
EECS 427

Lecture 18: Interconnects

Readings: 9.2-9.4

Reminders

- **Deadlines**
 - CAD8 is due Saturday 11/21 at 11:59 pm
- **Quiz 2 is on Wednesday 11/25**
 - Extended office hours this week
 - Sunday: noon–6 pm
 - Monday 3–3:30 pm and after 5 pm
 - Tuesday 3–6 pm
 - Half-lecture review in class on Monday 11/23
 - Sample Quiz 2 posted
- **What is remaining after Quiz 2**
 - 2.5 weeks to finish your project by 12/14

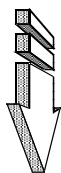
Last Time

- SRAM
- DRAM overview (3T and 1T)
- Memory reliability and yield
- Memory power reduction

Impact of Interconnect Parasitics

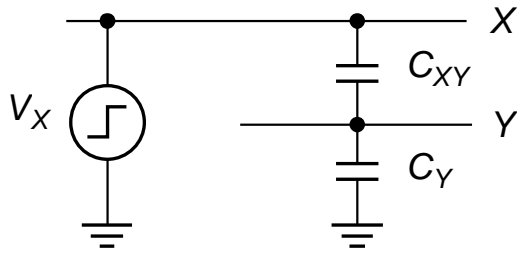
- **Reduce Robustness**
- **Affect Performance**
 - Increase delay
 - Increase power dissipation

Classes of Parasitics



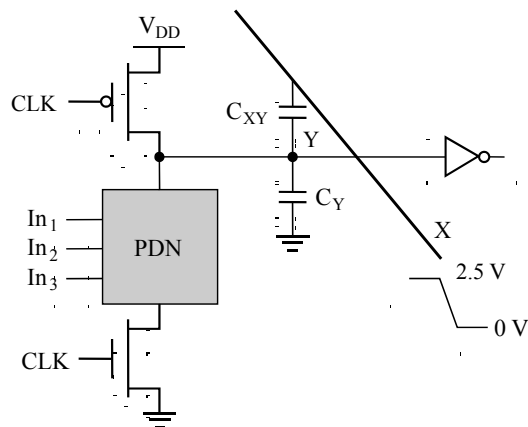
- **Capacitive**
- **Resistive**
- **Inductive**

Capacitive Cross Talk



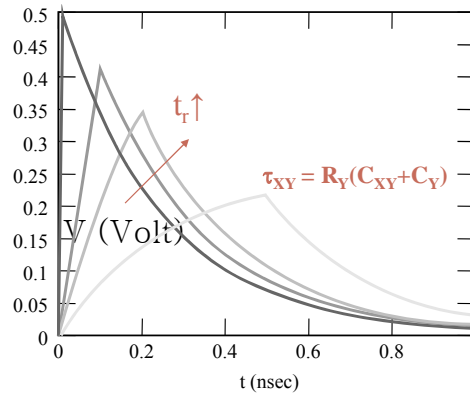
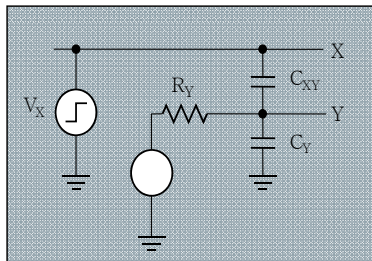
$$\Delta V_Y = \frac{C_{XY}}{C_Y + C_{XY}} \Delta V_X$$

