On-chip interconnect

Lecture

Professor Kjell Jeppson

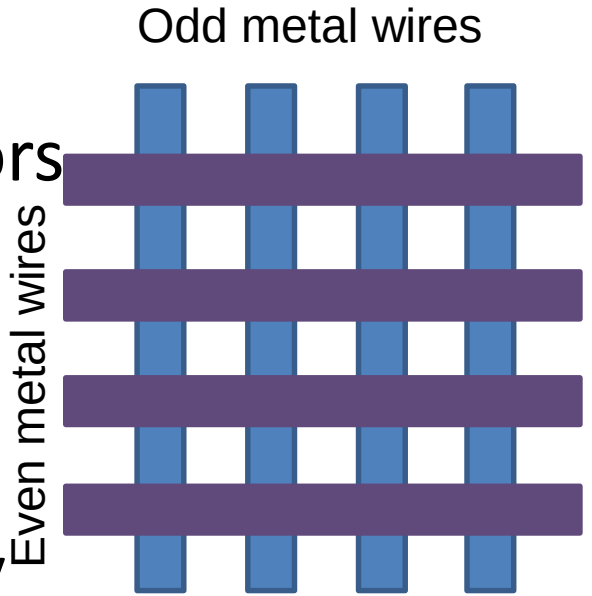
Chalmers University of Technology

Outline

- Introduction
- Interconnect Modeling
  - Wire Resistance
  - Wire Capacitance
- Wire RC Delay
- Elmore delay model (in separate videos)

Introduction

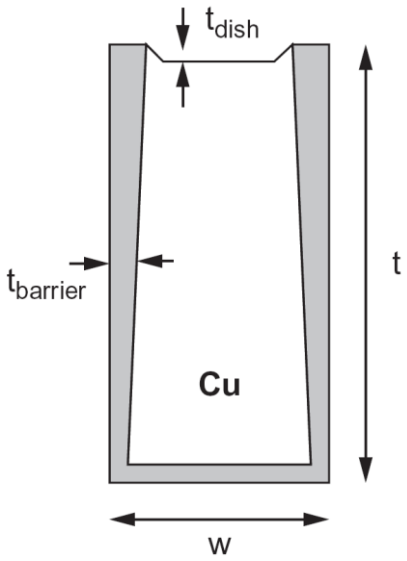
- Chips are mostly made of wires called *interconnect*
  - In stick diagram, wires set size
  - Transistors are little things under the wires
  - Many layers of wires
- Wires are as important as transistors
  - Speed
  - Power
  - Noise
- Alternating layers run orthogonally



Choice of metals

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

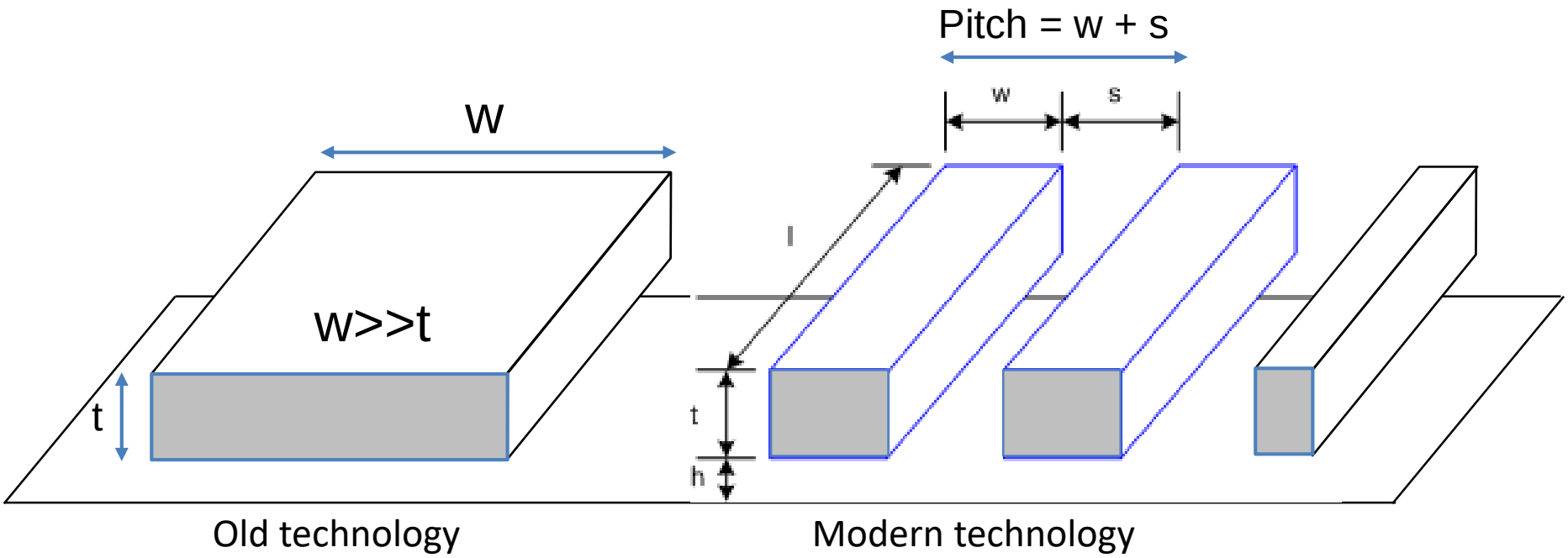
Metal	Bulk resistivity (μΩ·cm)
Silver (Ag)	1.6
Copper (Cu)	1.7
Gold (Au)	2.2
Aluminum (Al)	2.8
Tungsten (W)	5.3
Titanium (Ti)	43.0



