EECS151/251A
Introduction to Digital Design and ICs

Lecture 17: Wire & Energy

Sophia Shao

Universal Chiplet Interconnect Express (UCIe) Announced: Setting Standards For The Chiplet Ecosystem

To that end, today Intel, AMD, Arm, and all three leading-edge foundries are coming together to announce that they are forming a new and open standard for chiplet interconnects, which is aptly being named **Universal Chiplet Interconnect Express**, or **UCIe**. Taking significant inspiration from the very successful PCI-Express playbook, with UCIe the involved firms are creating a standard for connecting chiplets, with the goal of having a single set of standards that not only simplify the process for all involved, but lead the way towards full interoperability between chiplets from different manufacturers, allowing chips to mix-and-match chiplets as chip makers see fit. In other words, to make a complete and compatible ecosystem out of chiplets, much like today's ecosystem for PCIe-based expansion cards.

https://www.anandtech.com/show/17288/universal-chiplet-interconnect-express-ucie-announced-setting-standards-for-the-chiplet-ecosystem

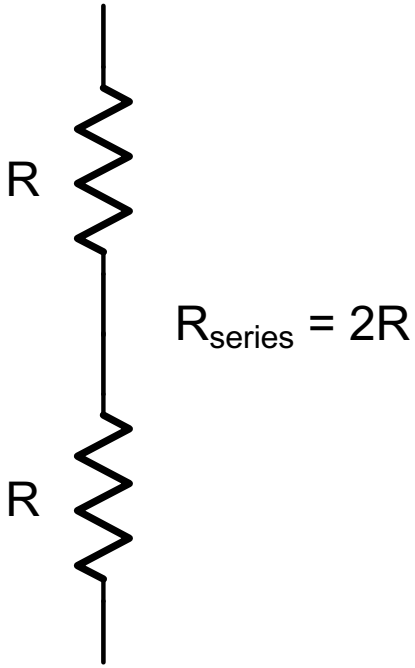
Review

- Logical Effort:
  - Logical effort is the ratio of the input capacitance to the input capacitance of a unit inverter delivering the same output current.
  - Only dependent on gate topology
  - To calculate LE and P, size transistors so that R is the same as Rinv, then
    - LE = Cin,gate / Cin,inv
    - P = Cp,gate / Cp,inv
- Multi-Stage Network:
  - $Delay = \sum_{i=1}^{N}(P_i + LE_i * FO_i)$
    - Stage Effort: $SE_i = LE_i * FO_i$
    - Path Fanout Effort: $FO_{path} = \frac{C_{load}}{C_{in}}$
    - Path Logical Effort: $LE_{path} = LE_1 * LE_2 \ldots * LE_N$
    - Path Effort: $PE = LE_{path} * FO_{path}$

- Wire
  - Overview
  - RC Tree Model
  - Elmore Delay
- Energy
  - Overview
  - Static & Dynamic Power