Capacitive Cross Talk Dynamic Node



3 x 1 μm overlap: 0.19 V disturbance

Capacitive Cross Talk Driven Node

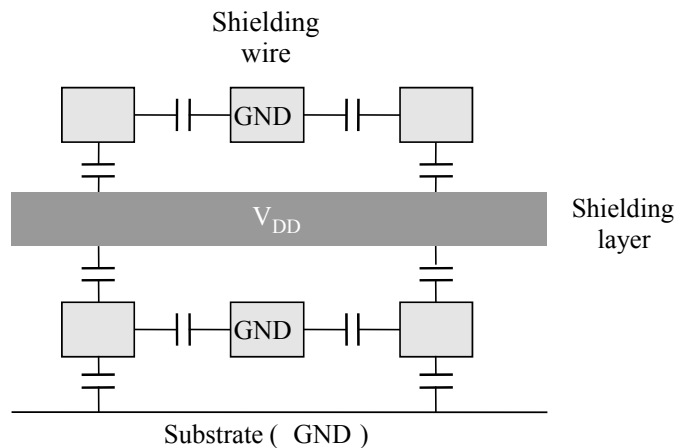


Keep time-constant smaller than rise time

Dealing with Capacitive Cross Talk

- Avoid floating nodes
- Protect sensitive nodes
- Make rise and fall times as large as possible
- Differential signaling
- Do not run wires together for a long distance
- Use shielding wires
- Use shielding layers

Shielding

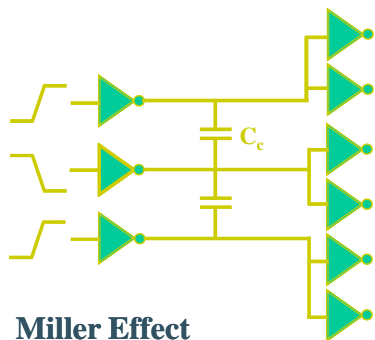


EECS 427 F09

Lecture 18

9

Cross Talk and Performance



- When neighboring lines switch in opposite direction of victim line, delay increases

DELAY DEPENDENT UPON ACTIVITY IN NEIGHBORING WIRES

Miller Effect

- Both terminals of capacitor are switched in opposite directions ($0 \rightarrow V_{dd}, V_{dd} \rightarrow 0$)
- Effective voltage is doubled and additional charge is needed (from $Q=CV$)

EECS 427 F09

Lecture 18

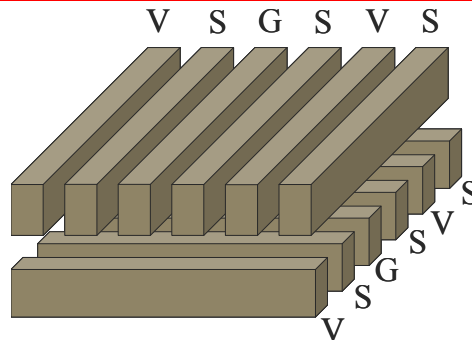
10

Impact of Cross Talk on Delay

bit $k-1$	bit k	bit $k+1$	Delay factor g
↑	↑	↑	1
↑	↑	—	$1+r$
↑	↑	↓	$1+2r$
—	↑	—	$1+2r$
—	↑	↓	$1+3r$
↓	↑	↓	$1+4r$

r is ratio between capacitance to GND and to neighbor

Structured Predictable Interconnect



Example: Dense Wire Fabric ([Sunil Kathri])

Trade-off:

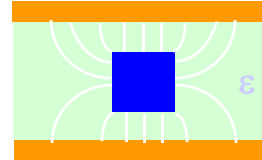
- Cross-coupling capacitance 40x lower, 2% delay variation
- Increase in area and overall capacitance

Also: FPGAs, VPGAs

Interconnect Projections

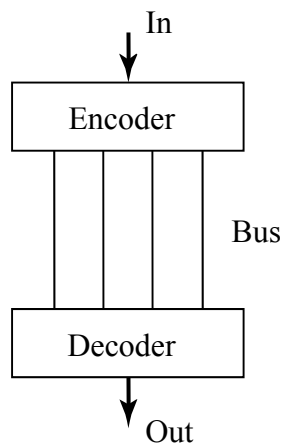
Low-k dielectrics

- Both *delay and power are reduced* by dropping interconnect capacitance
- Types of low-k materials include: inorganic (SiO_2), organic (Polyimides) and aerogels (ultra low-k)
- The numbers below are on the conservative side of the NRTS roadmap

