RC Extraction - Back Annotation - Timing

Ref:

PHYSICAL DESIGN ESSENTIALS by

Khosrow Golshan

Contents

- Interconnect Delay
- Elmore Delay Example
- Back Annotation

Introduction

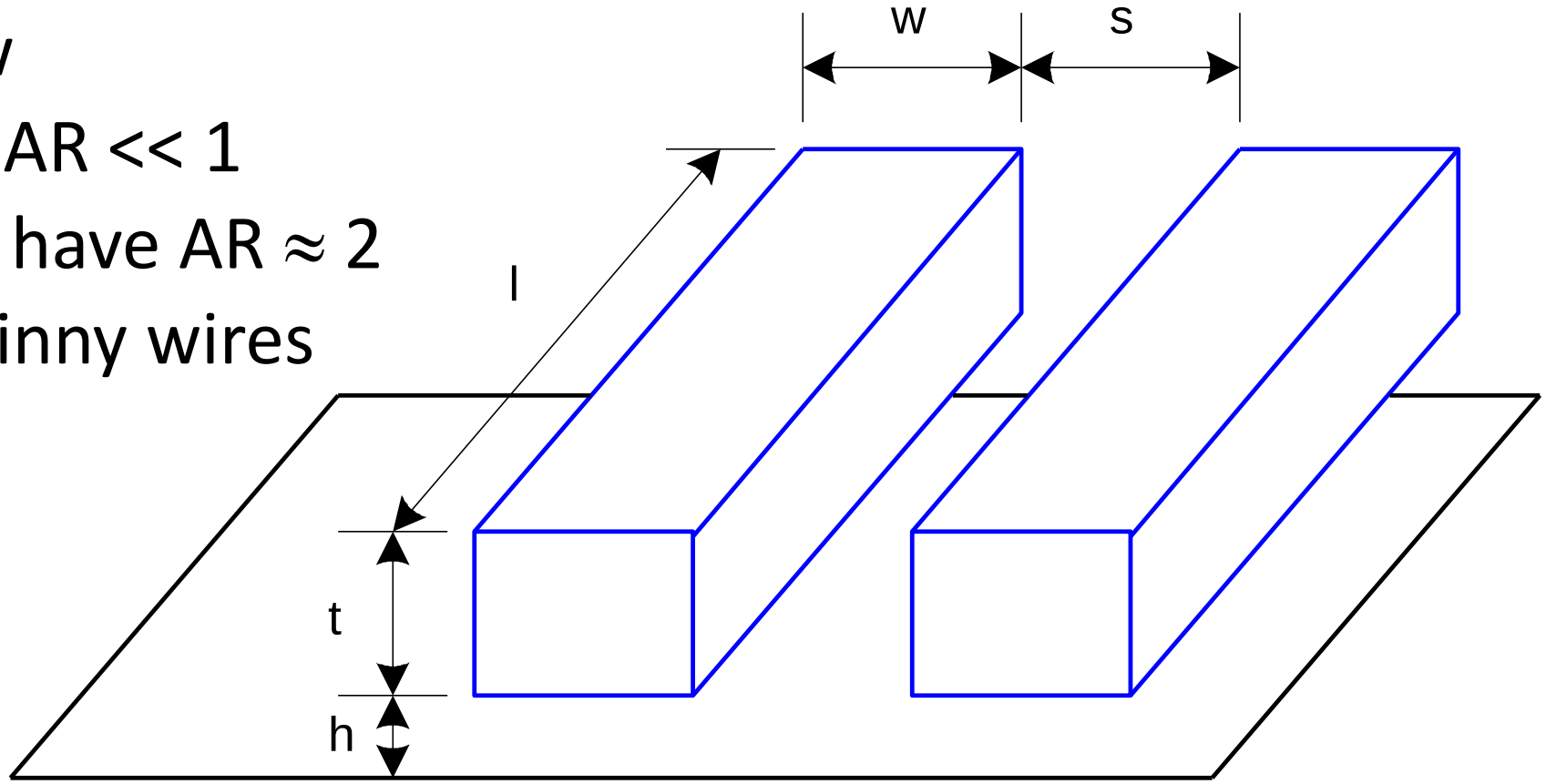
- Chips are mostly made of wires called *interconnect*
  - In stick diagram, wires set size
  - Transistors are little things under the wires
  - Many layers of wires
- Wires are as important as transistors
  - Speed
  - Power
  - Noise
- Alternating layers run orthogonally

Introduction

- Calculation of all routed net capacitances and resistances for the purpose of delay calculation, static timing analysis, circuit simulation, and signal integrity analysis
  - By analyzing each net in the design and
  - Taking into account the effects (such as dielectric stack) of the net's own topology and proximity to other nets
  - 3D models
  - Accurate results
- Resistance, capacitance, and inductance
  - Capacitance coefficient and sheet resistance
    - Based on measurements performed using test-keys from actual silicon data for best, nominal, and worst process conditions or corners
    - Change in the value due to variations in dimension

Wire Geometry

- Pitch = w + s
- Aspect ratio: AR = t/w
  - Old processes had AR << 1
  - Modern processes have AR ≈ 2
    - Pack in many skinny wires



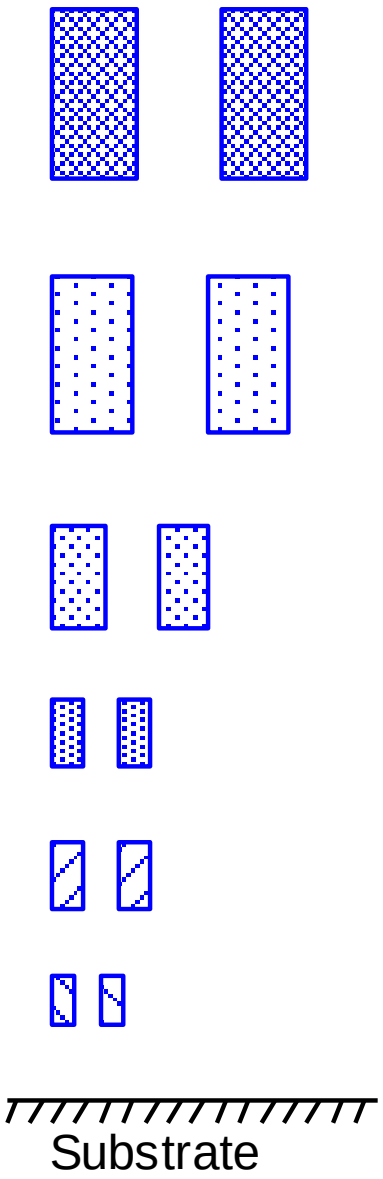
Layer Stack

- AMI 0.6 μm process has 3 metal layers
- Modern processes use 6-10+ metal layers
- Example:

  Intel 180 nm process

- M1: thin, narrow (< 3λ)
  - High density cells
- M2-M4: thicker
  - For longer wires
- M5-M6: thickest
  - For V_{DD}, GND, clk

Layer	T (nm)	W (nm)	S (nm)	AR
6	1720	860	860	2.0
	1000			
5	1600	800	800	2.0
	1000			
4	1080	540	540	2.0
	700			
3	700	320	320	2.2
	700			
2	700	320	320	2.2
	700			
1	480	250	250	1.9
	800			



Wire Resistance

- ρ = *resistivity* (Ω*m)

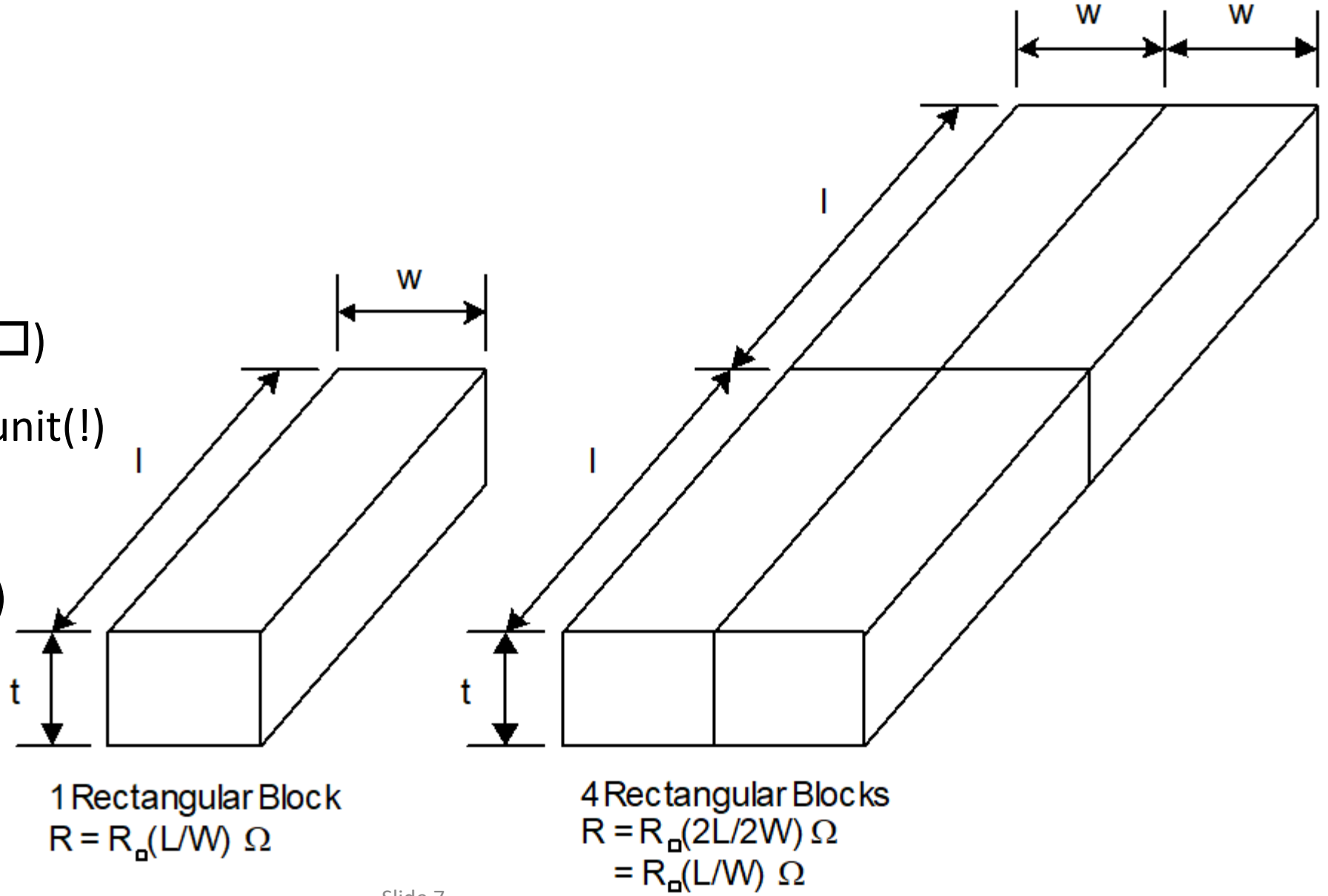
$$R = \frac{\rho}{t}\frac{l}{w}$$

$R_{\square}$ = *sheet resistance* (Ω/□)

□ is a dimensionless unit(!)