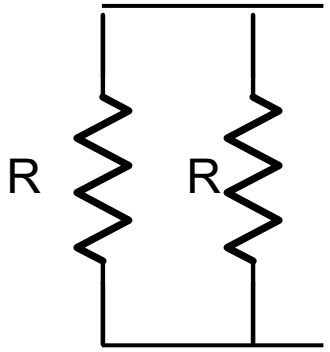
Resistance

- With two identical resistors, R



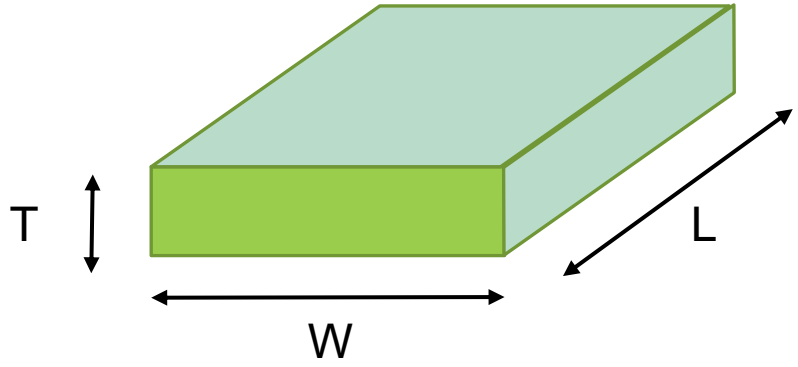
$R_{series} = 2R$

Equivalent to doubling length



$R_{parallel} = R/2$

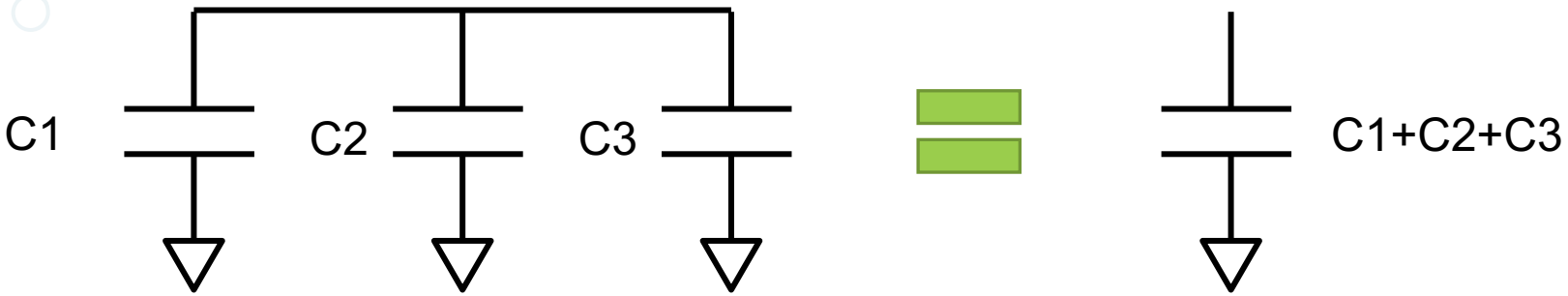
Equivalent to doubling width

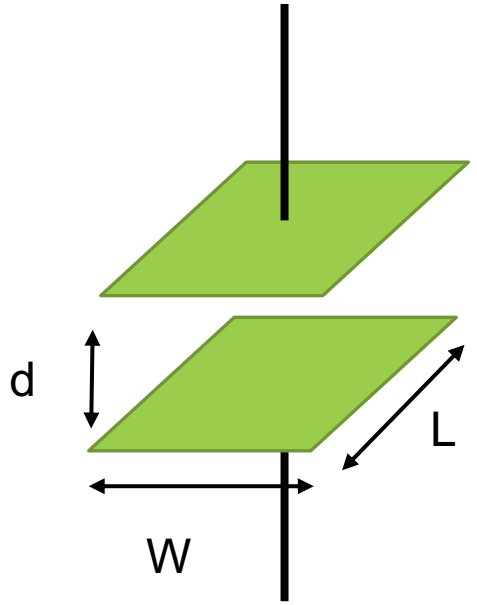


$$R = \rho \frac{L}{TW}$$

- In a planar process, designer controls W and L

Capacitance



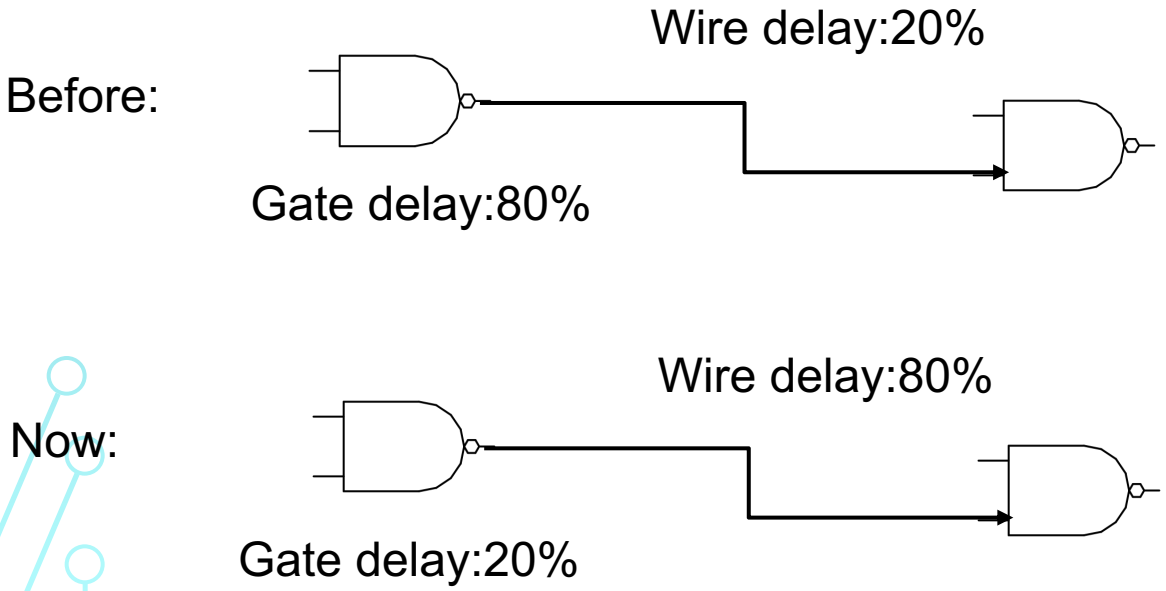


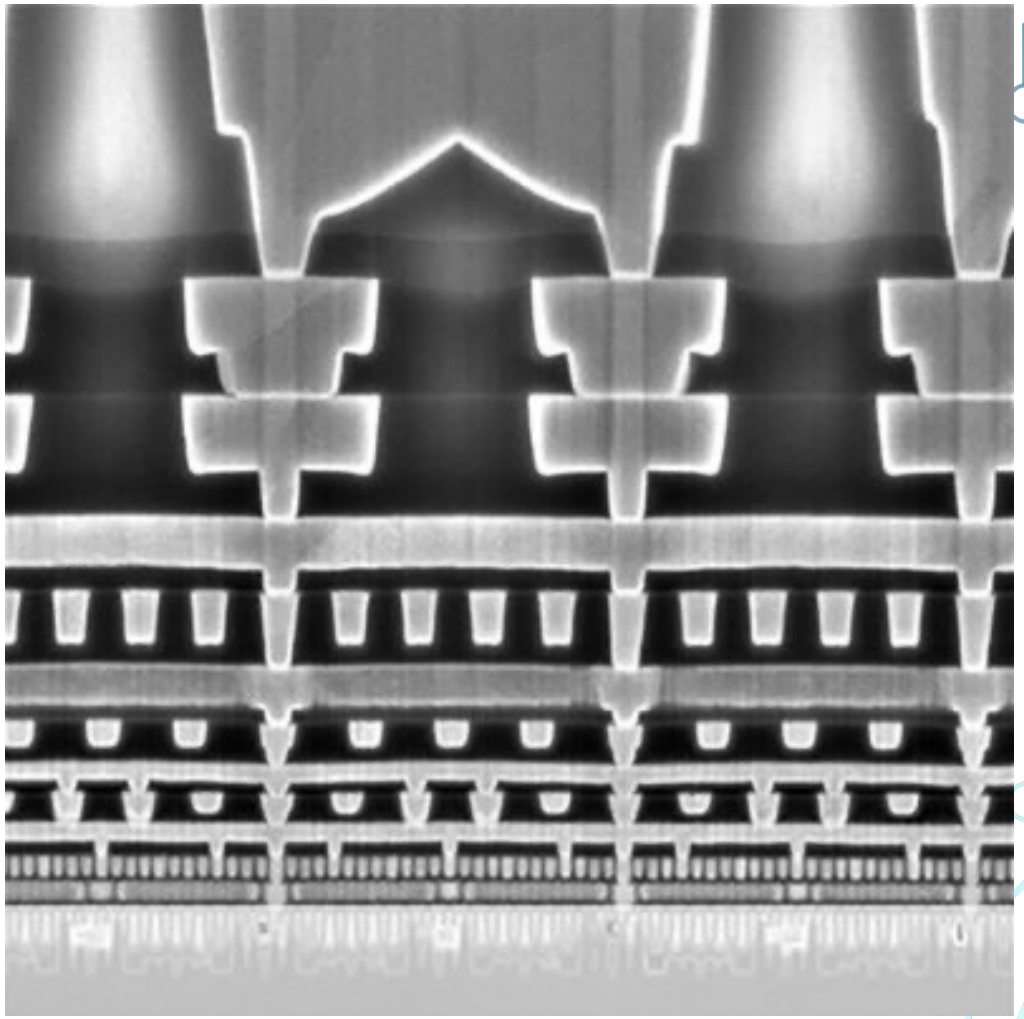
- Equivalent to increase the effective area of the capacitor plates.

$$C = \varepsilon \frac{LW}{d}$$

A modern technology is mostly wires

- Transistors are little things under the wires
- Many layers of wires
- Wires are as important as transistors
  - Speed and power