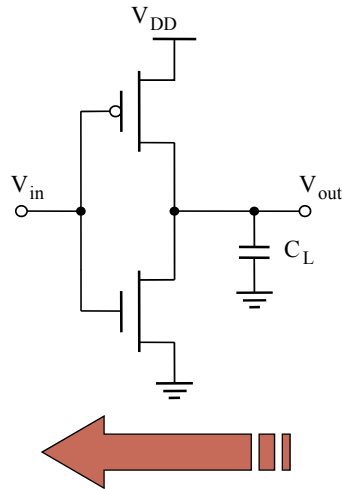


Generation	0.25 μm	0.18 μm	0.13 μm	0.1 μm	0.07 μm	0.05 μm
Dielectric Constant	3.3	2.7	2.3	2.0	1.8	1.5

Encoding Data Avoids Worst-Case Conditions



Driving Large Capacitances



$$t_p = \frac{C_L V_{swing}}{I_{av}}$$

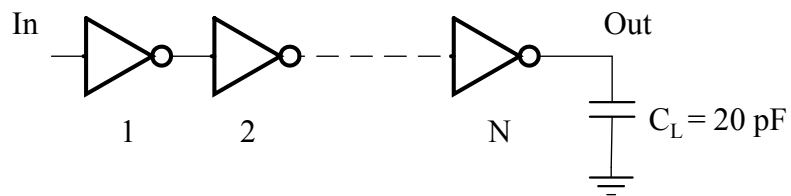
- Transistor Sizing
- Cascaded Buffers

EECS 427 F09

Lecture 18

15

Using Cascaded Buffers



0.25 μm process
 $C_{in} = 2.5 \text{ fF}$
 $t_{p0} = 30 \text{ ps}$

$F = C_L / C_{in} = 8000$
 $f_{opt} = 3.6 \text{ N} = 7$
 $t_p = 0.76 \text{ ns}$

(See Chapter 5)

EECS 427 F09

Lecture 18

16

Output Driver Design

Trade off Performance for Area and Energy

Given t_{pmax} find N and f

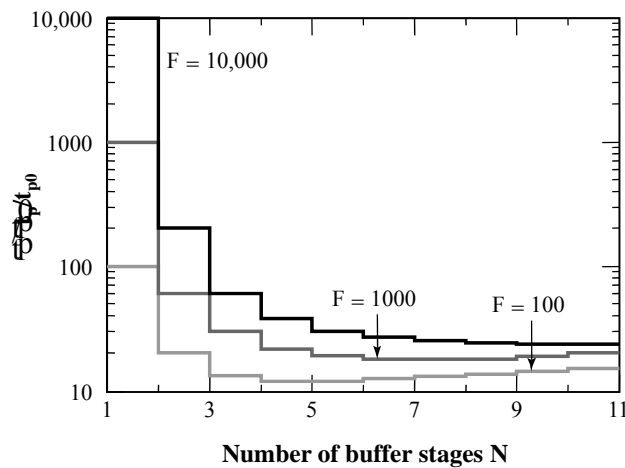
- Area

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

- Energy

$$E_{driver} = (1 + f + f^2 + \dots + f^{N-1})C_i V_{DD}^2 = \frac{F - 1}{f - 1} C_i V_{DD}^2 \approx \frac{C_L}{f - 1} V_{DD}^2$$

