ECE520 – VLSI Design

Lecture 8: Interconnect Manufacturing and Modeling

Payman Zarkesh-Ha

Office: ECE Bldg. 230B
Office hours: Wednesday 2:00-3:00PM or by appointment
E-mail: pzarkesh@unm.edu

Review of Last Lecture

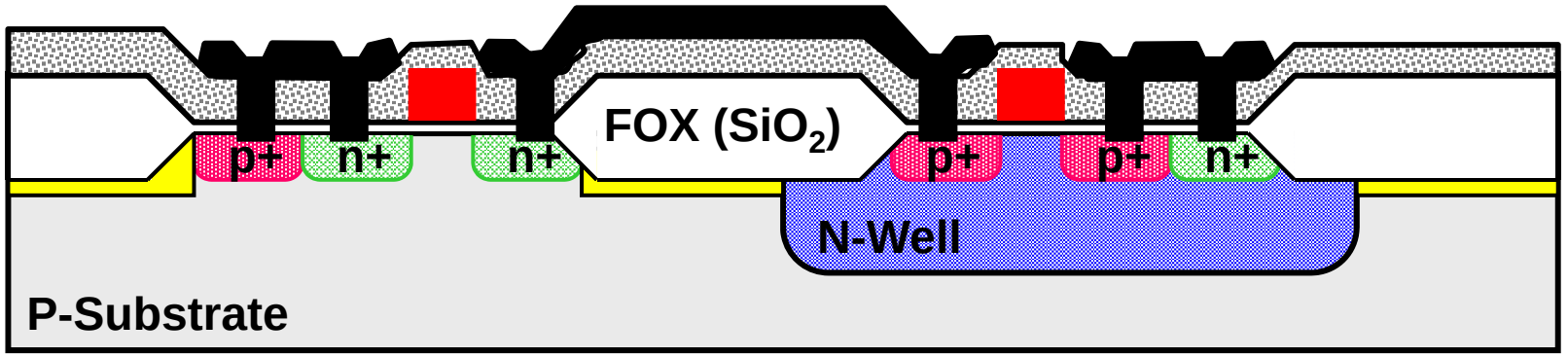
- **CMOS Manufacturing Process**
  - **Front-end Process**
- **Modern CMOS Process**
  - **Salisidation**
  - **Low Doped Drain (LDD)**
  - **Shallow Trench Isolation (STI)**

Today's Lecture

- **Interconnect Manufacturing Process**
  - **Back-end Process (Conventional)**
  - **Back-end Process (Modern – Dual Damascene)**
- **Interconnect Modeling**
  - **Parasitic Capacitance**
  - **Parasitic Resistance**
  - **Parasitic Inductance**
  - **Delay Estimation Techniques**