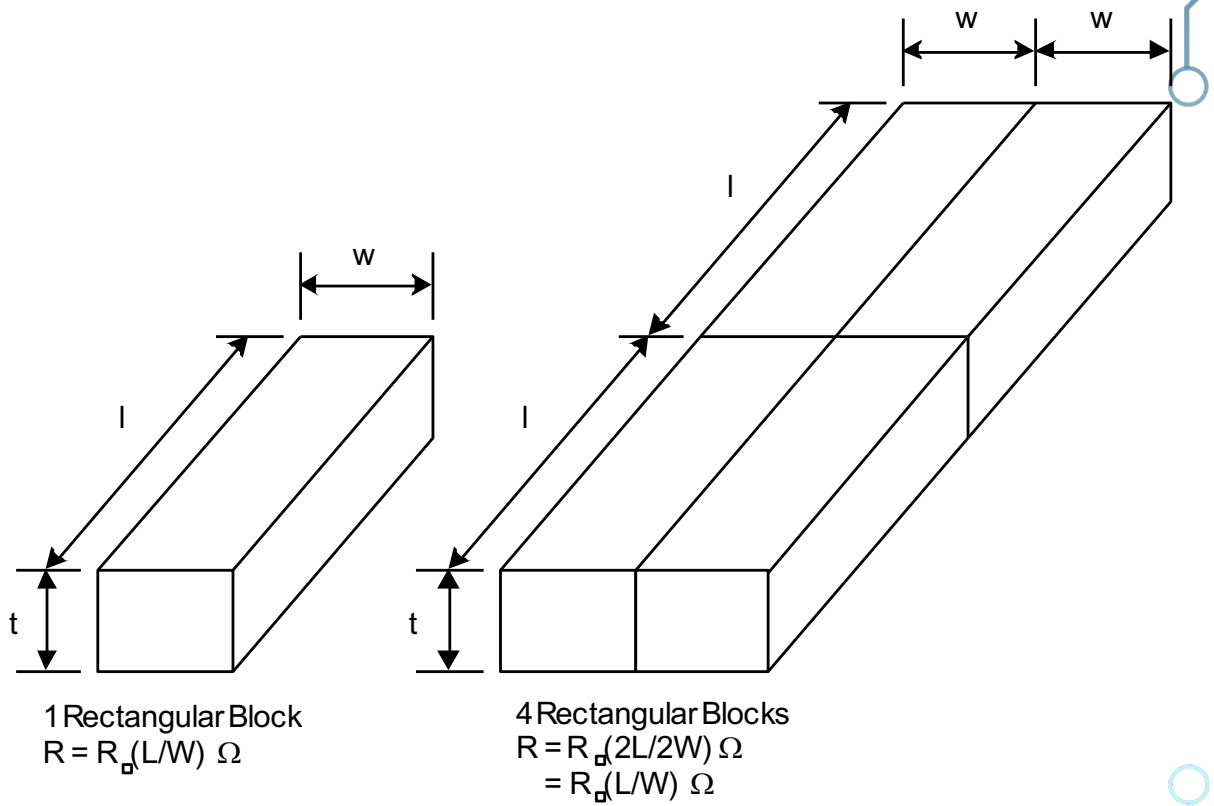


Wire Resistance

- ρ = *resistivity* (Ω*m)

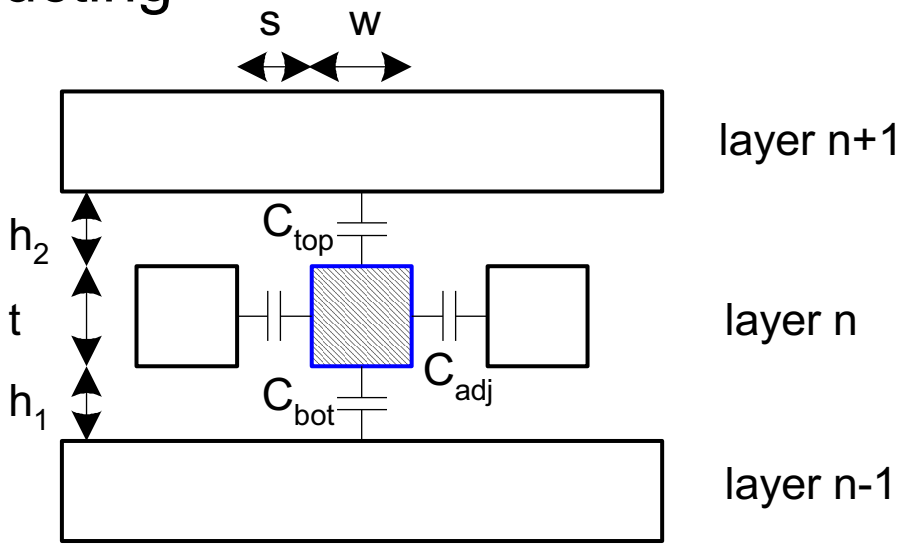
$$R = \frac{\rho}{t}\frac{l}{w} = R_{\square}\frac{l}{w}$$

- $R_{\square}$ = *sheet resistance* ($\Omega/\square$)
  - $\square$ is a dimensionless unit(!)
- Longer wire (L), higher R.



Wire Capacitance

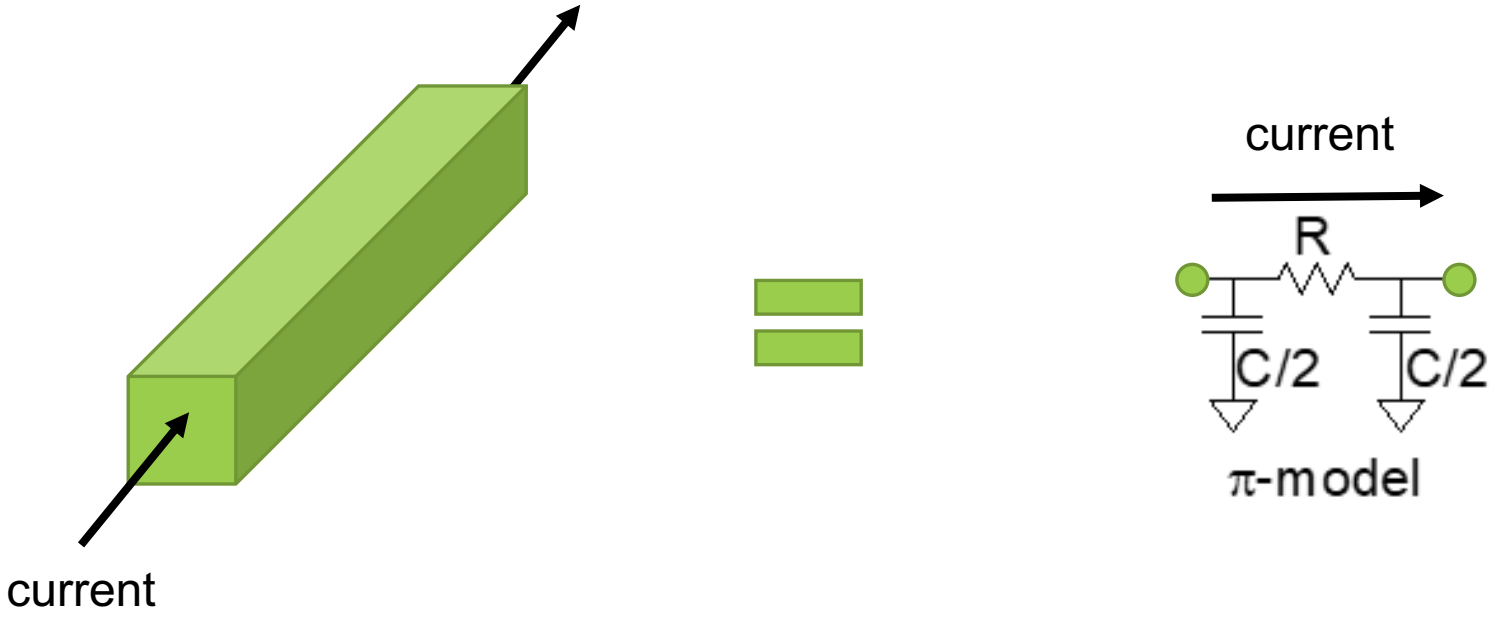
- Wire has capacitance per unit length
  - Capacitance exists between any pair of conducting surfaces.
  - In a multi-layer metal process:
    - To neighbors
    - To layers above and below
    - $C_{total} = C_{top} + C_{bot} + 2C_{adj}$