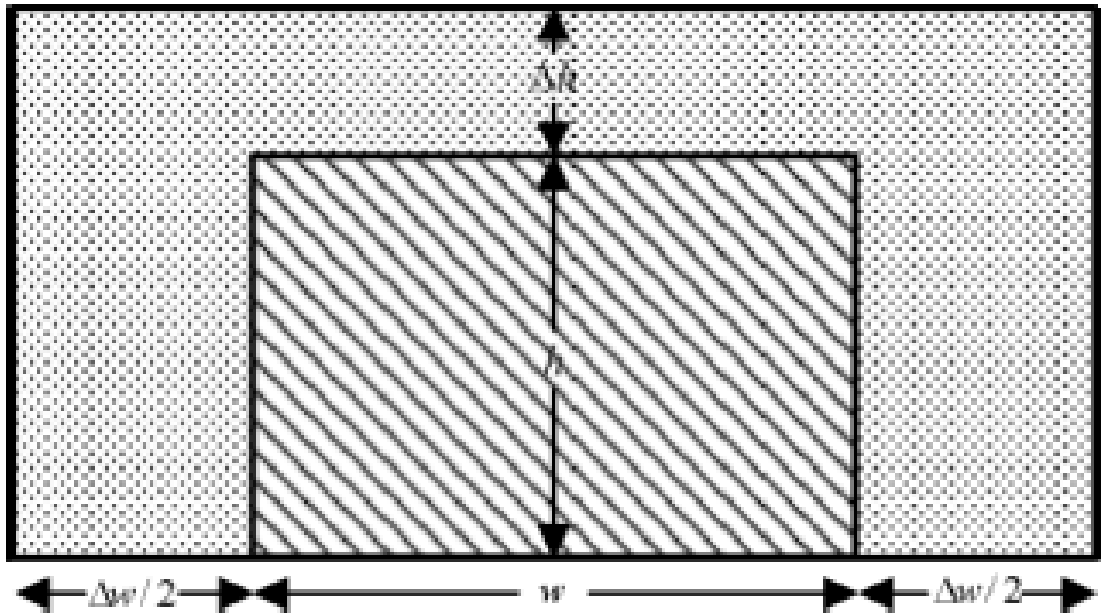
Count number of squares

R = $R_{\square}$ * (# of squares)



Resistance

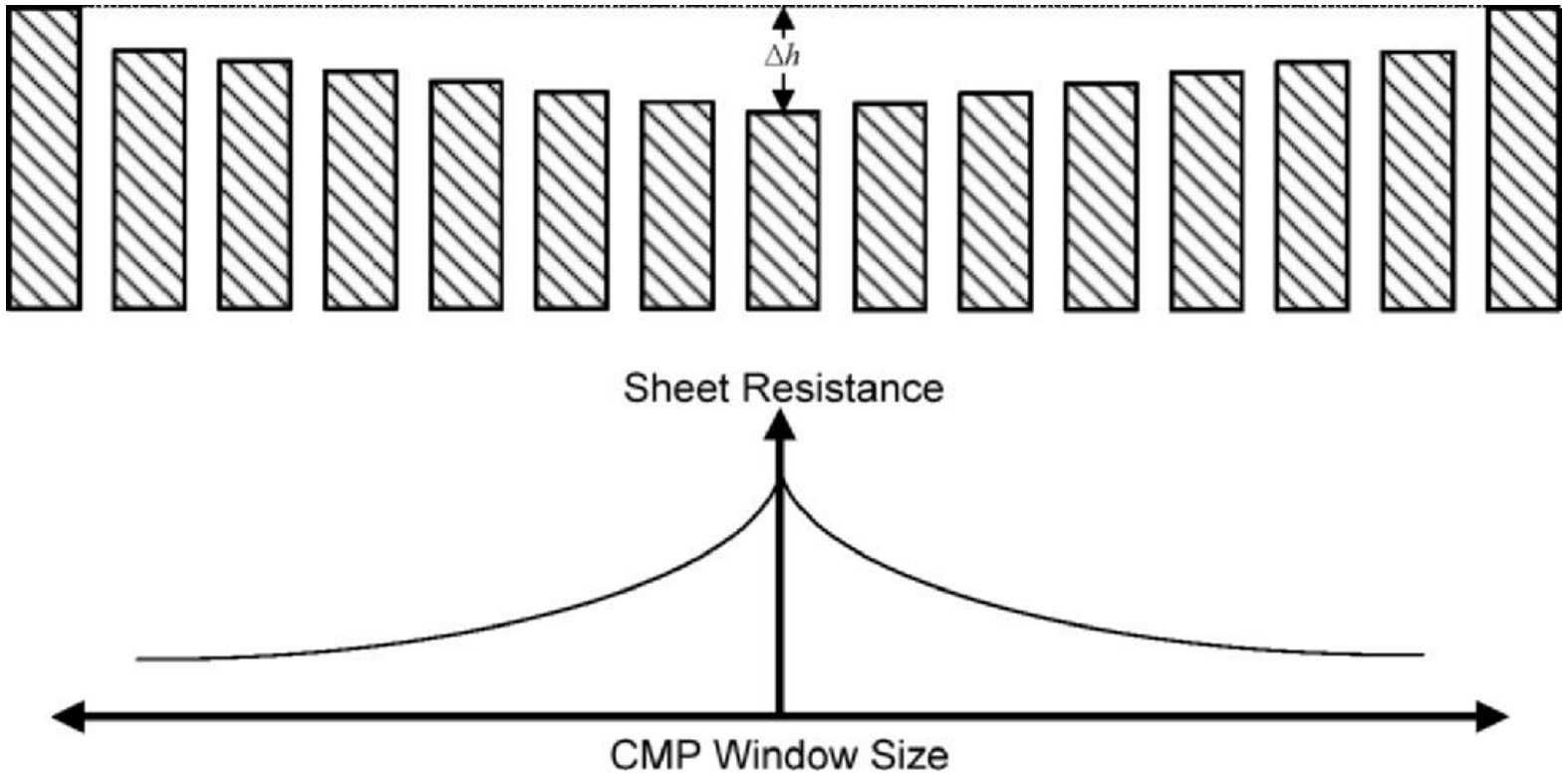
- Process-induced topology variations



- Variation in width is a result of metal layer patterning during the etching process
- Negative etch coefficient increases the width of the wire segment and reduces the spacing between wire segments
- Positive etch coefficient decreases the width of wire segment and increases the spacing between the wire segments

Resistance

- Change in the value of height due to
  - Metal layer and interlayer dielectric planarization
    - Eliminating irregular and discontinuous conditions between successive metal layers



Resistance

- Effect of temperature
- Increase with temperature due to the increase in electron collisions with metal atoms
- Temperature dependency of resistivity is given by
  - Fractional change in resistance - proportional to the temperature change

$$\frac{\Delta R}{\Delta T} = \beta R_0$$

where $\Delta T = (T - T_0)$ is the change in temperature T from its initial temperature T_0, $\Delta R = (R - R_0)$ corresponds to the change of resistance R from its initial resistance R_0, and proportionality constant β refers to Temperature Coefficient of Resistance (TCR).

Resistance

- Re-written as

$$\frac{R - R_0}{R_0} \equiv \beta(T - T_0)$$

$$R = R_0[1 + \beta(T - T_0)] .$$

Choice of Metals

- Until 180 nm generation, most wires were aluminum
- Modern processes often use copper
  - Cu atoms diffuse into silicon and damage FETs
  - Must be surrounded by a diffusion barrier

Metal	**Bulk resistivity (μΩ*cm)**
Silver (Ag)	1.6
Copper (Cu)	1.7
Gold (Au)	2.2
Aluminum (Al)	2.8
Tungsten (W)	5.3
Molybdenum (Mo)	5.3

Sheet Resistance

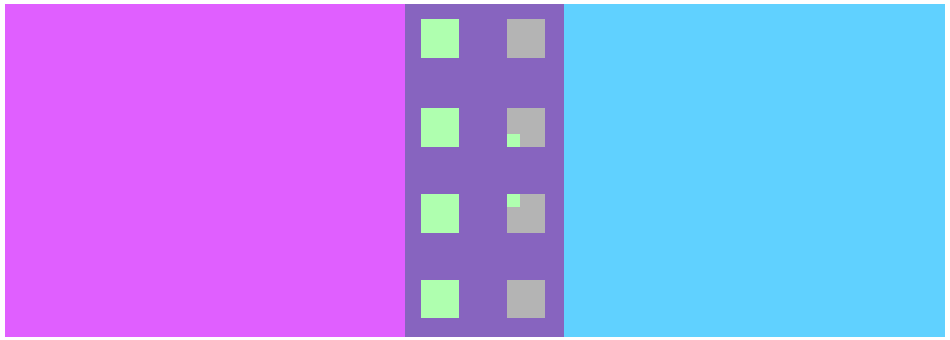
- Typical sheet resistances in 180 nm process

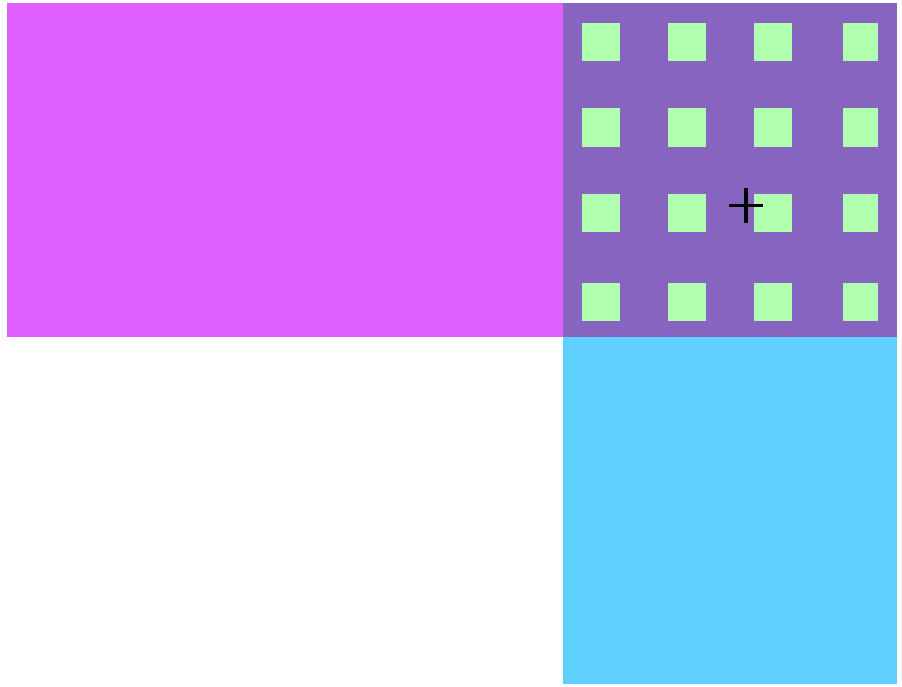
Layer	Sheet Resistance (Ω/□)
Diffusion (silicided)	3-10
Diffusion (no silicide)	50-200
Polysilicon (silicided)	3-10
Polysilicon (no silicide)	50-400
Metal1	0.08
Metal2	0.05
Metal3	0.05
Metal4	0.03
Metal5	0.02
Metal6	0.02

- Silicide – a compound that has silicon – electron diffuses quickly through it

Contacts Resistance

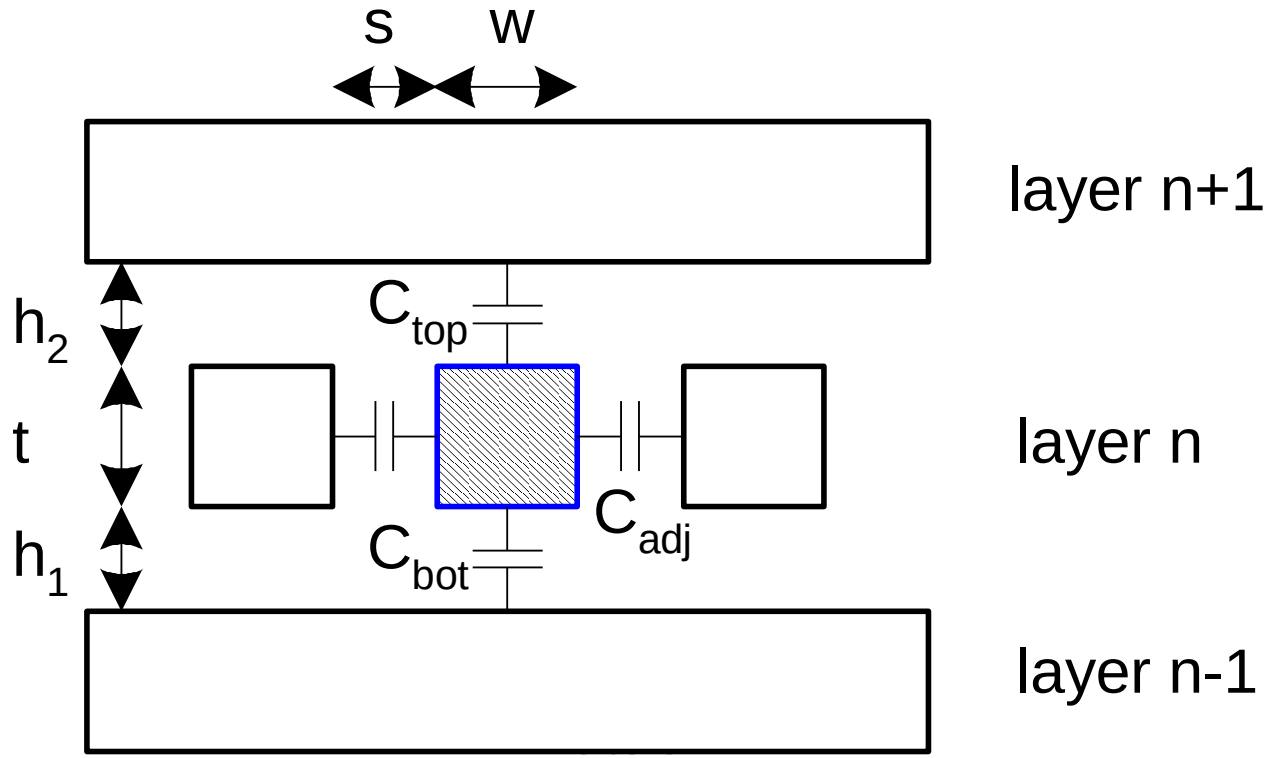
- Contacts and vias also have 2-20 Ω
- Use many contacts for lower R
  - Many small contacts for current crowding around periphery





Wire Capacitance

- Wire has capacitance per unit length
  - To neighbors
  - To layers above and below
- $C_{total} = C_{top} + C_{bot} + 2C_{adj}$