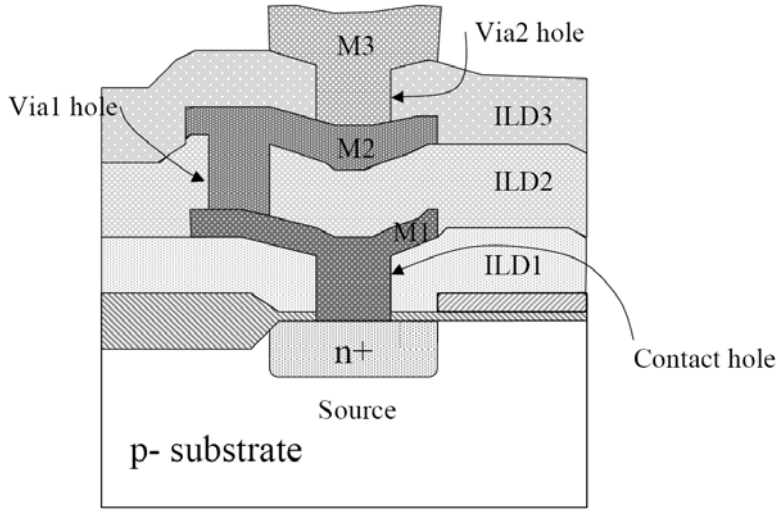
Conventional Interconnect Process

21. Deposit ILD (SiO_2)
22. Pattern and etch contact holes
23. Sputter on M1
24. Pattern and etch M1



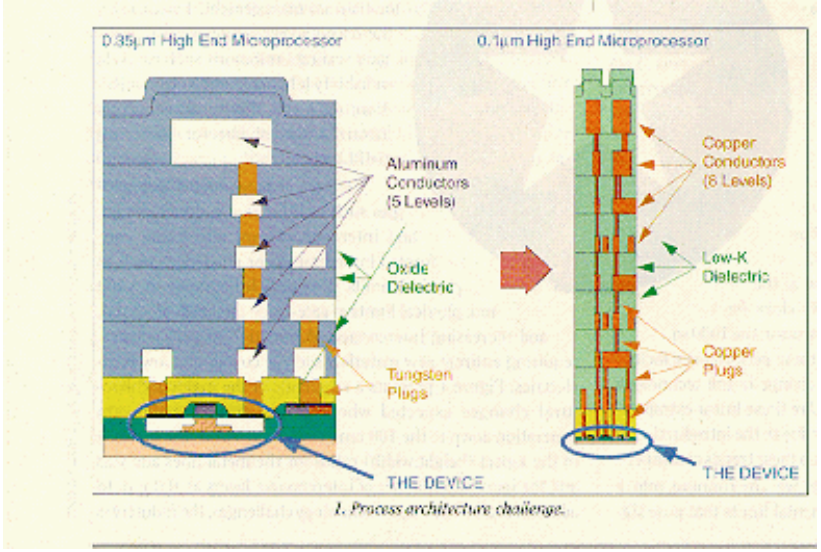
Problem with Conventional Process

- Adding more metal layers increases unevenness of the surface.
- Uneven surface (topography) causes yield problem with metal mask.
- Because of uneven surface, conventional interconnect process (reflow) requires several restriction on the layout design.
  - Stacking via is prohibited in conventional process.
  - The maximum number of metal layers is limited to 3.



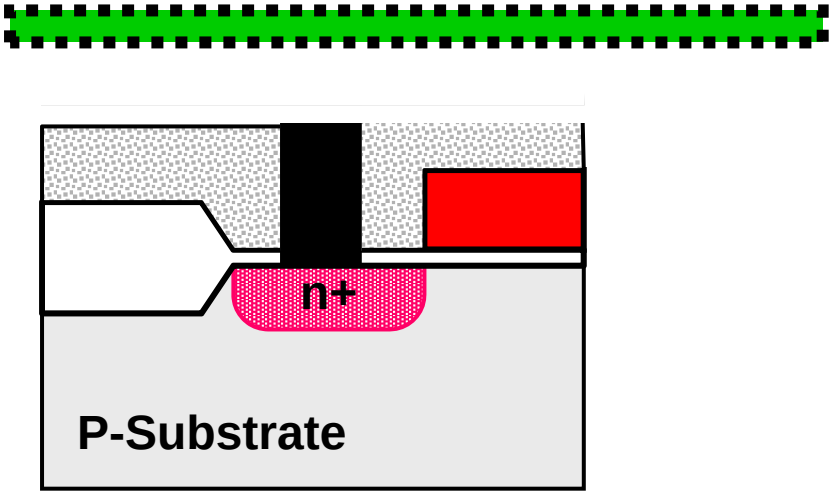
Problem with Conventional Process

- **Advanced VLSI technology requires many layers of interconnects**
  - **As transistor density/quantity increases there is more and more need for interconnect**
- **Planarization is therefore needed for back-end process in advanced VLSI technology.**



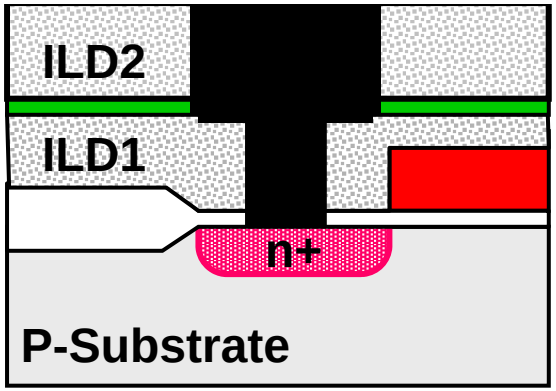
Planarization Process (Contact)

1. **Thick ILD (SiO_2) is laid down**
2. **Wafer's face is Chemically Mechanically Polished (CMP) in a slurry of etchents and aggregates**
3. **Contact holes are etched**
4. **Tungsten is sputtered on (forms tungsten plugs)**
5. **CMP is applied again (this time in a slurry that etches metal more so than SiO_2)**