- For simplicity
  - We just assume a single capacitance per wire.
  - $C = \varepsilon \frac{LW}{d}$
  - Longer wire (L), higher C.

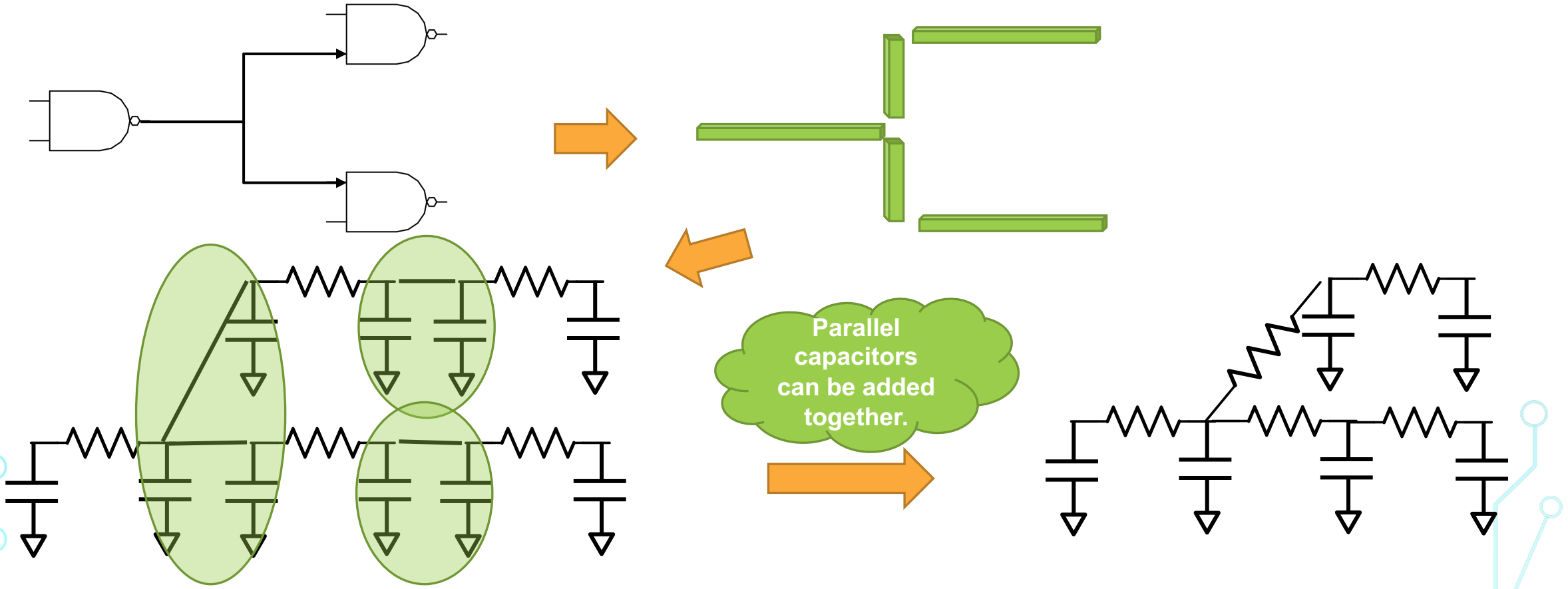
Wire RC Model

- Wire can be modeled as a RC circuit.
- Π (“pi) model
  - Account for the resistance R and the capacitance C of wire.
  - Simple model: only need 2 numbers.



