

Coping with Interconnect

Impact of Interconnect Parasitics

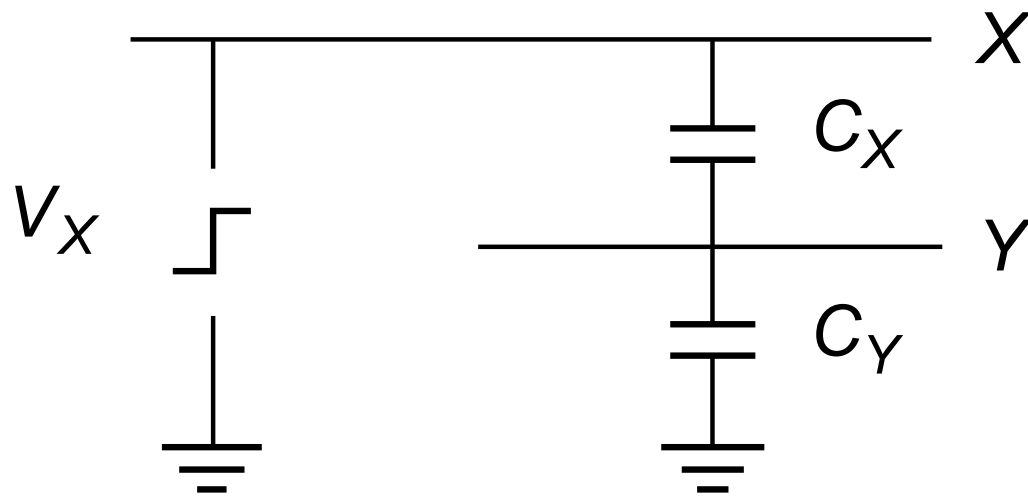
- **Reduce Robustness**
- **Affect Performance**
 - Increase delay
 - Increase power dissipation

Classes of Parasitics



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

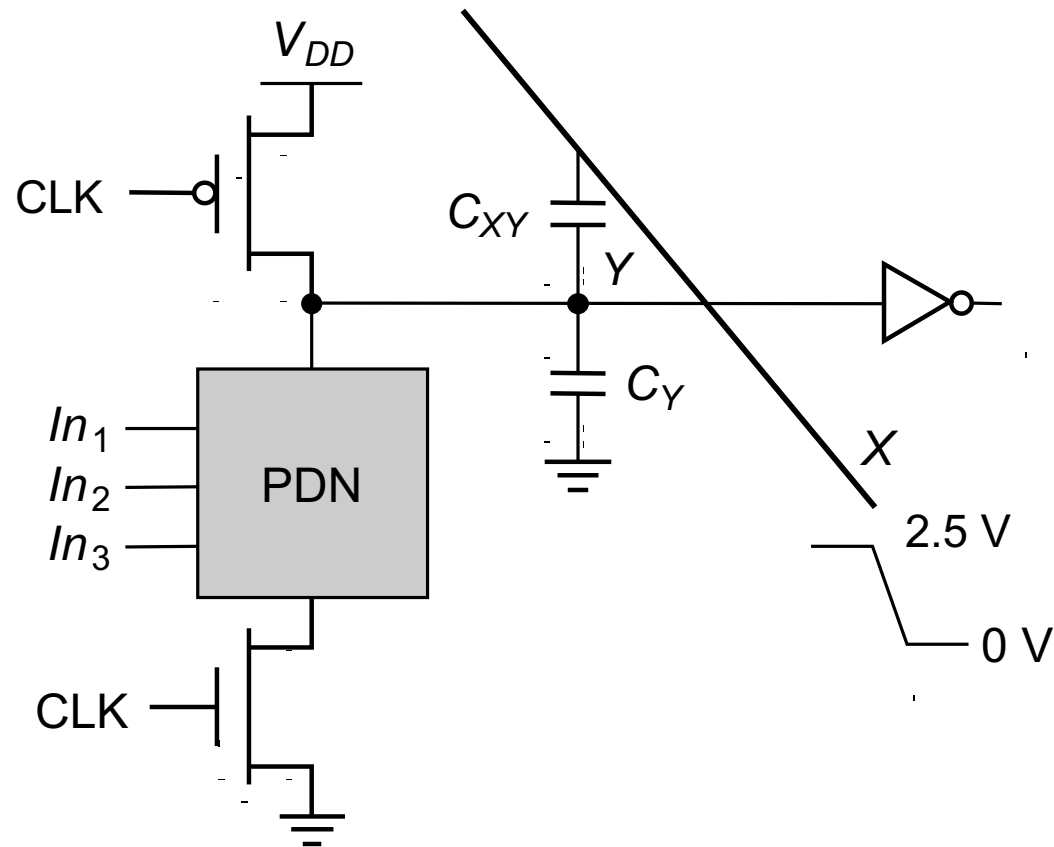
Capacitive Coupling (Coupling Cross Talk)



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

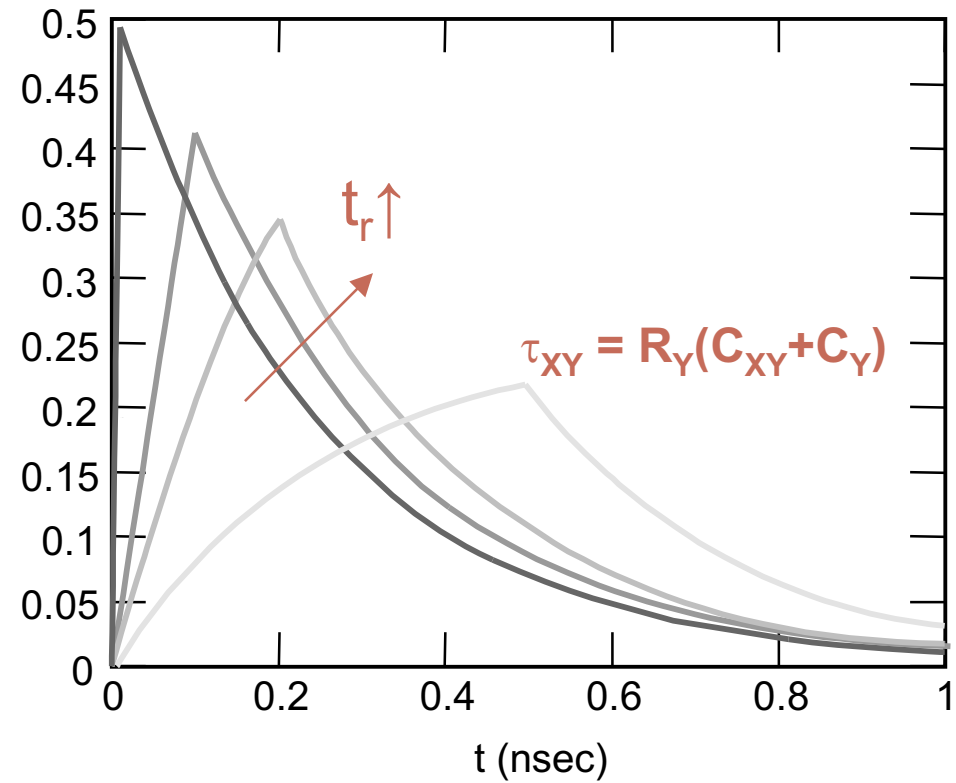
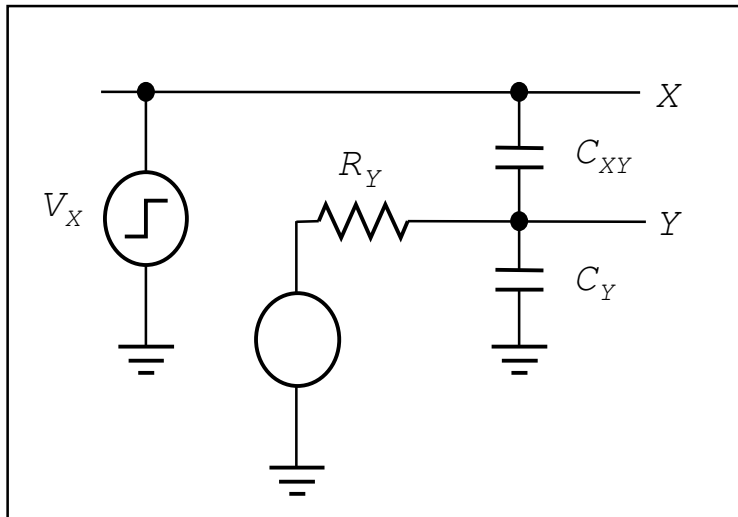
Capacitive Cross Talk