Capacitance

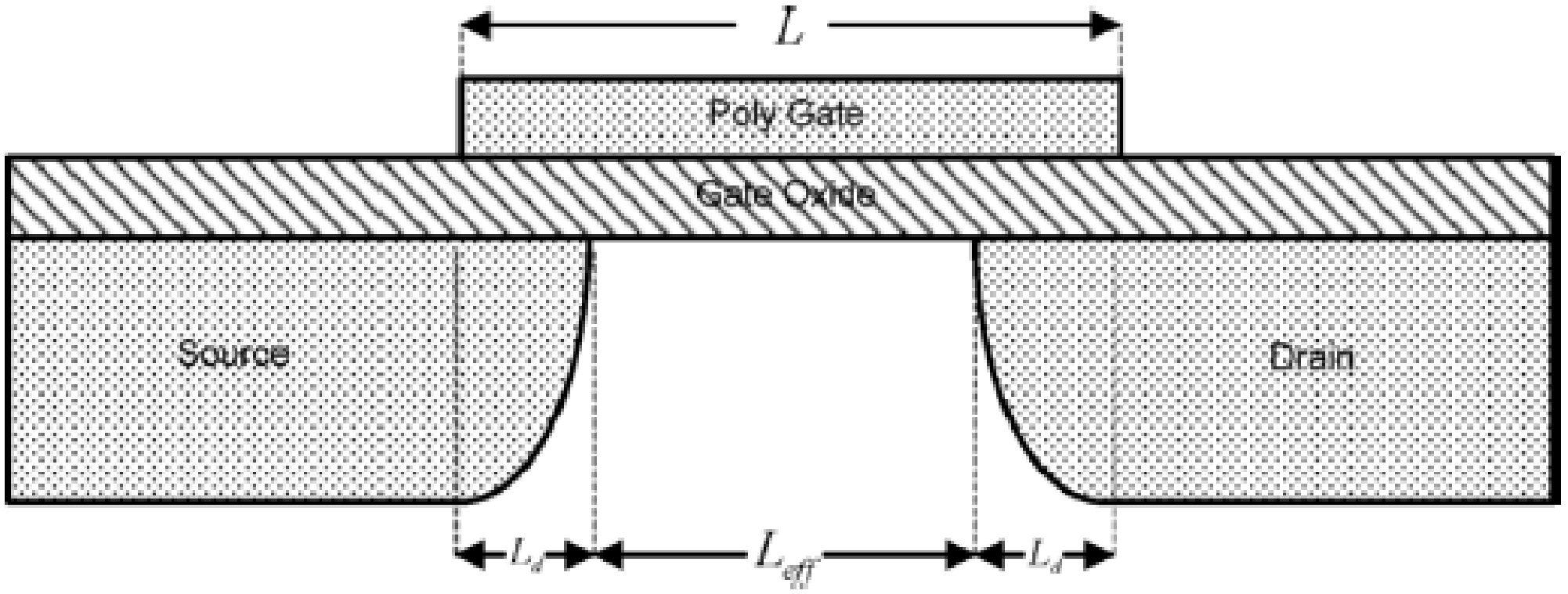
- Gate capacitance and routing segment capacitance
- Each transistor gate capacitance is the sum of capacitances between gate to MOSFET body, gate to source, and gate to drain
- Gate to MOSFET body (or substrate) capacitance with gate area of *A* defined as

$$C_{gs} = \left(\frac{\varepsilon_{ox} A}{T_{ox}}\right) = \left(\frac{\varepsilon_{ox} WL}{T_{ox}}\right)$$

where ε_{ox} is the dielectric constant of the gate oxide, T_{ox} is the gate oxide thickness, L is the transistor drawn gate length, and W is the transistor drawn gate width. The term $\varepsilon_{ox} / T_{ox}$ expresses the oxide capacitance.

Capacitance

- Lateral diffusion



Capacitance

- Variations

$$C_{gs} = \frac{\varepsilon_{ox} W (L - 2L_d)}{T_{ox}}$$

$$C_s, C_d = \frac{\varepsilon_{ox} W L_d}{T_{ox}} ;$$

where C_s and C_d are gate-to-source and gate-to-drain capacitance.

Capacitance

- The routing segment capacitance is extracted after final routing.
- During the routing segments' capacitance extraction, the following types of capacitance are considered:

- Area capacitance
- Fringe capacitance
- Sidewall capacitance

Capacitance

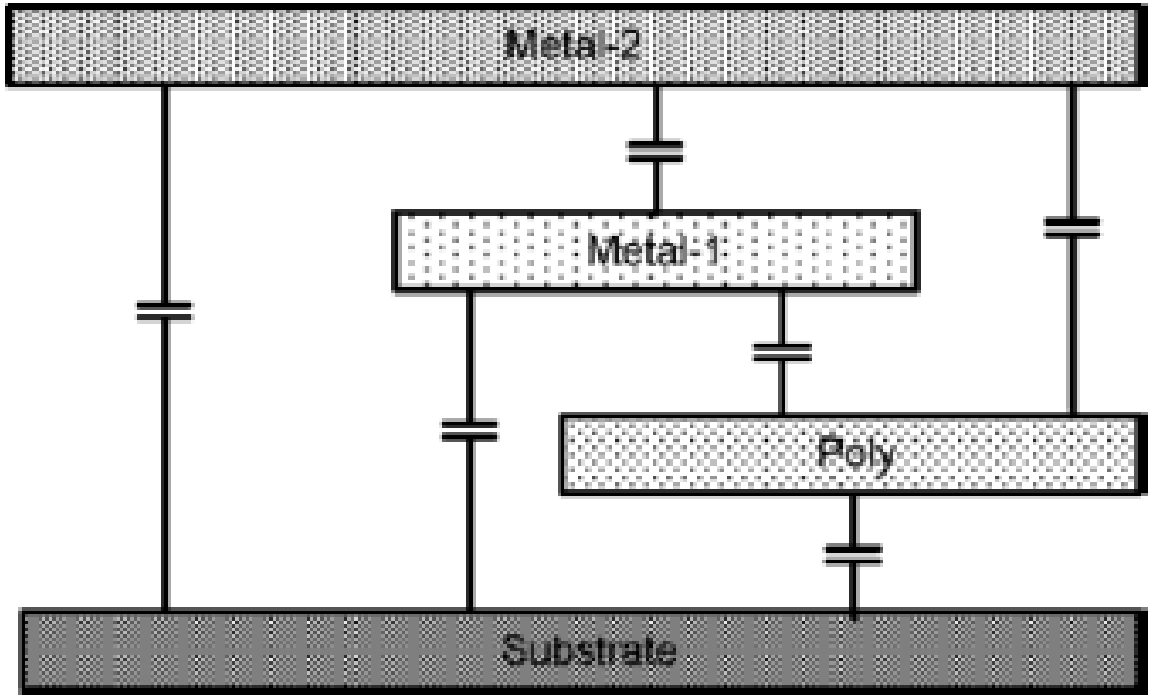
- Area capacitance
  - Routing segment capacitance to substrate and to other exposed routing segments above and below

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

where C_p is the area or plate capacitance, ε is the permittivity of the dielectric, L and W are the overlapping length and width of the routing segment, d is the distance between the routing segments and corresponds to interlayer dielectric, or ILD, thickness, and the term $(\varepsilon W / d)$ refers to capacitance per unit length.

Capacitance

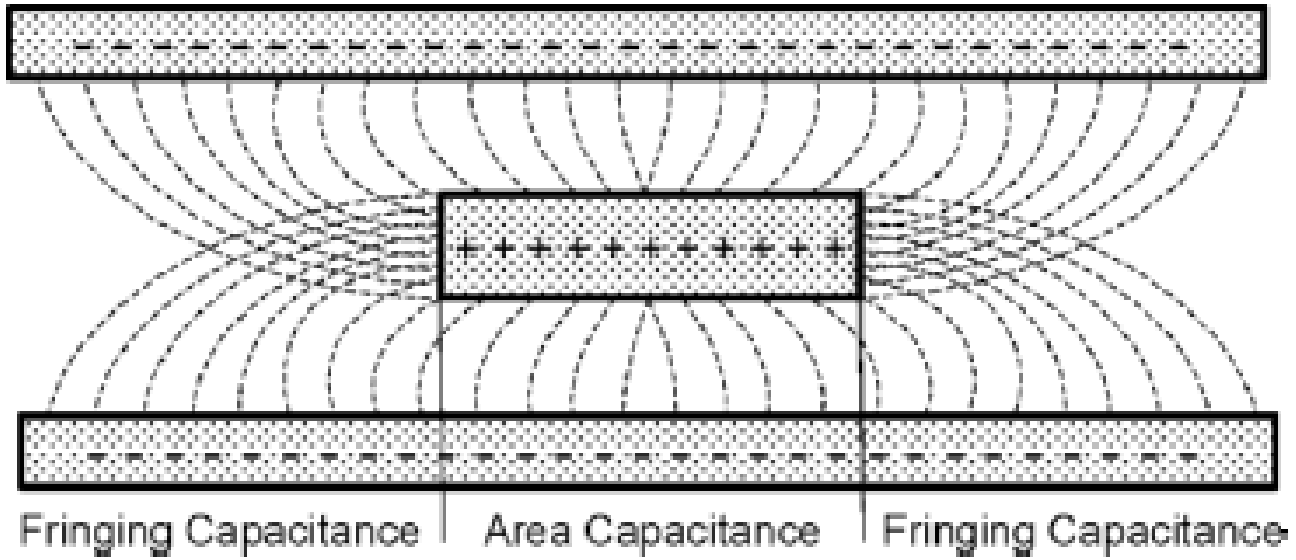
- Area Capacitance of Two Metal Route



- Interlayer dielectric thickness and corresponding capacitance per unit area for various layers are defined by semiconductor foundries for different process corners

Capacitance

- Fringing capacitance is from the sidewall of routing segments to the substrate or other routing segment surfaces
  - Electrostatic field theory



Electrical Field Flux Between Three Conducting Layers

Capacitance

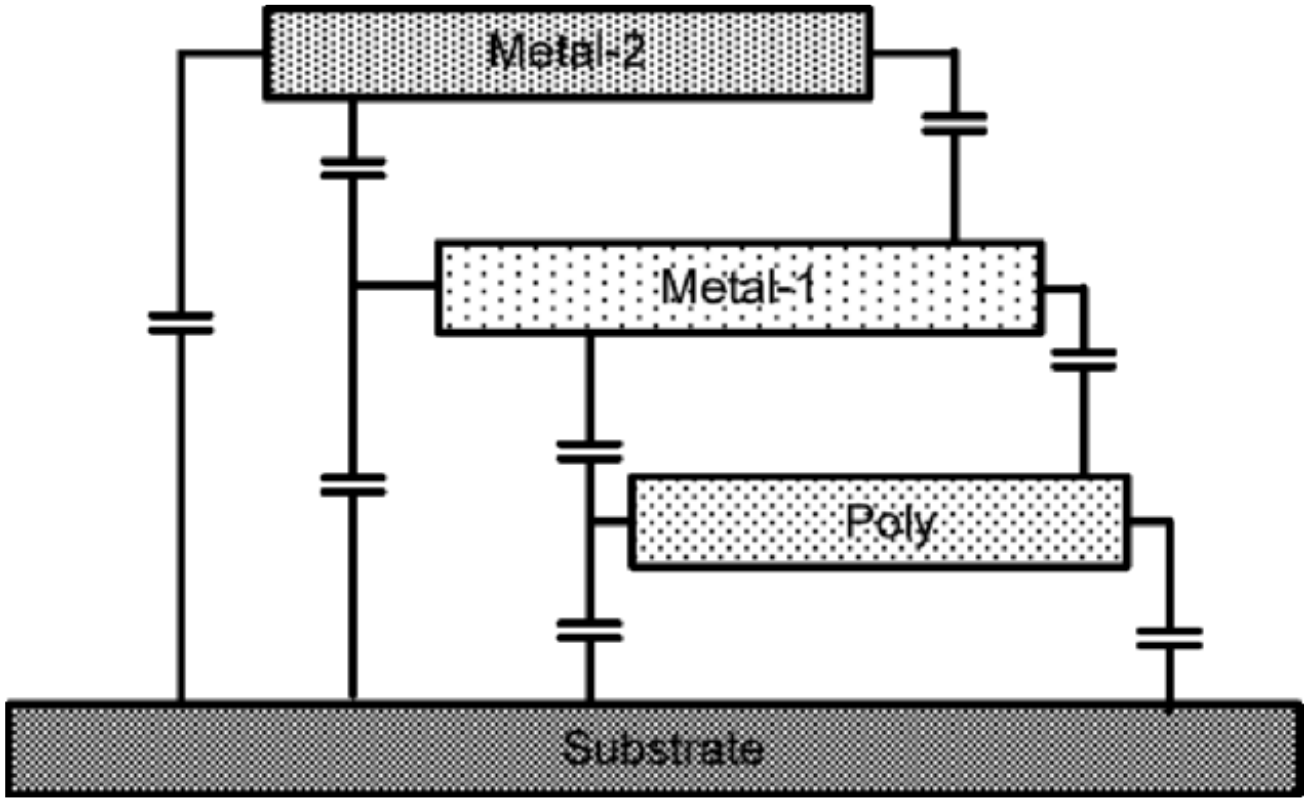
- Frequently used methods for fringing capacitance extraction
  - Analytic approximation
  - Empirical expression

$$C_f = \frac{2\pi\varepsilon L}{\ln\{1+\frac{2d}{h}+[\frac{2d}{h}(\frac{2d}{h}+2)]^{1/2}\}} \quad \text{for } W \geq \frac{d}{2} \qquad C_f = \varepsilon L[0.77+1.06(\frac{W}{d})^{0.25}+1.06(\frac{h}{d})^{0.5}]$$

C_f is the fringing capacitance, ε is the permittivity of the dielectric,
L is the length of the routing layer,
h is the layer thickness, d is the dielectric thickness, and
W is the width of the routing layer.