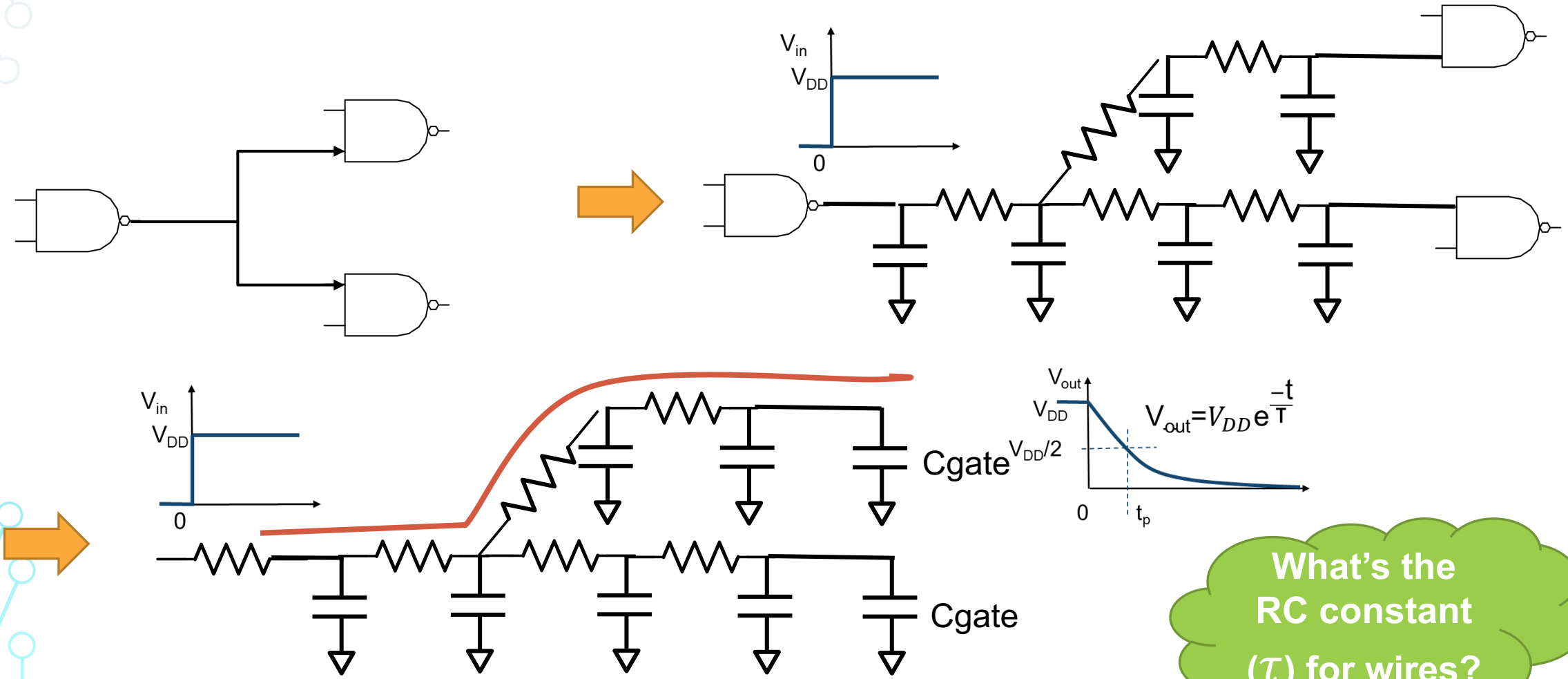
RC Tree

- What we had was just one segment of wire.
- In “reality”, each wire of the design has its own RC model.



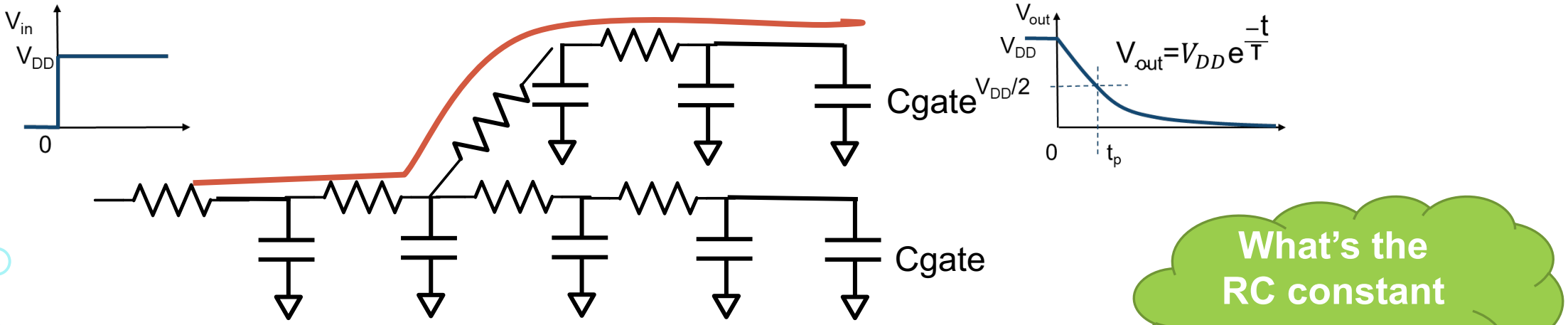
RC Tree Delay

- How long does it take for a signal to propagate through wires (i.e., RC trees)?



Elmore Delay

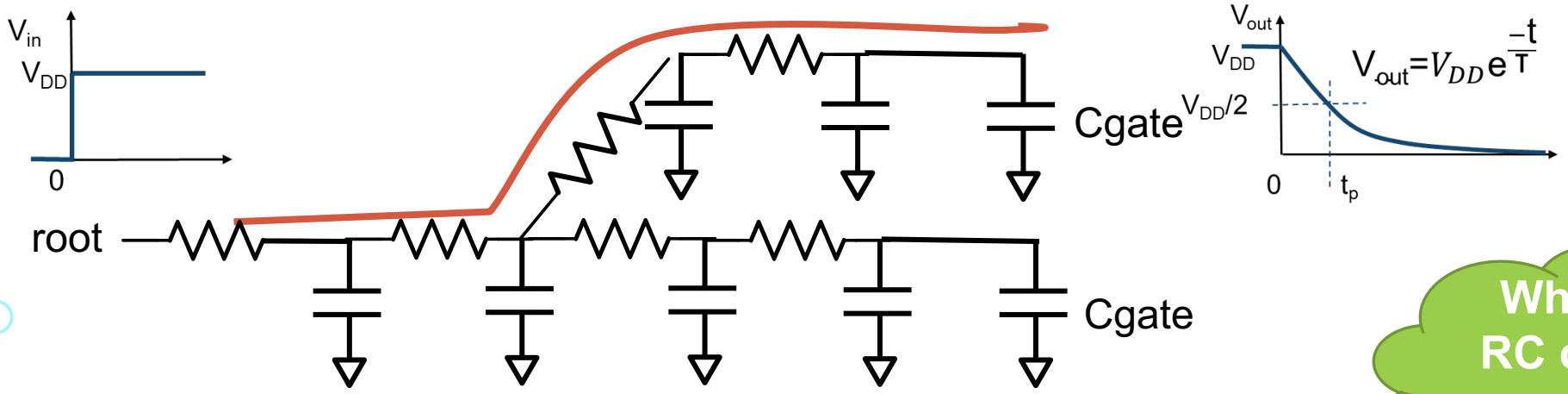
- A simple **approximation** to calculate the delay through an RC tree.
  - Originally proposed in the 40s.
  - Resurrected in the 80s for RC trees in integrated circuits
  - Works fairly well for delay calculation.
- We can use Elmore Delay to calculate τ.



What's the RC constant (τ) for wires?

Elmore Delay τ Calculation: Tree Walking

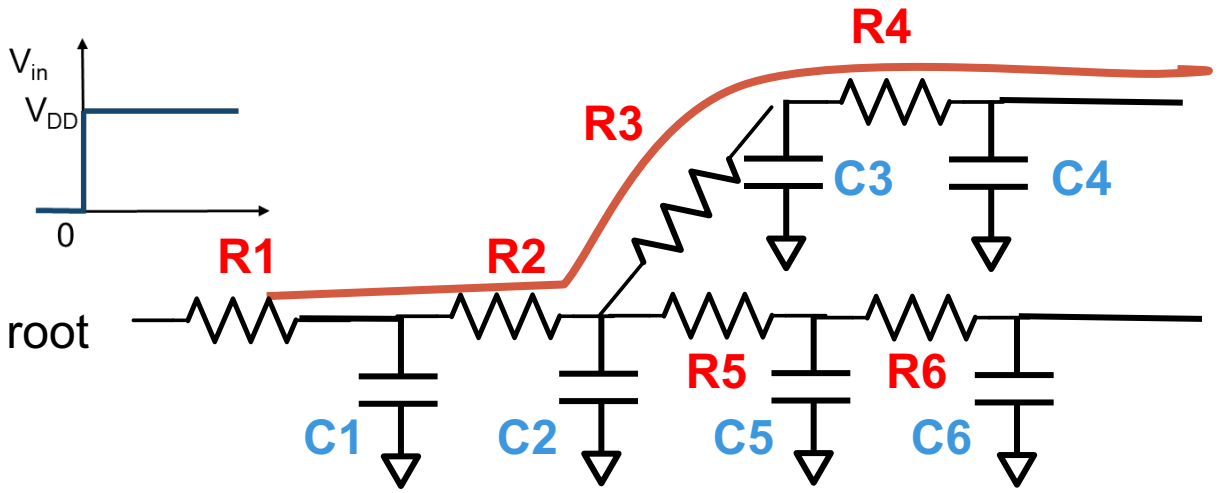
- Start from the root of the tree to the leaf of the tree that we want to calculate delay
  - Start: τ = 0;
  - Whenever we see a resistor along the way, $\tau \mathrel{+}= R * \sum(all\ capacitors\ downstream)$
  - "Downstream capacitors": any capacitors that are reachable by the current through R

