

EEC 116 Lecture #9: Coping With Interconnect

**Rajeevan Amirtharajah
University of California, Davis**

Announcements

- **Lab 5 continues this week**
- **Homework 5 due Nov. 18**
- **Midterm curve being determined, will be posted shortly**

Outline

- **Review: Wire Models**
- **Minimum Delay Sizing**
- **Wires: Rabaey Ch. 4 and Ch. 9 (Kang & Leblebici, 6.5-6.6)**

Wire Models

Complexity



- **Ideal Short Circuit**
- **Lumped Capacitor (C)**
- **Lumped RC**
- **Distributed RC**
 - Ladder Filter (series R, shunt C)
- **Distributed LC**
 - Lossless Transmission Line (series L, shunt C)
- **Distributed RLGC**
 - Lossy Transmission Line (series L, R; shunt G, C)

Interconnect Models: Regions of Applicability

- **For highest speed applications, wire must be treated as a transmission line**
 - Includes distributed series resistance, inductance, capacitance, and shunt conductance (*RLGC*)
- **Many applications it is sufficient to use lumped capacitance (*C*) or distributed series resistance-capacitance model (*RC*)**
- **Valid model depends on ratio of rise/fall times to time-of-flight along wire**
 - l : wire length
 - v : propagation velocity (speed of light)
 - l/v : time-of-flight on wire

Interconnect Models: Regions of Applicability

- **Transmission line modeling (inductance significant):**

$$t_{\text{rise}} (t_{\text{fall}}) < 2.5 \times (l / v)$$

- **Either transmission line or lumped modeling:**

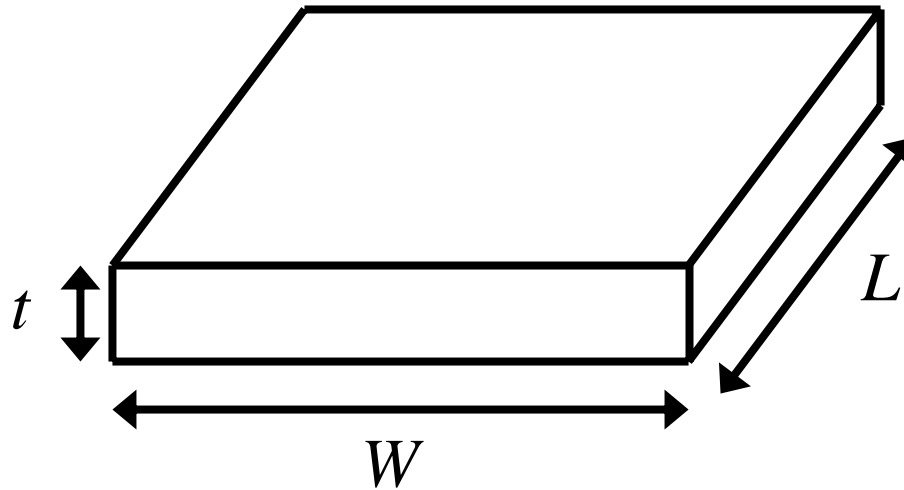
$$2.5 \times (l / v) < t_{\text{rise}} (t_{\text{fall}}) < 5 \times (l / v)$$

- **Lumped modeling:**

$$t_{\text{rise}} (t_{\text{fall}}) > 5 \times (l / v)$$

Resistance

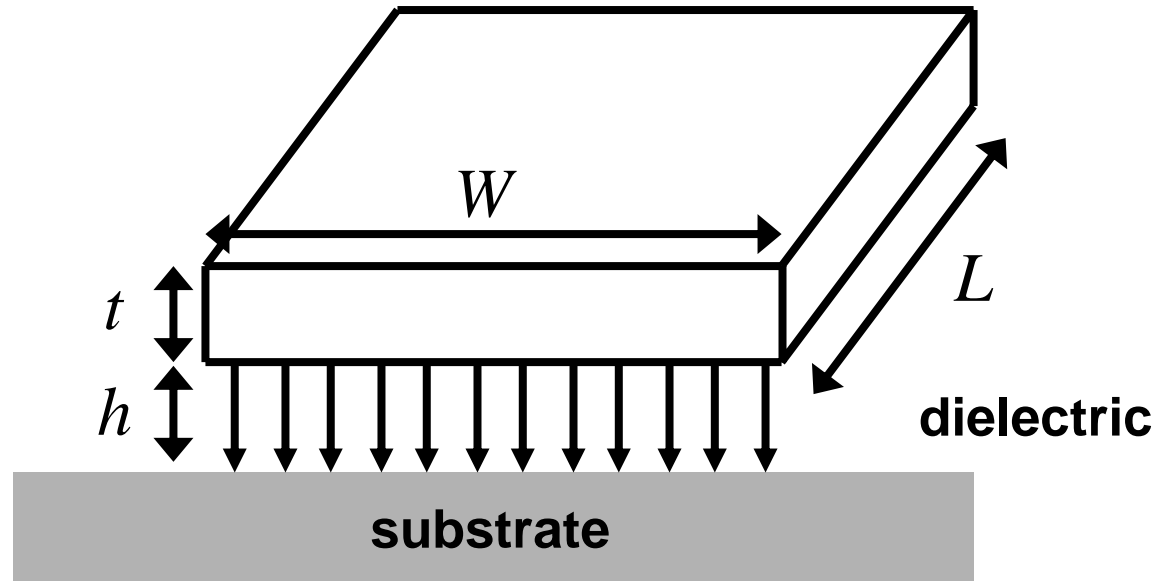
- Resistance proportional to length and inversely proportional to cross section
- Depends on material constant resistivity ρ ($\Omega\text{-m}$)



$$R = \frac{\rho L}{A} = \frac{\rho L}{tW} = R_{sq} \frac{L}{W} \quad R_{sq} = \frac{\rho}{t}$$

Parallel-Plate Capacitance

- **Width large compared to dielectric thickness, height small compared to width: E field lines orthogonal to substrate**



$$C = \frac{\epsilon_r}{h} WL$$

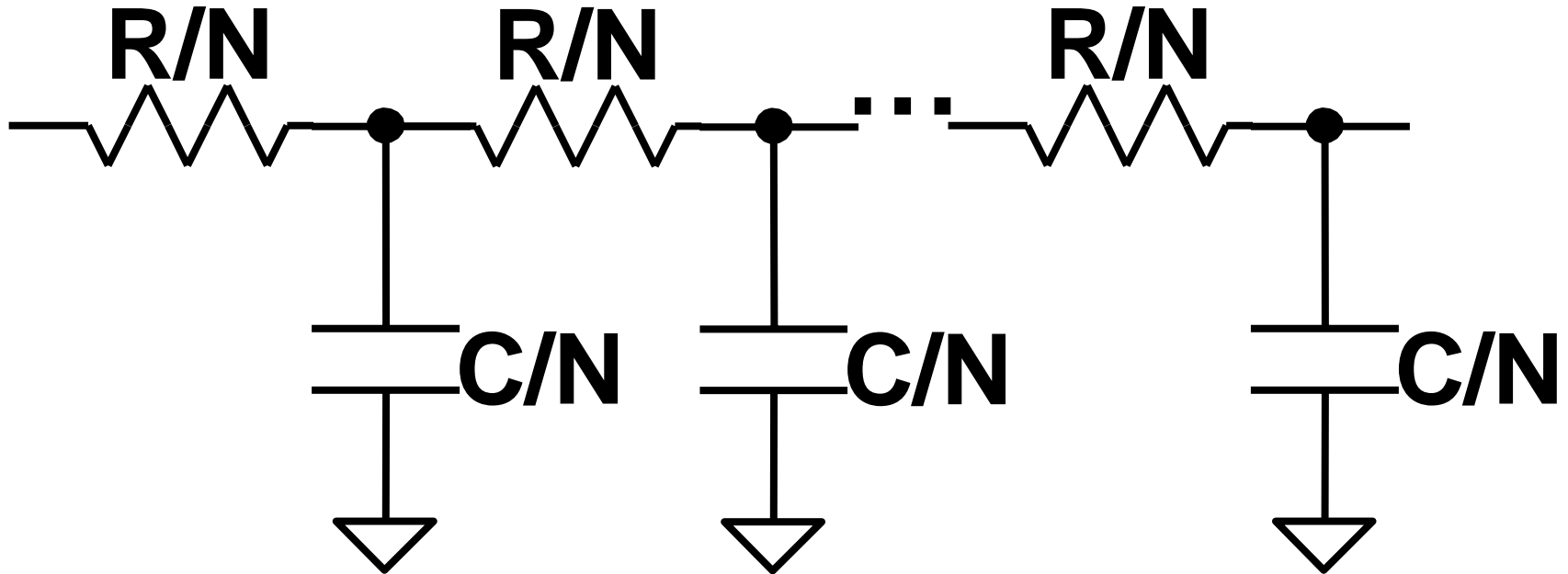
Total Capacitance Model

- **Total capacitance per unit length is parallel-plate (area) term plus fringing-field term:**

$$C = C_{pp} + C_{fringe} = \frac{\epsilon_r}{h} \left(W - \frac{t}{2} \right) + \frac{2\pi\epsilon_r}{\log(2h/t + 1)}$$

- **Model is simple and works fairly well (Rabaey, 2nd ed.)**
 - More sophisticated numerical models also available
- **Process models often give both area and fringing (also known as sidewall) capacitance numbers per unit length of wire for each interconnect layer**