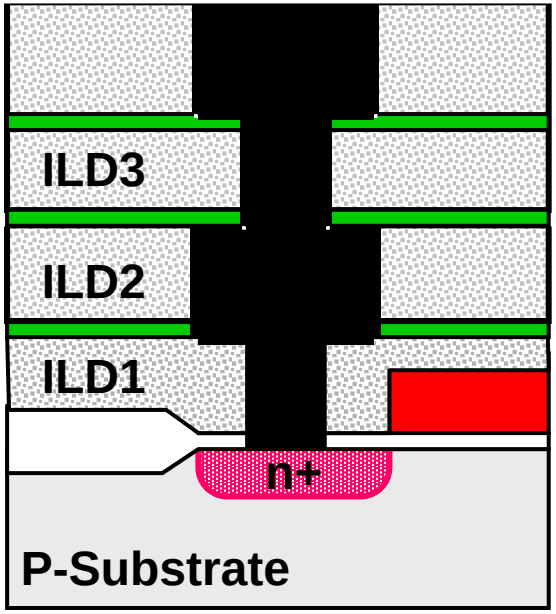
Dual Damascene Copper Metallization (M1)

1. Contact holes are filled with tungsten and plannarized.
2. An etch stop layer and ILD2 are deposited
3. The M1 pattern is etched in the ILD2
4. A barrier Cu seed layer is sputtered next
5. Copper is plated onto the surface of the wafer
6. CMP removes copper and barrier metal except in the M1 trench.



Dual Damascene Copper Metallization (M2)

1. After M1 CMP, a nitride etch stop and the ILD3 layers are deposited.
2. Next the Via1 holes are patterned and etched
3. The Via1 holes are filled with a sacrificial layer similar to photoresist (SLAM)
4. The M2 is patterned and etched into the ILD3
5. The SLAM is removed and a barrier/seed layer is deposited.
6. Copper plating and CMP finishes off the M2 process