Layer stack

- AMS 0.35 μm process has 3 metal layers
  - M1 for within-cell routing
  - M2/M3 for vertical/horizontal routing between cells
- Modern processes use 6-10+ metal layers
  - M1: thin, narrow (< 1.5*minimum feature size)
    - High density wiring in cells
  - Mid layers: thick, wide
    - Global interconnect
  - Top layers: THICK, WIDE
    - For V_{DD}, GND, clk

Wire geometry



Today: pack in many skinny wires!

For long skinny wires resistance cannot be neglected since cross sectional area shrinks with feature size, wire length stays the same or increases.

Hence: wire resistances can no longer be neglected!

Wire geometry

- Long skinny wires
  - wire resistance cannot be neglected

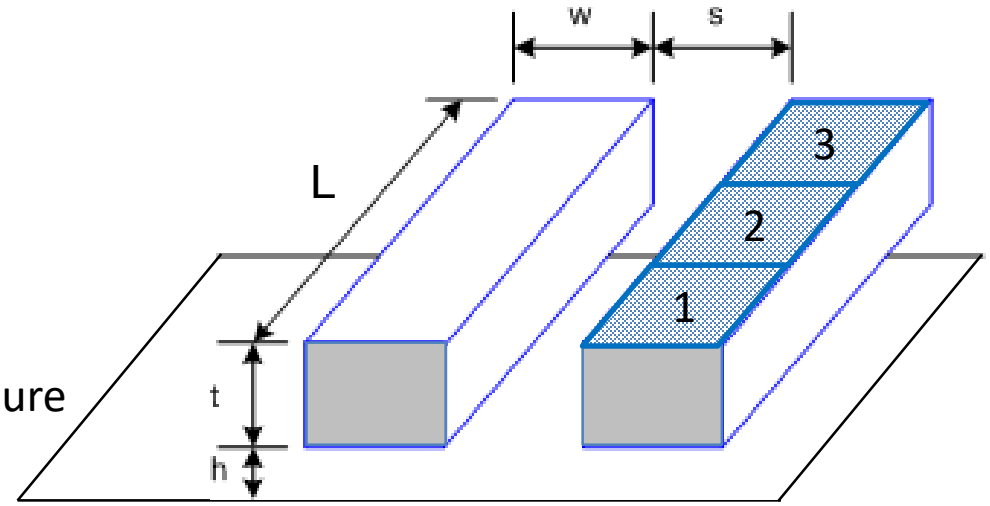
$$R = \rho \frac{L}{A} = rL$$

- Wire length is increasing with large chips
- Wire cross sectional area shrinks with feature size
- Wires can no longer be modeled as capacitances alone
- We need improved wire models that considers wire resistance along with wire capacitance!
- Model must be distributed between at least two circuit nodes!