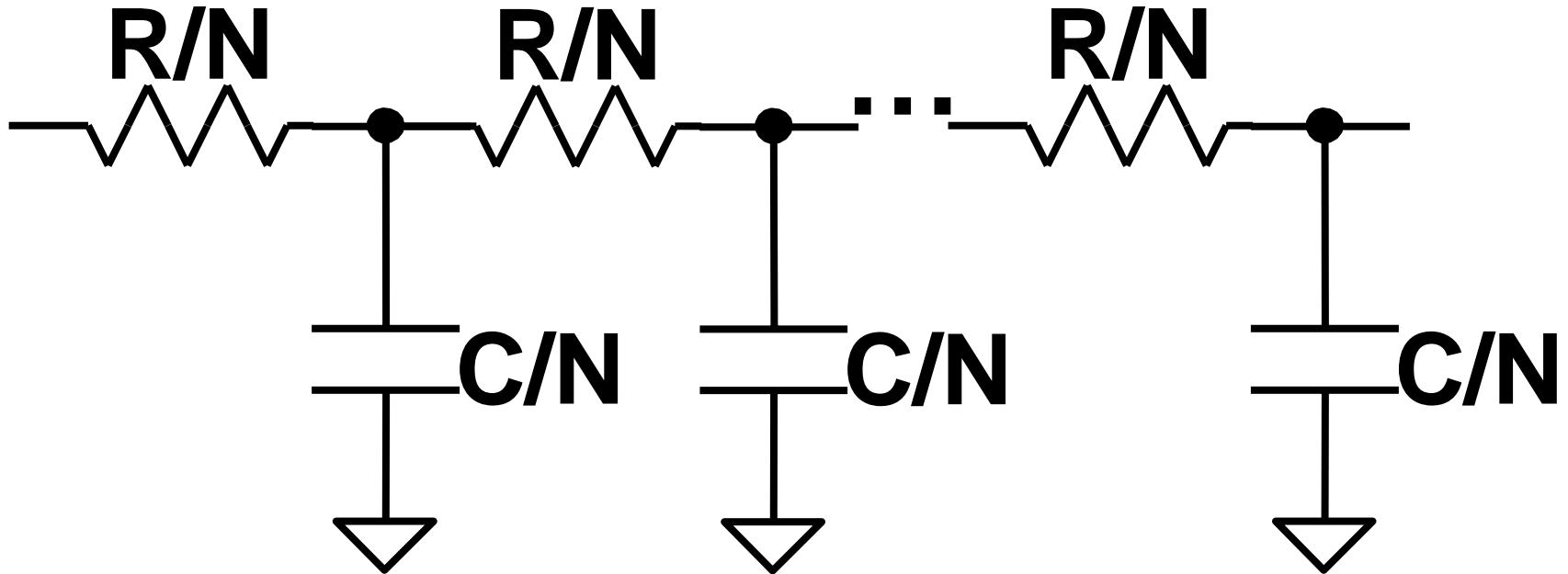
RC Ladder Network Delay



- Elmore delay approximation for *RC* ladder network:

$$\tau_{Di} = \sum_{i=1}^N C_i \sum_{j=1}^i R_{ji} = \sum_{i=1}^N C_i R_{ii} = RC \frac{N+1}{2N}$$

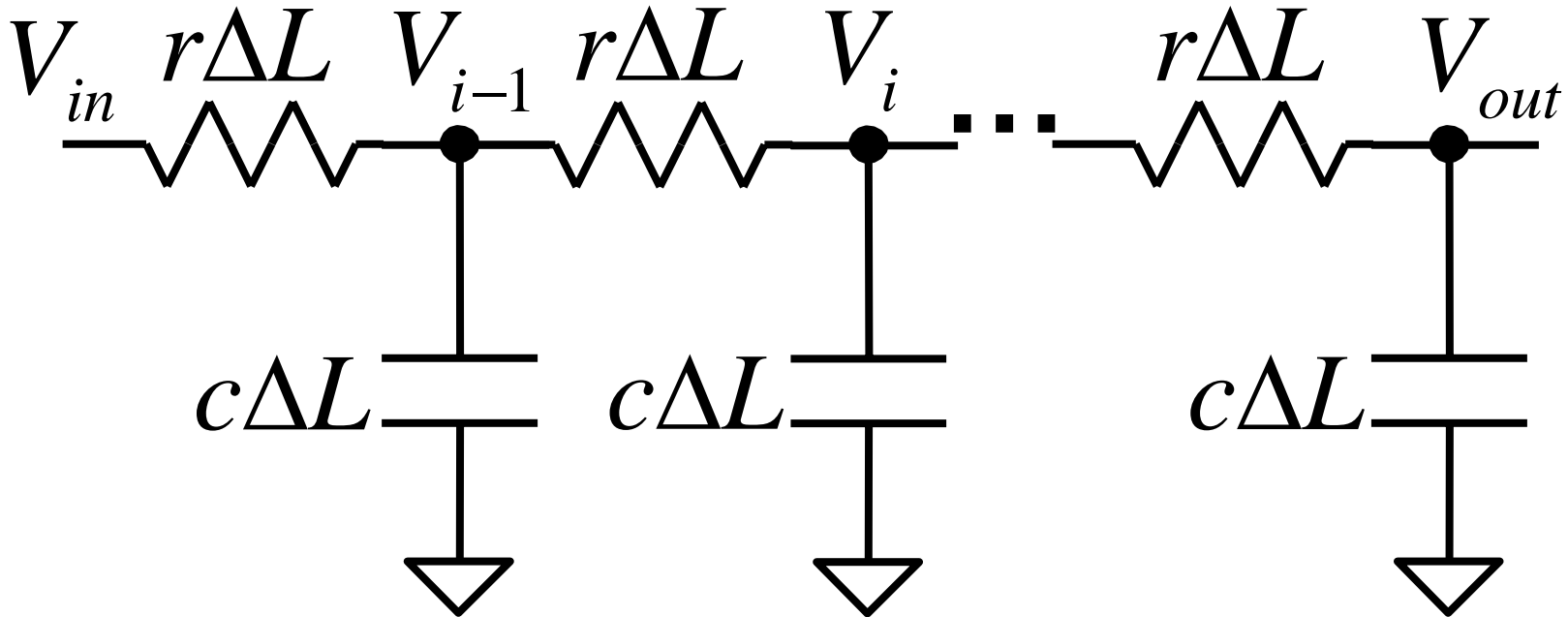
RC Ladder Network Delay



- Elmore delay approximation for *RC* ladder network:

$$t_{DN} = \frac{RC}{2} = \frac{rcL^2}{2} \quad \text{as } N \rightarrow \infty$$