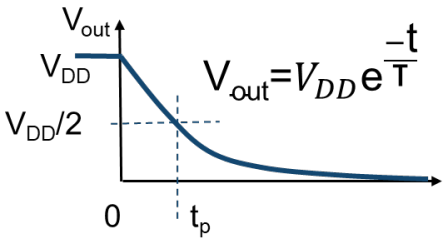


What's the RC constant (τ) for wires?

Elmore Delay τ Calculation: Tree Walking

- $\tau = R1(C1 + C2 + C3 + C4 + C5 + C6)$
  $+ R2(C2 + C3 + C4 + C5 + C6)$
  $+ R3(C3 + C4)$
  $+ R4(C4)$

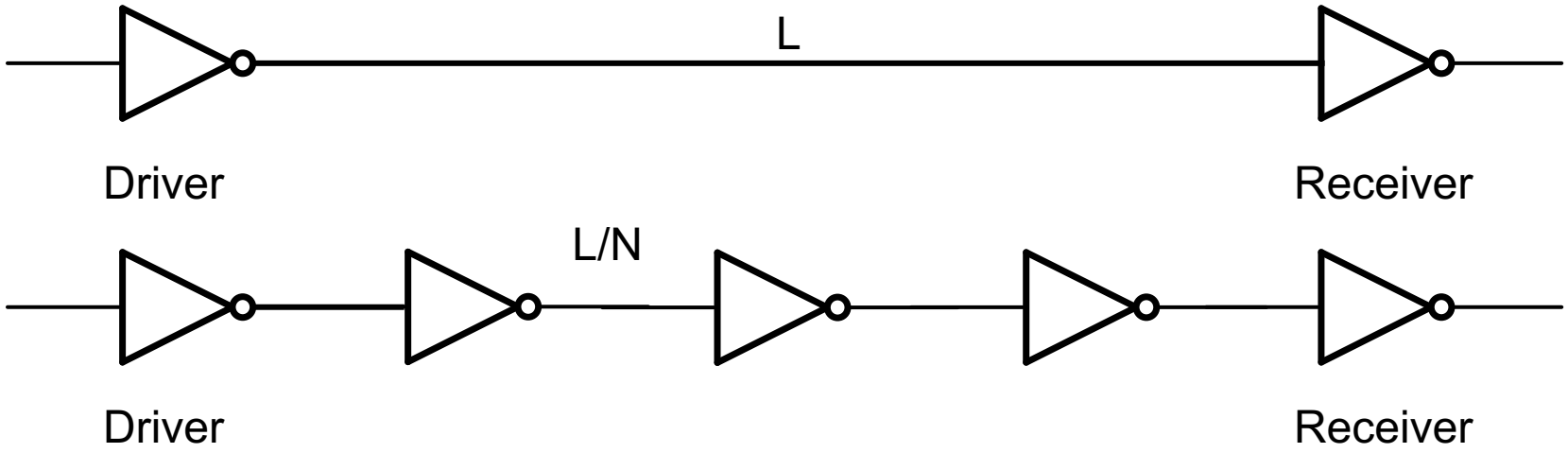




What's the RC constant (τ) for wires?

Wire Delay

- Instead of having a long wire of L, if we split the wire into N segments and insert an inverter to drive the wire, how would the delay change between the driver node and receiver node? (Ignore the inverter delay)