Dynamic Node



3 x 1 μm overlap: 0.19 V disturbance

Capacitive Cross Talk Driven Node

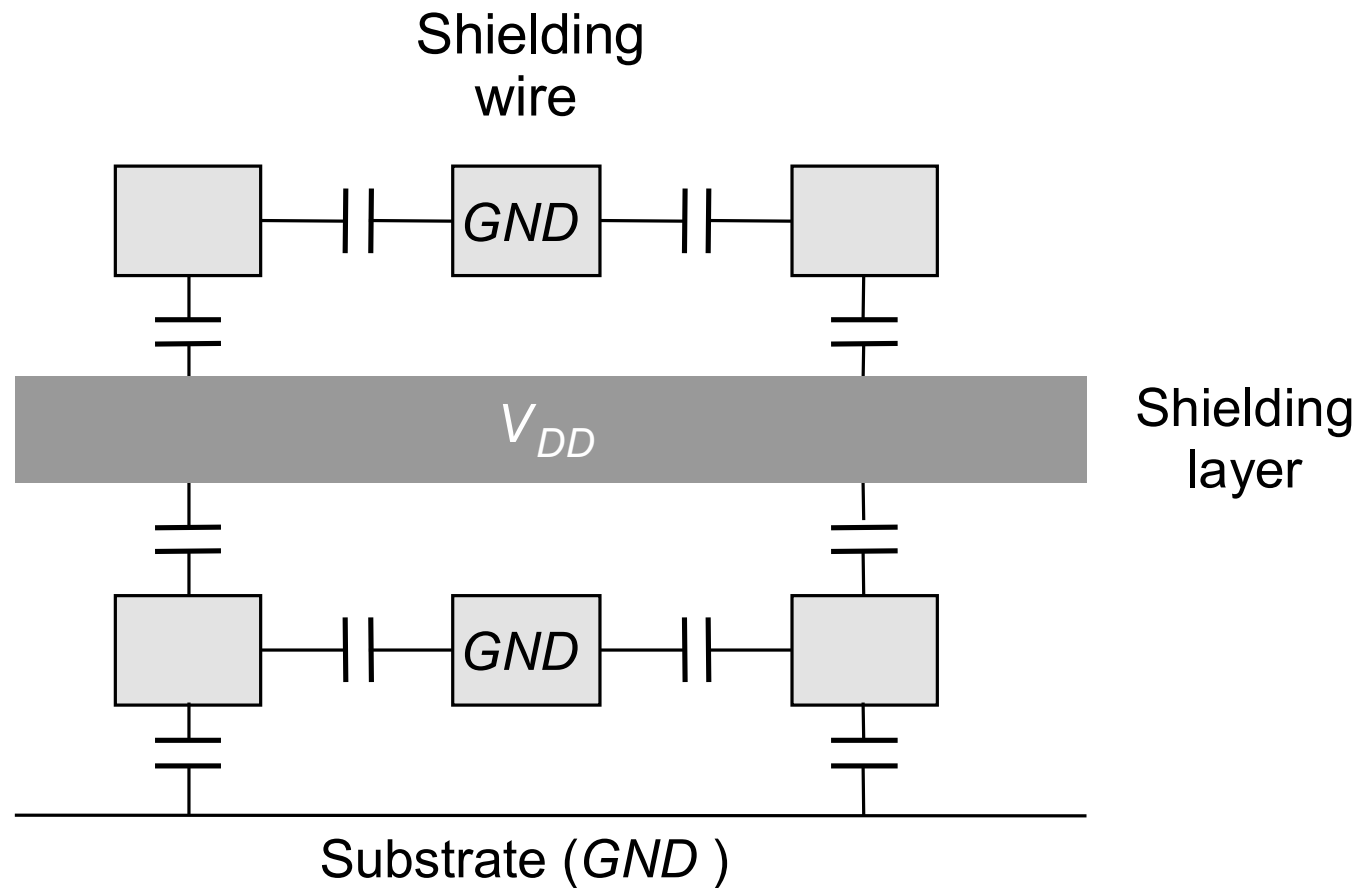


Keep time-constant smaller than rise time

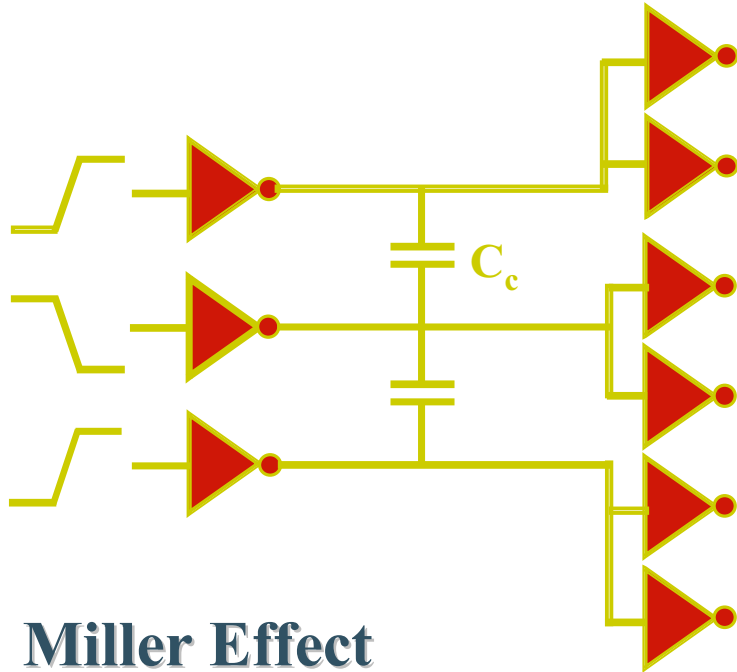
Dealing with Capacitive Cross Talk

- ❑ 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



Cross Talk and Performance



Miller Effect

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

DELAY DEPENDENT UPON
ACTIVITY IN NEIGHBORING
WIRES

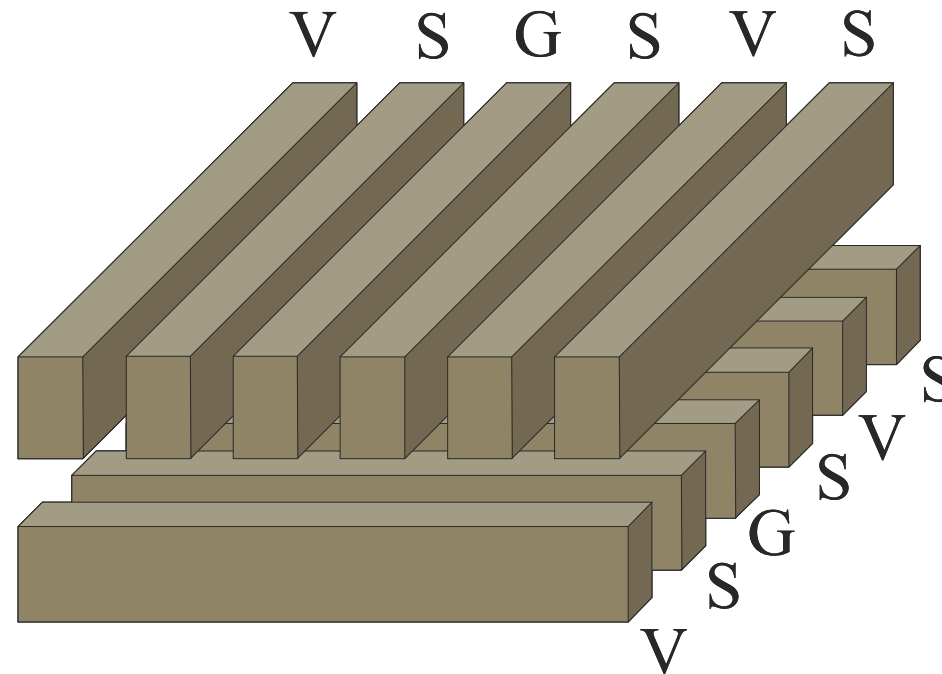
- Both terminals of capacitor are switched in opposite directions
($0 \rightarrow V_{dd}$, $V_{dd} \rightarrow 0$)
- Effective voltage is doubled and additional charge is needed
(from $Q=CV$)

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:

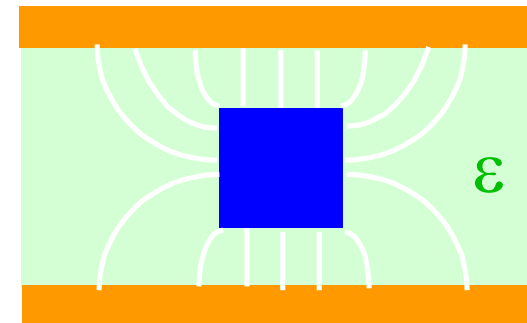
- Cross-coupling capacitance 40x lower, 2% delay variation
- Increase in area and overall capacitance