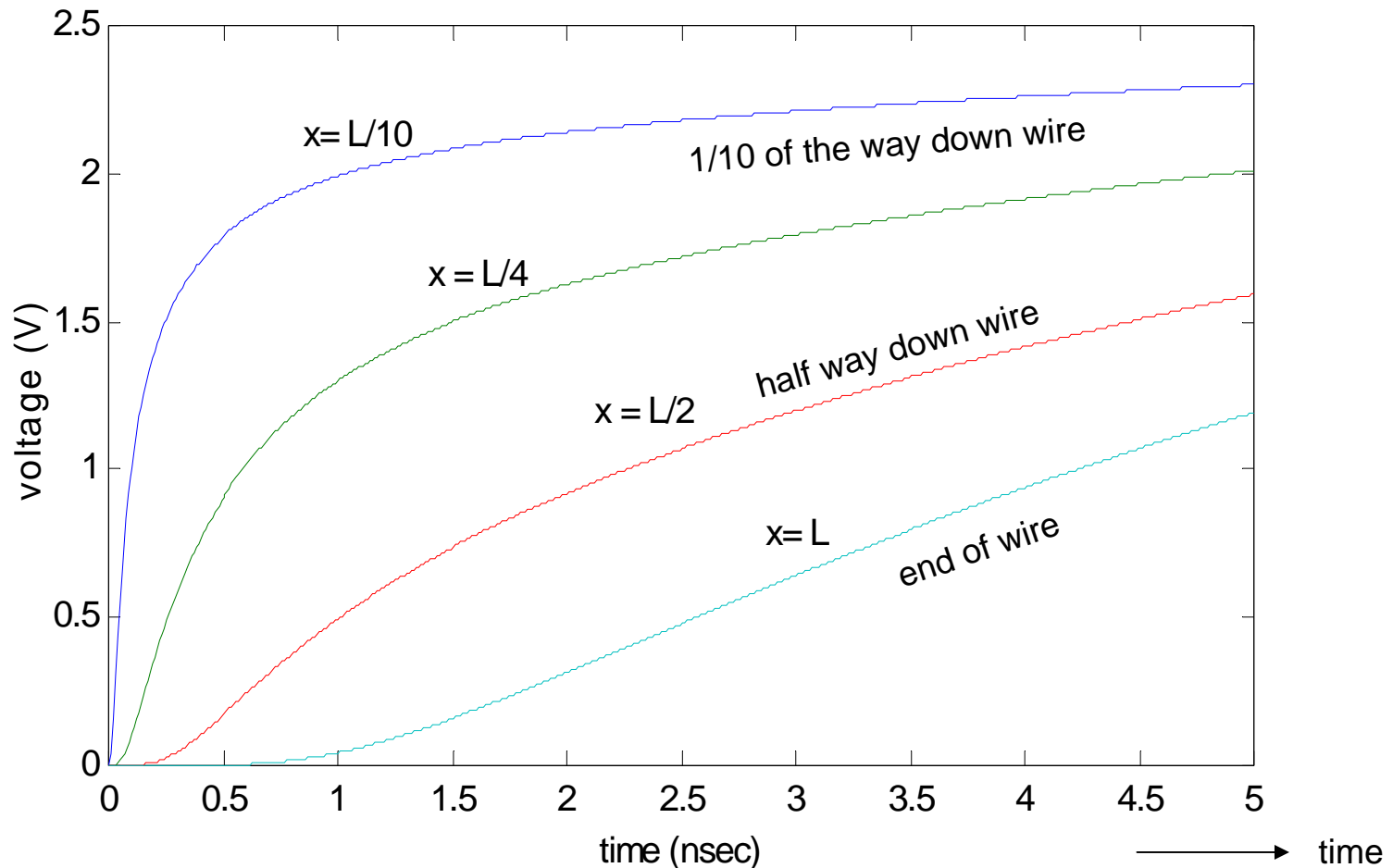
Distributed RC Model



- **Differential equation at i th node (from KCL):**

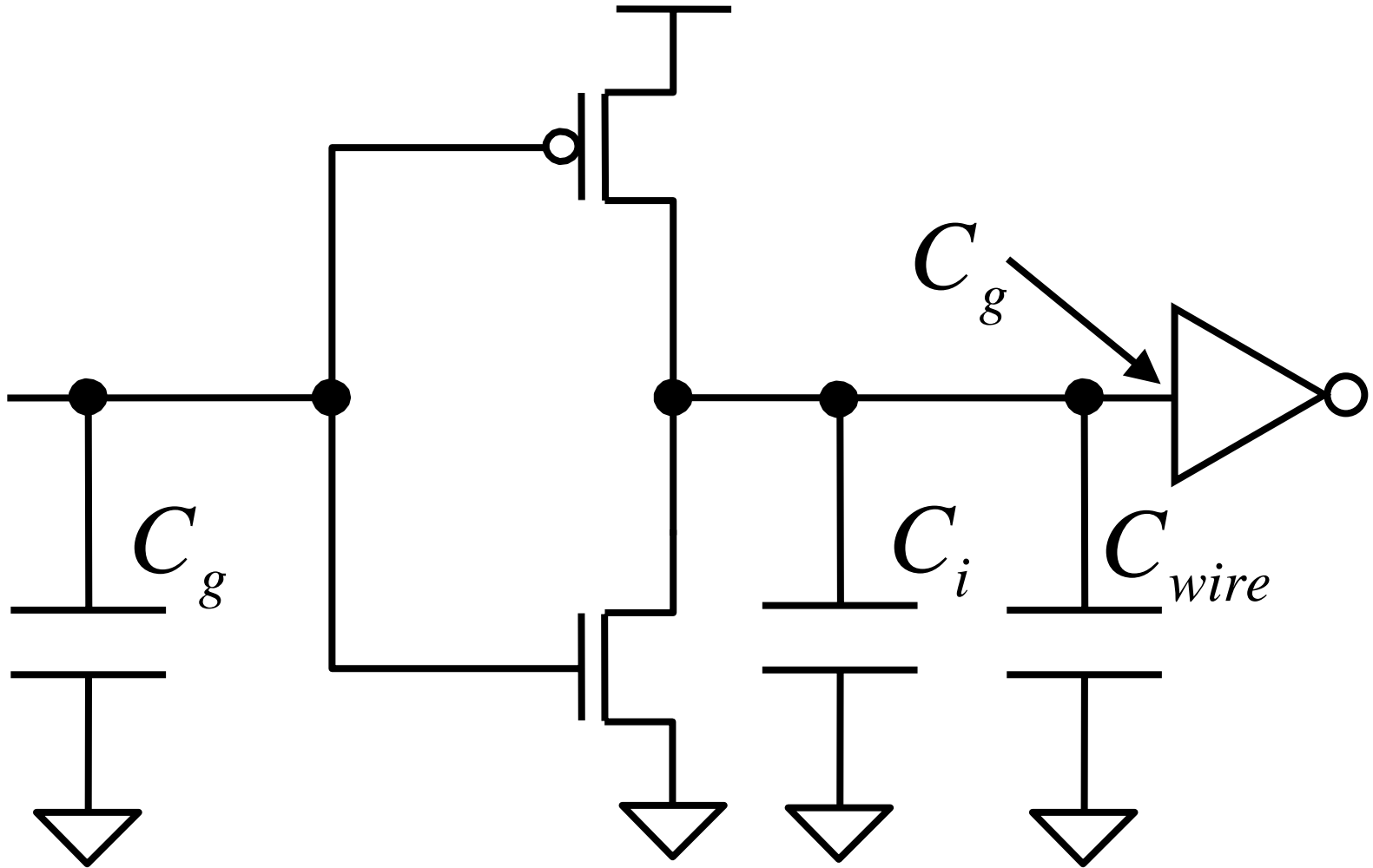
$$c\Delta L \frac{\partial V_i}{\partial t} = \frac{(V_{i+1} - V_i) + (V_{i-1} - V_i)}{r\Delta L}$$

Distributed RC Wire Step Response

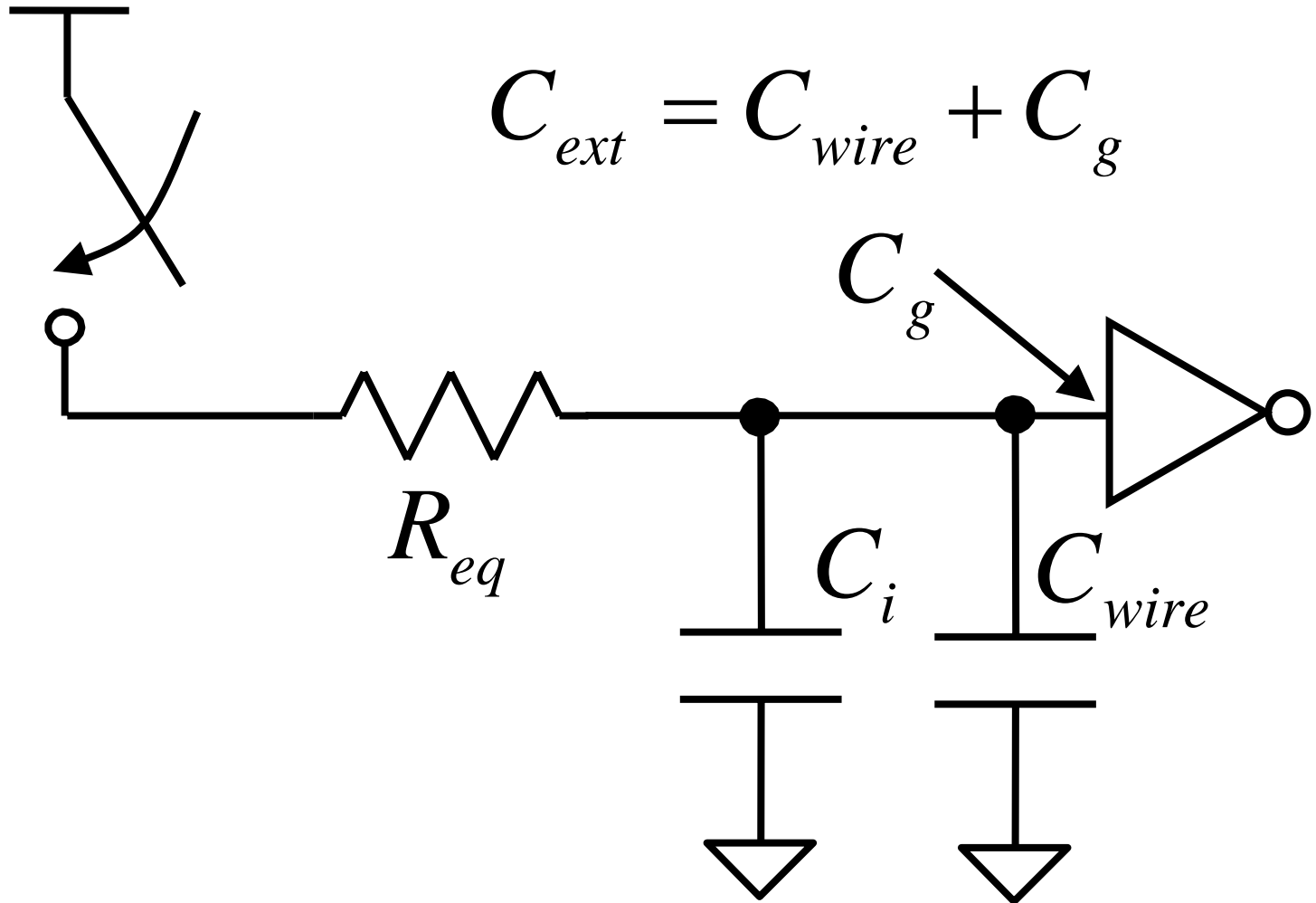


- **Step-response of an RC Wire as a function of time and space (Fig. 4-15, p. 157)**

Intrinsic (Self-Load) and Extrinsic Capacitance



RC Switch Model for Inverter Sizing



- **Model delay using ideal switch and resistor for MOSFET**

Unloaded Inverter Delay

- Estimate delay using ideal switch and resistor model (RC time constant):

$$t_{pd} \propto R_{eq} (C_i + C_{ext})$$

$$\propto R_{eq} C_i (1 + C_{ext} / C_i)$$

$$\propto t_{p0} (1 + C_{ext} / C_i)$$

- Define intrinsic inverter delay (with fudge factor):

$$t_{p0} = 0.69 R_{eq} C_i$$

- C_i consists of source / drain and overlap capacitance

Fastest Loaded Inverter Sizing

- **Decrease delay by enlarging transistor (increases current, decreases R_{eq}) by factor S :**

$$t_{pd} = 0.69 \frac{R_{eq}}{S} SC_i \left(1 + \frac{C_{ext}}{SC_i} \right)$$

$$t_{pd} = t_{p0} \left(1 + \frac{C_{ext}}{SC_i} \right)$$

- **Intrinsic delay independent of sizing**
- **Infinite S yields fastest gate (eliminates external load), reducing delay to intrinsic in the limit**