Delay as a Function of F and N



Output Driver Design

0.25 μm process, $C_L = 20 \text{ pF}$

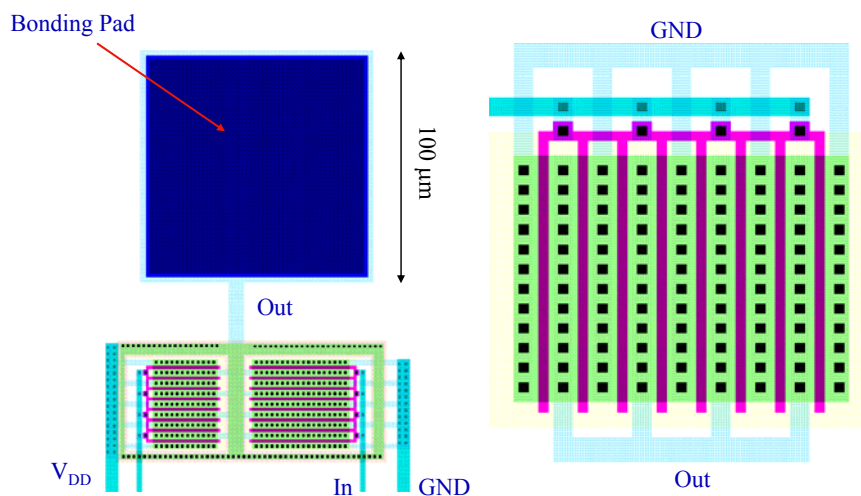
Transistor Sizes for optimally-sized cascaded buffer $t_p = 0.76 \text{ ns}$

Stage	1	2	3	4	5	6	7
W_n (μm)	0.375	1.35	4.86	17.5	63	226.8	816.5
W_p (μm)	0.71	2.56	9.2	33.1	119.2	429.3	1545.5

Transistor Sizes of redesigned cascaded buffer $t_p = 1.8 \text{ ns}$

Stage	1	2	3
W_n (μm)	0.375	7.5	150
W_p (μm)	0.71	14.4	284

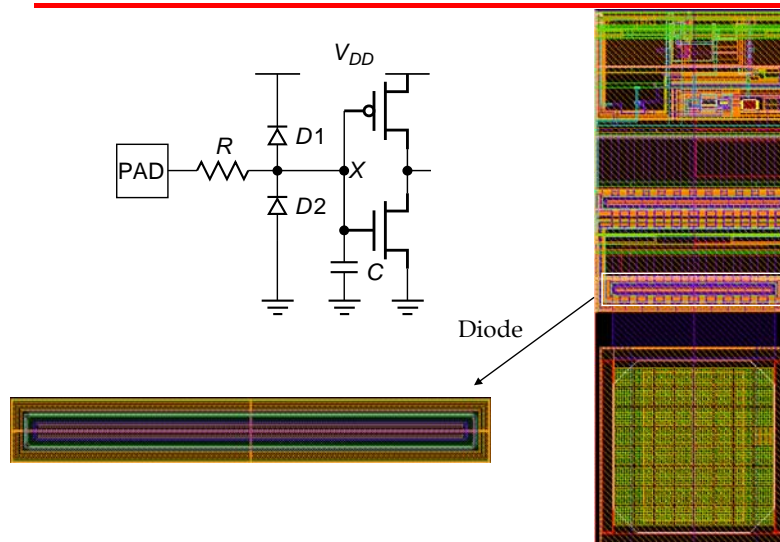
Bonding Pad Design



ESD Protection

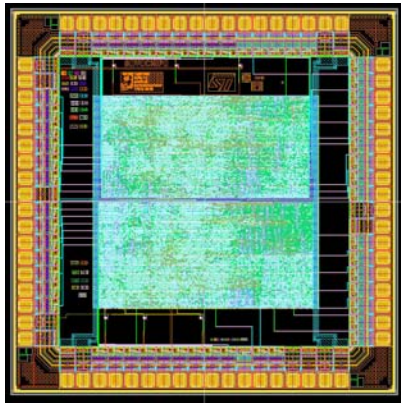
- When a chip is connected to a board, there is unknown (potentially large) static voltage difference
- Equalizing potentials requires (large) charge flow through the pads
- Diodes sink this charge into the substrate – need guard rings to pick it up.