Also: FPGAs, VPGAs

Interconnect Projections

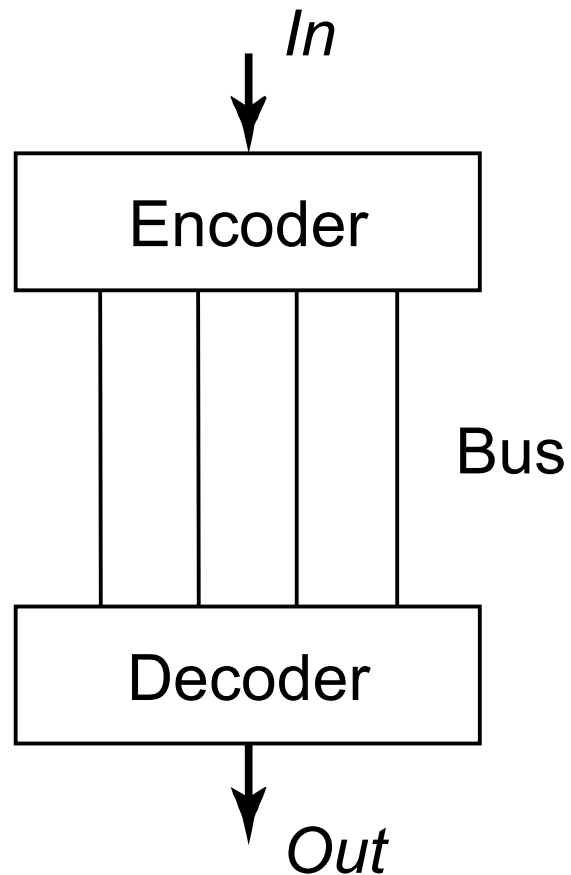
Low-k dielectrics

- 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

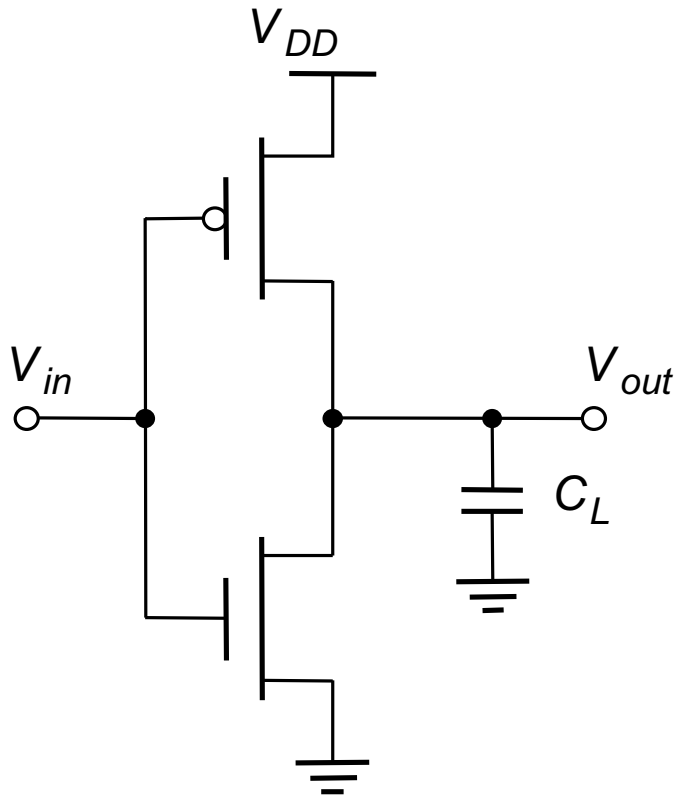


Generation	0.25 μm	0.18 μm	0.13 μm	0.1 μm	0.07 μm	0.05 μm
Dielectric Constant	3.3	2.7	2.3	2.0	1.8	1.5

Encoding Data Avoids Worst-Case Conditions



Driving Large Capacitances

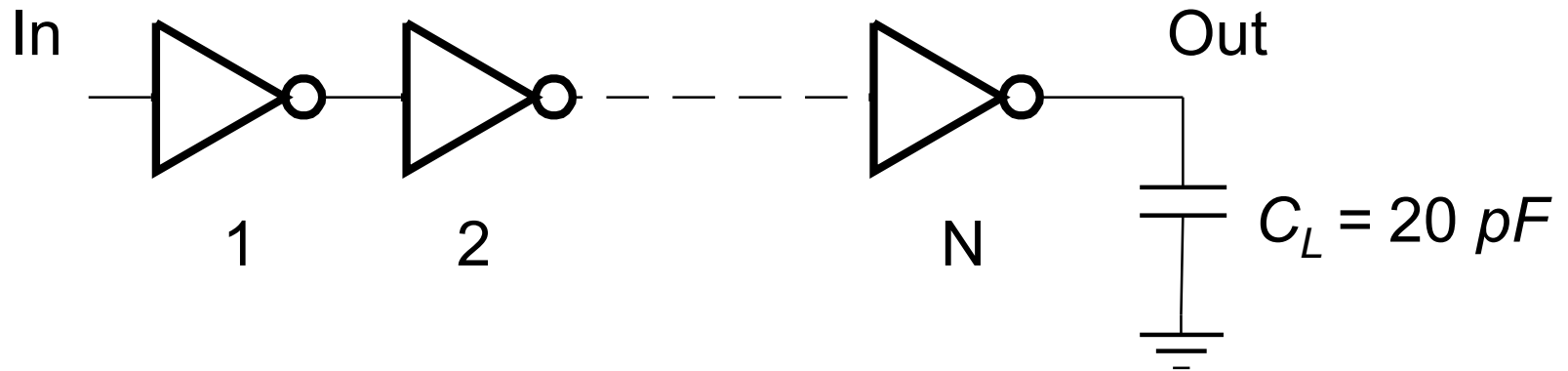


$$t_p = \frac{C_L V_{swing}}{I_{av}}$$

- Transistor Sizing
- Cascaded Buffers



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)

Output Driver Design

Trade off Performance for Area and Energy

Given t_{pmax} find N and f

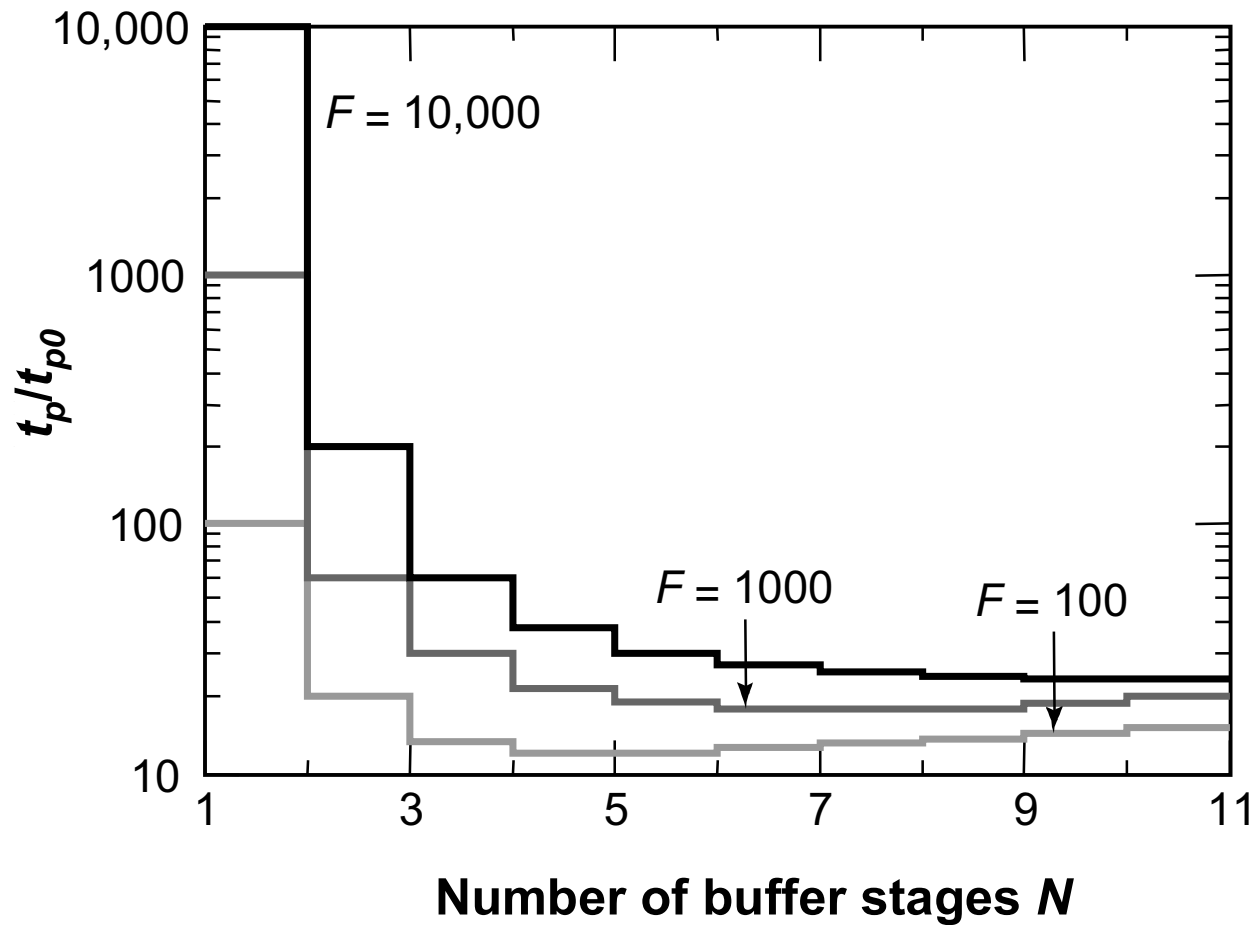
□ Area

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

□ 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, $C_L = 20 \text{ pF}$

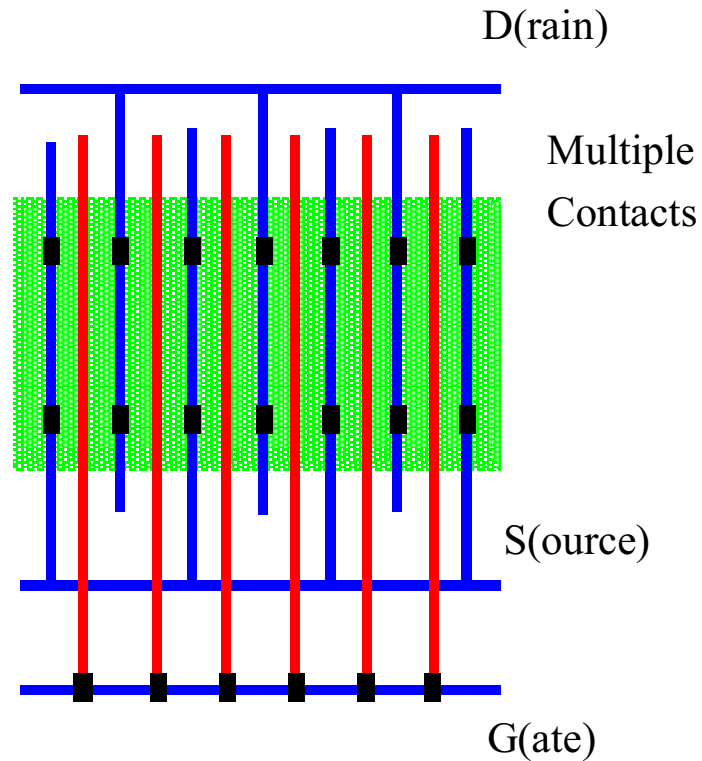
Transistor Sizes for optimally-sized cascaded buffer $t_p = 0.76 \text{ ns}$