Relating Self-Load to Gate Capacitance

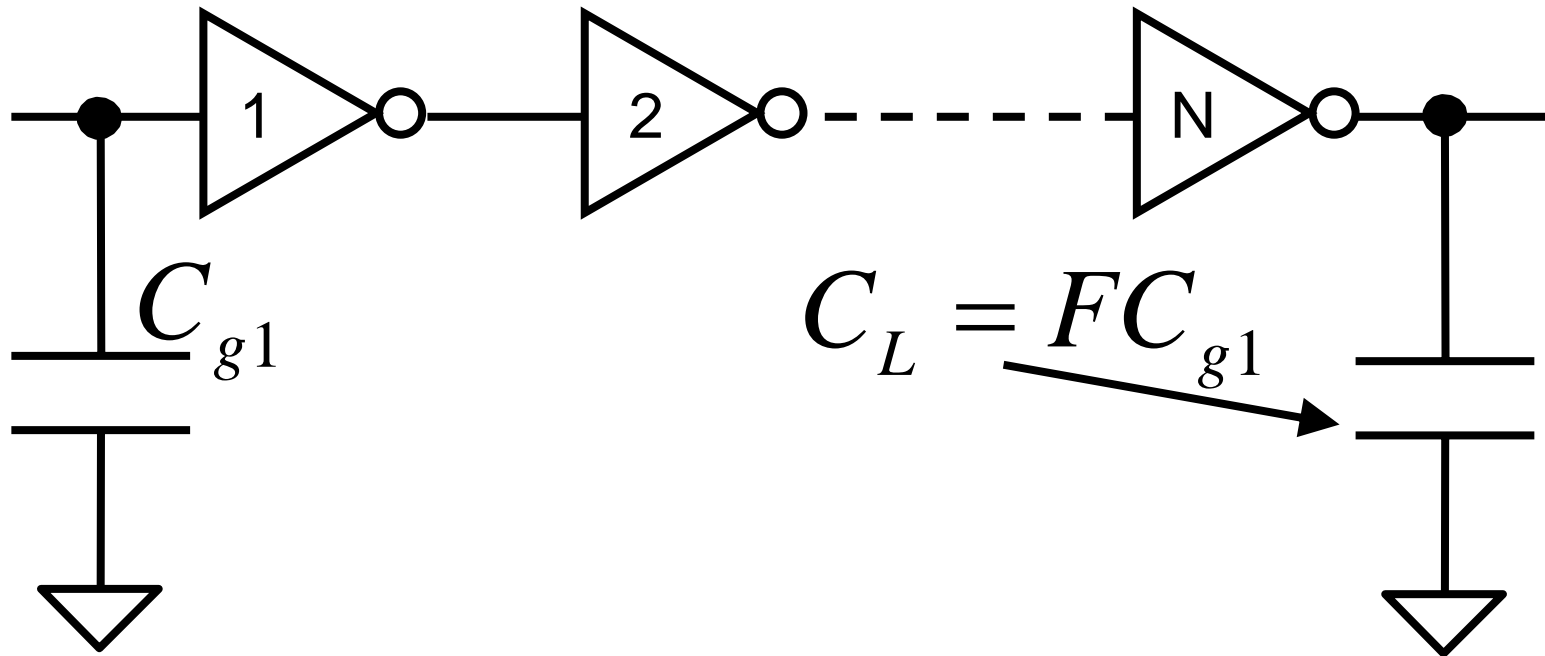
- Increasing transistor sizing enlarges self-load and gate input capacitance
- Convenient to relate them by a constant factor γ (γ around 1 in submicron processes)

$$C_i = \gamma C_g$$

$$t_{pd} = t_{p0} \left(1 + \frac{C_{ext}}{\gamma C_g} \right) = t_{p0} (1 + f / \gamma)$$

- f is effective fanout of gate
- Delay depends only on ratio between external load capacitance and input capacitance

Inverter Chain Sizing for Minimum Delay



- Using inverter sizing, want to minimize delay of driving large load C_L
- Optimize using equivalent resistance delay equation derived in previous slides

Total Inverter Chain Delay

- Delay of the j th inverter stage is (ignoring wiring):

$$t_{pd,j} = t_{p0} \left(1 + \frac{C_{g,j+1}}{\gamma C_{g,j}} \right) = t_{p0} \left(1 + \frac{f_j}{\gamma} \right)$$

- Total delay is:

$$T_{pd} = \sum_{j=1}^N t_{pd,j} = t_{p0} \sum_{j=1}^N \left(1 + \frac{C_{g,j+1}}{\gamma C_{g,j}} \right)$$

where

$$C_{g,N+1} = C_L = FC_{g,1}$$

Optimal Inverter Sizing for Minimum Delay

- **Minimize delay by taking partial derivatives wrt $C_{g,j}$, set them equal to 0**
 - N-1 equations in N unknowns
 - Solution for jth inverter is geometric mean of its neighbors sizing:

$$C_{g,j} = \sqrt{C_{g,j-1} C_{g,j+1}}$$

- **Implies each inverter has constant scale-up factor f_j :**

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

- **Minimum delay:** $T_{pd} = N t_{p0} \left(1 + \sqrt[N]{F} / \gamma \right)$

Optimal Inverter Stages for Minimum Delay

- **Delay trade off in the number of stages N**
 - Too many stages, intrinsic delay term dominates
 - Too few stages, extrinsic delay term due to fanout ratio dominates
- **Taking derivative of T_{pd} wrt N and setting equal to zero yields equation to solve for scale up factor for optimal number of stages:**

$$\gamma + \sqrt[N]{F} - \frac{\sqrt[N]{F} \ln F}{N} = 0$$

Optimal Inverter Stages for Minimum Delay

- Taking derivative of T_{pd} wrt N and setting equal to zero yields scale up factor for optimal number of stages:

$$f = e^{\left(1 + \frac{\gamma}{f}\right)}$$

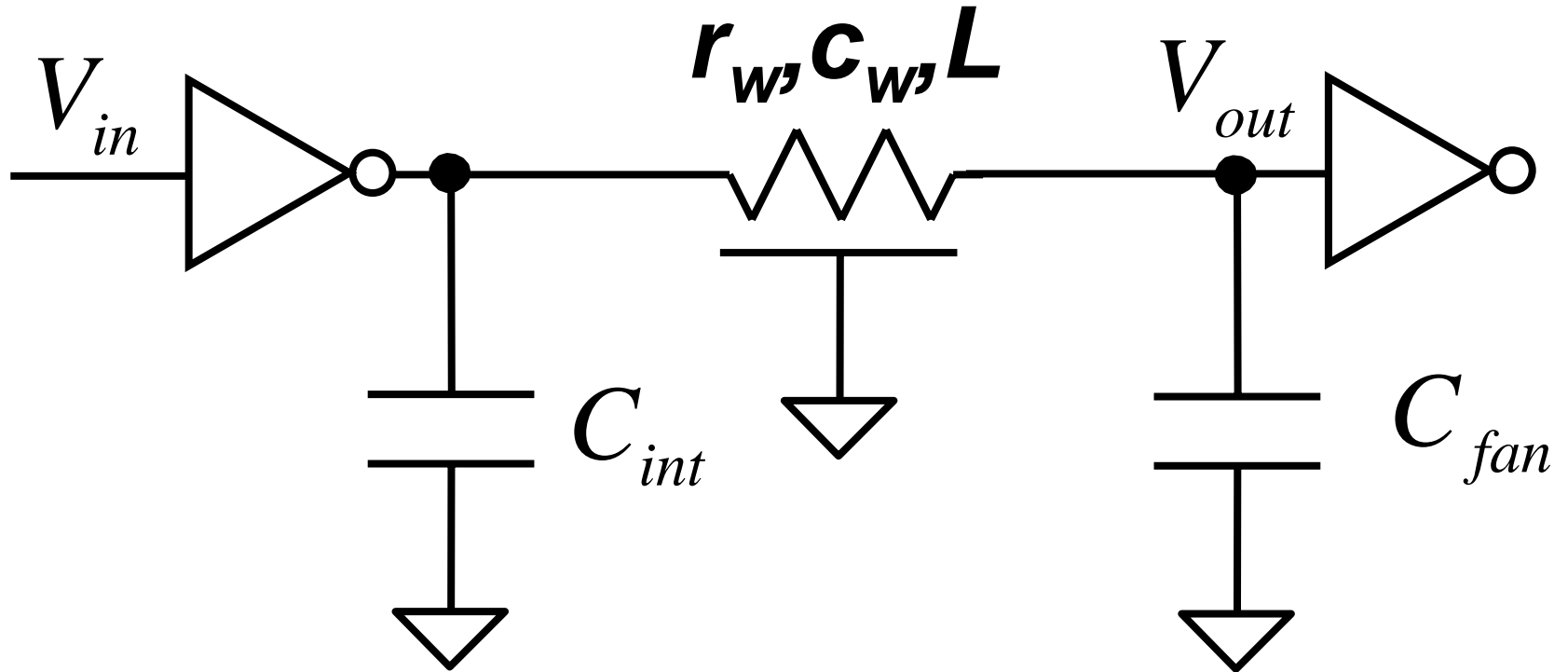
- Closed form solution when $\gamma = 0$, $N = \ln(F)$

$$f = e = 2.71828$$

- For more typical case of $\gamma = 1$, $f = 3.6$

- Often choose $f = 4$

Long Interconnect Delay



- Calculate delay by applying Elmore delay expression and switch RC model for driving inverter

Long Interconnect Delay

- **Plugging in:**

$$t_{pd} = 0.69R_{dr}C_{int} + \\ (0.69R_{dr} + 0.38R_w)C_w + \\ 0.69(R_{dr} + R_w)C_{fan}$$

- **R_{dr} : driver resistance**
- **C_{int} : driver intrinsic capacitance**
- **R_w : total wire resistance**
- **C_w : total wire capacitance**
- **C_{fan} : input capacitance of fanout gate**

Long Interconnect Delay

- Rearranging terms:

$$t_{pd} = 0.69R_{dr}(C_{int} + C_{fan}) + \\ 0.69(R_{dr}c_w + r_w C_{fan})L + \\ 0.38(r_w c_w)L^2$$