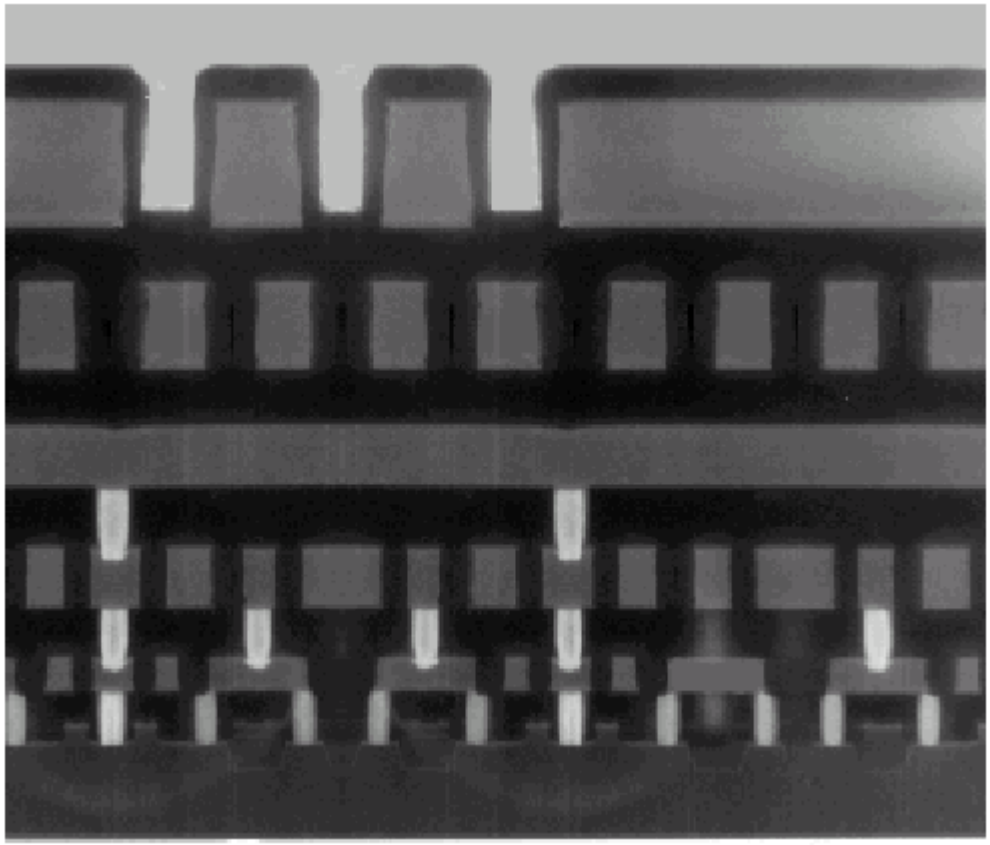


Example: Intel 0.25um backend process

5 metal layers
Ti/Al - Cu/Ti/TiN
Polysilicon dielectric

LAYER	PITCH	THICK	A.R.
Isolation	0.67	0.40	-
Polysilicon	0.64	0.25	-
Metal 1	0.64	0.48	1.5
Metal 2	0.93	0.90	1.9
Metal 3	0.93	0.90	1.9
Metal 4	1.60	1.33	1.7
Metal 5	2.56	1.90	1.5
	µm	µm	

Layer pitch, thickness and aspect ratio

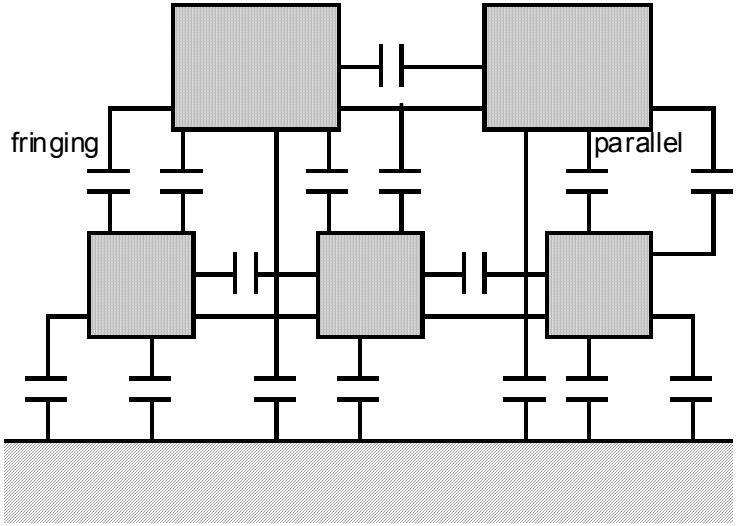


Interconnect Modeling

- **Interconnect parasitics**
  - **reduce reliability (crosstalk noise)**
  - **affect performance and power consumption**
- **Interconnect Modeling**
  - **Parasitic Capacitance**
  - **Parasitic Resistance**
  - **Parasitic Inductance**

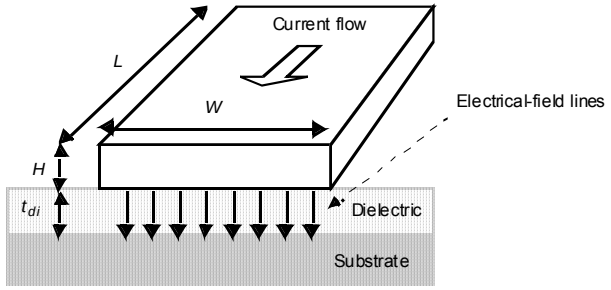
Interconnect Capacitance

- Extraction of interconnect capacitance in modern VLSI technology is very complicated because of
  - Non-homogenous dielectric (etch stop, barrier liner, etc.)
  - Complex pattern of neighboring interconnects (need 3D modeling)
- There are two types of capacitances:
  - Ground capacitances
  - Coupling capacitances



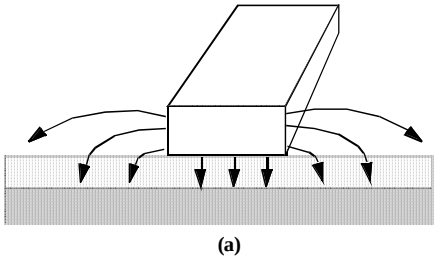
Interconnect Capacitance Modeling

- Often simple parallel plate model is used for hand calculation

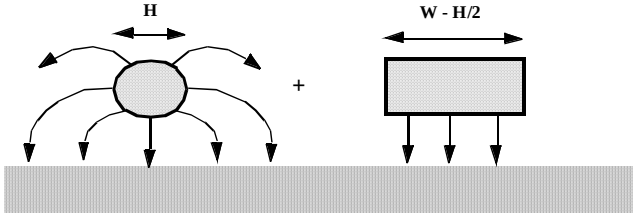


$$c_{int} = \frac{\varepsilon_{di}}{t_{di}} WL$$

- Capacitance model that includes fringing effect is complicated. However, the following simplified model can also be used