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

How to Design Large Transistors

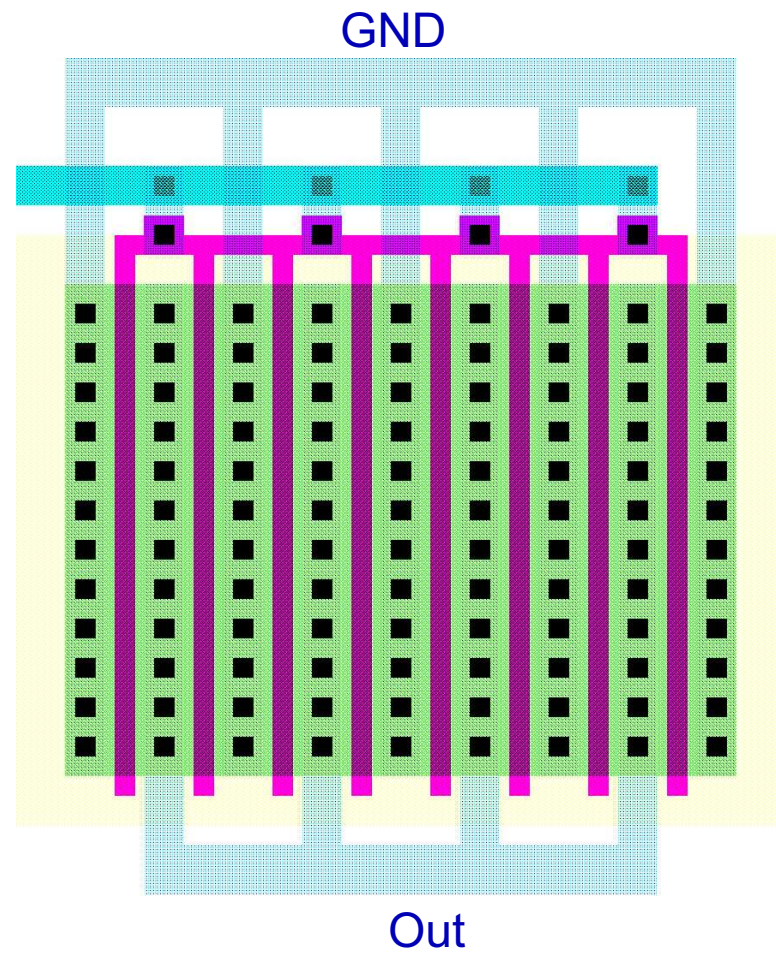
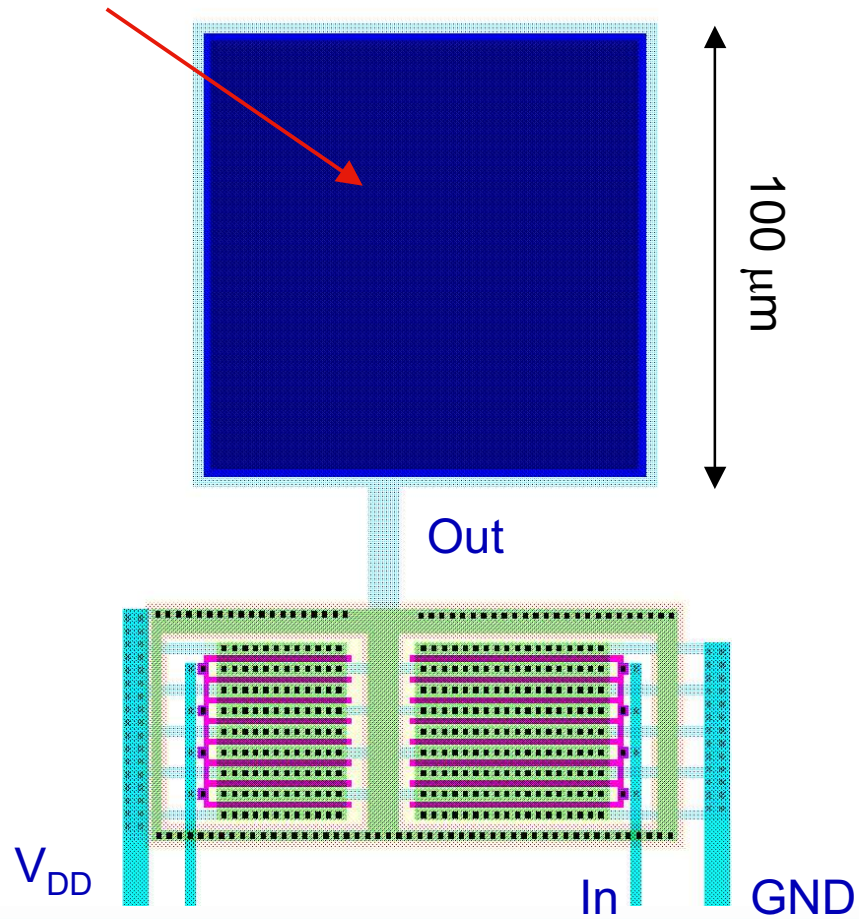


Reduces diffusion capacitance
Reduces gate resistance

small transistors in parallel

Bonding Pad Design

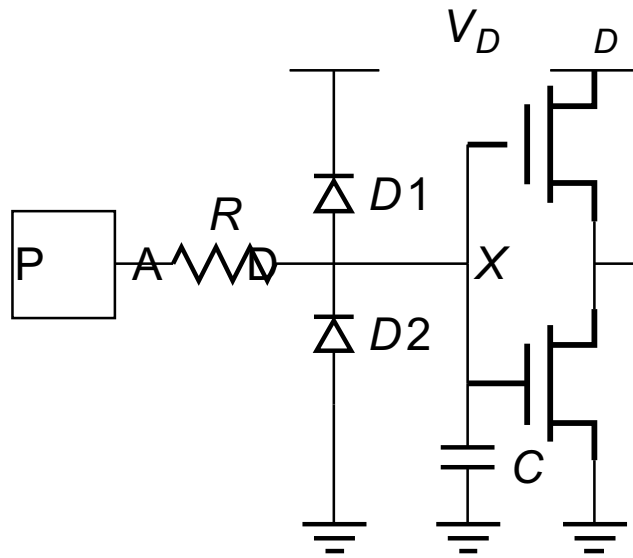
Bonding Pad



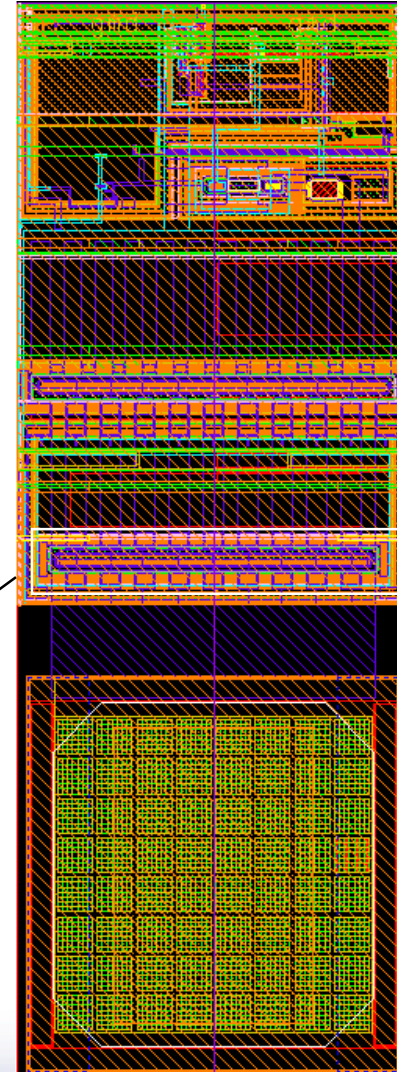
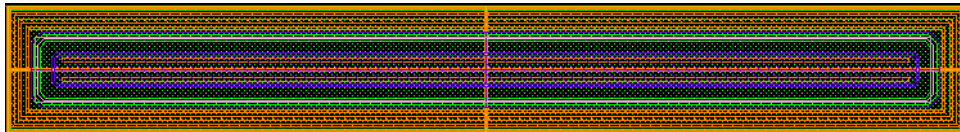
ESD Protection

- ❑ 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.