ESD Protection

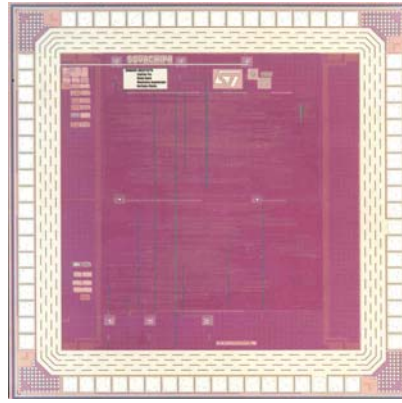


Pad Frame

Layout



Die Photo

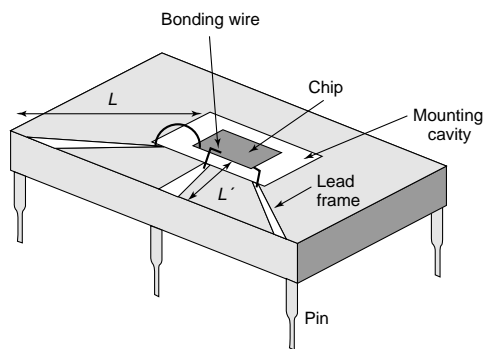


EECS 427 F09

Lecture 18

23

Chip Packaging



- Bond wires ($\sim 25\mu\text{m}$) are used to connect the package to the chip

- Pads are arranged in a frame around the chip

- Pads are relatively large ($\sim 100\mu\text{m}$ in $0.25\mu\text{m}$ technology), with large pitch ($100\mu\text{m}$)

- Many chips areas are 'pad limited'

EECS 427 F09

Lecture 18

24

Chip Packaging

- An alternative is 'flip-chip':
 - Pads are distributed around the chip
 - The soldering balls are placed on pads
 - The chip is 'flipped' onto the package
 - Can have many more pads

Reducing the swing

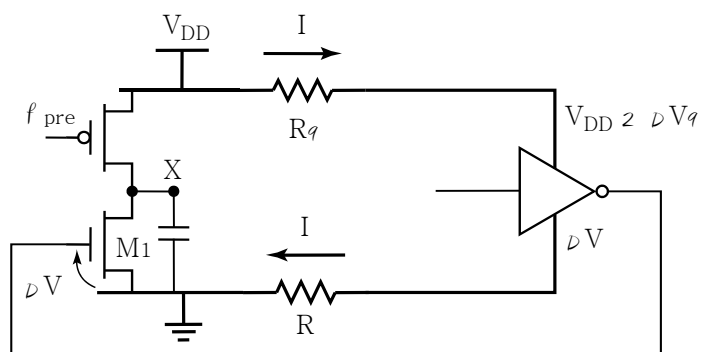
$$t_{pHL} = \frac{C_L V_{swing}/2}{I_{av}}$$

- Reducing the swing potentially yields linear reduction in delay
- Also results in reduction in power dissipation
- Delay penalty is paid by the receiver
- Requires use of "sense amplifier" to restore signal level
- Frequently designed differentially (e.g. LVDS)

Impact of Resistance

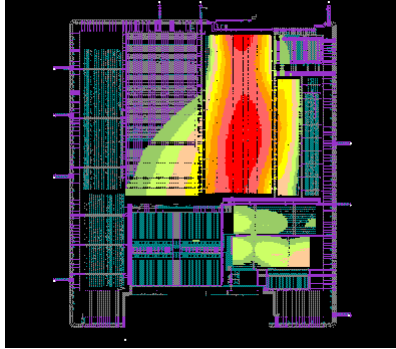
- We have already learned how to drive RC interconnect
- Impact of resistance is commonly seen in power supply distribution:
 - IR drop
 - Voltage variations
- Power supply is distributed to minimize the IR drop and the change in current due to switching of gates