Wire resistance

$$R = \rho \frac{L}{A} = rL$$

$$\text{Area} = w \times t \Rightarrow R = \rho \frac{L}{w \times t}$$



All wires in certain layer has
the same thickness, *t,* hence
sheet resistivity is a convenient measure

$$R_S = \frac{\rho}{t}$$ in ohms per square, Ω/□

The ”number of squares” in the direction of current flow is L/W

If we assign a certain width, W_0, to all wires in certain layer
they will all have the same resistance per unit length, $r=R_S/W_0$

Example

- Estimate the sheet resistance of a 220 nm thick copper wire if the resistivity of the thin copper film is 22 nΩ·m.

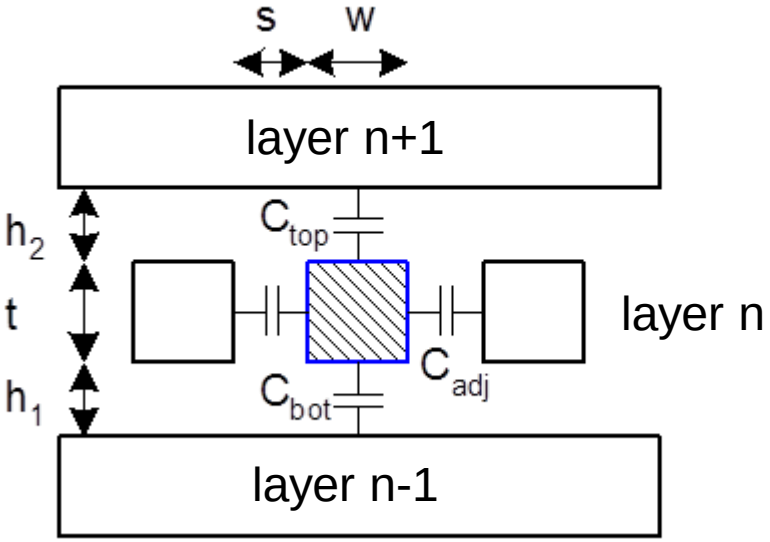
$$R_S = \frac{\rho}{t} = \frac{22\ \text{n}\Omega\cdot\text{m}}{220\ \text{nm}} = 0.10\ \Omega/\square$$

- Find the total resistance if the wire is 0.125 μm wide and 1 mm long. (Ignore the barrier layer)

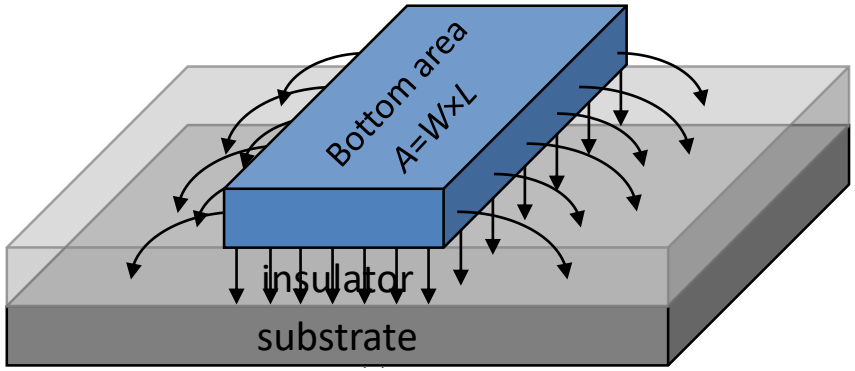
$$R = \underbrace{0.10\ \Omega/\square}_{R_S} \times \underbrace{\frac{1000\ \mu\text{m}}{0.125\ \mu\text{m}}}_{L/W = 8000\ \text{"squares"}} = 800\ \Omega$$