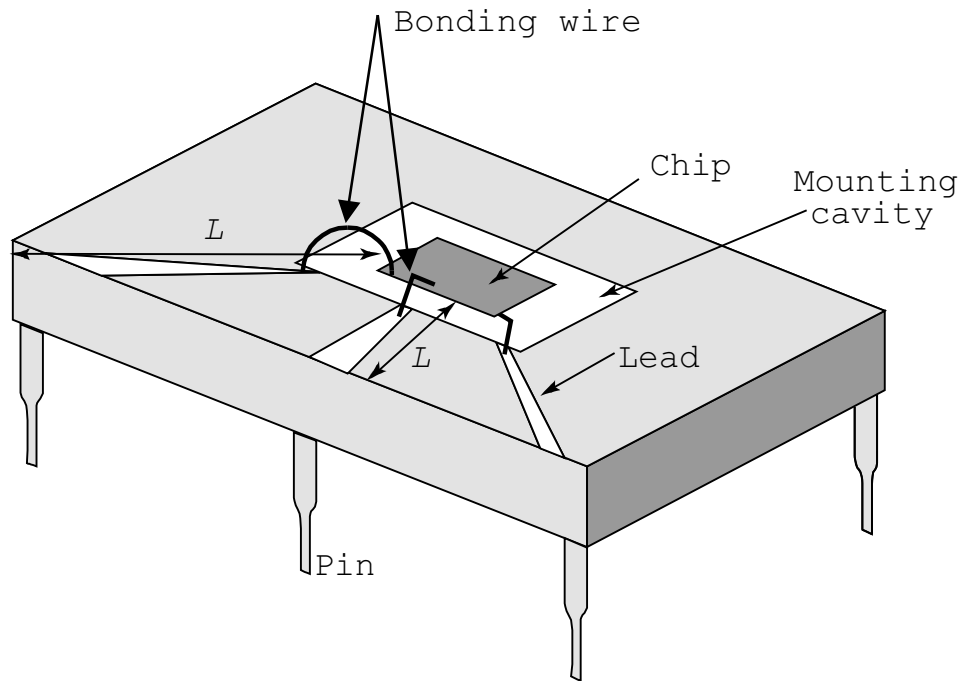
ESD Protection



Diode



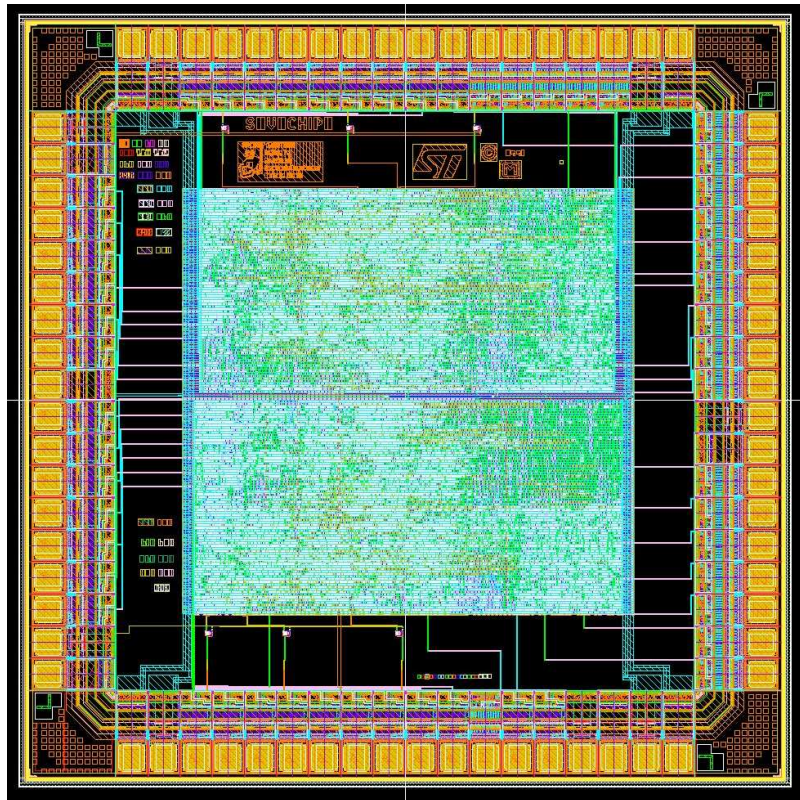
Chip Packaging



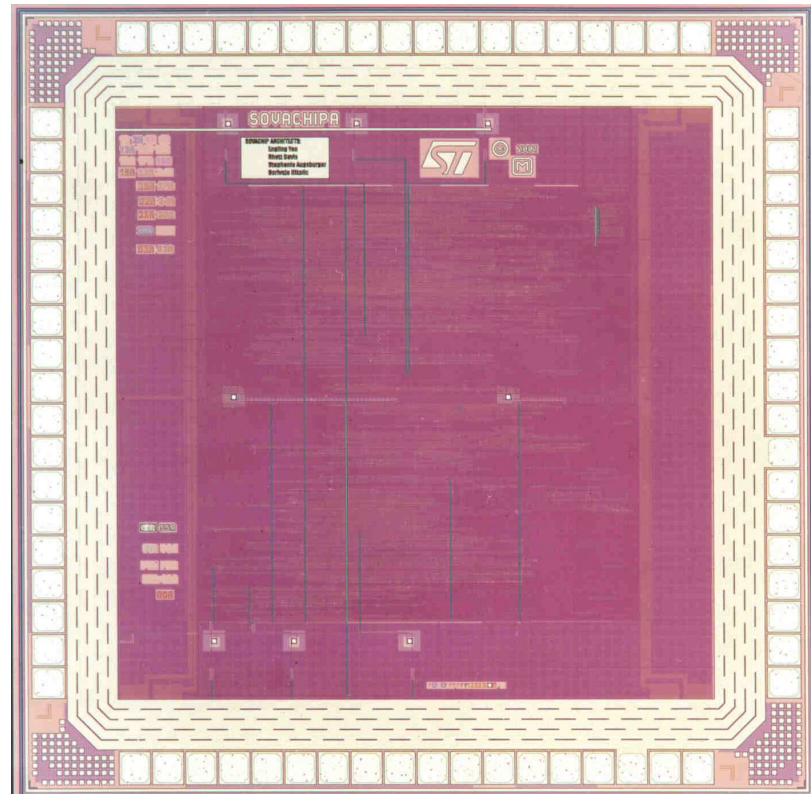
- Bond wires ($\sim 25\mu\text{m}$) are used to connect the package to the chip
- Pads are arranged in a frame around the chip
- Pads are relatively large ($\sim 100\mu\text{m}$ in $0.25\mu\text{m}$ technology), with large pitch ($100\mu\text{m}$)
- Many chips areas are 'pad limited'