- R_{dr} : driver resistance
- C_{int} : driver intrinsic capacitance
- r_w : per unit length wire resistance
- c_w : per unit length wire capacitance
- C_{fan} : input capacitance of fanout gate

Optimal Inverter Sizing for Delay Constraint

- **Problem:** given a maximum propagation delay time $t_{p,max}$, find number of stages N and scaleup factor f s.t. overall area is minimized (i.e., find a solution that sets $T_{pd} = t_{p,max}$)

$$T_{pd} = Nt_{p0}F^{1/N} \geq t_{p,max}$$

- **Solve numerically using integer programming (see Figure 9-8, Rabaey p. 455)**
 - N between 1 and 10 for $F = 100-10^4$ and $t_{p,max}/t_{p0}$ between 10 and 10^4

Driver Area for Delay Constraint Sizing

- Driver area A_{dr} can be derived as function of minimum sized inverter area A_{min} :

$$A_{dr} = (1 + f + f^2 + \dots + f^{N-1}) A_{min}$$

- Summing the power series yields:

$$A_{dr} = \left(\frac{f^N - 1}{f - 1} \right) A_{min} = \frac{F - 1}{f - 1} A_{min}$$

- Area inversely proportional to scaleup factor f

Driver Power for Delay Constraint Sizing

- Driver power P_{dr} can be derived as function of minimum sized inverter intrinsic capacitance C_i :

$$P_{dr} = (1 + f + f^2 + \dots + f^{N-1}) C_i V_{DD}^2 f_{dr}$$

- Summing the power series yields:

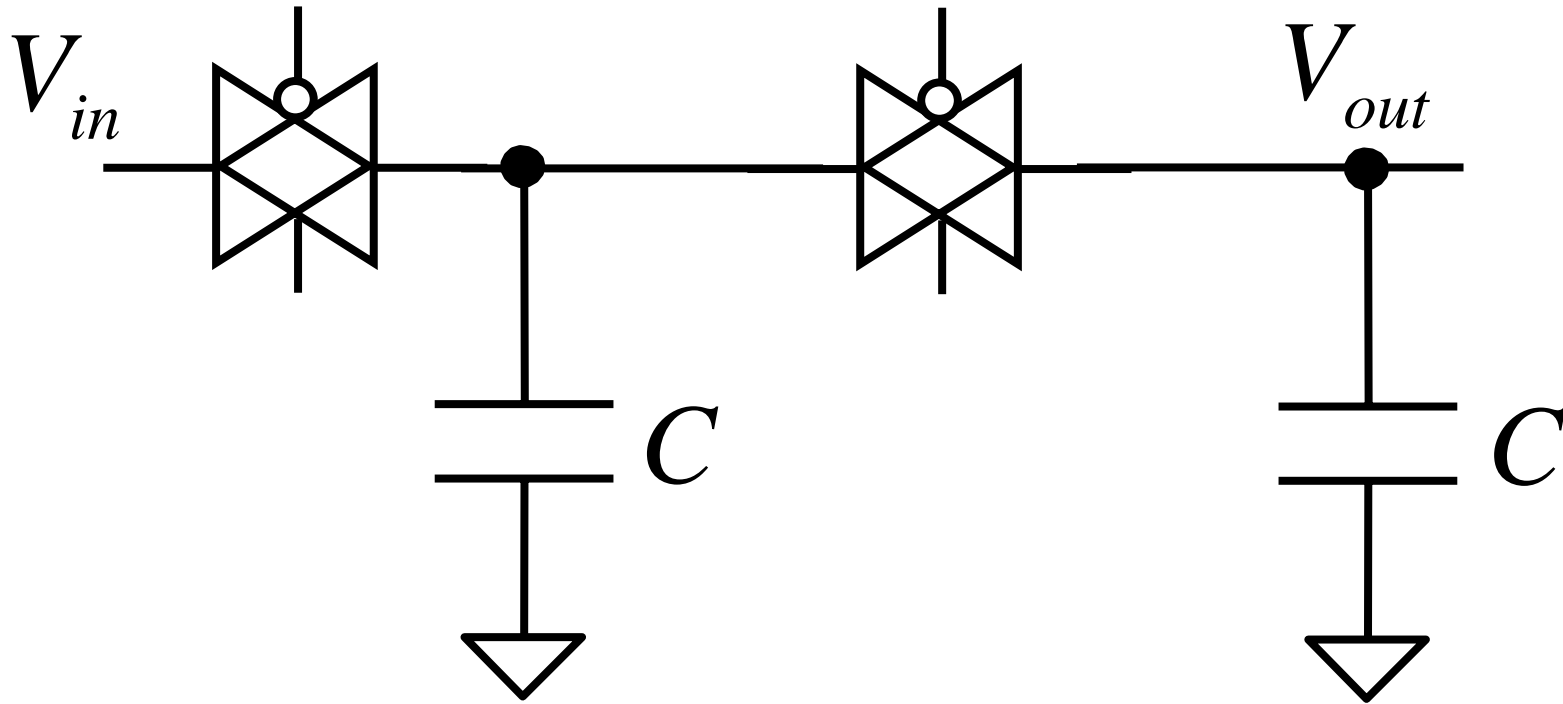
$$P_{dr} = \frac{F - 1}{f - 1} C_i V_{DD}^2 f_{dr} \approx \frac{C_L}{f - 1} V_{DD}^2 f_{dr}$$

- Power inversely proportional to scaleup factor f , but f_{dr} constrained by $t_{p,max}$

Example

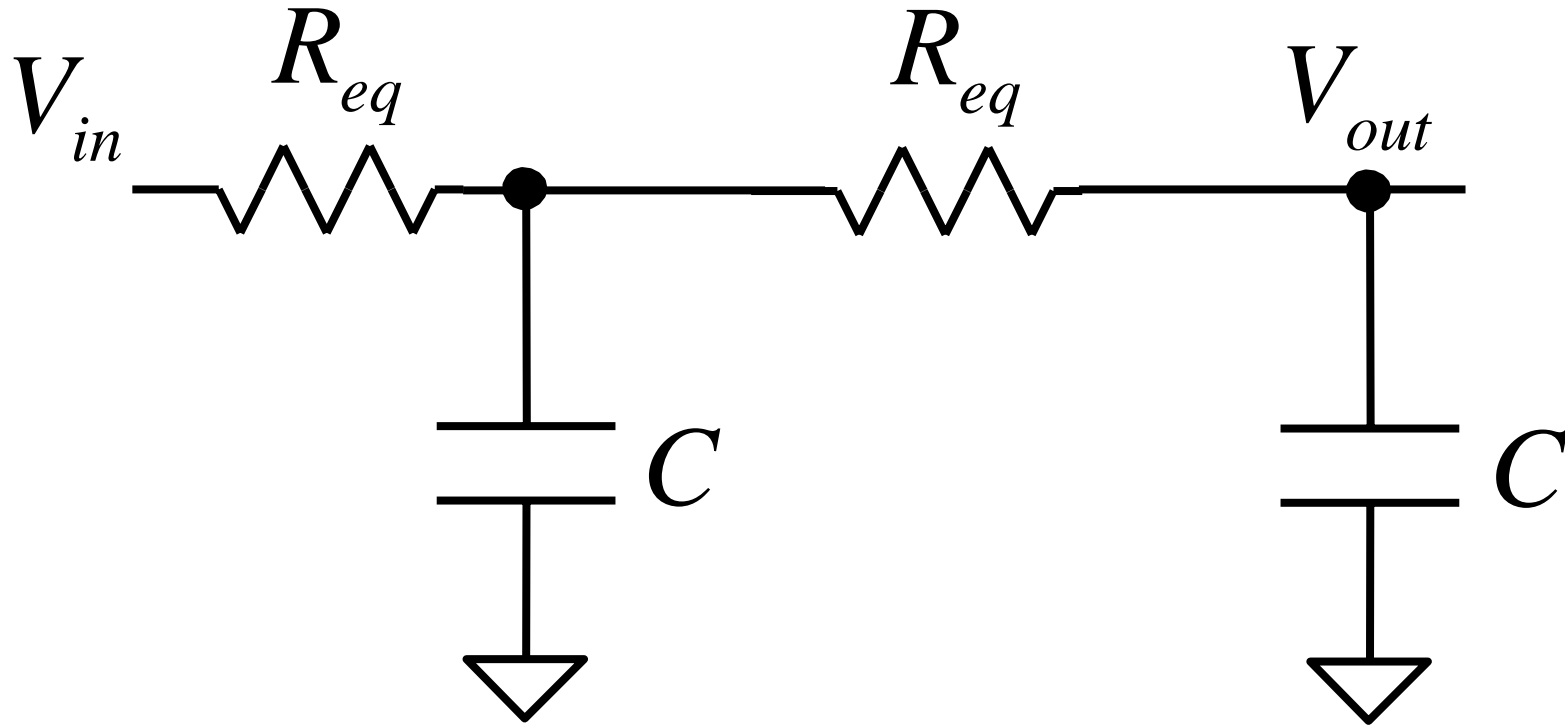
- **Off-chip Capacitor Driver Sizing**

Long Transmission Gate Chain Delay



- Calculate delay for N transmission gates in series by applying Elmore delay expression and switch RC model for transmission gate