(a)



(b)

$$C_{int} = C_{PP} + C_{Fringe} \approx \frac{(w - H/2)\varepsilon_{di}}{t_{di}} + \frac{2\pi\varepsilon_{di}}{Ln(t_{di}/H)}$$

Capacitance of Interconnect Layers

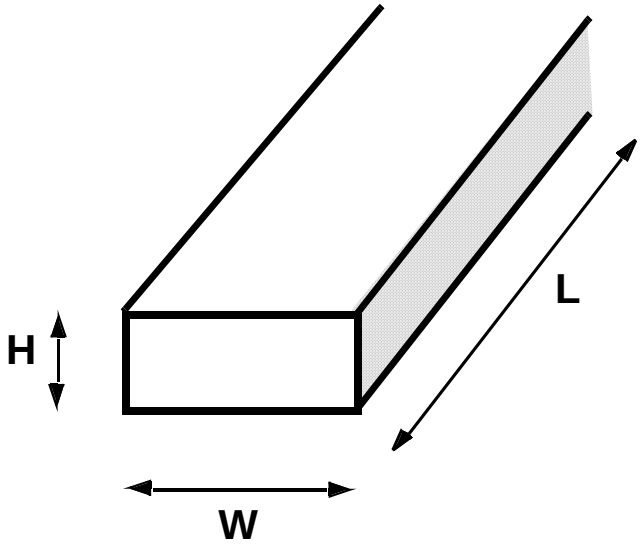
- To intentionally implement a capacitor in the chip, the layer is decided based on the value of the capacitor

	Field	Active	Poly	Al1	Al2	Al3	Al4
Poly	88						
	54						
Al1	30	41	57				
	40	47	54				
Al2	13	15	17	36			
	25	27	29	45			
Al3	8.9	9.4	10	15	41		
	18	19	20	27	49		
Al4	6.5	6.8	7	8.9	15	35	
	14	15	15	18	27	45	
Al5	5.2	5.4	5.4	6.6	9.1	14	38
	12	12	12	14	19	27	52

Area Capacitance
Fringe Capacitance

Interconnect Resistance

- Extraction of interconnect capacitance is simpler except for special cases (test chips where accurate resistance of a pattern is needed)
- Sheet resistance is an easy method of resistance measurement in layout (Only the metal aspect ratio is needed, no thickness information)



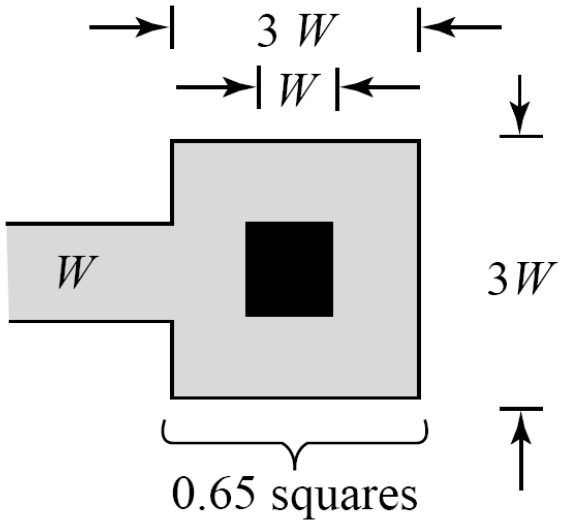
$$R = \frac{\rho L}{WH}$$

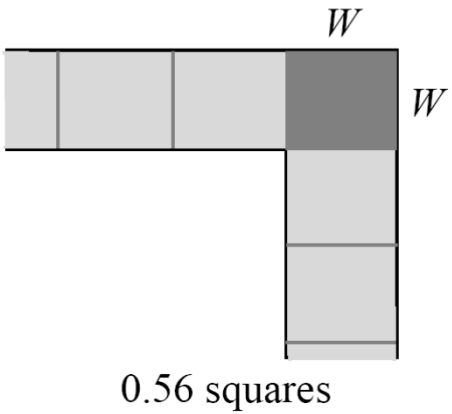
$$R = \left(\frac{\rho}{H}\right)\frac{L}{W} = R_{\square}\frac{L}{W}$$