- Wires 125 nm wide have a resistivity per unit length of 800 Ω/mm

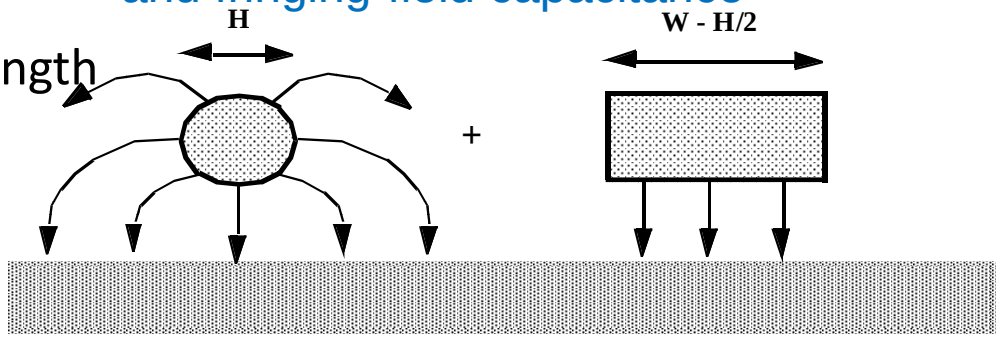
Wire capacitance



Parallel plate capacitance



and fringing-field capacitance



Wires have a capacitance *c* per unit length

- to neighbors in the same layer
- to layers above and below

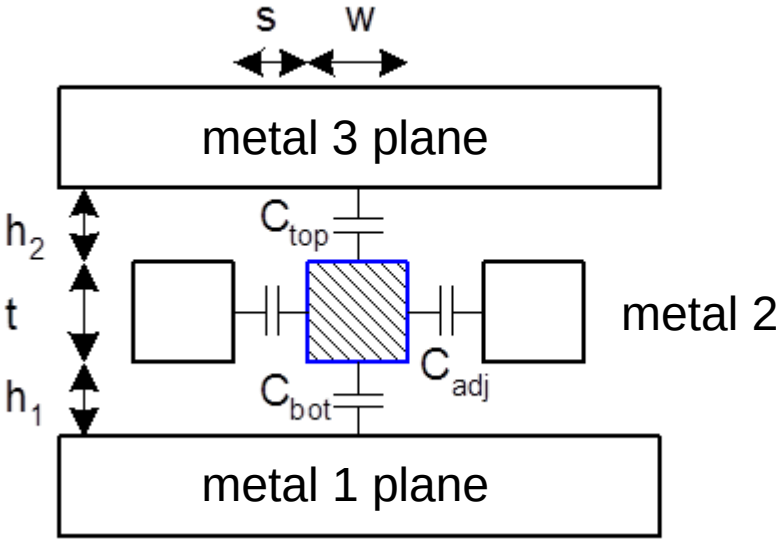
Parallel-plate capacitance equation

- $C_{pp} = \varepsilon A/h$, where top/bottom area $A=W\times L$
- $\varepsilon=\kappa\varepsilon_0$, $\varepsilon_0=8.85\cdot10^{-12}$ F/m in vacuum
- SiO_2 permittivity is $\kappa\approx4$
- low-kappa materials have $\kappa<3$

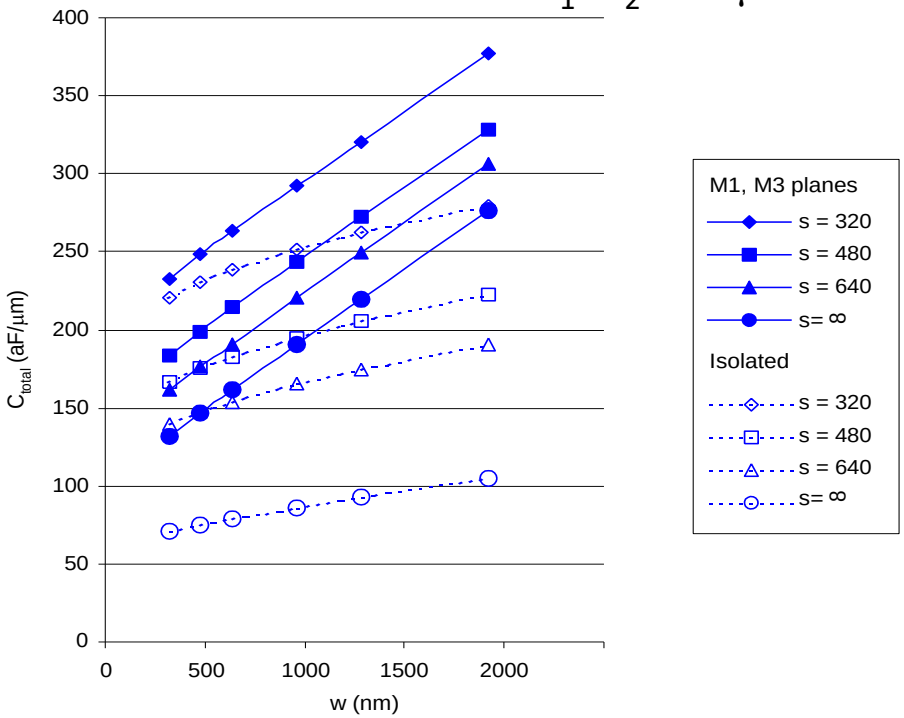
$$c_{wire} = c_{pp} + c_{fringe} = \frac{w\varepsilon_{di}}{t_{di}} + \frac{2\pi\varepsilon_{di}}{\log(t_{di}/H)}$$