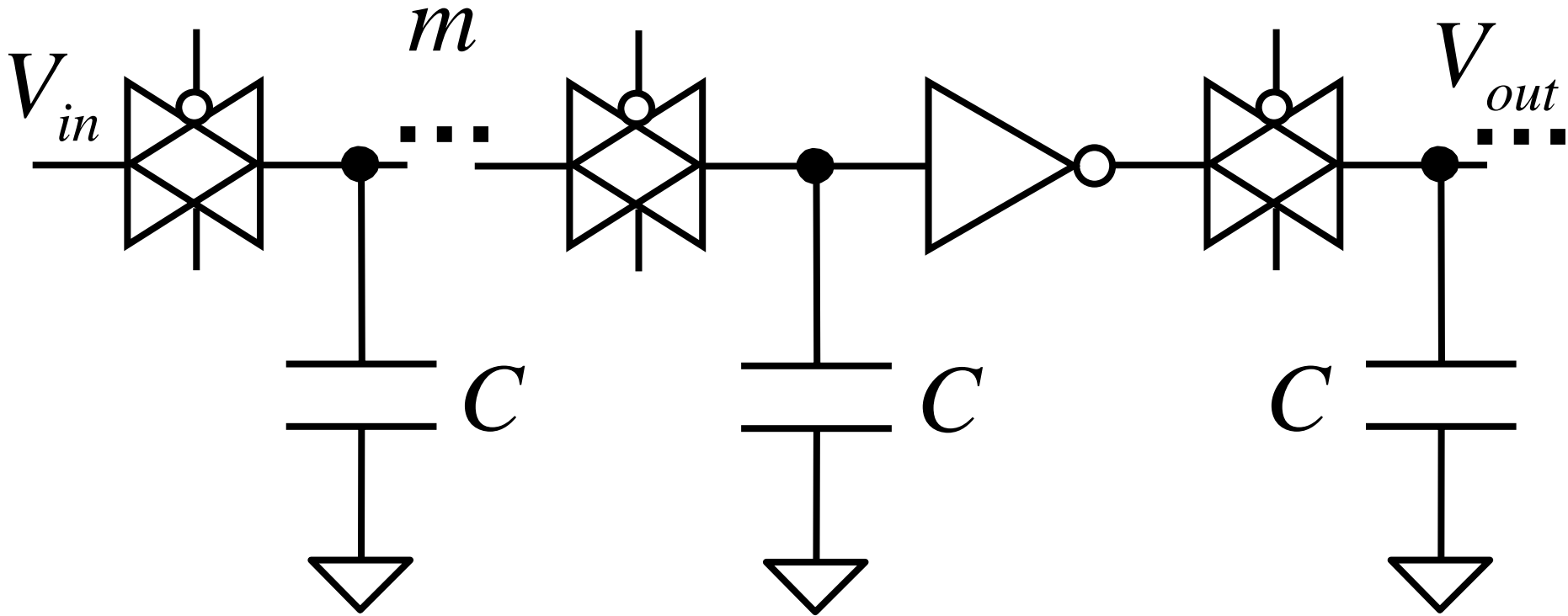
Long Transmission Gate Chain Delay



- Elmore delay expression:

$$\tau_D = 0.69 \sum_{k=1}^N CR_{eq} = 0.69 R_{eq} C \frac{N(N+1)}{2}$$

Buffered Transmission Gate Chain



- **Insert buffers (inverters) every m transmission gates to reduce delay**

Buffered Transmission Gate Chain Delay

- Assume inverter has delay t_{inv} inserted every m transmission gates
- Plugging in:

$$\tau_D = 0.69 \left[\frac{N}{m} CR_{eq} \frac{m(m+1)}{2} \right] + \left(\frac{N}{m} - 1 \right) t_{inv}$$

$$\tau_D = 0.69 \left[CR_{eq} \frac{N(m+1)}{2} \right] + \left(\frac{N}{m} - 1 \right) t_{inv}$$

- Buffering results in linear dependence on number of switches N instead of quadratic

Optimal Buffer Insertion for Minimum Delay

- Minimize delay by taking partial derivative wrt m , set equal to 0:

$$m_{opt} = 1.7 \sqrt{\frac{t_{inv}}{R_{eq} C}}$$

- Buffer insertion period depends on ratio of inverter delay and transmission gate switch RC delay
- Find minimum delay by plugging in m_{opt}

Minimum Buffered Transmission Gate Delay

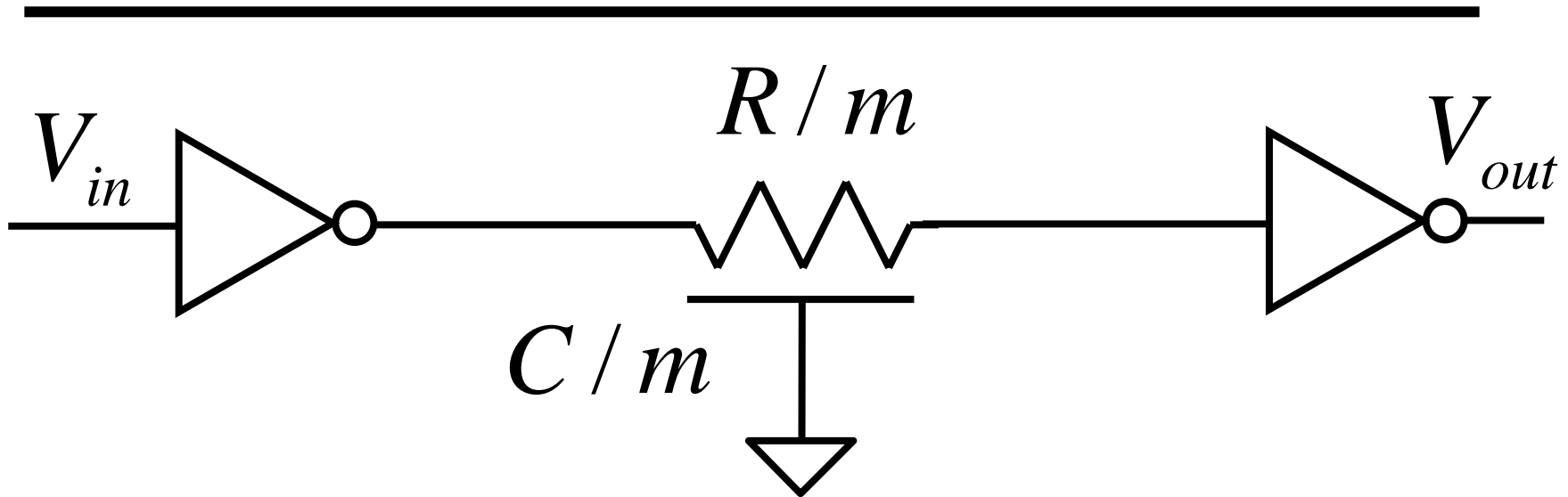
$$\tau_D = 0.69 \left[CR_{eq} \frac{N(m_{opt} + 1)}{2} \right] + \left(\frac{N}{m_{opt}} - 1 \right) t_{inv}$$

$$\tau_D = 0.69 \left[CR_{eq} \frac{N \left(1.7 \sqrt{\frac{t_{inv}}{CR_{eq}}} + 1 \right)}{2} \right] + \left(\frac{N}{1.7 \sqrt{\frac{t_{inv}}{CR_{eq}}}} - 1 \right) t_{inv}$$

$$\tau_D \approx 0.69 \left[\frac{1.7N}{2} \sqrt{t_{inv} CR_{eq}} \right] + \frac{N \sqrt{t_{inv} CR_{eq}}}{1.7}$$

- **Geometric mean dependence should look familiar!**

Buffered Long Wire



- Insert inverters (repeaters) m times in a long wire of total resistance R and total capacitance C
- Assume inverters have delay t_{inv} , wire of length L and per unit length resistance r and capacitance c
- Find optimum number of repeaters m_{opt} as above

Optimal Repeater Insertion for Minimum Delay

- Minimize delay by taking partial derivative wrt m , set equal to 0:

$$m_{opt} = L \sqrt{\frac{0.38rc}{t_{inv}}} = \sqrt{\frac{t_{wire}}{t_{inv}}}$$

- t_{wire} is delay of unbuffered wire
- Corresponding minimum delay by plugging in m_{opt}

$$\tau_D = 2\sqrt{t_{wire}t_{inv}}$$

- Optimal delay found when each wire segment delay equals inverter delay

Long Interconnect Delay With Repeater Size

- Taking optimal repeater sizing (S) into account:

$$\tau_D = m \left(\begin{array}{l} 0.69 \frac{R_{dr}}{S} \left[S \gamma C_{dr} + \frac{c_w L}{m} + S C_{dr} \right] + \\ 0.69 r_w \left(\frac{L}{m} \right) (S C_{dr}) + 0.38 (r_w c_w) \left(\frac{L}{m} \right)^2 \end{array} \right)$$

- R_{dr} : minimum-sized driver resistance
- C_{dr} : minimum-sized driver intrinsic capacitance
- r_w : per unit length wire resistance
- c_w : per unit length wire capacitance

