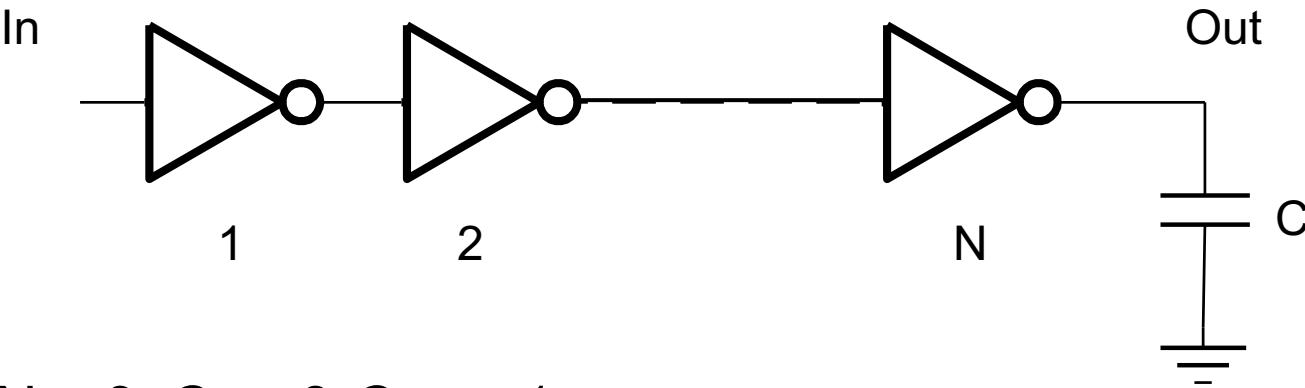
Example: Minimizing the delay of an inverter chain



- $N = 3, C_L = 8, C_{in,1} = 1$

- Delay is minimized when each stage has the same fanout.
 - $f = \sqrt[N]{F} = \sqrt[N]{C_L/C_{in,1}}$
- **Minimum path delay:**
 - $D = Nf + N = N \sqrt[N]{F} + N$

Example: Minimizing the delay of an inverter chain



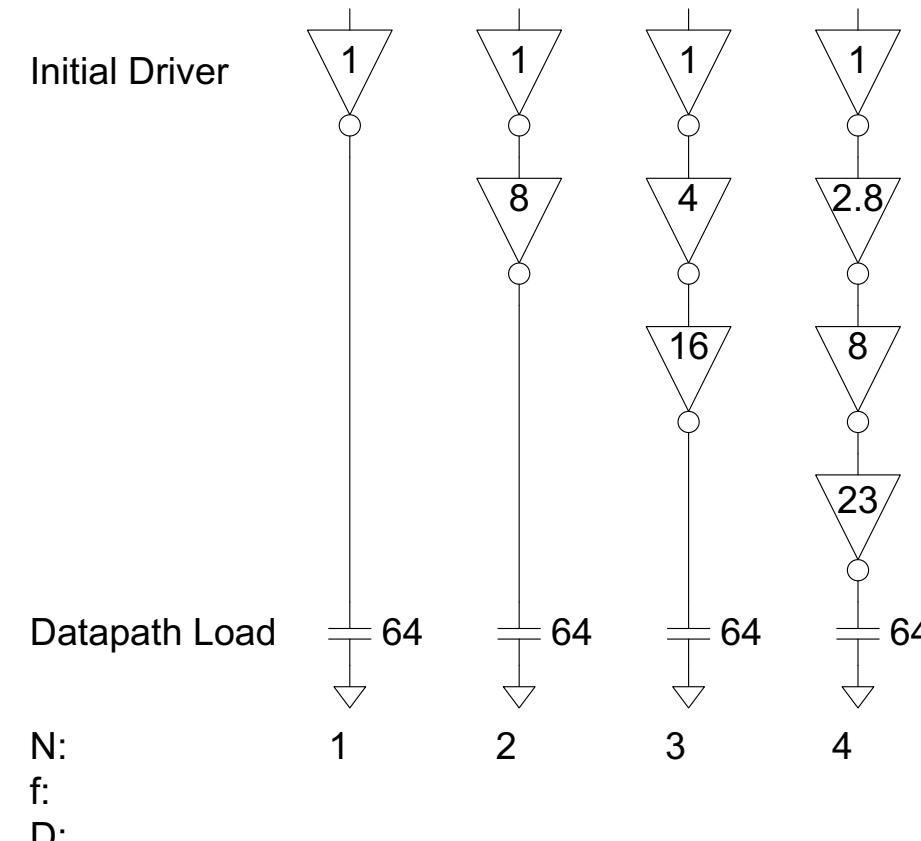
- $N = 3, C_L = 8, C_{in,1} = 1$
- To minimize the delay, $C_L/C_{in,1}$ has to be evenly distributed across $N=3$ stages.
 - $f = \sqrt[3]{8} = 2$
 - Recall $f_1 = C_{in,2}/C_{in,1}, f_2 = C_{in,3}/C_{in,2}, f_3 = C_L/C_{in,3}$
- **Size inverters from the back**
 - $C_{in,3} = C_L/f_3 = 8/2 = 4$
 - $C_{in,2} = C_{in,3}/f_2 = 4/2 = 2$

- Delay is minimized when each stage has the same fanout.
 - $f = \sqrt[N]{F} = \sqrt[N]{C_L/C_{in,1}}$
- **Minimum path delay:**
 - $D = Nf + N = N \sqrt[N]{F} + N$

Example: Best Number of Stages

- How many stages should a path use?
 - Minimizing number of stages is not always fastest
- Example: drive 64-unit load with inverters