Note: Cap is per unit length

Metal2 Capacitance Data



$t=h_1=h_2=0.7$ μm



Wire has capacitance *c* per unit length

- to neighbors in the same layer
- to layers above and below

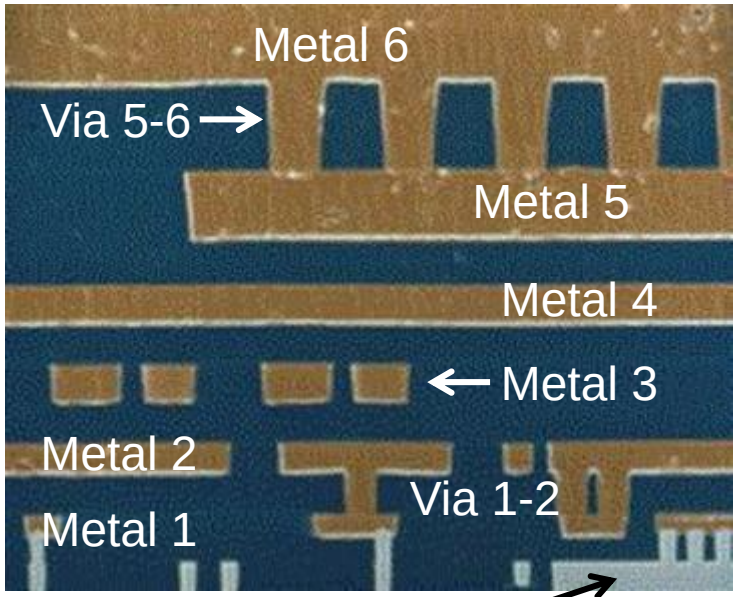
- Wires typically have capacitances about ~ 0.2 fF/μm, i. e. 200 fF/mm
- Compare with the 1.2 fF/μm for the MOSFET gate capacitances