Optimal Repeater Design for Minimum Delay

- Minimize delay in usual fashion:

$$m_{opt} = L \sqrt{\frac{0.38 r_w c_w}{0.69 R_{dr} C_{dr} (\gamma + 1)}} = \sqrt{\frac{t_{wire}}{t_{p0} (1 + 1/\gamma)}}$$

$$\tau_D = (1.38 + 1.02 \sqrt{1 + \gamma}) L \sqrt{R_{dr} C_{dr} r_w c_w}$$

$$S_{opt} = \sqrt{\frac{R_{dr} c_w}{r_w C_{dr}}}$$

- $t_{p0}(1+1/\gamma)$ is delay of fanout of 1 ($f = 1$) inverter

Optimal Repeater Design for Minimum Delay

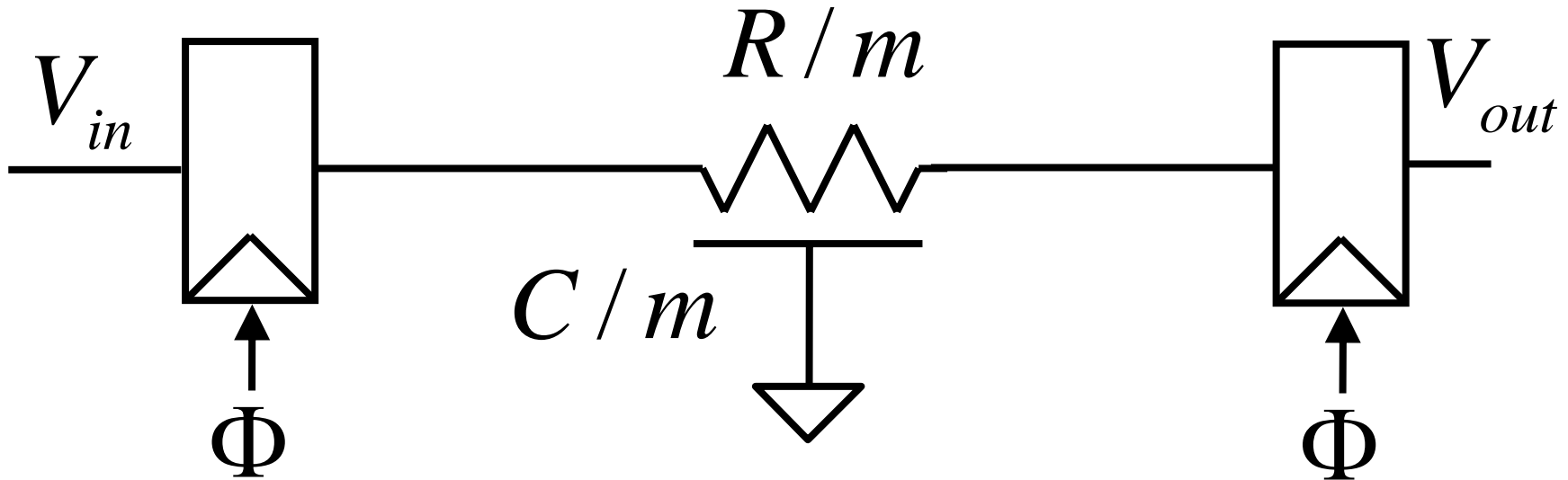
- Inserting repeaters linearizes delay dependence on length
- Optimal wire segment length exists for given technology and interconnect layer (Critical Length L_{crit}):

$$L_{crit} = \frac{L}{m_{opt}} = \sqrt{\frac{t_{p0} (1 + 1/\gamma)}{0.38 r_w c_w}}$$

- Delay of a segment of critical length:

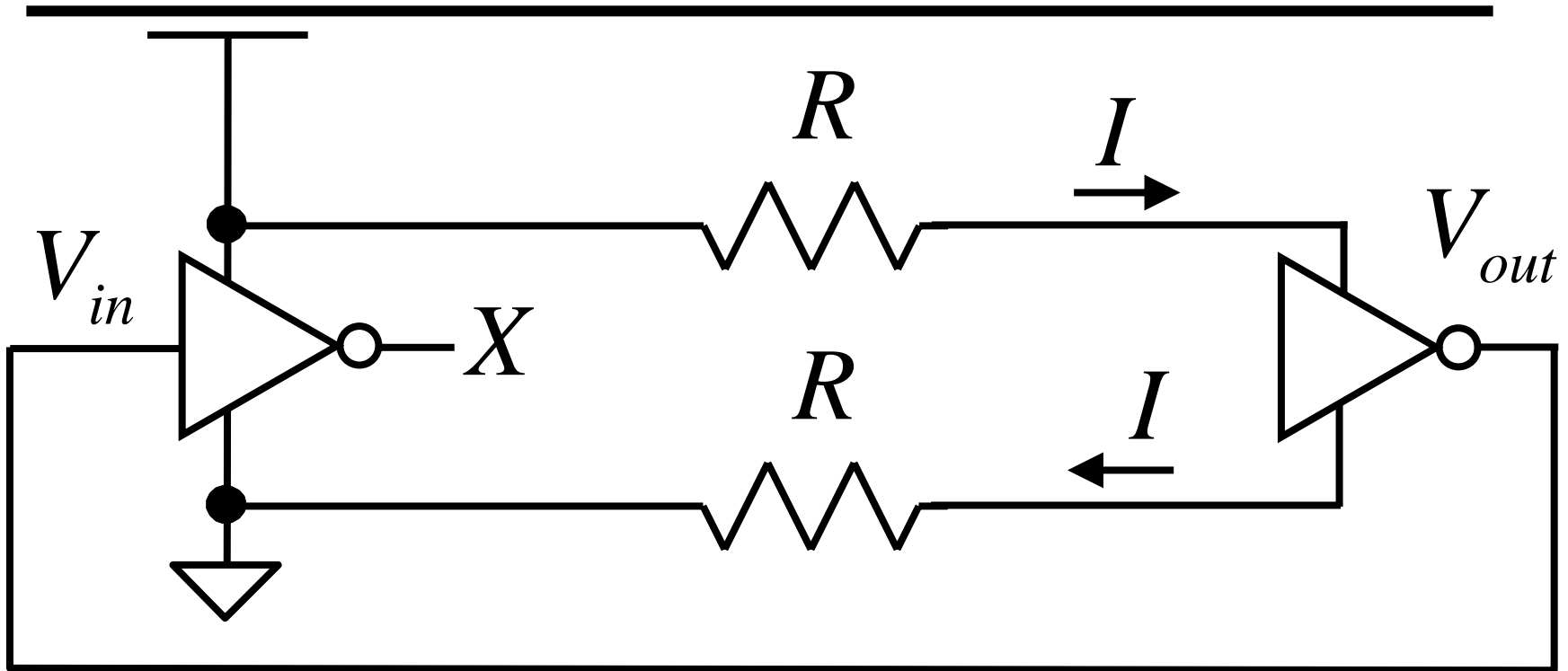
$$\tau_{D,crit} = \frac{\tau_D}{m_{opt}} = 2 \left(1 + \sqrt{\frac{0.69}{0.38(1 + \gamma)}} \right) t_{p0} (1 + 1/\gamma)$$

Pipelined Long Wire



- If minimum delay is longer than clock cycle (which can be the case for global wires), then registers can be inserted to pipeline the wire
- Latency increases or stays the same, but throughput of wire increases
- One of many architecture-level options for long wires

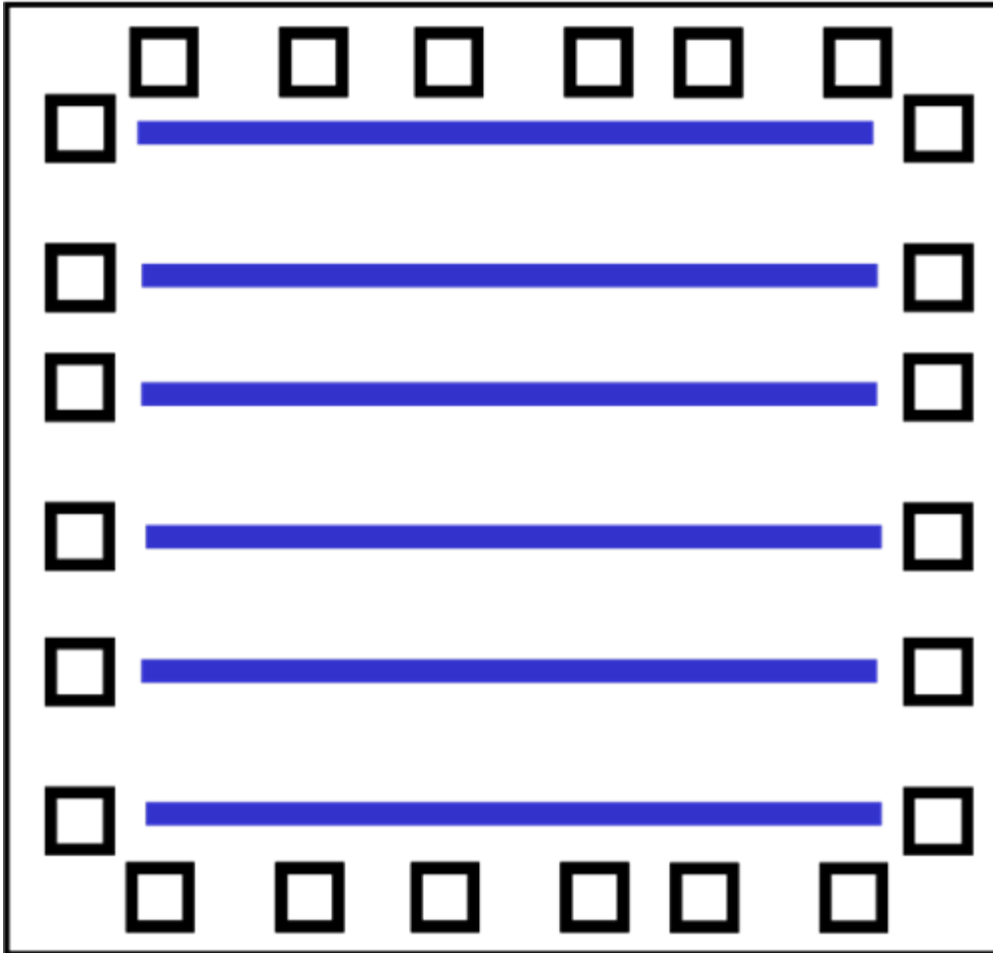
Power Distribution Network Resistance



- Finite resistance of power/ground lines (total resistance R) causes voltage drops which degrade noise margins:

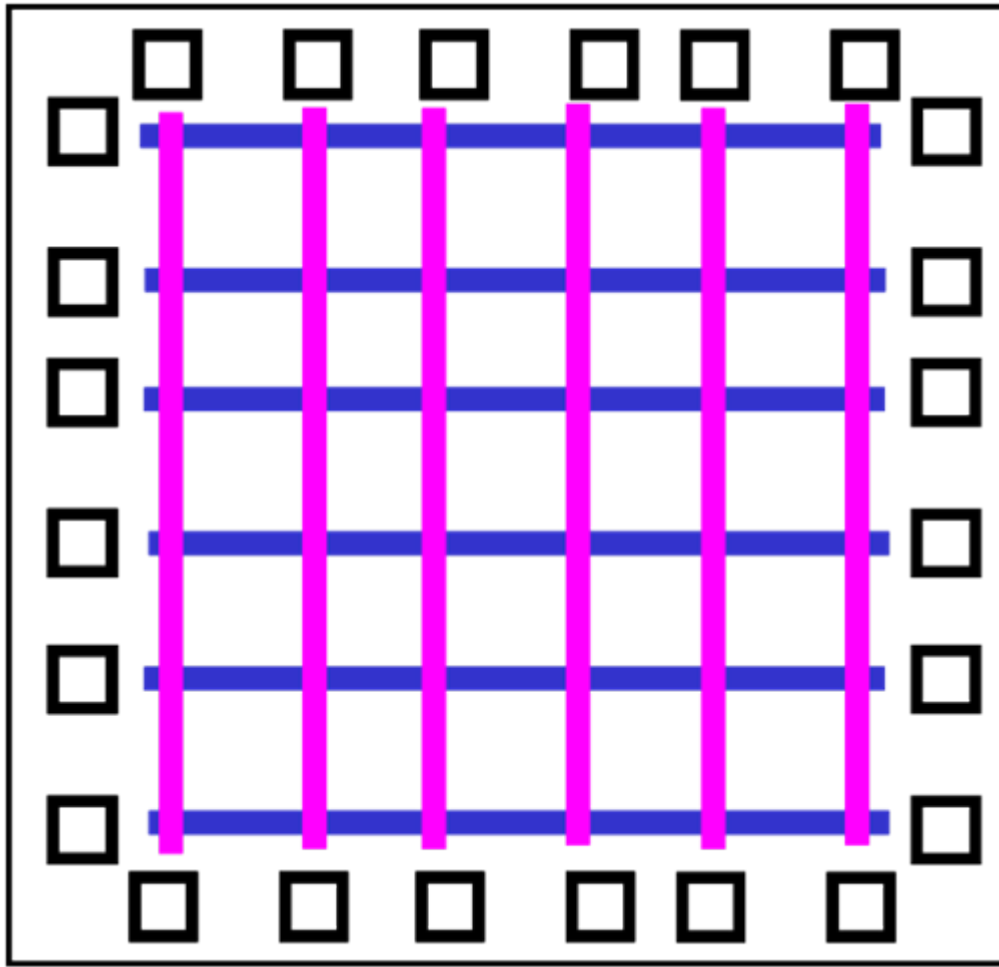
$$\Delta V < V_{out} < V_{DD} - \Delta V$$

Single Layer Power Grid



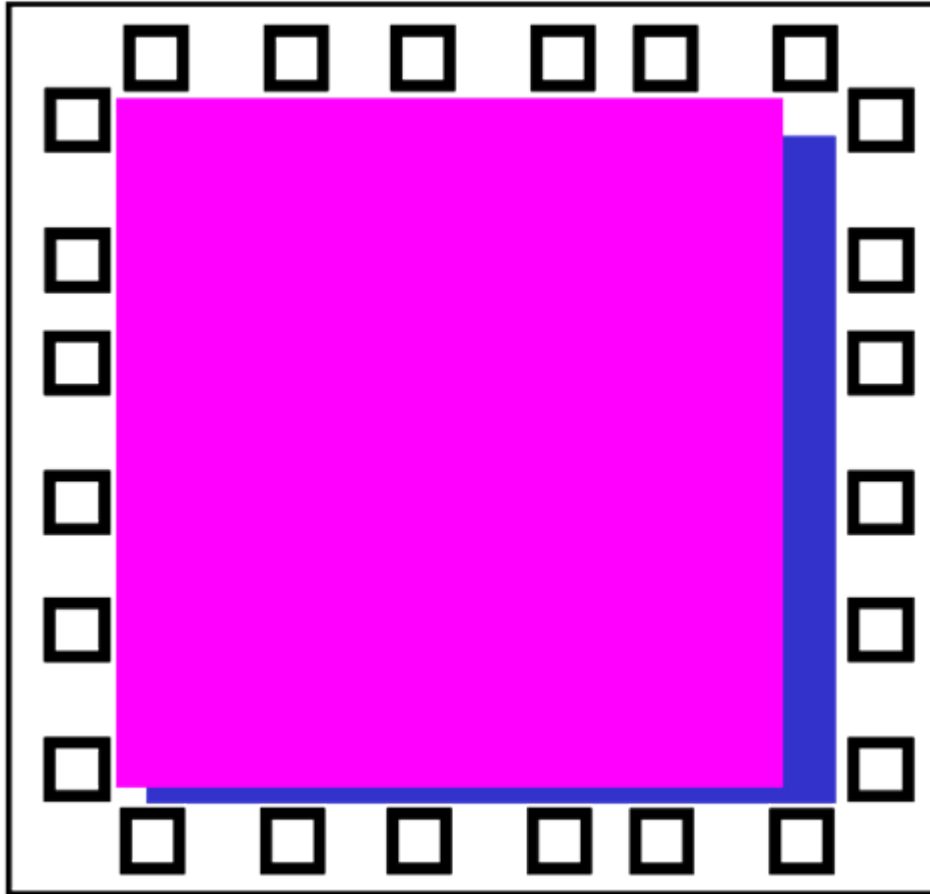
- Power routed vertically or horizontally on same layer
- VDD/GND brought in from two edges of chip
- Local power grids strapped to this grid and routed to lower metal layers

Dual Layer Power Grid



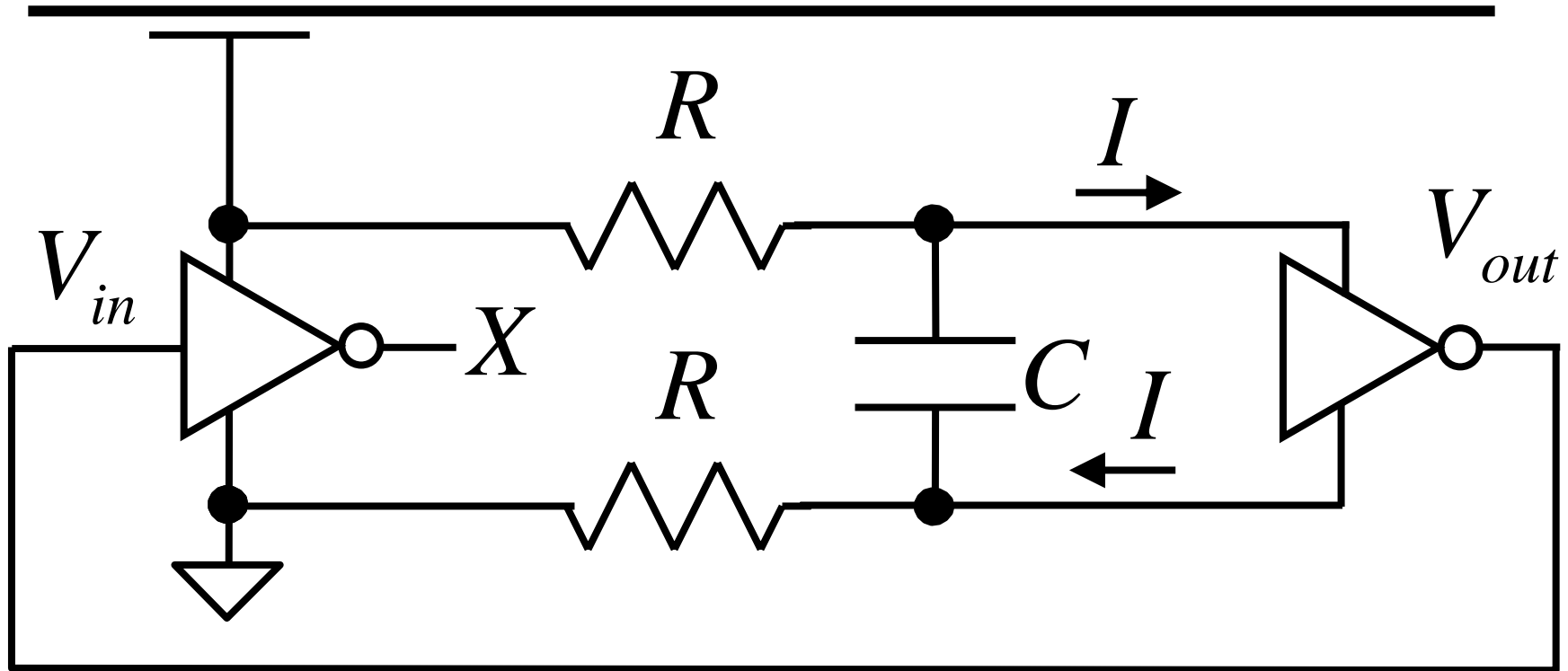
- Power routed vertically and horizontally on two layers
- VDD/GND brought in from all four edges of chip
- Local power grids strapped to this grid and routed to lower metal layers
- Can occupy 90% of two layers

Power Planes



- Devote two layers to power
- VDD/GND brought in from all four edges of chip
- Drastically reduces power supply resistance
- Can shield signal layers from crosstalk
- Need enough layers for routing

Power Distribution Bypass Capacitance



- **Local supply bypass capacitance provides low impedance path for high frequency (switching) currents to flow, reducing drops on output voltage**

Next Topic: Memories

- **Memory principles and circuits**
 - ROM: Read Only Memory
 - RWM (Read/Write Memory) or RAM (Random Access Memory)
 - DRAM, SRAM
 - Nonvolatile memories (Flash, PROM, EEPROM)