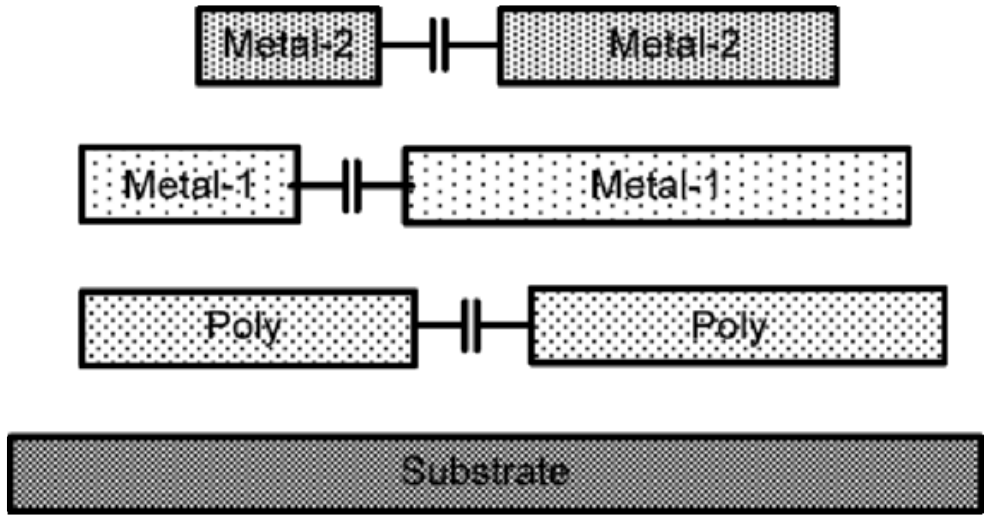
Capacitance

- Simplified fringing capacitance of different conducting layers



Capacitance

- Sidewall capacitance is due to the coupling between the sidewalls of routing segments on the same layer and is similar to the area capacitance



- It is the function of distance separating the routing segments and the length that they extend in parallel

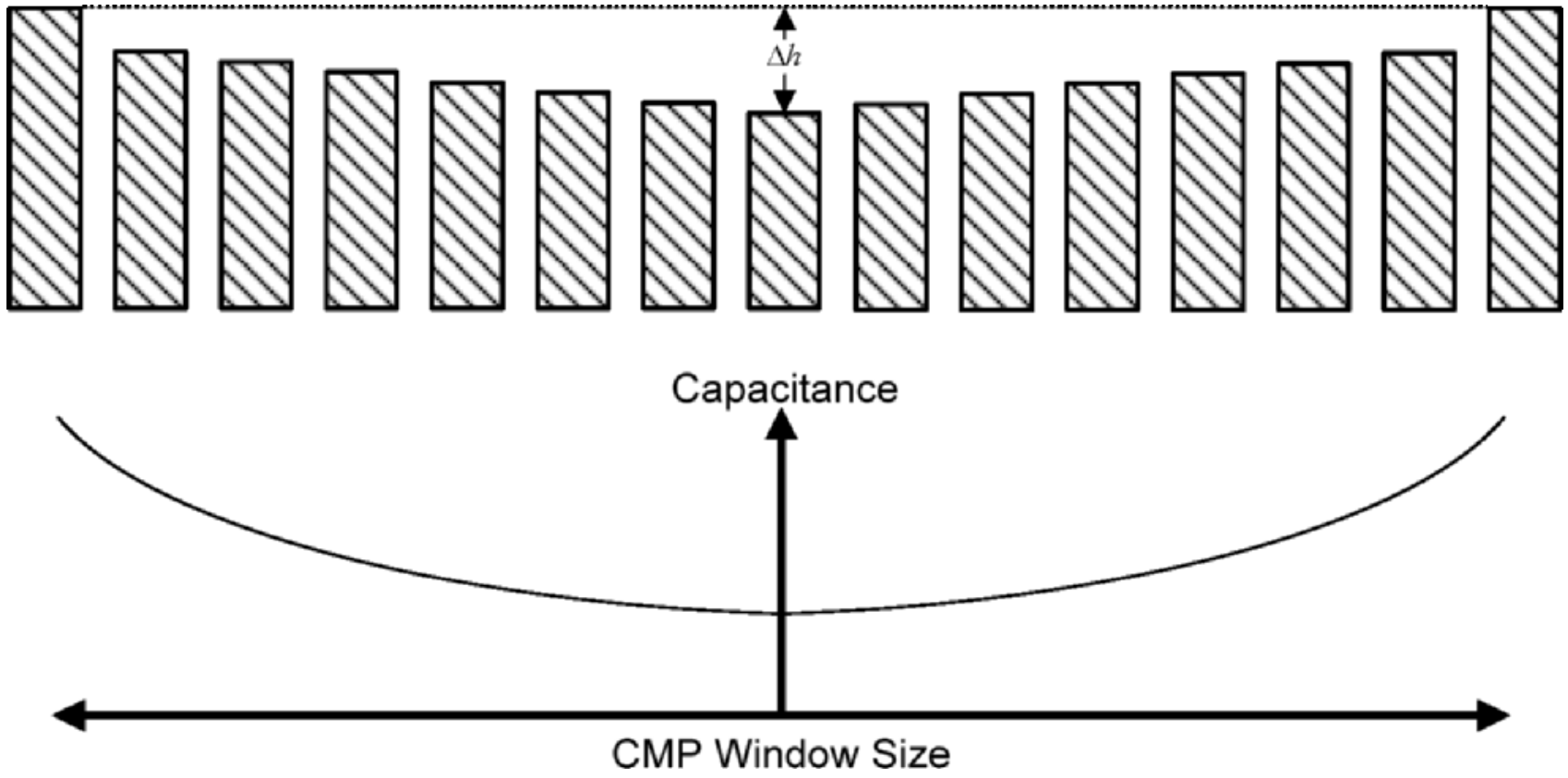
$$C_s = \varepsilon \frac{Lh}{s}$$

where C_s corresponds to the sidewall capacitance, ε is the permittivity of the dielectric, L is the length of routing segment, h is the thickness of routing layer, and s is the separation between two adjacent routing layer segments.

Capacitance

- Dramatic increase in the value of sidewall capacitance due to
  - Shrinking feature size
  - Height dominates over width
- As *'s'* increases value of sidewall capacitance approaches zero
  - Increasing spacing between routing tracks in the place-and-route technology file
- Capacitance values also subjected to process-induced variations
  - Variation in metal thickness affects fringing and sidewall capacitance
  - Variation in the dielectrics' thickness impacts the area capacitance value

Capacitance



Effect of CMP on Capacitance Value

Capacitance

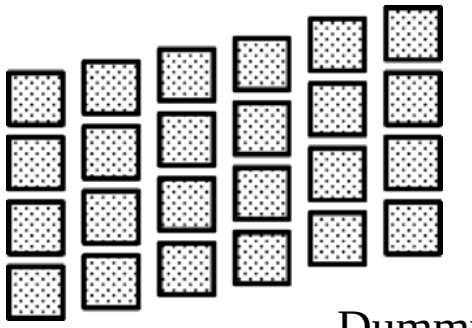
- In the final interconnect capacitance calculation, the sum of all transistor gate capacitances associated with such interconnect must be added to the interconnect capacitance

$$C_{total} = \sum_{i=1}^{n} (C_p + C_f + C_s)_i + \sum_{j=1}^{k} (C_g)_j$$

where C_{total} is the interconnect total capacitance, C_p is the area capacitance, C_f is the fringing capacitance, C_s is the sidewall capacitance of each segment of the interconnect, and C_g is the transistor's input gate and junction capacitance connected to that interconnect.

Capacitance

- With the advent of copper interconnect and its susceptibility to topographical thickness variation due to the CMP process, semiconductor manufacturers require the use of a dummy fill
- Dummy fills are isolated islands (e.g. metal), and are used to create more uniform density that leads to a planar die surface
- In the portion of the die, surfaces that are outside of a specified density range will have dummy fills inserted to meet the targeted density



Dummy Fill Pattern

Capacitance Trends

- Parallel plate equation: $C = \varepsilon A/d$
  - Wires are not parallel plates, but obey trends
  - Increasing area (W, t) increases capacitance
  - Increasing distance (s, h) decreases capacitance
- Dielectric constant
  - $\varepsilon = k\varepsilon_0$
- $\varepsilon_0 = 8.85 \times 10^{-14}$ F/cm
- k = 3.9 for SiO_2
- Processes are starting to use low-k dielectrics

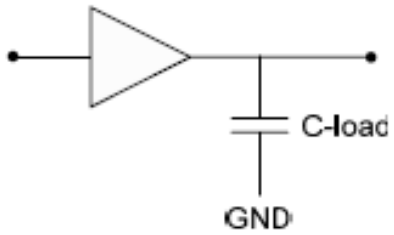
Diffusion & Polysilicon

- Diffusion capacitance is very high (about 2 fF/μm)
  - Comparable to gate capacitance
  - Diffusion also has high resistance
  - Avoid using diffusion *runners* for wires!
- Polysilicon has lower C but high R
  - Use for transistor gates
  - Occasionally for very short wires between gates

Models to represent inter-connection parasitics

- Lumped capacitance model
- Lumped resistance and capacitance (RC) model
- Distributed resistance and capacitance (π)model
- Distributed resistance, capacitance, and inductance (RCL) model

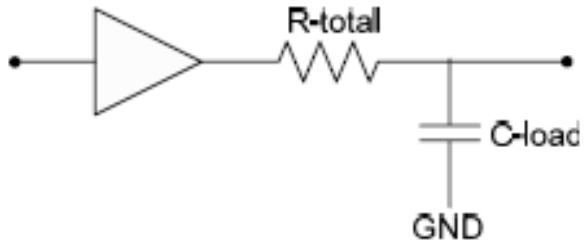
- Lumped capacitance is a single order approximation and considers only the total capacitance value of interconnection while ignoring the resistance value
  - wire delay negligible
- adequate to calculate the propagation delay through the gate



- Simple model - some layout synthesis tools still use it to estimate the capacitive loading effects during initial placement

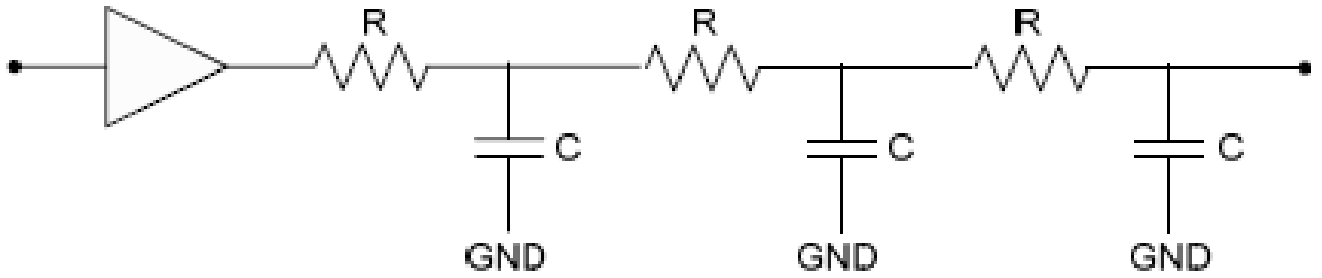
Models to represent inter-connection parasitics

- RC model is a second-order approximation
  - takes into account the effect of loading capacitance as well as the total wire resistance of interconnections
- Here, the sum of wire capacitances is used to compute the cell delay
  - based on the characterized model from the library and
  - the product of lumped interconnect capacitance and resistance is used to compute the output slope
- Using lumped RC to calculate interconnection delay is considered pessimistic and inaccurate for long interconnects