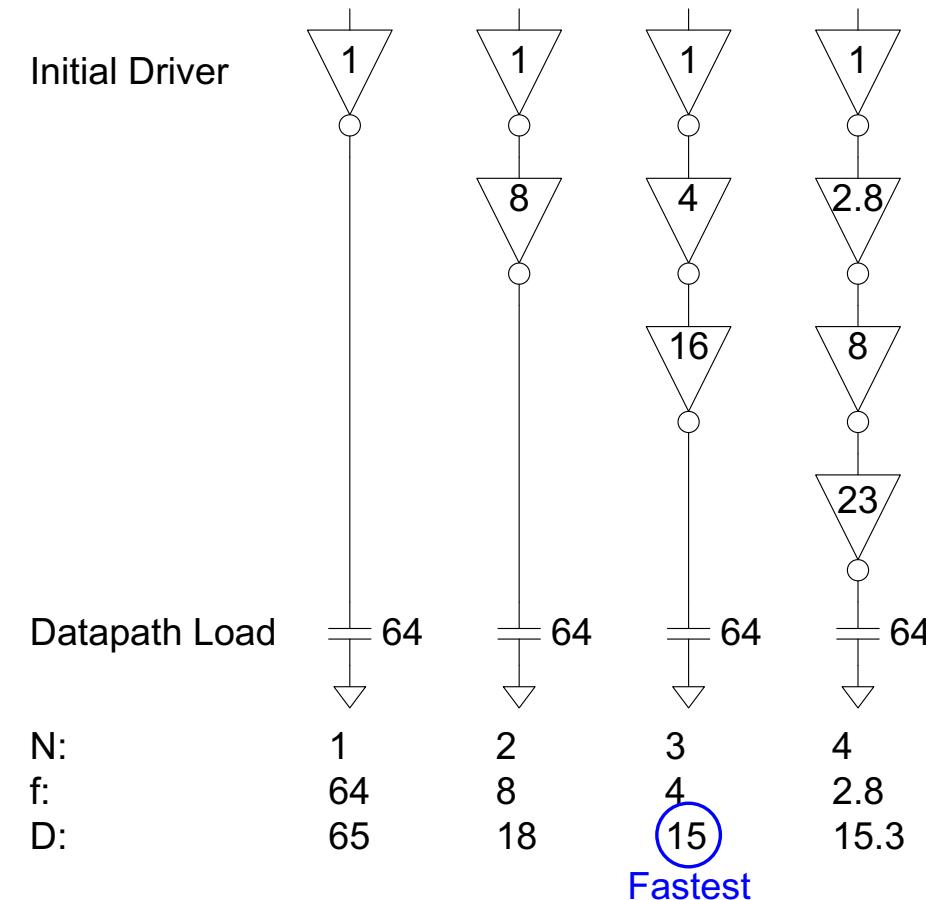
$$\begin{aligned} D &= NF^{1/N} + N \\ &= N(64)^{1/N} + N \end{aligned}$$



Example: Best Number of Stages

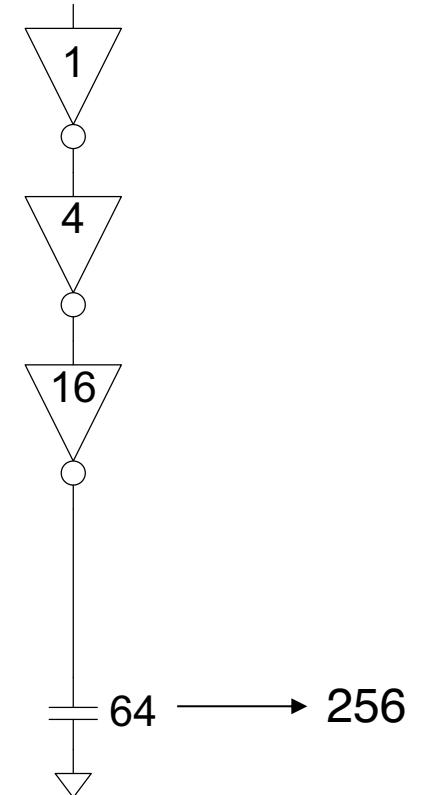
- How many stages should a path use?
 - Minimizing number of stages is not always fastest
- Example: drive 64-unit load with inverters

$$\begin{aligned} D &= NF^{1/N} + N \\ &= N(64)^{1/N} + N \end{aligned}$$



Quiz: Inverter Chain

- If we increase the load from 64 to 256, which of the following choices will have the smallest delay?
 - A. Remove the size 16 inverter.
 - B. Add a size 64 inverter.
 - C. Do nothing.



Inverter and Inverter Chain Recap

- Inverter Delay

- $t_p = \ln 2 * R_{eq} C_{in} (1 + C_L / C_{in}) = \tau_{INV} (1 + f)$

- $\tau_{INV} = \ln 2 * R_{eq} C_{in}$

- $Fanout = f = C_L / C_{in}$

- Normalized Delay to τ_{INV}

- $D(\text{inv}) = 1 + f$



- Inverter Chain

- Path Delay Delay = $\sum(1 + f_i) = N + \sum f_i$

- Path Fanout Effort $F = \prod f_i = C_L / C_{in,1}$

- Size the inverters to minimize the delay of an inverter chain

- Every inverter stage has the same effort delay, i.e., $f_i = \sqrt[N]{F}$

- The size of each inverter stage can be determined by working backward

- $C_{in,i} = \frac{C_{L,i}}{f_i}$