Pad Frame

Layout



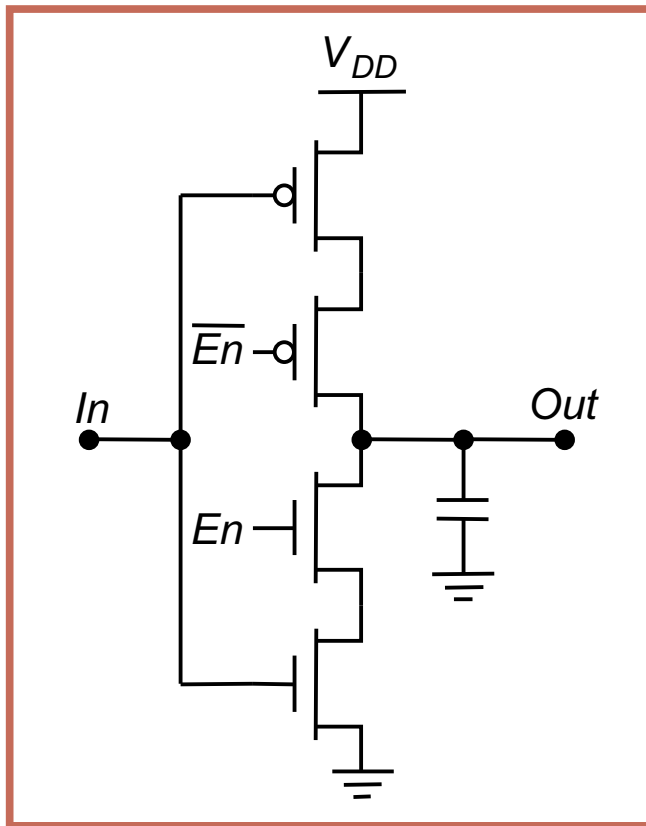
Die Photo



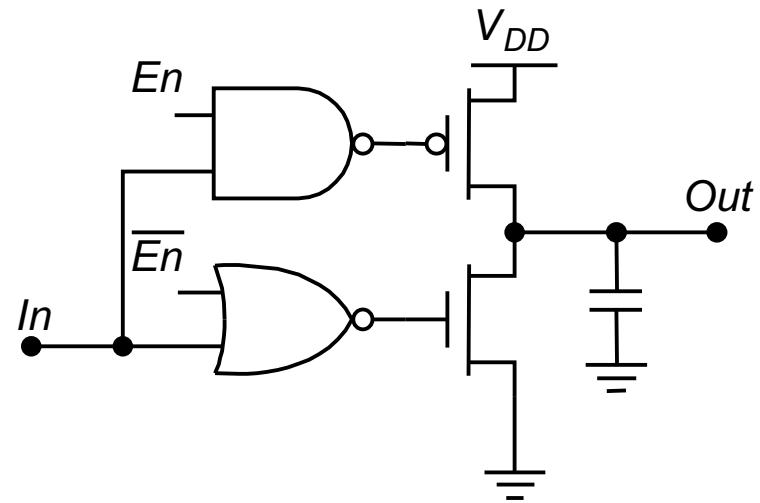
Chip Packaging

- 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

Tristate Buffers



$$Out = In \cdot En + Z \cdot \overline{En}$$



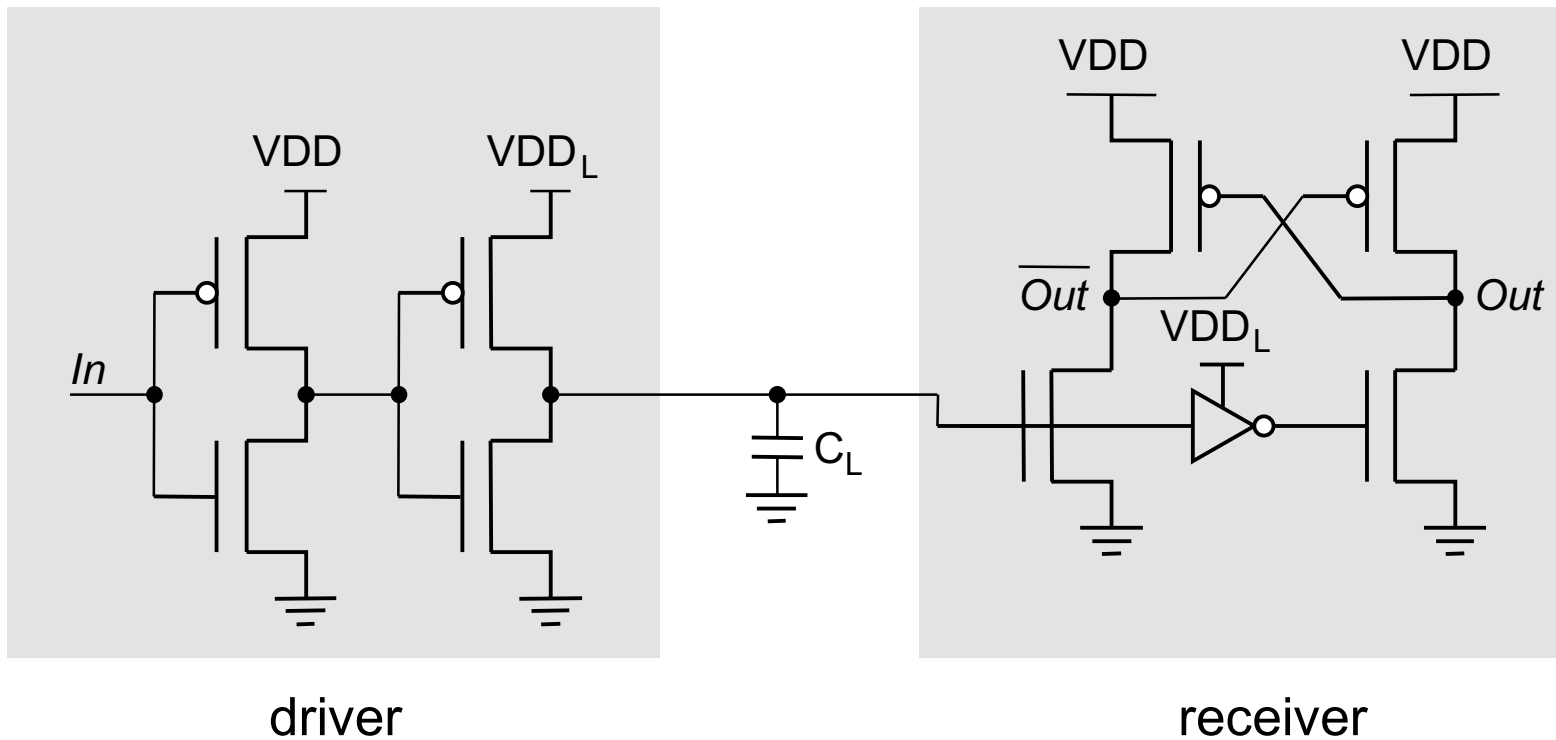
Increased output drive

Reducing the swing

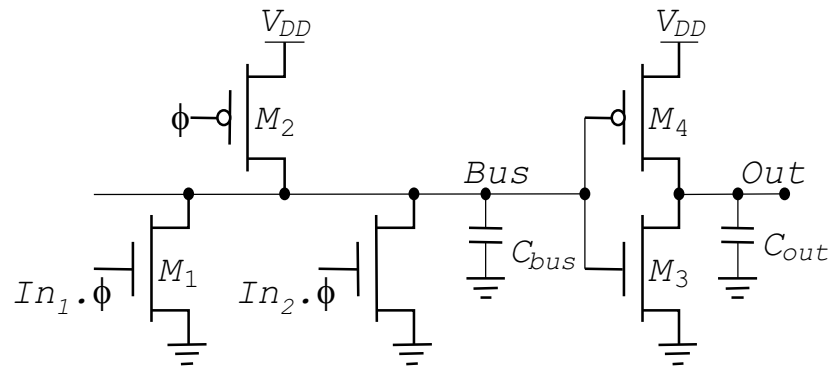
$$t_{pHL} = \frac{C_L V_{swing}/2}{I_{av}}$$

- ❑ 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)

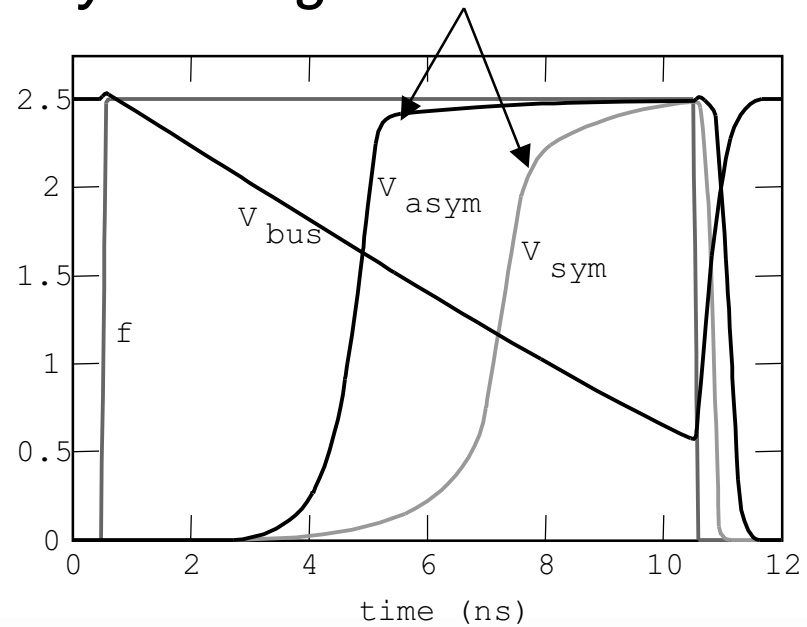
Single-Ended Static Driver and Receiver



Dynamic Reduced Swing Network



*Improve Performance
By shifting invert threshold*



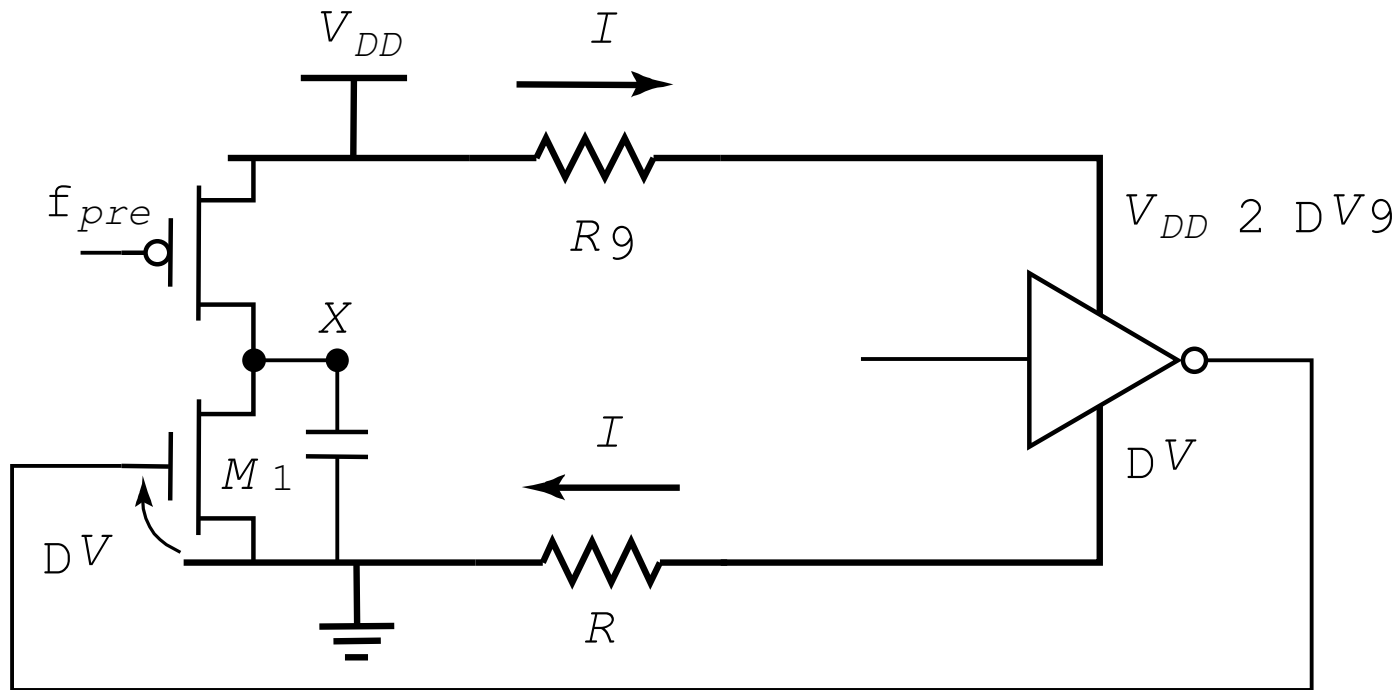
INTERCONNECT

Dealing with Resistance

Impact of Resistance

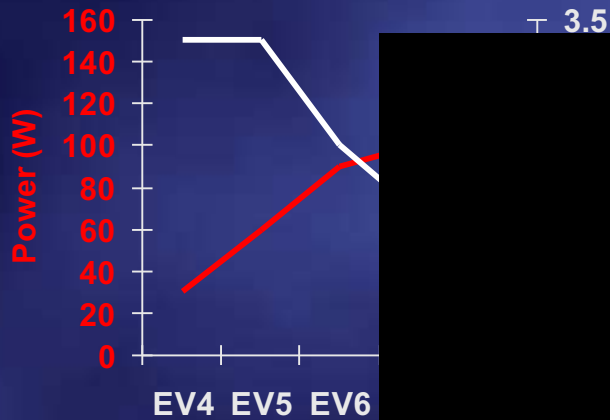
- ❑ We have already learned how to drive RC interconnect
- ❑ Impact of resistance is commonly seen in power supply distribution:
 - IR drop
 - Voltage variations
- ❑ Power supply is distributed to minimize the IR drop and the change in current due to switching of gates

RI Introduced Noise

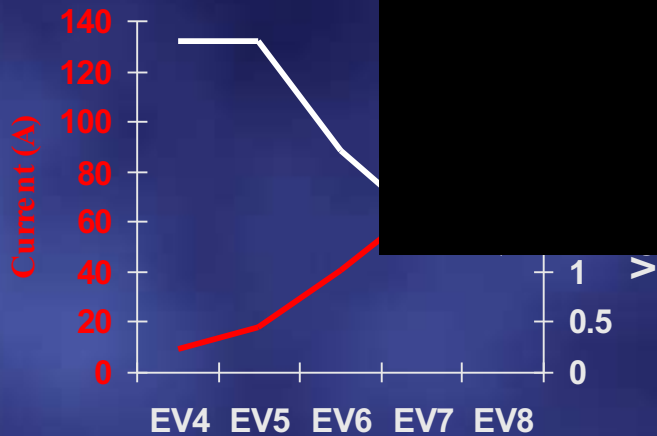


Power Dissipation Trends

Power Dissipation



Supply C



- ◆ Power consumption is increasing

Technology needed

is increasing faster!