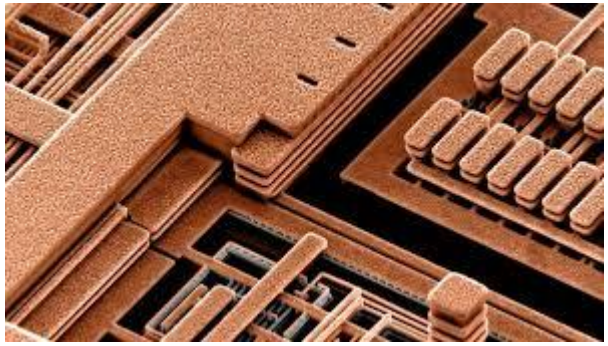
Wire RC delay

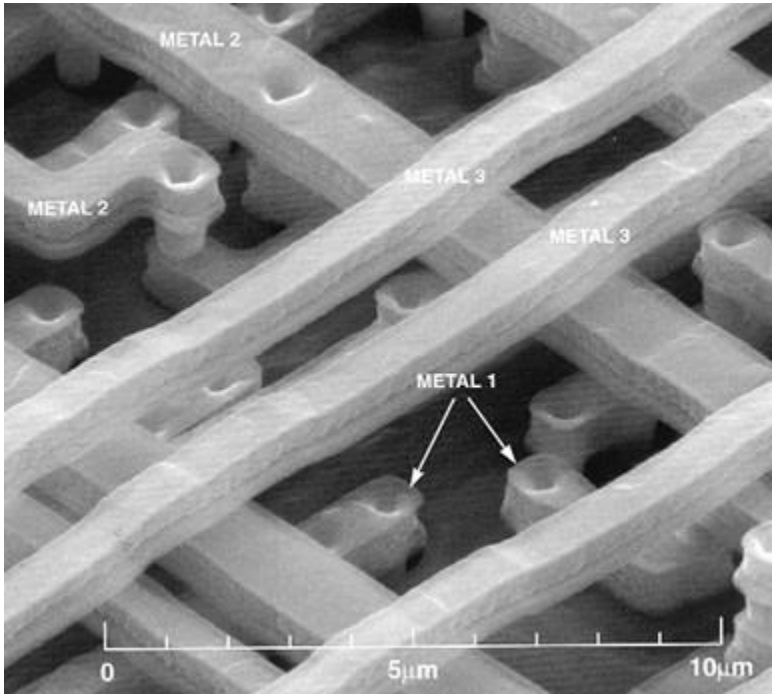
- Wire RC product increases as L^2 with wire length L
- For our 1 mm example wire, the *RC* product was 160 ps
  - *r* is wire resistance per unit length ~800 Ω/mm (0.8 kΩ /mm)
  - *c* is wire capacitance per unit length ~200 fF/mm
- In comparison, the *RC* time constant of an inverter with a FO4 was previously found to be 36 ps (5×7.2 ps)
  - FO4 delay = 0.7×36 = 25 ps

$$RC = rcL^2$$

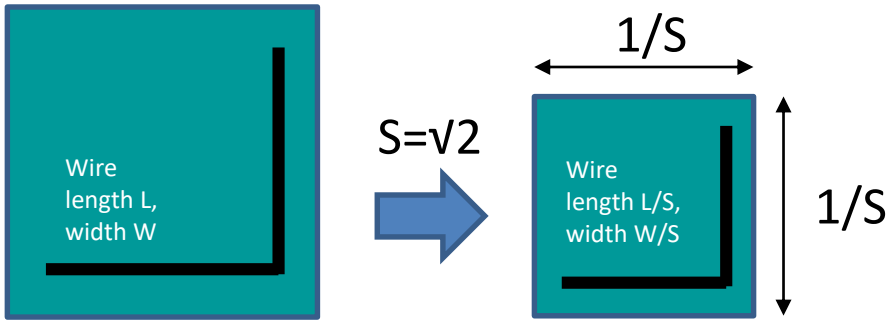
Modern interconnect







Wire delay scaling – Local wires



stays constant

$$R = R_S \frac{L/S}{W/S}, \quad R_S = \frac{\rho}{t}$$

- For local wires crossing the same amount of circuitry
  - Resistance stays roughly constant
    - Aspect ratio does not change
    - Sheet resistivity does not change if wire height stays large and/or change material to copper
  - Capacitance decreases by scaling factor
    - Cap/unit length stays constant, while length decreases

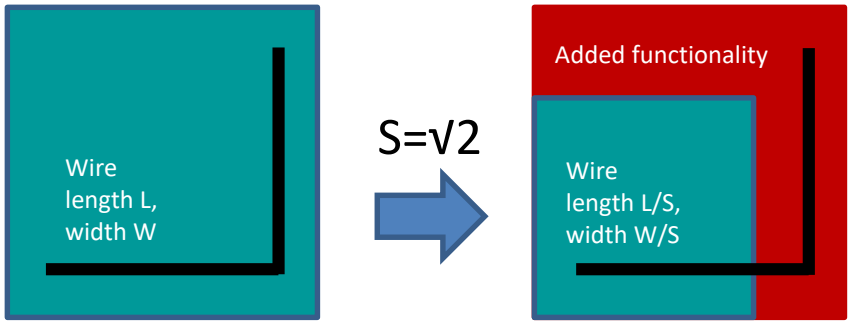
stays constant

$$C = (WC_{ox})\, L/S$$

- Hence, wire delay tracks gate delay ~1/S

From Mark Horowitz at Design Automation Conference 2000

Wire delay scaling – Global wires



stays constant

$$R = R_S \frac{L}{W/S} \sim S$$

- For global wires crossing the whole chip
  - Resistance grows linearly (with scaling factor)
  - Capacitance stays fixed
    - Cap/unit length stays constant, as does wire length