Models to represent inter-connection parasitics

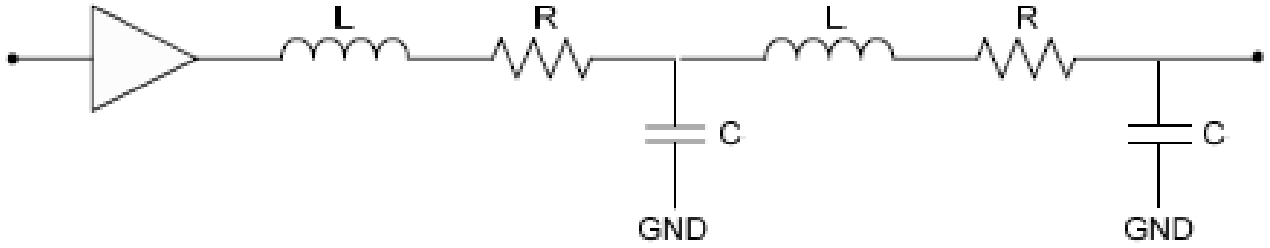
- Distributed resistance and capacitance - third order interconnect delay approximation
- Interconnections are segmented into a series of resistor and capacitor network resembling a transmission line



- Effective for computing wire delays of very highly resistive interconnection networks because
  - as wire resistance increases it has a tendency to shield the actual wire capacitance in the presence of driving gate resistance

Models to represent inter-connection parasitics

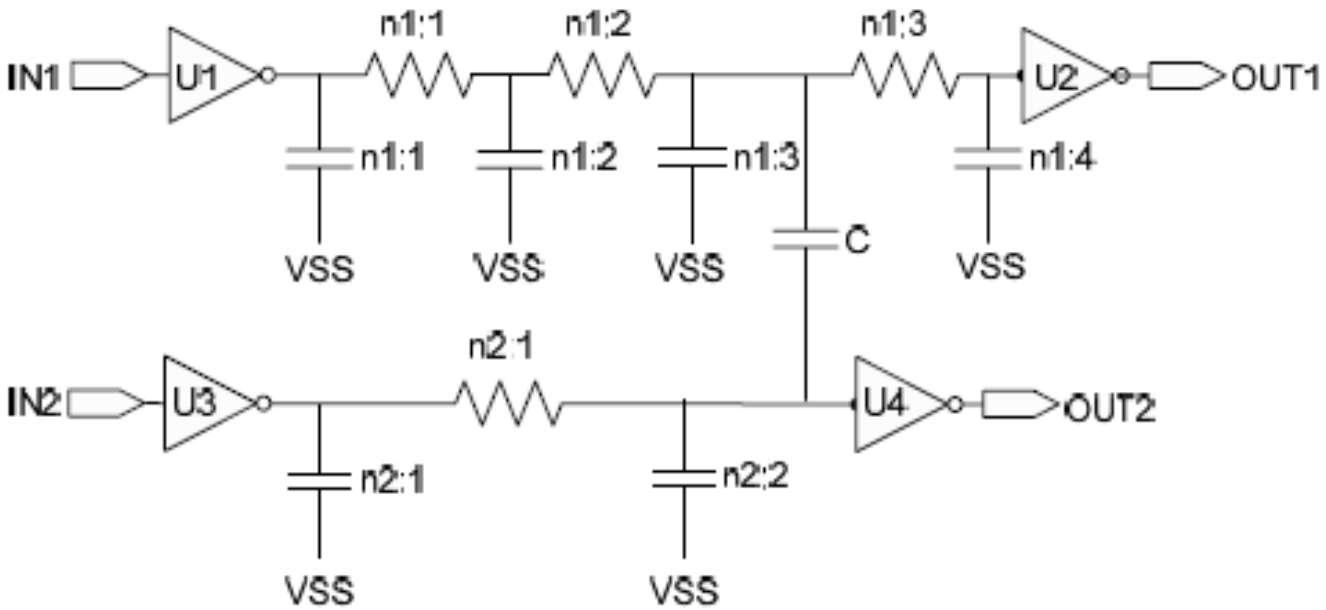
- The RCL model accounts for resistance, capacitance, and inductance and is a fourth-order interconnect approximation
- Wire segment inductance
- Mutual inductance between long segments that run in parallel within the same layer



- Important in analyzing the effect of wire inductance
- Timing and signal integrity analysis
- This is mainly due to the fact that as resistance per unit length increases significantly, so does the line inductance
- It is important to note that the presence of inductance not only increases signal propagation delay, but also has a tendency to cause voltage overshoot, which results in a reduction of signal rise time, which in turn, is one cause for the increase of crosstalk noise

Models to represent inter-connection parasitics

- Standard Parasitic Extended Format (SPEF)
- RC parasitic capacitance and resistance values extracted per net based on their actual geometry and layer width and spacing information
- Most commonly used formats used to import and export
- Institute of Electrical and Electronics Engineers (IEEE) standard



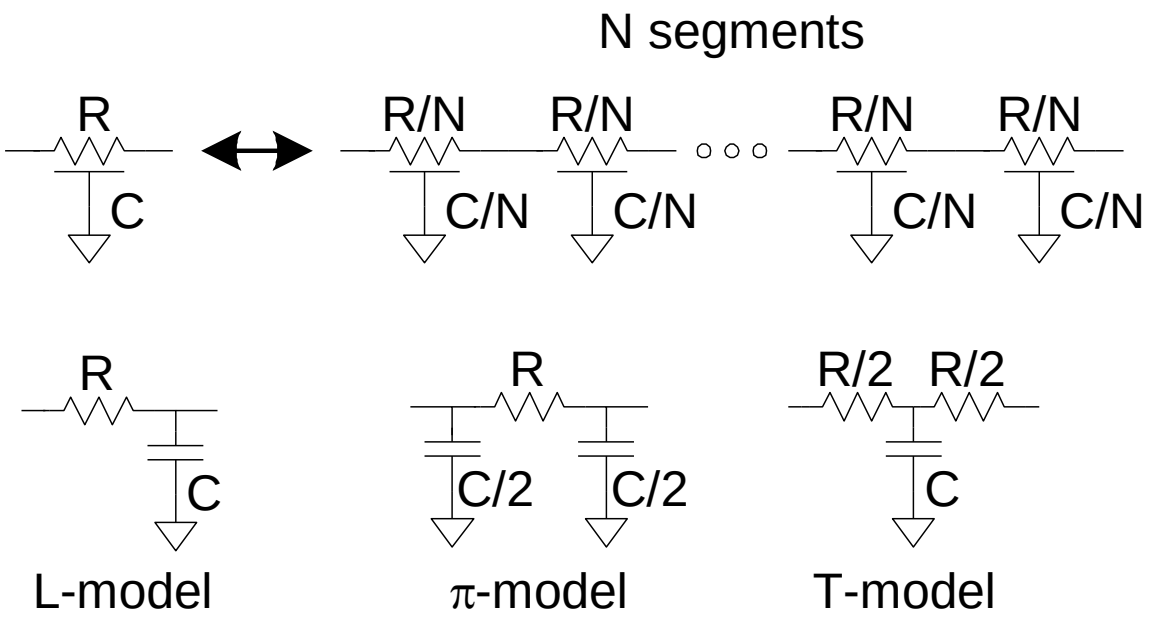
Circuit Presentation

```
*SPEF "IEEE 1481-1998"
*DESIGN "SAMPLE_SPEF"
*DATE "Thu Dec 22 12:12:24
2004"
*VENDOR    "None"
*PROGRAM   "None"
*VERSION   "1.1.0"
*DESIGN_FLOW   "PIN_CAP
NONE" "FULL_CONNECTIVITY"
"ROUTING_CONFIDENCE 90"
*DIVIDER /
*DELIMITER :
*BUS_DELIMITER [ ]
*T_UNIT  1 PS
*C_UNIT  1 FF
*R_UNIT  1 OHM
*L_UNIT  1 HENRY


*POWER_NETS VDD
*GROUND_NETS VSS

*PORTS
IN1 0 *L 0 *S 0 0
IN2 0 *L 0 *S 0 0
OUT1 0 *L 0 *S 0 0
OUT2 0 *L 0 *S 0 0

*D_NET n1 7.1
*CONN
*I U1:Z I *L 0 D INV
*I U2:A I *L 10

*CAP
1 n1:1 0.2
2 n1:2 0.3
3 n1:3 0.4
4 n1:4 0.1
5 n1:3 n2:2 0.07
```

```
*RES
1 n1:1 n1:2 2.2
2 n1:2 n1:3 8.0
3 n1:3 n1:4 3.7
*END


*D_NET n2 11.7
*CONN
*I U3:Z I *L 0 D INV
*I U4:A I *L 10.0

*CAP
1 n1:1 0.9
2 n1:2 0.8
3 n2:2 n1:3 0.07

*RES
1 n2:1 n1:2 5.0
*END

*D_NET IN1 15.1
*CONN
*I U1:A *L 10
*P IN1 O *L 0

*CAP
1 IN1:1 5.1
*END

*D_NET IN2 17.1
*CONN
*I U3:A *L 10
*P IN2 O *L 0

*CAP
1 IN2:1 7.1

*END
```

Lumped Element Models

- Wires are a distributed system
  - Approximate with lumped element models
- 3-segment π-model is accurate to 3% in simulation
- L-model needs 100 segments for same accuracy!
- Use single segment π-model for Elmore delay

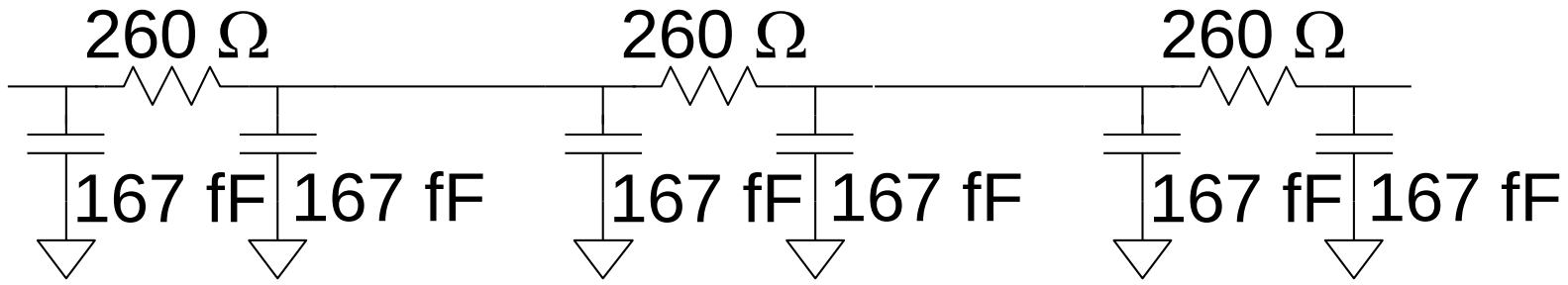


Example

- Metal2 wire in 180 nm process
  - 5 mm long
  - 0.32 μm wide
- Construct a 3-segment π-model
  - $R_{\square}$ =
  - $C_{permicron}$ =