A: Increase

B: Decrease

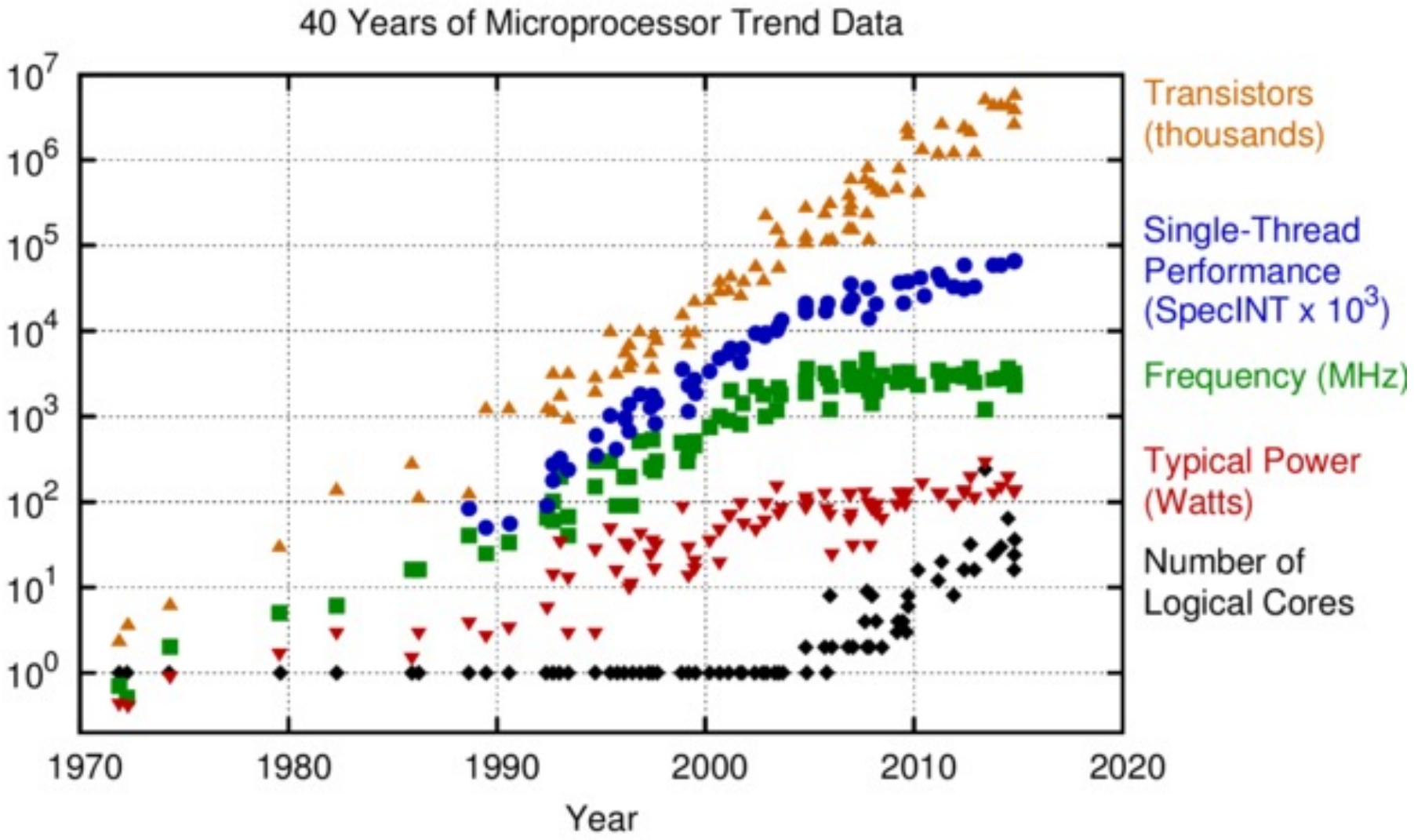
C: Doesn't change

Administrivia

- Enjoy Spring Break!
- Project starts this week.
  - Outstanding Designer Award!

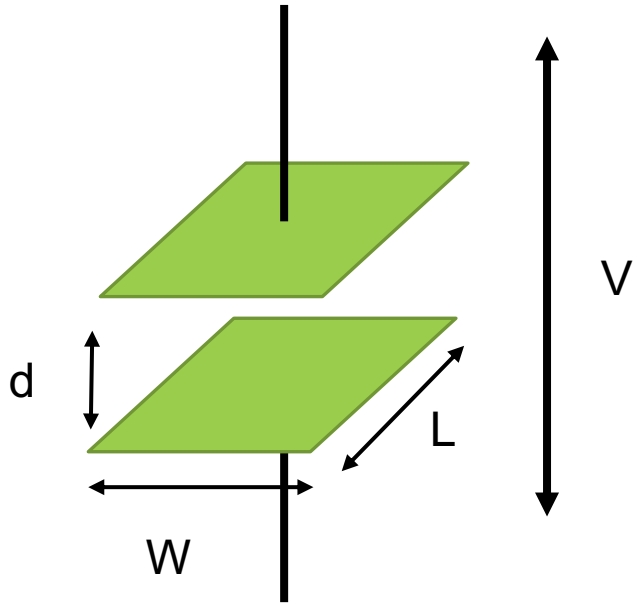
- Wire
  - Overview
  - RC Tree Model
  - Elmore Delay
- Energy
  - Overview
  - Static & Dynamic Power

Processor Frequency Scaling



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2015 by K. Rupp

Power and Energy



$$C = \varepsilon \frac{LW}{d}$$

$$E_C = \int_0^{\infty} I(t)\,V(t)dt = \int_0^{\infty} C\frac{dV}{dt}V(t)dt$$

$$= C\int_0^{V_C} V(t)\,dV = \frac{1}{2}CV_C^2$$

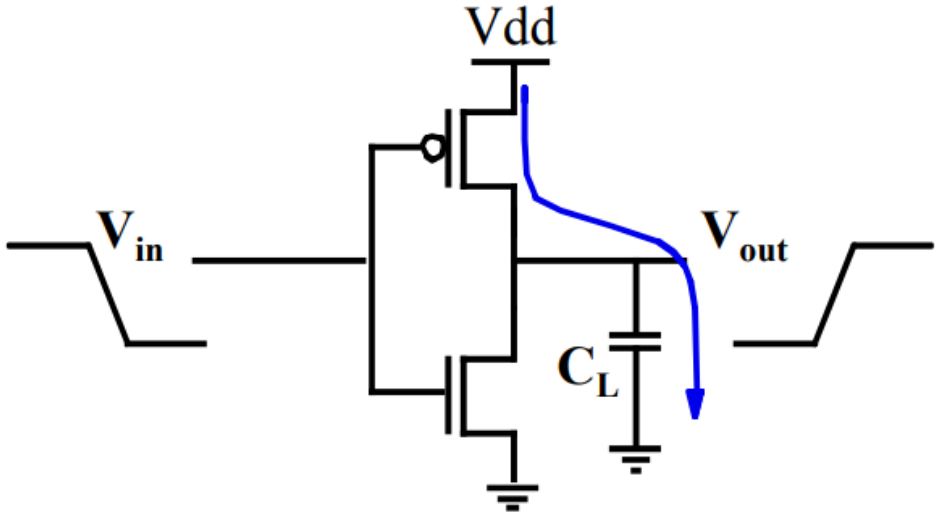
$$P_C = \frac{E_C}{T} = \frac{1}{2}CV_C^2 f$$

Where does power go in CMOS?

- Dynamic Power Consumption
  - Charging and discharging capacitors
- Short-Circuit Currents
  - Short-circuit path between supply rails during switching
- Leakage (Static) Power
  - Leaky transistors

#1: Dynamic Power Consumption

- To reduce dynamic power:
  - Reduce C_L, Vdd, f, and activity factor



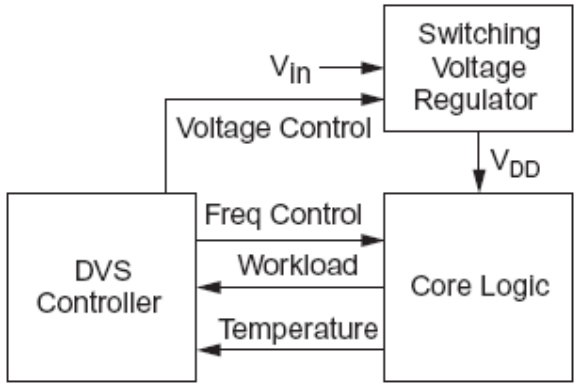
$$\frac{Energy}{transition} = C_L * V_{dd}^2$$

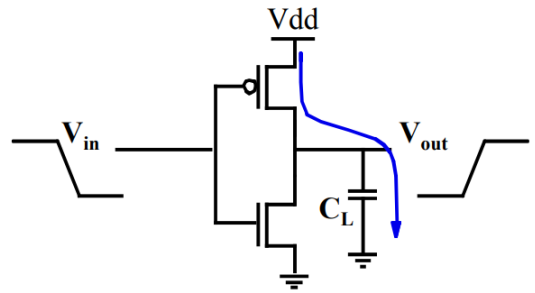
$$Power = \frac{Energy}{transition} * f_{sw} = \alpha * C_L * V_{dd}^2 * f_{clk}$$

α: activity factor
- Clock α = 1
- Depends on gates and data pattern
- Typical α = 0.1