Power will be a major

concern. Distribution are critical

to power growth

g

inner (Cu)

s

- Clock system design

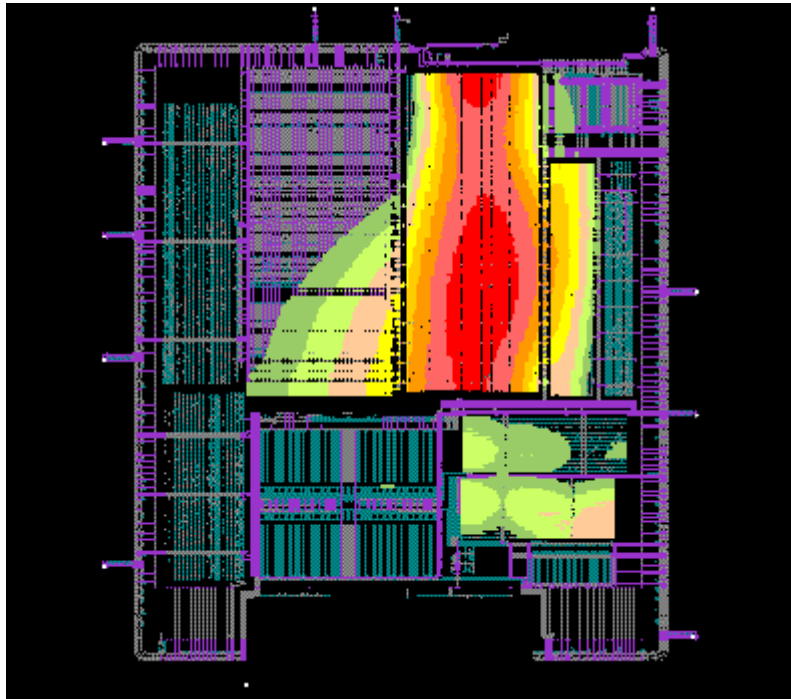
- micro-architecture

COMPAQ

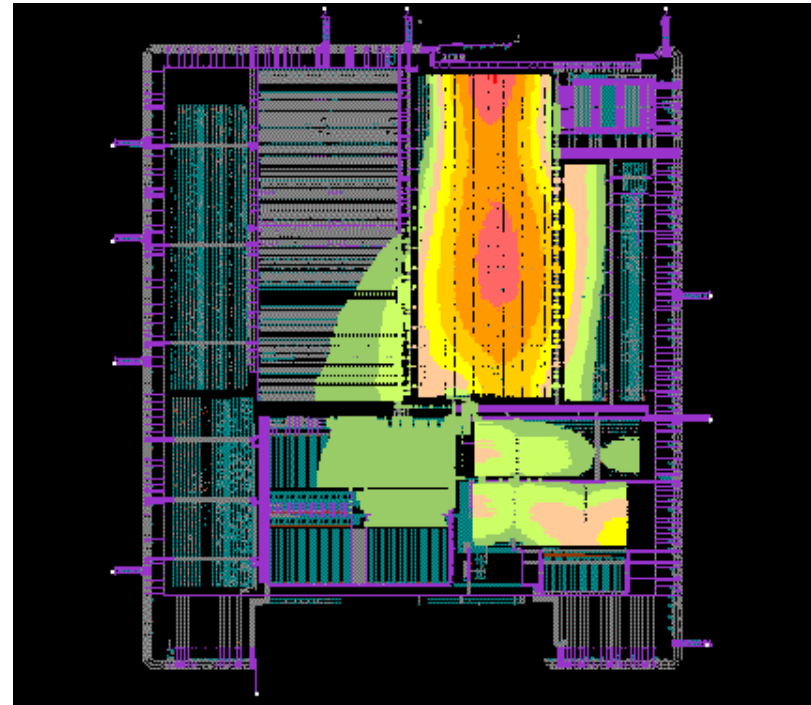
ASP DAC 2000

Resistance and the Power Distribution Problem

Before



After

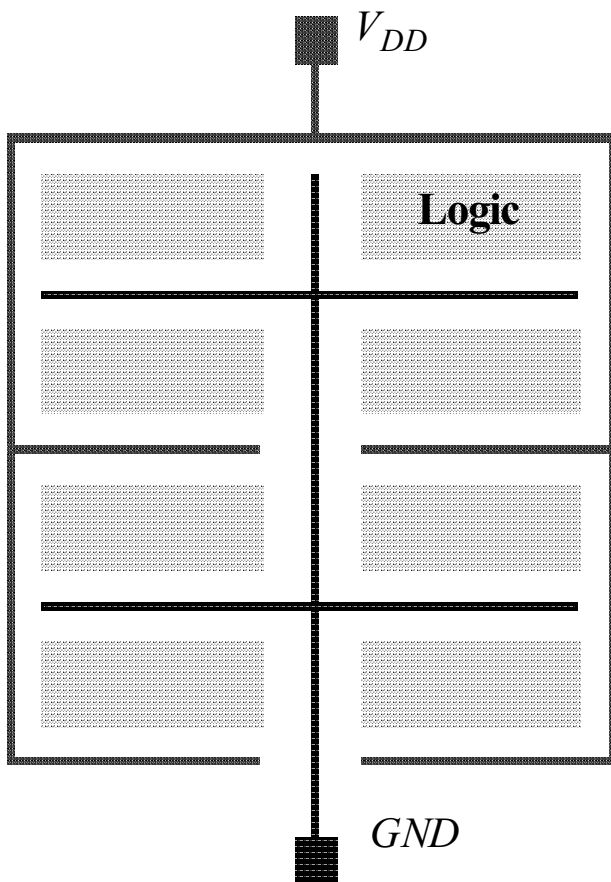


- **Requires fast and accurate peak current prediction**
- **Heavily influenced by packaging technology**