Sheet Resistance

Sheet Resistance

- Measurement shows that the effective number of squares of the "dog bone" style contact region is 0.65 and for a 90º corner is 0.56





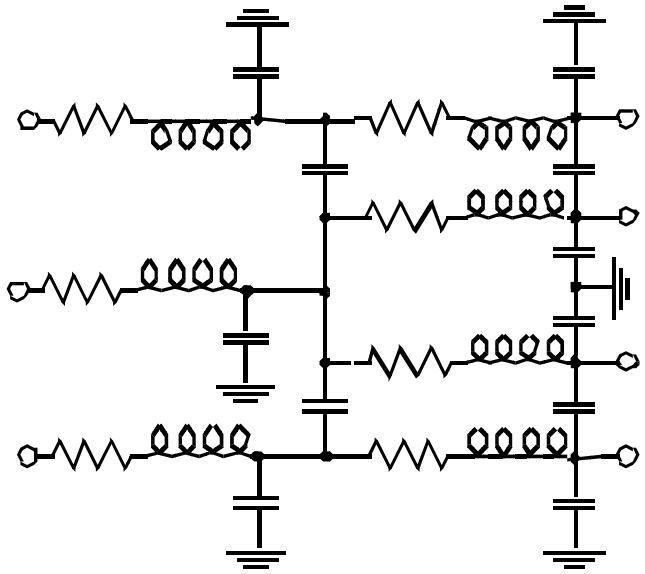
Sheet Resistance of Materials

- To intentionally implement a resistor in the chip, the material is decided based on the value of the resistor

Material	Sheet Resistance (Ω/❑)
n- or p-well diffusion	1000 – 1500
n^+, p^+ diffusion	50 – 150
n^+, p^+ diffusion with silicide	3 – 5
n^+, p^+ polysilicon	150 – 200
n^+, p^+ polysilicon with silicide	4 – 5
Aluminum	0.05 – 0.1

Interconnect Inductance

- Extraction and modeling of interconnect inductance is extremely hard because of
  - Non-identified return path
  - Unlike capacitance, the effect of inductance goes beyond nearest neighbors
- It is used only for specific nets such as clock and power supply interconnects
- Has not yet been used by industry for timing analysis



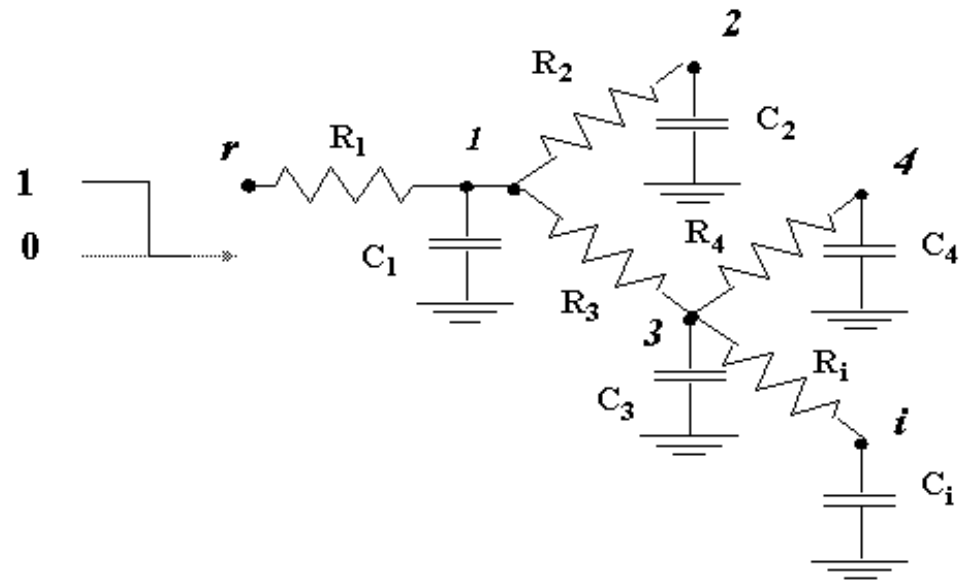
Dealing with Interconnect Parasitics

- **Reduce interconnect Capacitance**
  - Use better dielectric material (low-K dielectric)
  - Reduce wire-length (efficient layout)
  - Increase wire spacing
- **Reduce Interconnect Resistance**
  - Use better conductor material (Copper)
  - Reduce wire-length (efficient layout)
  - Increase wire width
- **Reduce Interconnect Inductance**
  - Use proper return path
  - Slow down the ramp time