RI Introduced Noise

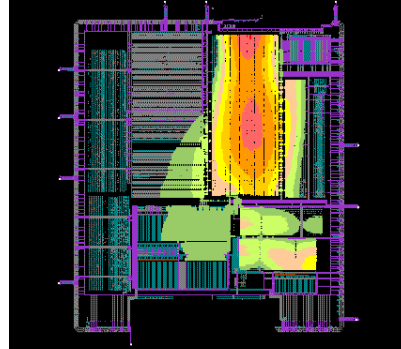


Resistance and the Power Distribution Problem

Before



After

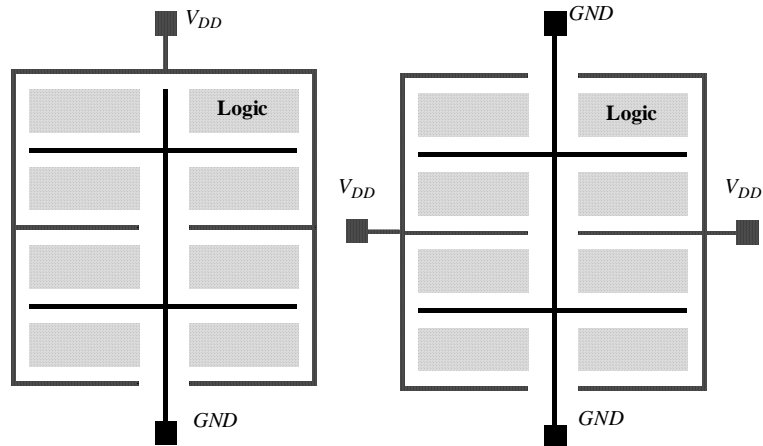


- Requires fast and accurate peak current prediction
- Heavily influenced by packaging technology

Power Distribution

- Low-level distribution is in Metal 1
- Power has to be 'strapped' in higher layers of metal.
- The spacing is set by IR drop, electromigration, inductive effects
- Always use multiple contacts on straps

Power and Ground Distribution



(a) Finger-shaped network

(b) Network with multiple supply pins

EECS 427 F09

Lecture 18

31

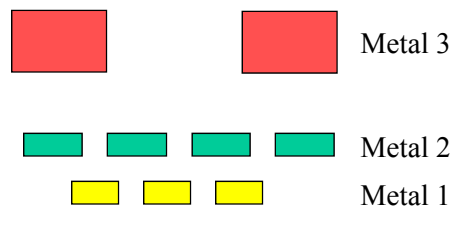
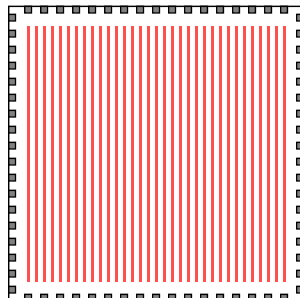
3 Metal Layer Approach (EV4)

3rd "coarse and thick" metal layer added to the technology for EV4 design

Power supplied from two sides of the die via 3rd metal layer

2nd metal layer used to form power grid

90% of 3rd metal layer used for power/clock routing



EECS 427 F09

Courtesy Compaq
Lecture 18

32

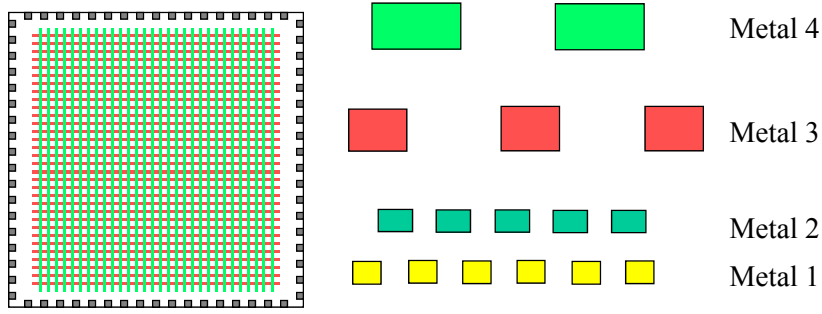
4 Metal Layers Approach (EV5)