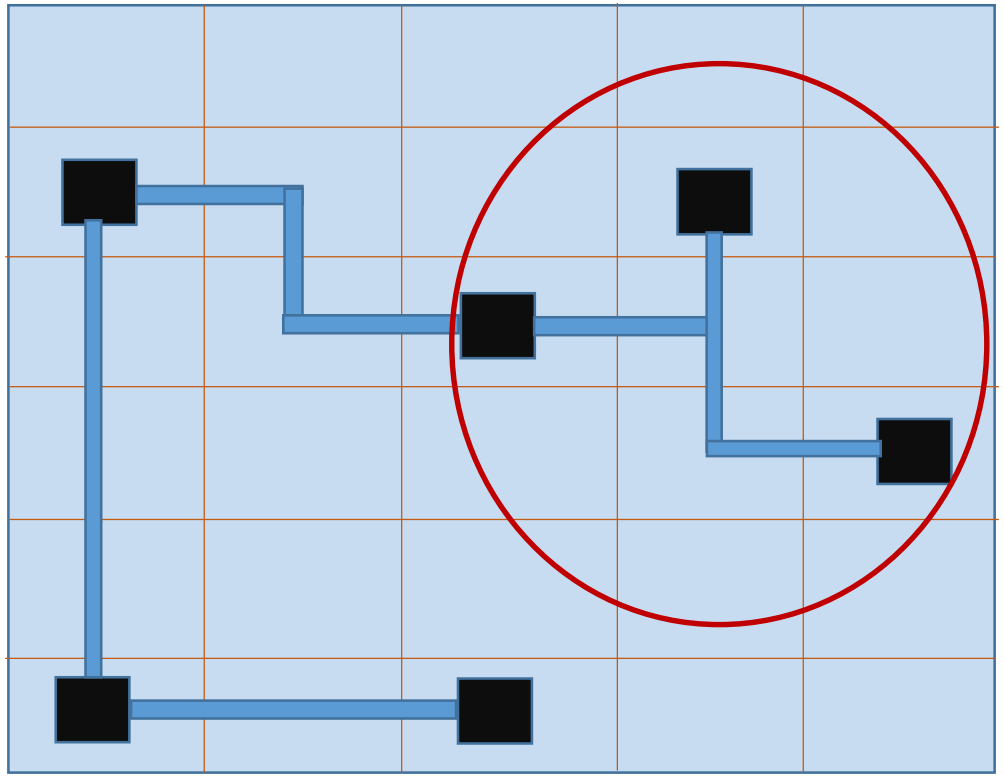
Example

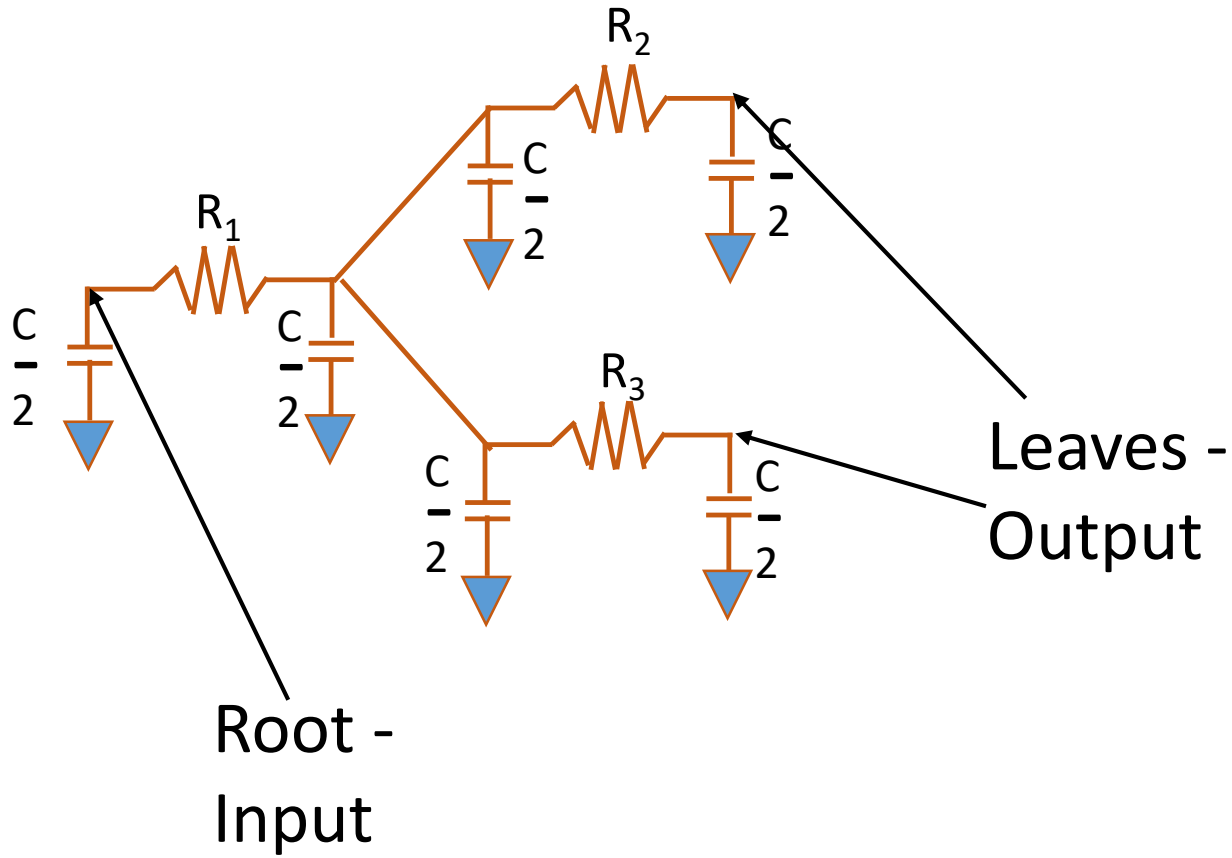
- Metal2 wire in 180 nm process
  - 5 mm long
  - 0.32 μm wide
- Construct a 3-segment π-model
  - $R_{\square}$ = 0.05 Ω/□ => R = 781 Ω
  - $C_{permicron}$ = 0.2 fF/μm => C = 1 pF



RC Tree for Interconnect Models

- RC Tree – Consists of tree of resistors without loops and capacitors "hanging off" in-between tree nodes

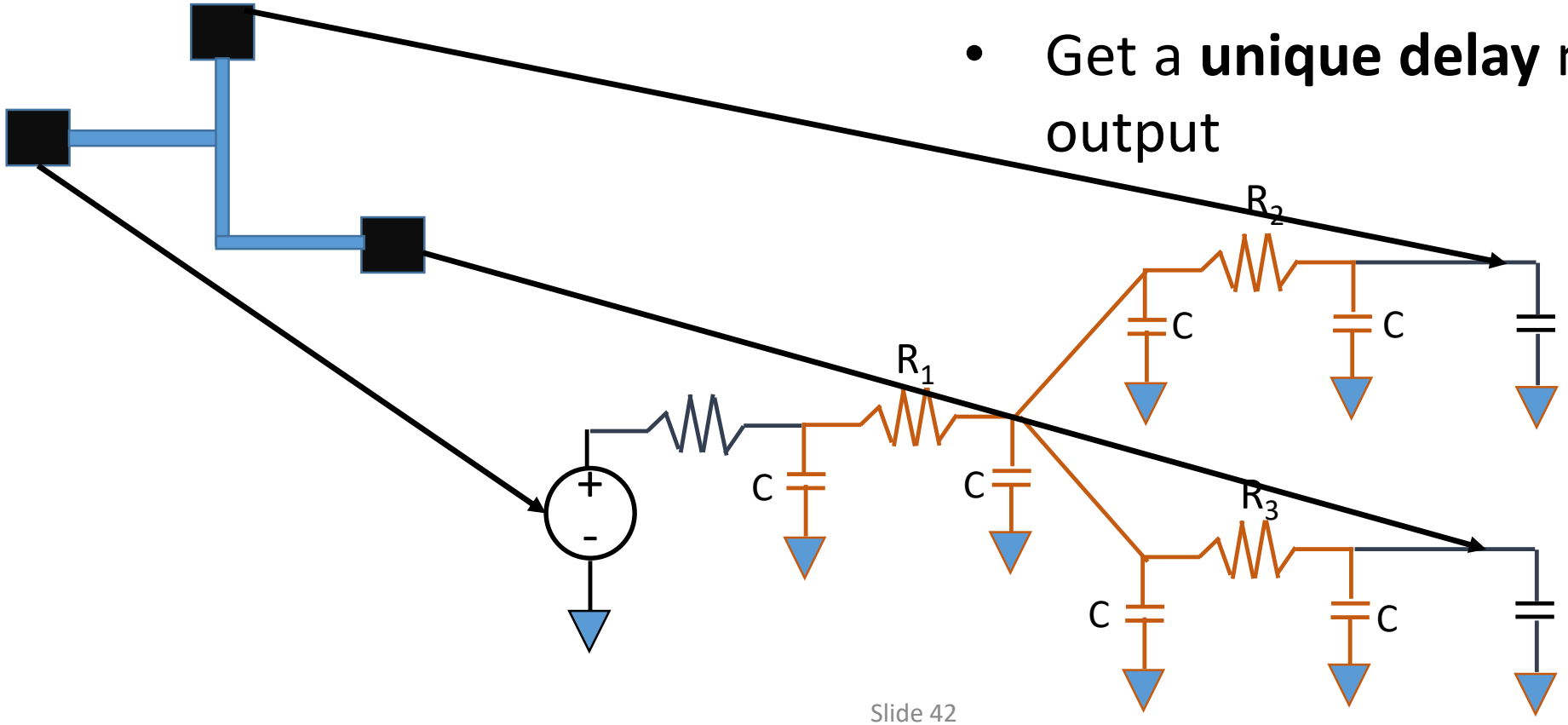




RC Tree for Interconnect Models

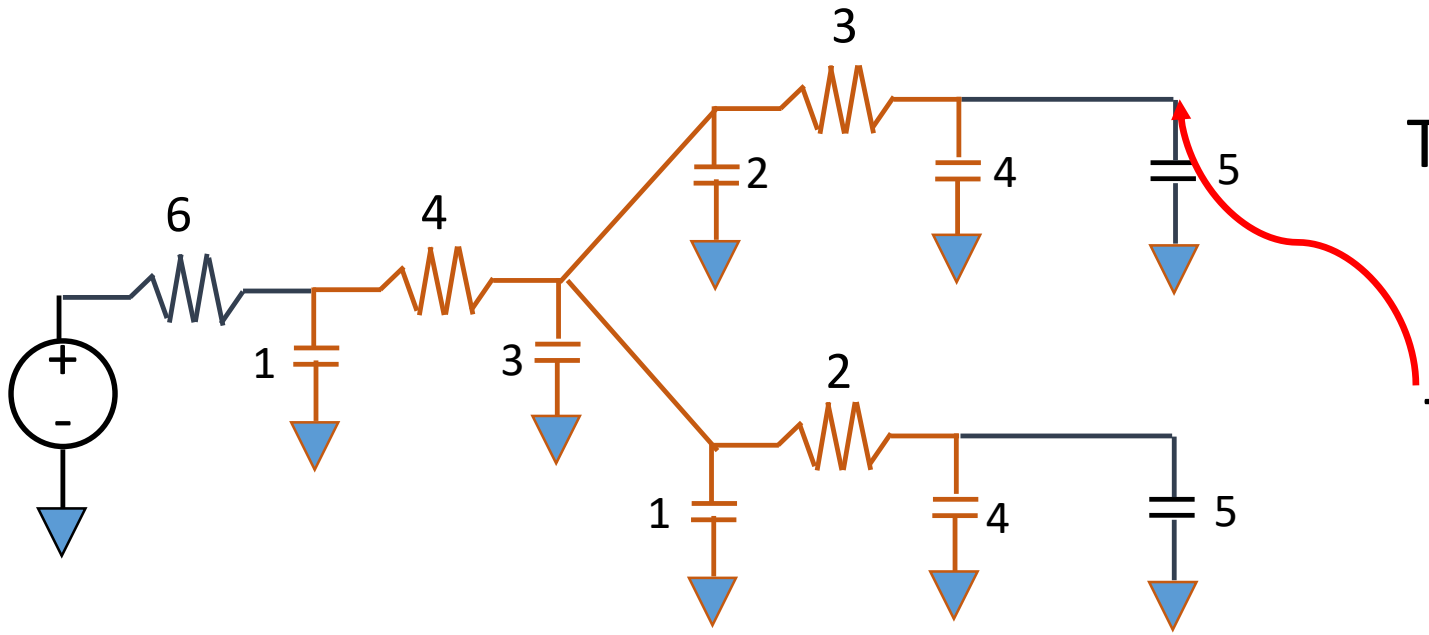
- Root - Voltage source + resistor (driving gate)
- Leaf - Capacitor as load (driven gate)

- Use tree to compute a **delay** number for each output of tree
- Get a **unique delay** number for each output



Elmore Delay

- Simple approximation to the delay through an RC network
- $T_{delay} = \sum R_i * C_i$
- Find the path from Root to Leaf through resistors and all capacitors below them