Delay Estimation Techniques

- **SPICE Simulation**
  - Very slow – not practical for chip level analysis
  - Good for specific nets such as clocks or critical path
- **Asymptotic Waveform Evaluation (AWE)**
  - Is an industry standard for delay estimation
  - uses moment matching to determine a set of low frequency dominant poles that approximate the transient response
- **Elmore Delay Analysis**
  - Uses only the first moment (dominant pole)
  - Can be used for first order approximation in a complicated RC tree

Elmore Delay in Lumped RC Tree

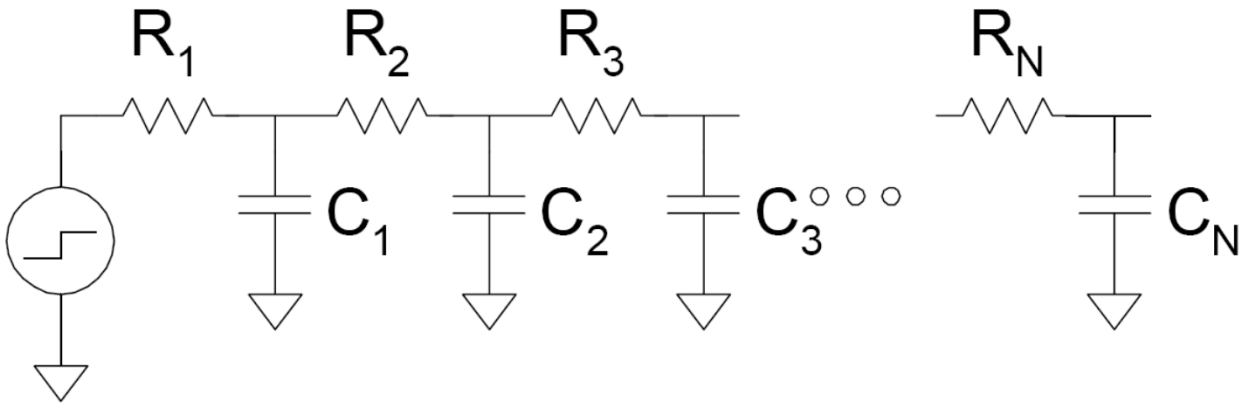


$$R_{ik} = \sum R_j \Rightarrow (R_j \in [path(s \to i) \cap path(s \to k)])$$

$$\tau_{Di} = \sum_{k=1}^{N} C_k R_{ik}$$

$$\tau_{Di} = R_1C_1 + R_1C_2 + (R_1 + R_3)C_3 + (R_1 + R_3)C_4 + (R_1 + R_3 + R_i)C_i$$

Example 1: Elmore Delay in RC Ladder



$$\tau_{Di} = \sum_{k=1}^{N} R_{ki} C_k$$
$$= R_1 C_1 + (R_1 + R_2) C_2 + \ldots + (R_1 + R_2 + \ldots + R_N) C_N$$