4th "coarse and thick" metal layer added to the technology for EV5 design

Power supplied from four sides of the die

Grid strapping done all in coarse metal

90% of 3rd and 4th metals used for power/clock routing



EECS 427 F09

Courtesy Compaq
Lecture 18

33

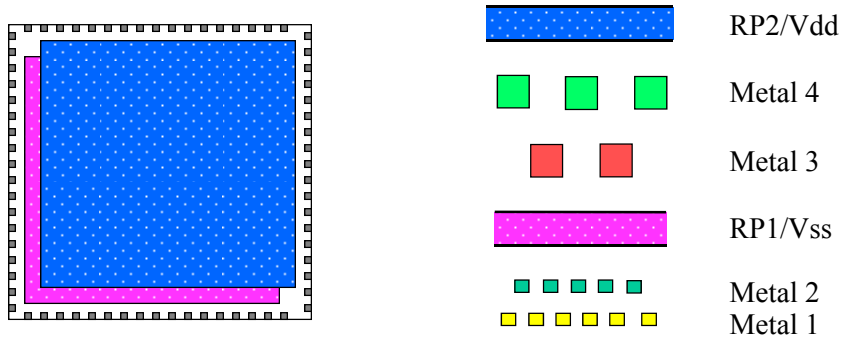
6 Metal Layer Approach – EV6

2 reference plane metal layers added to the technology for EV6 design

Solid planes dedicated to Vdd/Vss

Significantly lowers resistance of grid

Lowers on-chip inductance

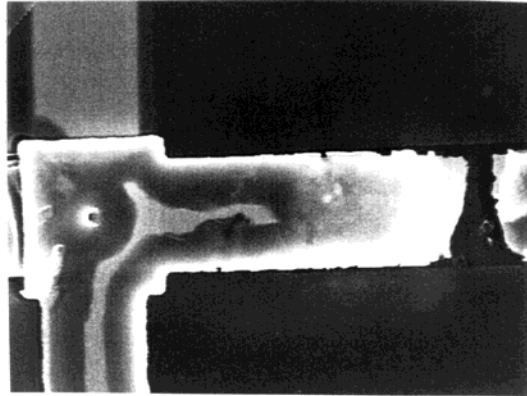


EECS 427 F09

Courtesy Compaq
Lecture 18

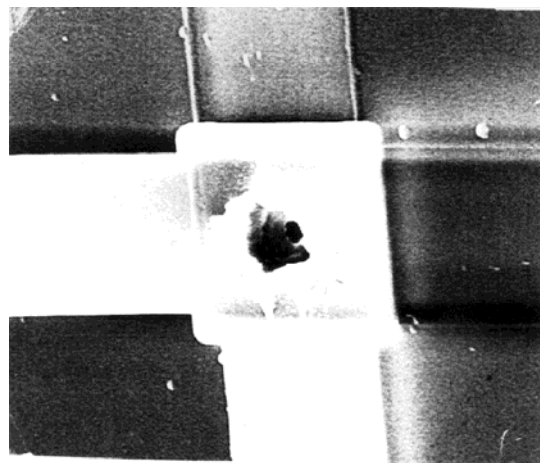
34

Electromigration (1)

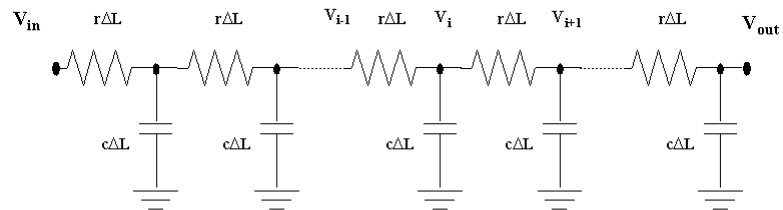


Limits dc-current to $1 \text{ mA}/\mu\text{m}$

Electromigration (2)



The Elmore Delay RC Chain



$$\tau_N = \sum_{i=1}^N R_i \sum_{j=i}^N C_j = \sum_{i=1}^N C_i \sum_{j=1}^i R_j$$

37

Wire Model

Assume: Wire modeled by N equal-length segments

$$\tau_{DN} = \left(\frac{L}{N}\right)^2 (rc + 2rc + \dots + Nrc) = (rcL^2) \frac{N(N+1)}{2N^2} = RC \frac{N+1}{2N}$$

For large values of N :

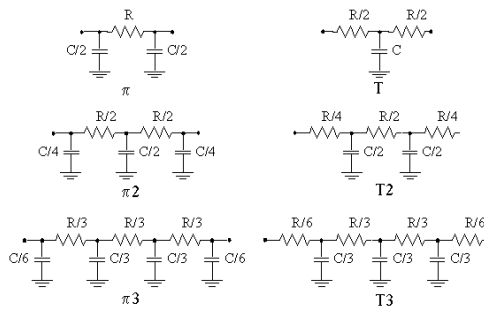
$$\tau_{DN} = \frac{RC}{2} = \frac{rcL^2}{2}$$