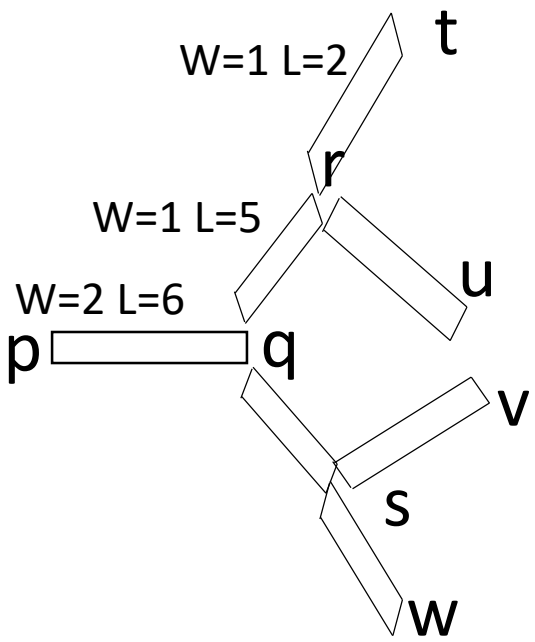


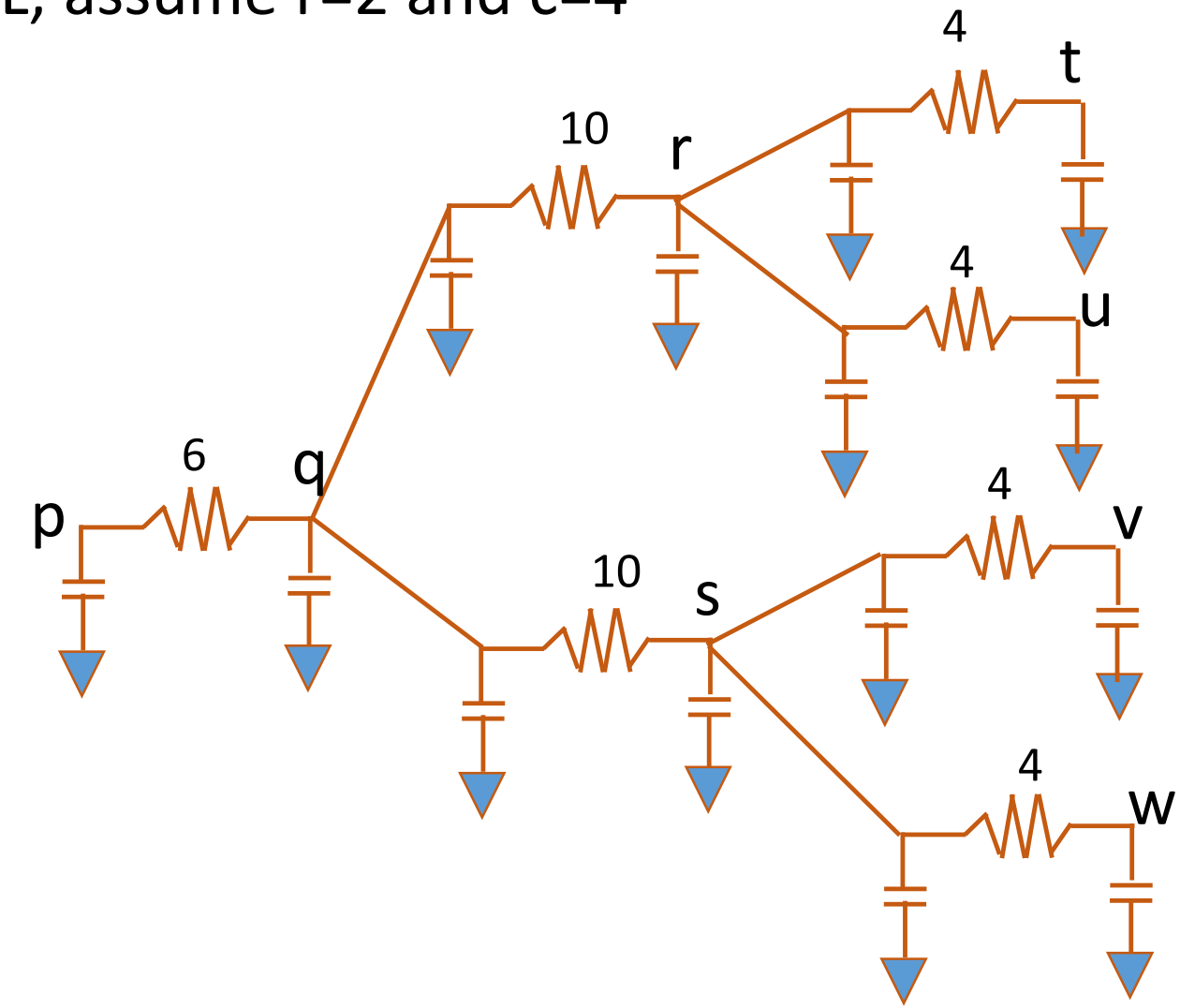
$T_{delay} = 6*(1+3+2+4+5+1+4+5)$
$+4*(3+2+4+5+1+4+5)$
$+3(4+5)$

$T_{delay} = 273$

Elmore Delay

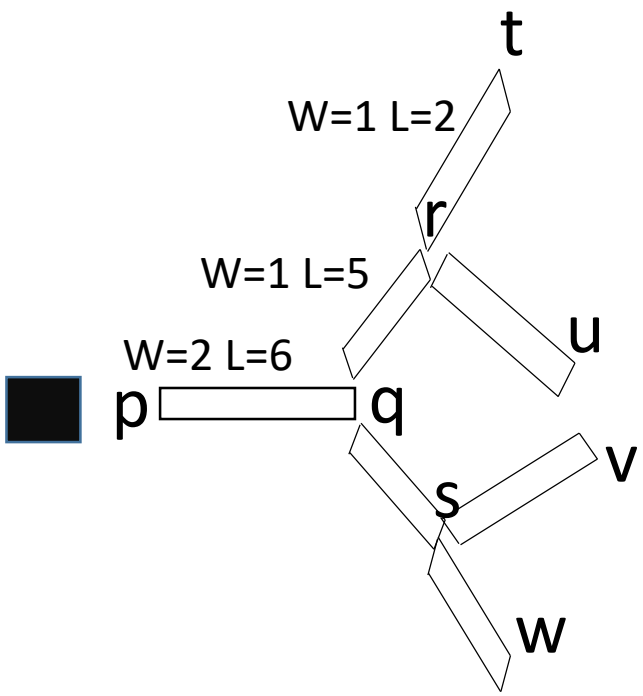
- Example
- Consider R=r*L/W and C=c*W*L, assume r=2 and c=4

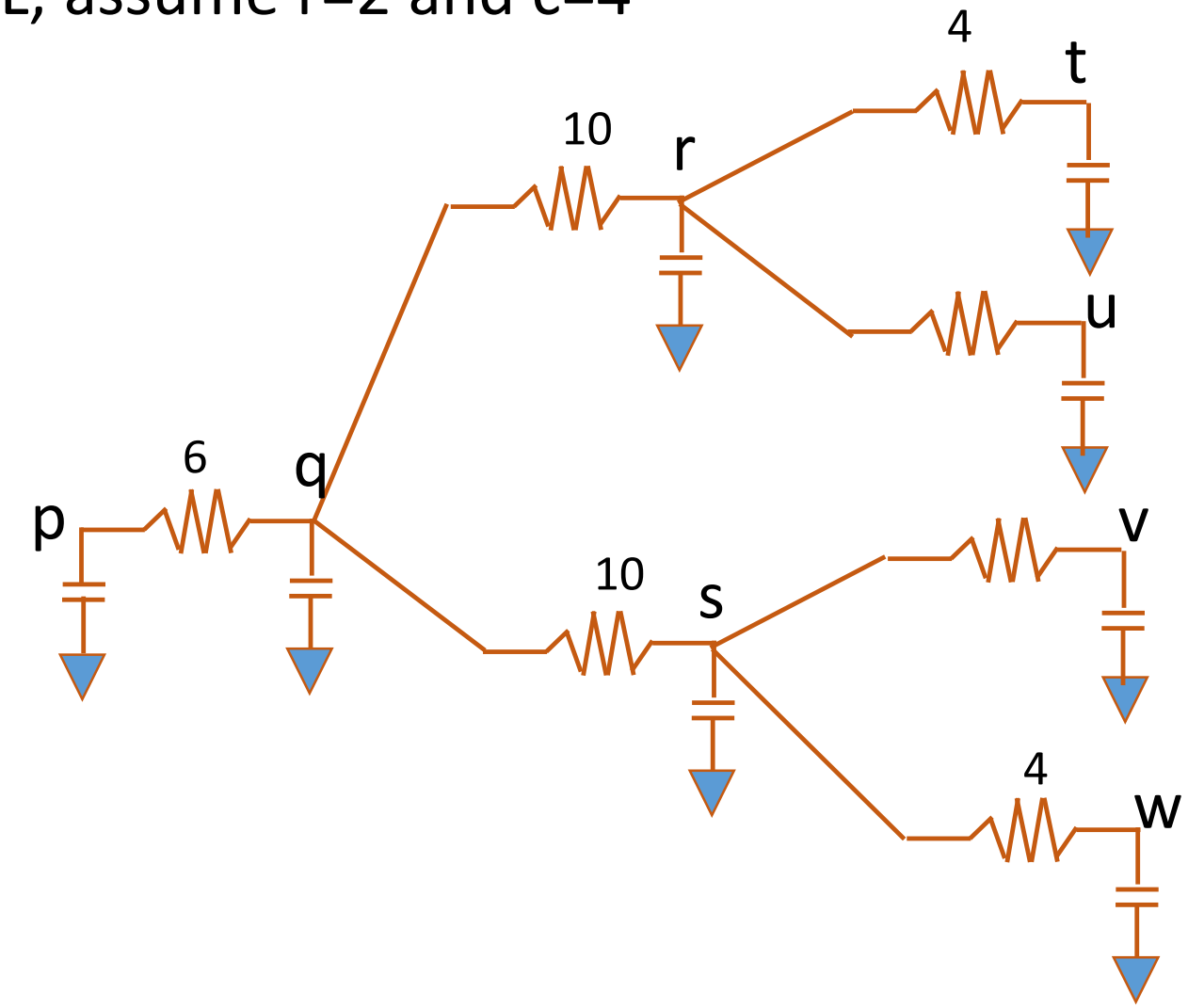




Elmore Delay

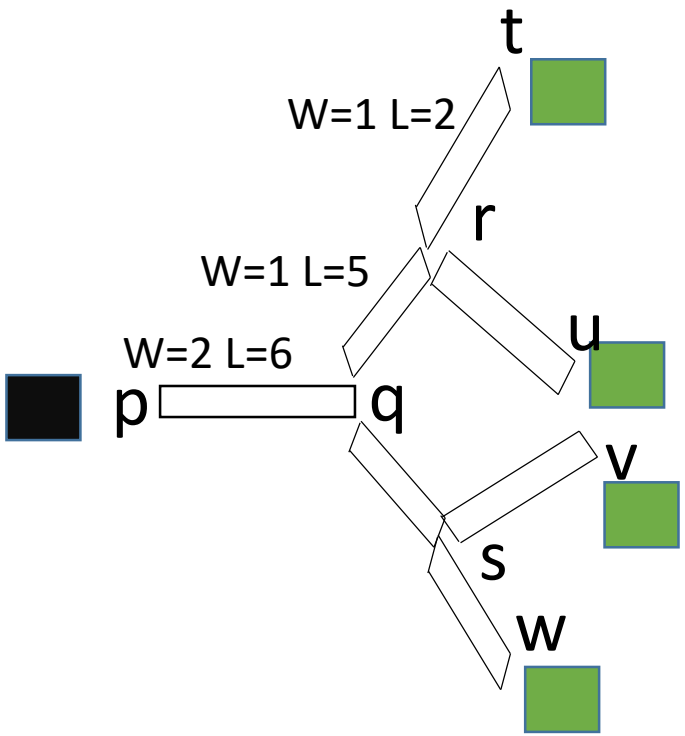
- Example
- Consider R=r*L/W and C=c*W*L, assume r=2 and c=4