Example 2: Another Way of Elmore Delay

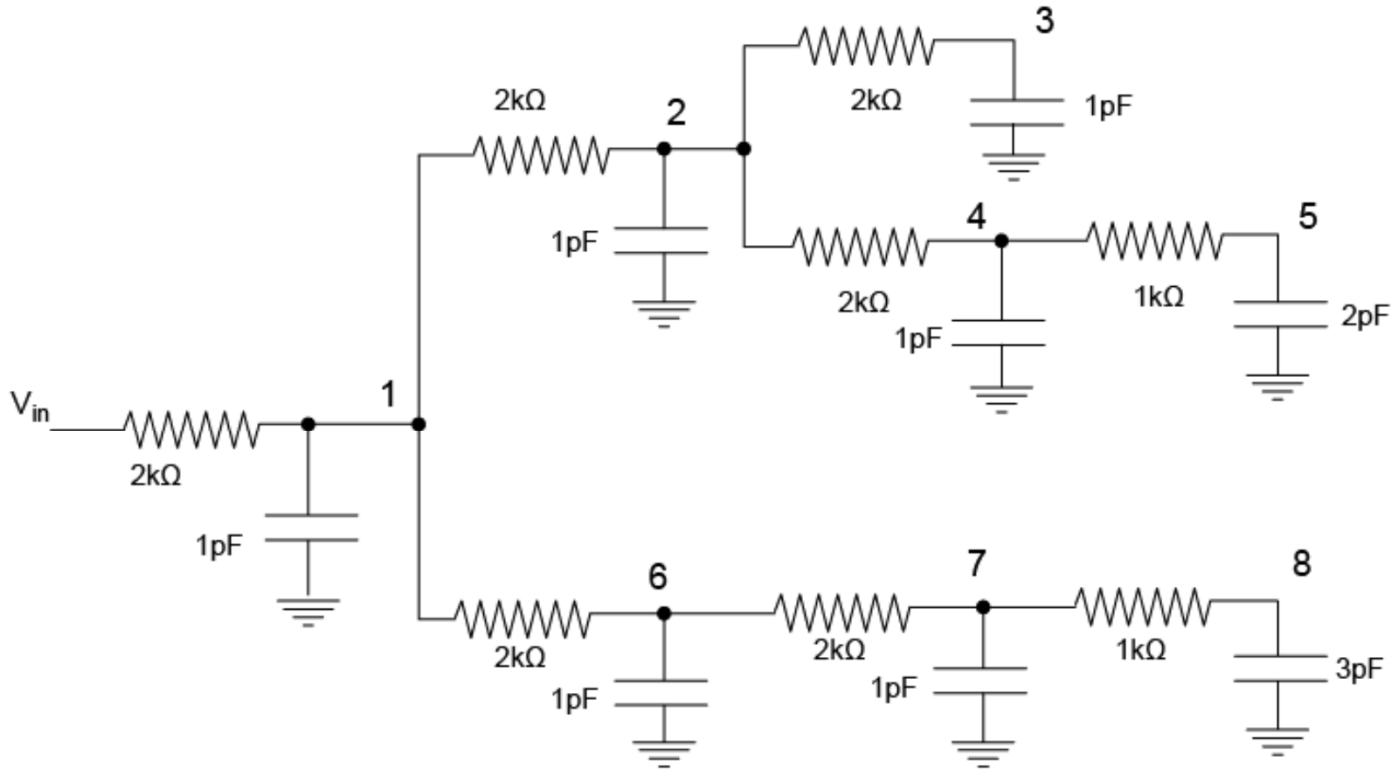
$\tau_{D1} = 2k\Omega \times (1pF + 1pF + 1pF + 1pF + 2pF + 1pF + 1pF + 3pF) = 22$ ns

$\tau_{D4} = \tau_{D1} + 2k\Omega \times (1pF + 1pF + 1pF + 2pF) + 2k\Omega \times (1pF + 2pF) = 38$ ns

$\tau_{D5} = \tau_{D4} + 1k\Omega \times 2pF = 40$ ns

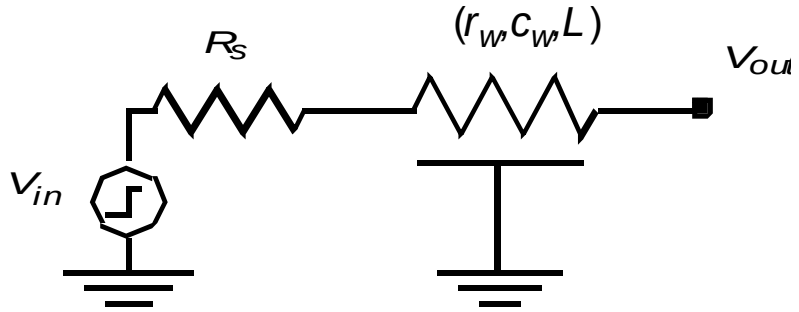
$\tau_{D6} = \tau_{D1} + 2k\Omega \times (1pF + 1pF + 3pF) = 32$ ns

$\tau_{D8} = \tau_{D6} + 2k\Omega \times (1pF + 3pF) + 1k\Omega \times 3pF = 43$ ns



Simplified Delay Equations

Voltage Range	Lumped RC-network	Distributed RC-network
0→50% (t_p)	0.69 RC	0.38 RC
0→63% (τ)	RC	0.5 RC
10%→90% (t_r)	2.2 RC	0.9 RC



$$\tau_D = R_s C_w + \frac{R_w C_w}{2} = R_s C_w + 0.5 r_w c_w L^2$$

$$t_p = 0.69 R_s C_w + 0.38 R_w C_w$$