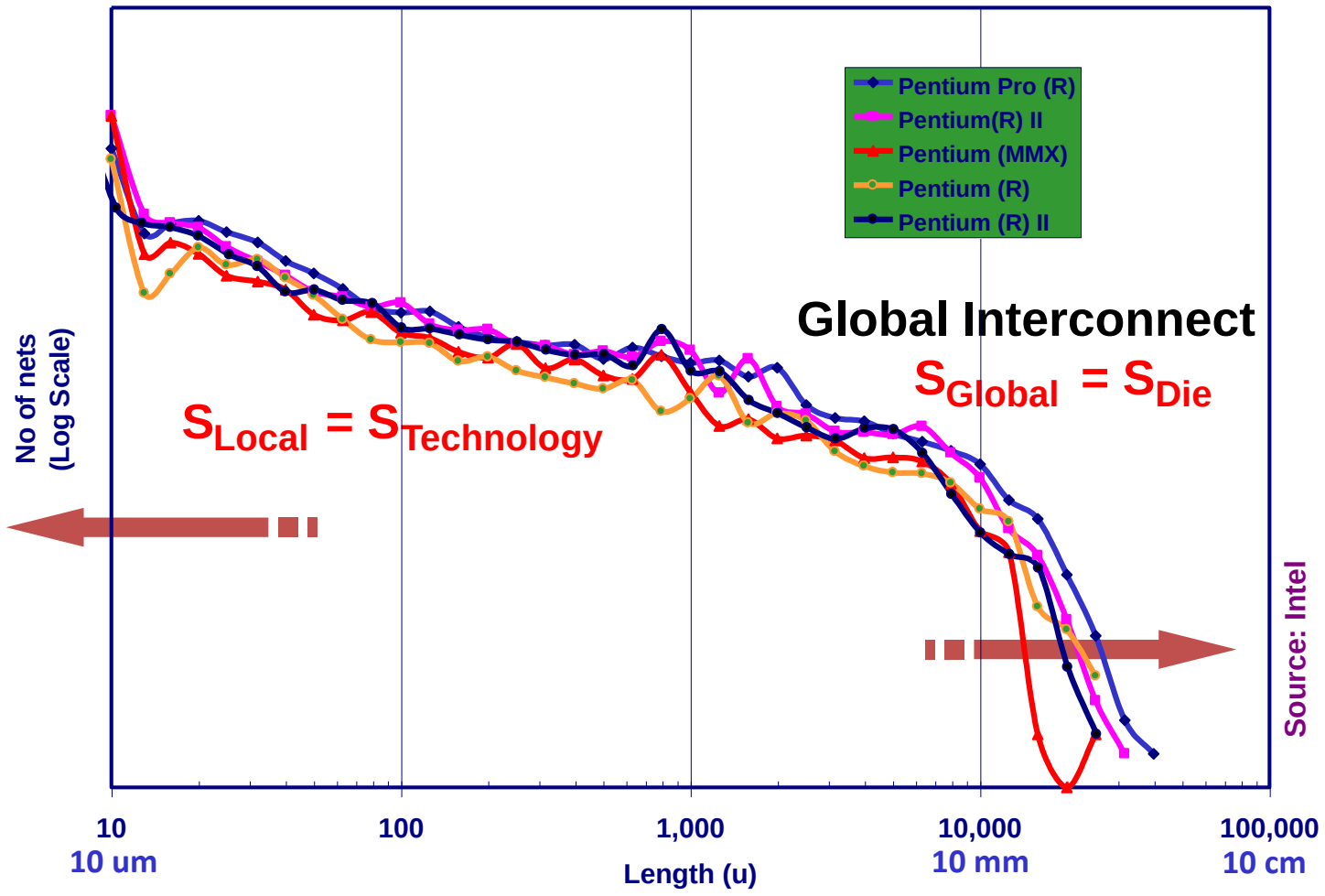
stays constant

$$C = (WC_{ox})L$$

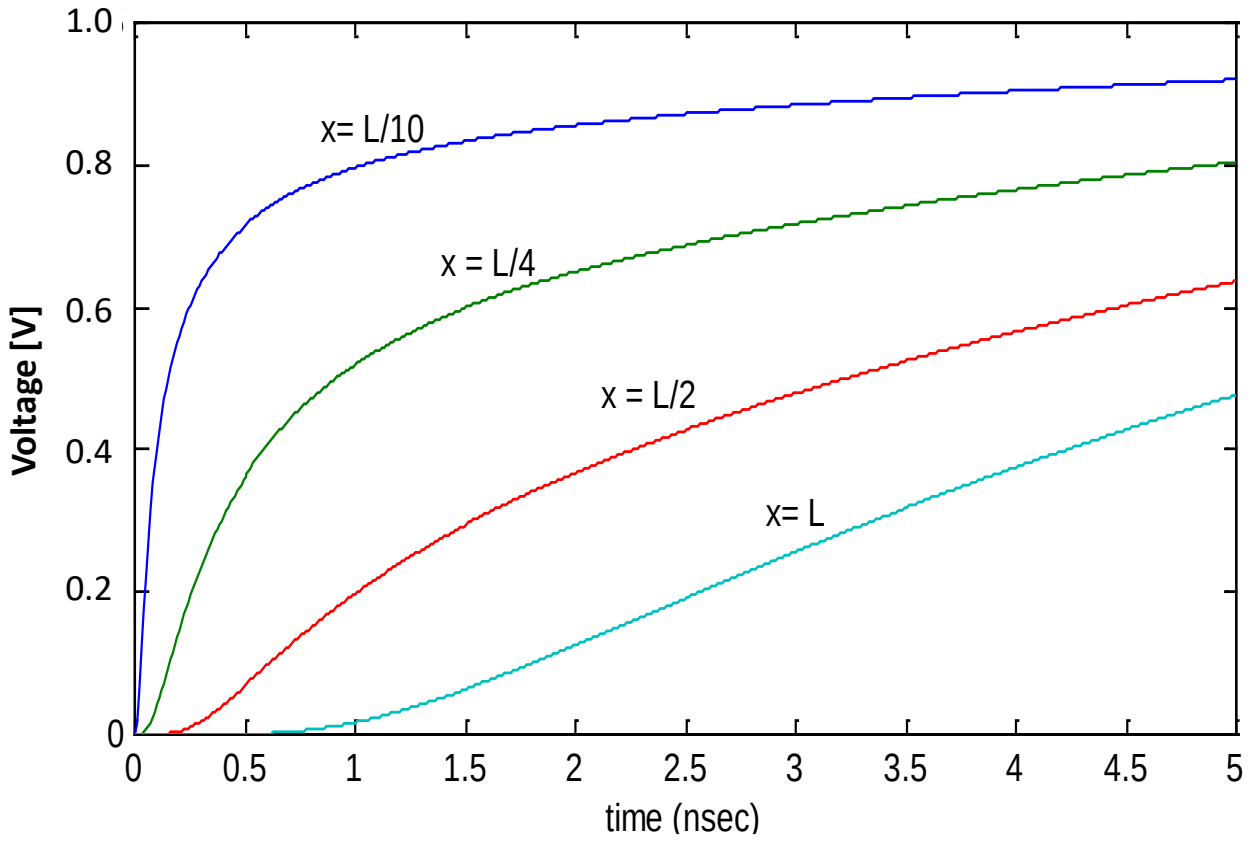
- Two opposite trends:
  - Wire delay increases (~S) - gate delay decreases (~1/S)

From Mark Horowitz at Design Automation Conference 2000

Modern Interconnect



Step-response to a rising input voltage along an RC wire as a function of time and wire length



Summary

We have

- discussed the importance of accurate wire modeling considering not only wire capacitance but also wire resistance as wires get longer and skinnier
- defined the concept of sheet resistance in ohms per square
- had a look at typical on-chip wire length distributions
- had a look at wire capacitance dependence on the surrounding wiring on top, below, and along sidewalls

Thanks a lot for listening!