#1: Dynamic Power Consumption

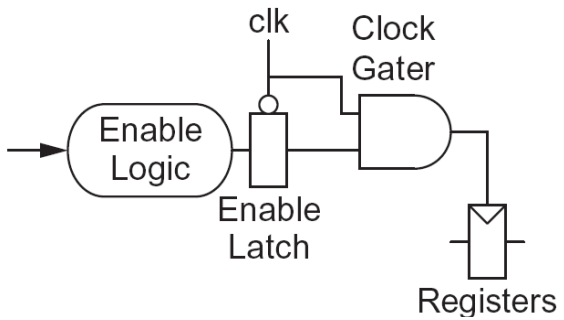
- Voltage and frequency scaling (lower Vdd, f)





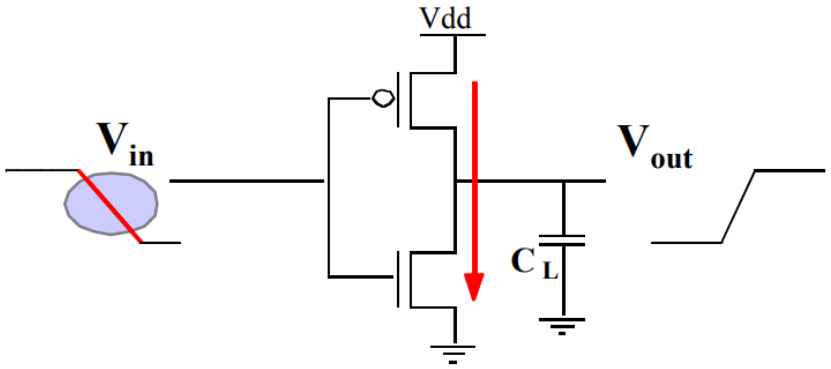
$$Power = \alpha * C_L * V_{dd}^2 * f_{clk}$$

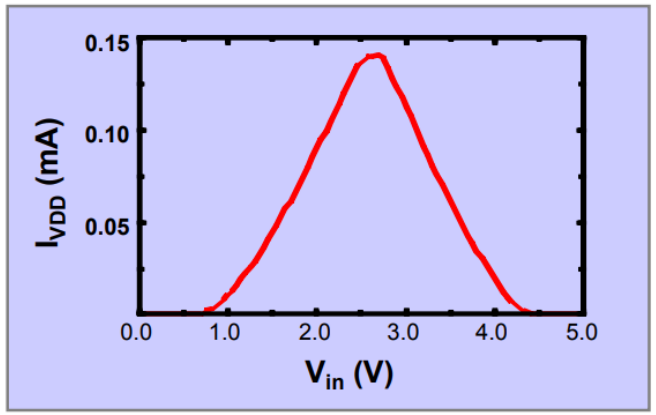
- Reduce capacitance (lower C_L)
  - Gate: minimize device sizes w/o significantly hurting perf.
  - Wire: shorten long wires
- Reduce activity factor (lower α)
  - Clock gating



#2: Short-circuit Current

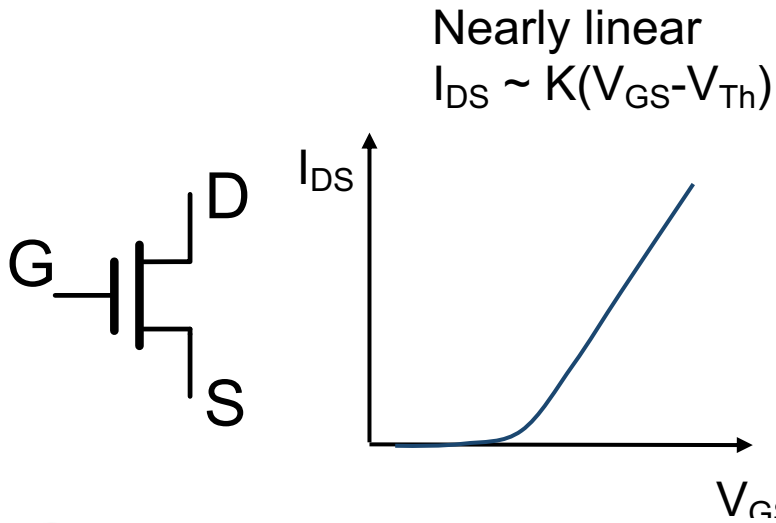
- During transition, both PUD and PDN are on.
- Similar to dynamic power, related to switching activity.
- Typically small, <10% of total power consumption
- Becomes less important in advanced technology
  - Threshold voltages do not scale as fast as supply voltage.
  - If $V_{dd} < V_{thn} + |V_{thp}|$, no short circuit.

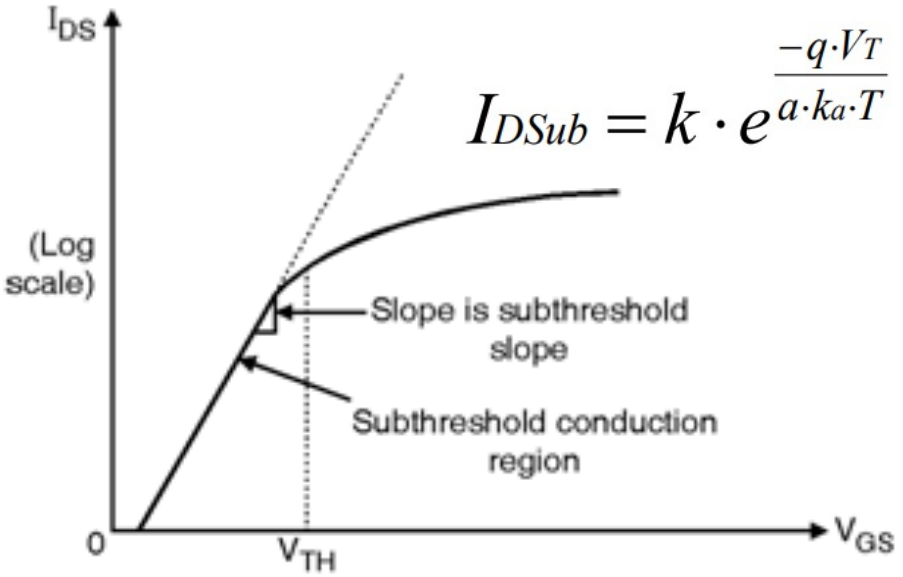




#3 Leakage (Static) Power

- Power is consumed with a chip is not switching.
  - "Off" is not really Off.
- Subthreshold leakage currents grow exponentially with increases in temperature and decreases in threshold voltage.
  - But threshold voltage scaling is key to circuit performance.
- To reduce leakage power
  - Power gating: Turn OFF power to blocks when they are idle to save leakage
  - Use high-Vt cells.





I_{DS} Vs V_{GS} characteristics in log scale

Summary

- Wire also contributes to delay, especially in modern technology.
- We can use RC model to capture wire delay as well.
- Energy becomes an increasingly important optimization goal.
  - Dynamic Energy
  - Static Energy