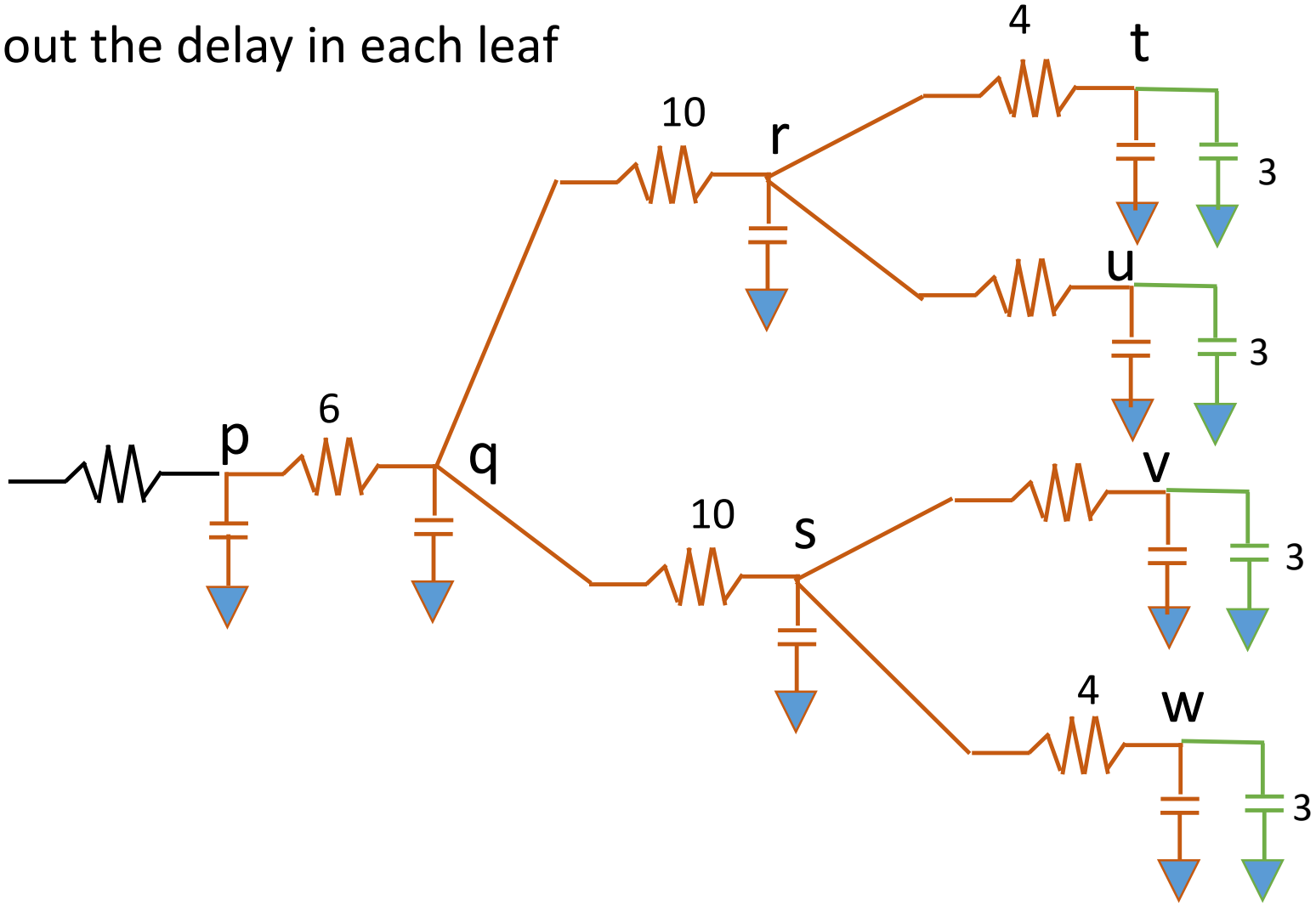




Elmore Delay

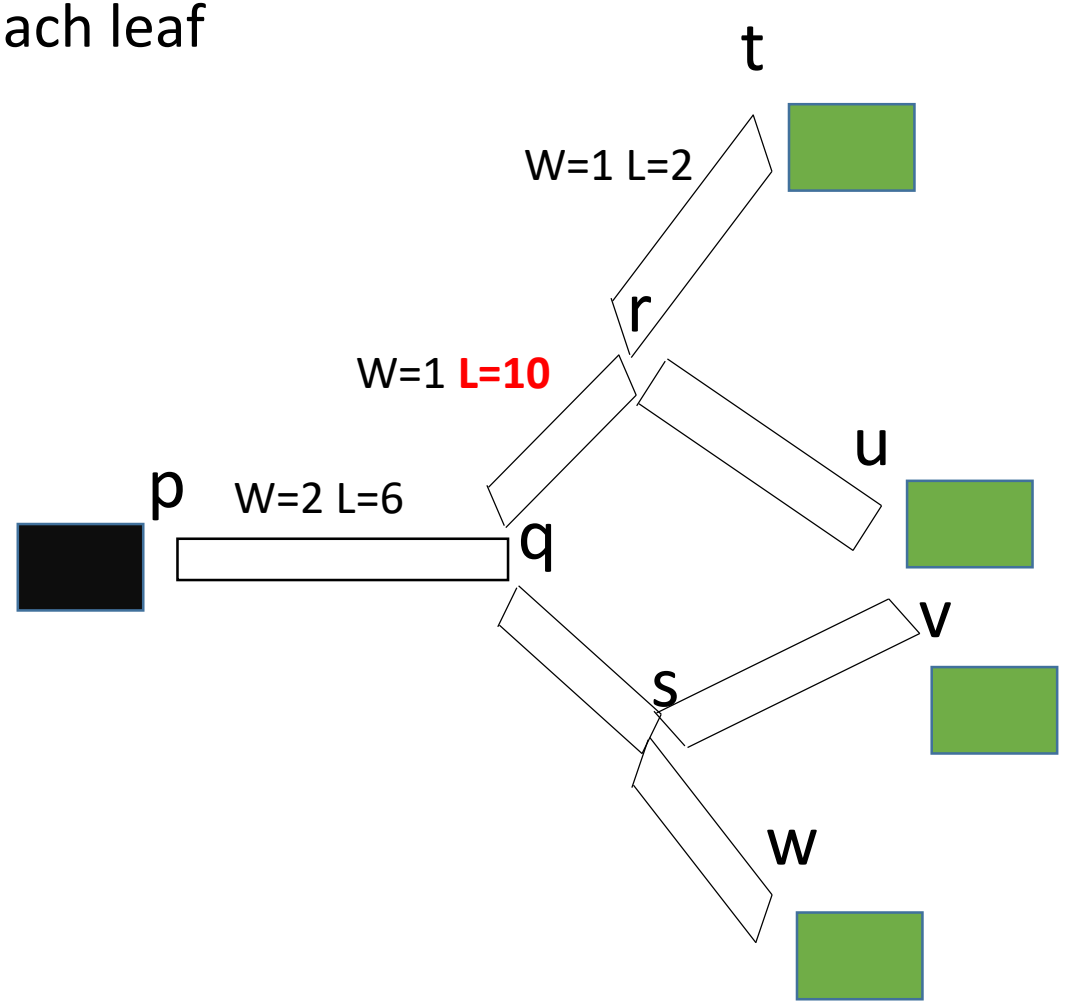
- Example - 1
- Consider R=r*L/W and C=c*W*L, assume r=2 and c=4
- Add driver and driven gates - Find out the delay in each leaf





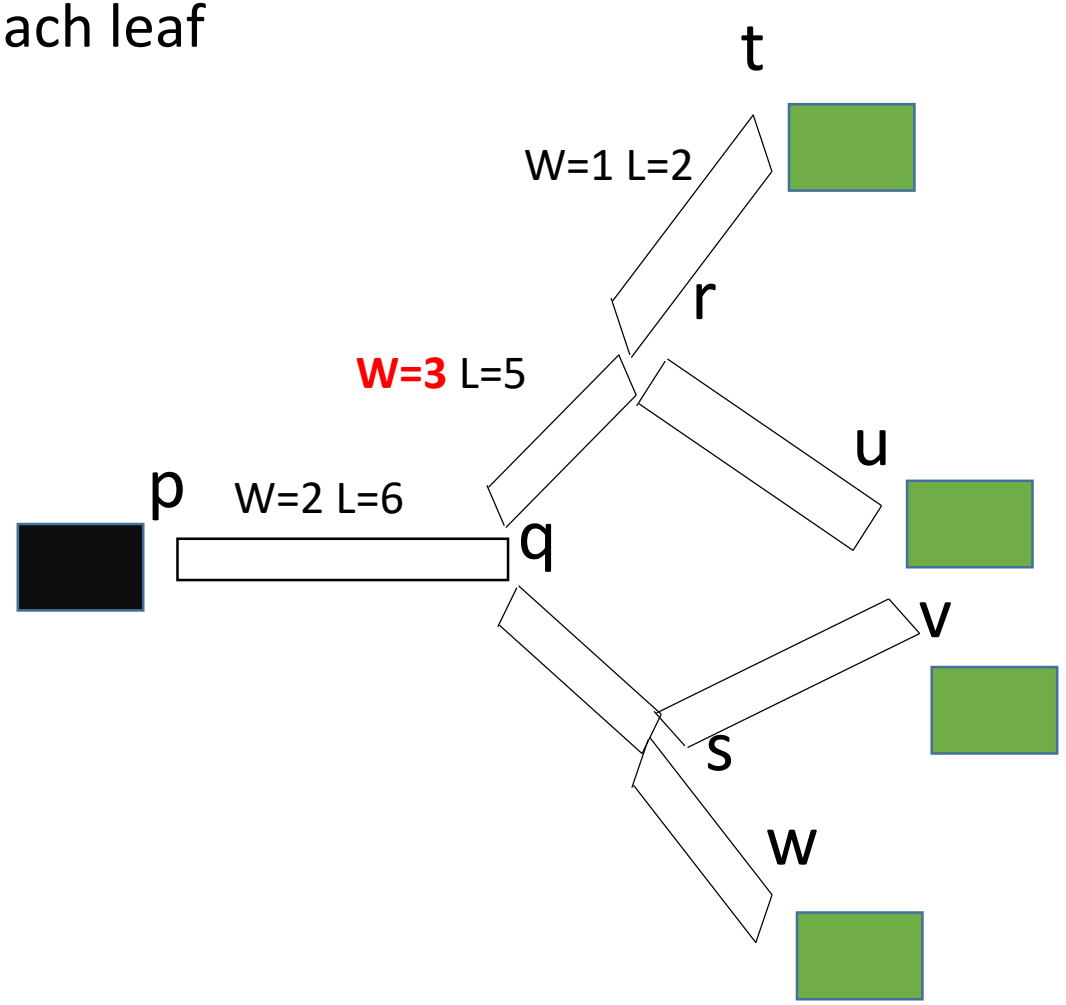
Elmore Delay

- Example - 2
- Consider R=r*L/W and C=c*W*L, assume r=2 and c=4
- Find out the delay in each leaf



Elmore Delay

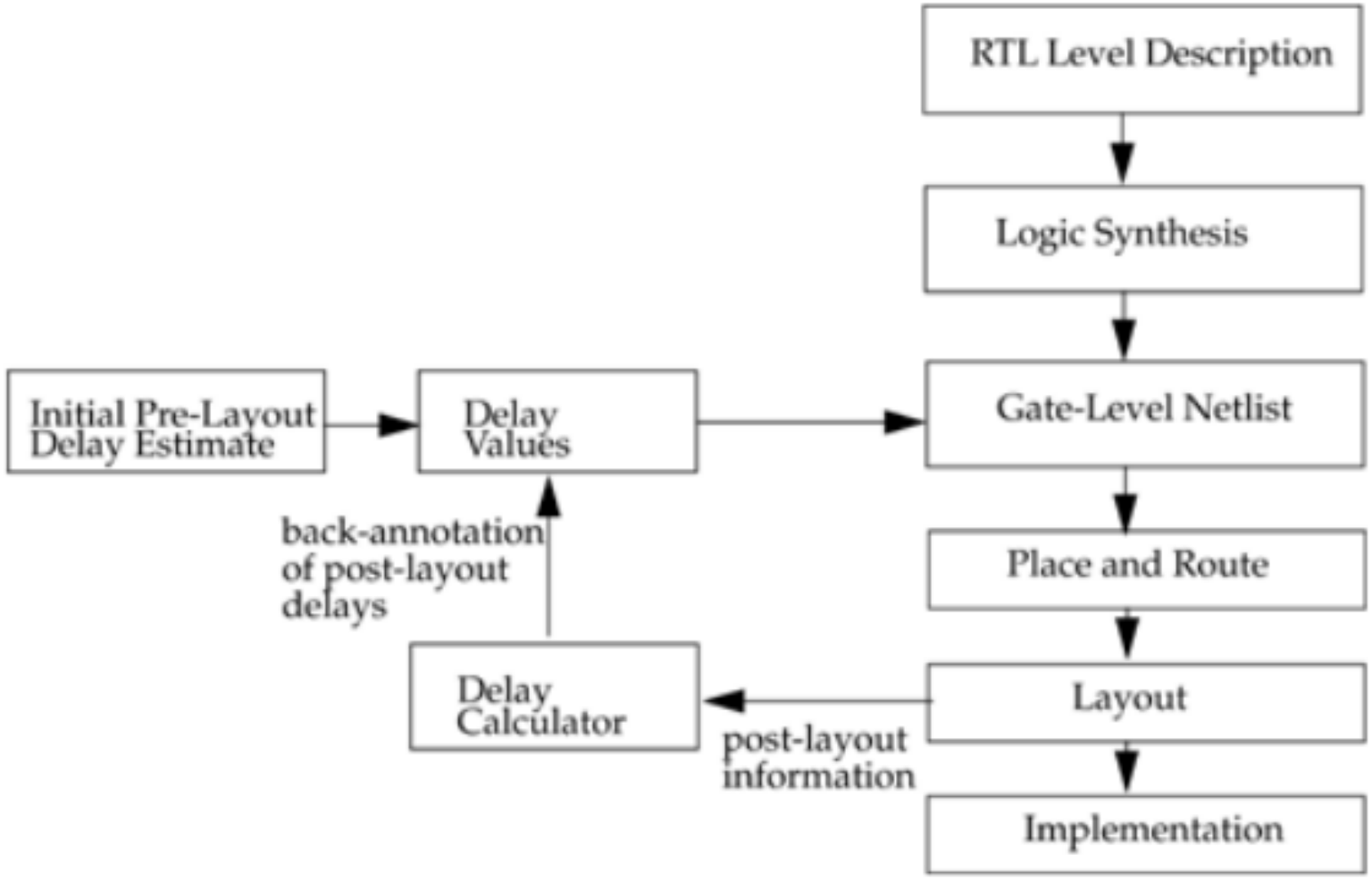
- Example - 3
- Consider R=r*L/W and C=c*W*L, assume r=2 and c=4
- Find out the delay in each leaf



Back Annotation

1. The designer writes the RTL description and then performs functional simulation.
2. The RTL description is converted to a gate-level netlist by a logic synthesis tool.
3. The designer obtains pre-layout estimates of delays in the chip by using a delay calculator and information about the IC fabrication process.
   - Then, the designer does timing simulation or static timing verification of the gate-level netlist, using these preliminary values to check that the gate level netlist meets timing constraints.
4. The gate-level netlist is then converted to layout by a place and route tool. The post-layout delay values are computed from the resistance (R) and capacitance (C) information in the layout.
   - The R and C information is extracted from factors such as geometry and IC fabrication process.
5. The post-layout delay values are back-annotated to modify the delay estimates for the gate level netlist.
   - Timing simulation or static timing verification is run again on the gate-level netlist to check if timing constraints are still satisfied.
6. If design changes are required to meet the timing constraints, the designer has to go back to the RTL level, optimize the design for timing, and then repeat Step 2 through Step 5.

Flow of back annotation



- A standard format called the Standard Delay Format (SDF) is popularly used for back-annotation.
- Details of delay back-annotation can be obtained from the IEEE Standard Verilog Hardware Description Language document