38

RC-Model

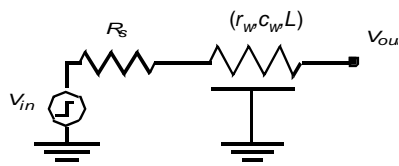
Voltage Range	Lumped RC-network	Distributed RC-network
0→50% (t_p)	0.69 RC	0.38 RC
0→63% (τ)	RC	0.5 RC
10%→90% (t_r)	2.2 RC	0.9 RC

Step Response of Lumped and Distributed RC Networks:
Points of Interest.



39

Driving an RC-line



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

40

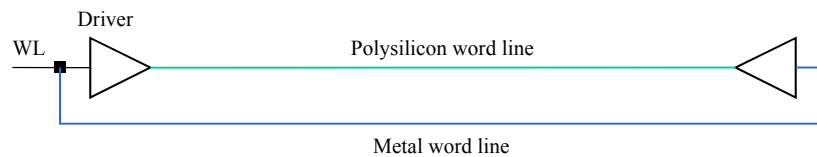
The Global Wire Problem

$$T_d = 0.377R_w C_w + 0.693(R_d C_{out} + R_d C_w + R_w C_{out})$$

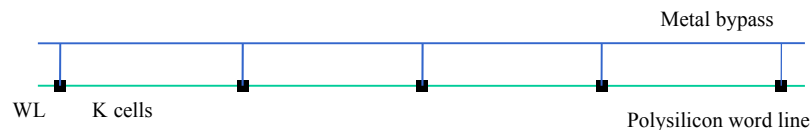
Challenges

- No further improvements to be expected after the introduction of Copper (superconducting, optical?)
- Design solutions
 - Use of fat wires
 - Insert repeaters — but might become prohibitive (power, area)
 - Efficient chip floorplanning
- Towards “communication-based” design
 - How to deal with latency?
 - Is synchronicity an absolute necessity?

Using Bypasses

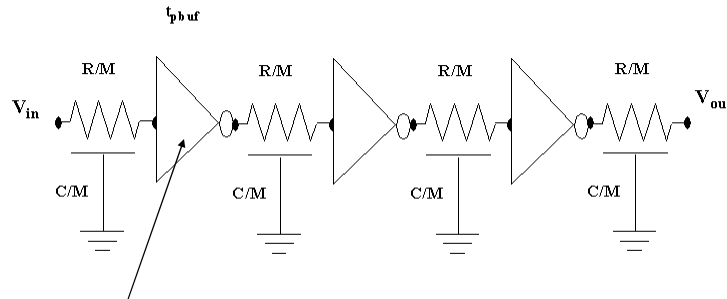


Driving a word line from both sides



Using a metal bypass

Reducing RC-delay



Repeater

$$M = L \sqrt{\frac{0.38rc}{t_{buf}}} \quad (\text{chapter 5})$$

EECS 427 F09

Lecture 18

43

Repeater Insertion (Revisited)

Taking the repeater loading into account

$$m_{opt} = L \sqrt{\frac{0.38rc}{0.69R_d C_d (\gamma + 1)}} = \sqrt{\frac{t_{pwire(unbuffered)}}{t_{p1}}}$$

$$s_{opt} = \sqrt{\frac{R_d c}{r C_d}}$$

For a given technology and a given interconnect layer, there exists an optimal length of the wire segments between repeaters. The delay of these wire segments is independent of the routing layer!

$$L_{crit} = \frac{L}{m_{opt}} = \sqrt{\frac{t_{p1}}{0.38rc}} \quad t_{p,crit} = \frac{t_{p,min}}{m_{opt}} = 2 \left(1 + \sqrt{\frac{0.69}{0.38(1+\gamma)}} \right) t_{p1}$$