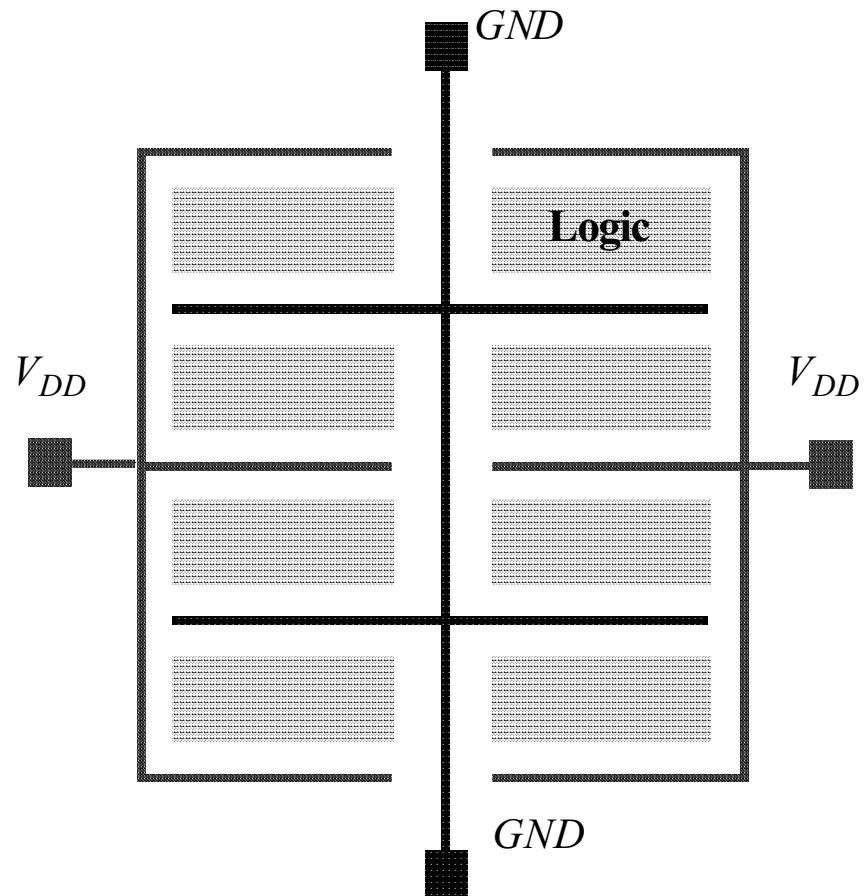
Power Distribution

- ❑ 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

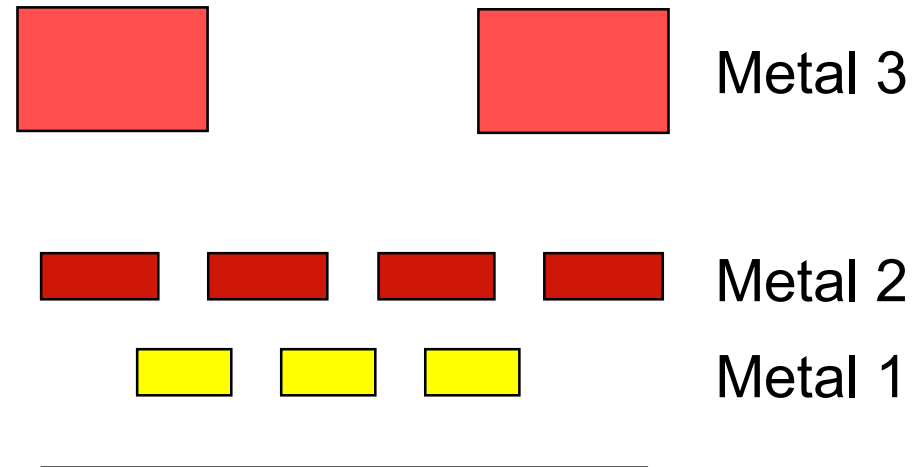
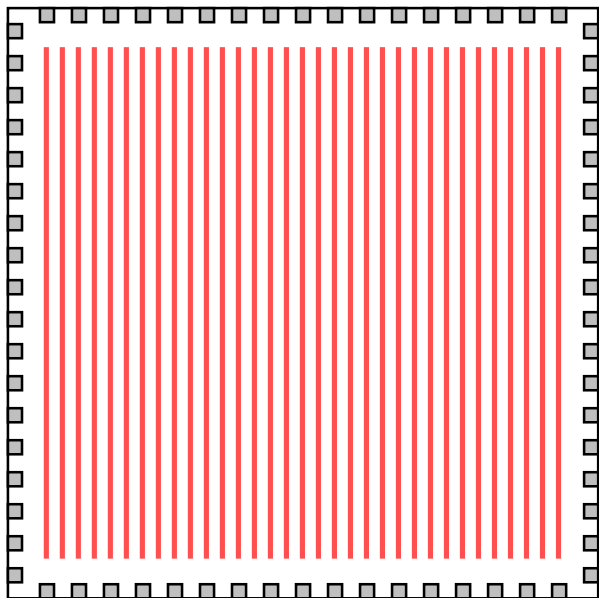
3 Metal Layer Approach (EV4)

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



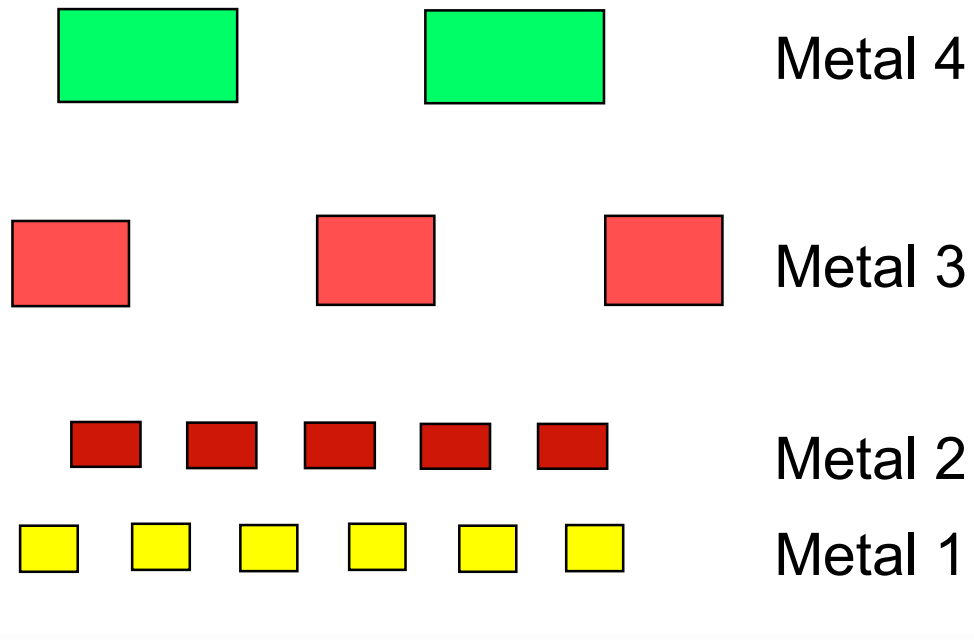
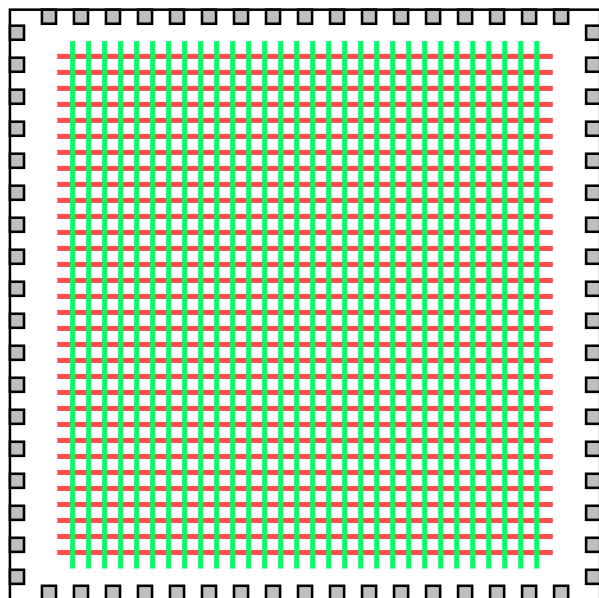
4 Metal Layers Approach (EV5)

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



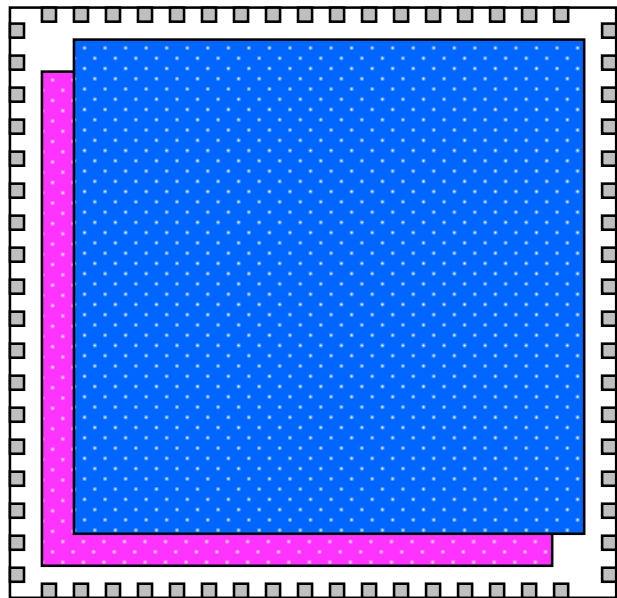
6 Metal Layer Approach – EV6

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



RP2/Vdd



Metal 4



Metal 3



RP1/Vss

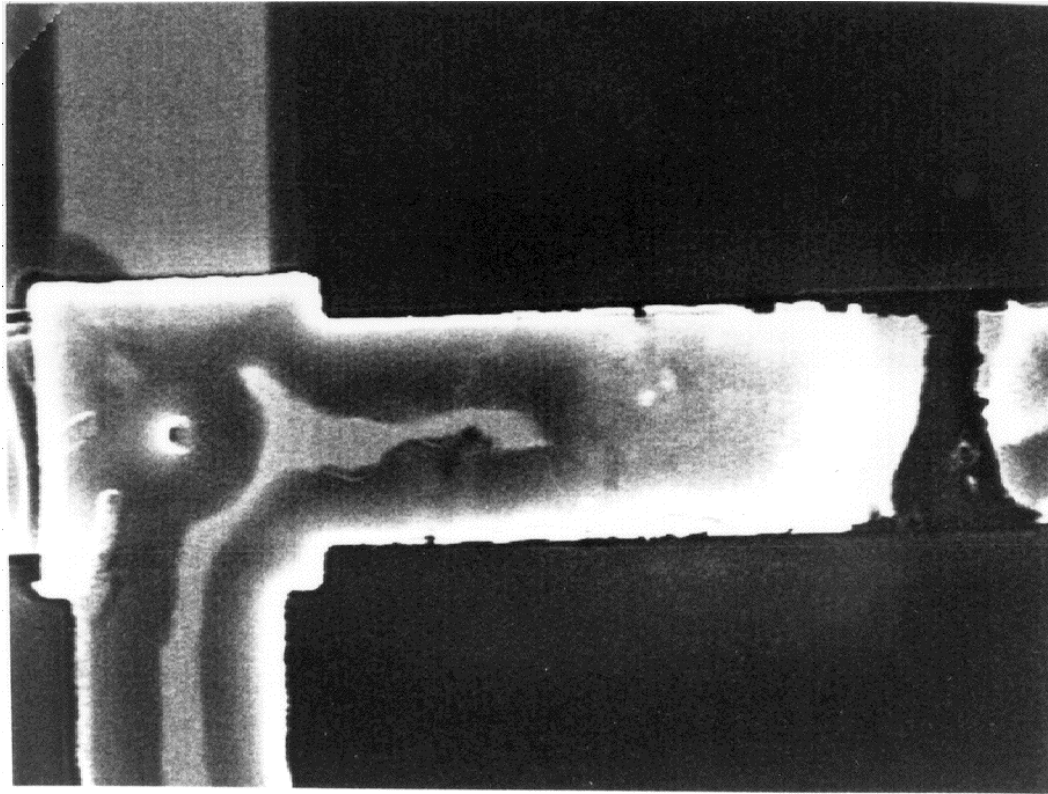


Metal 2



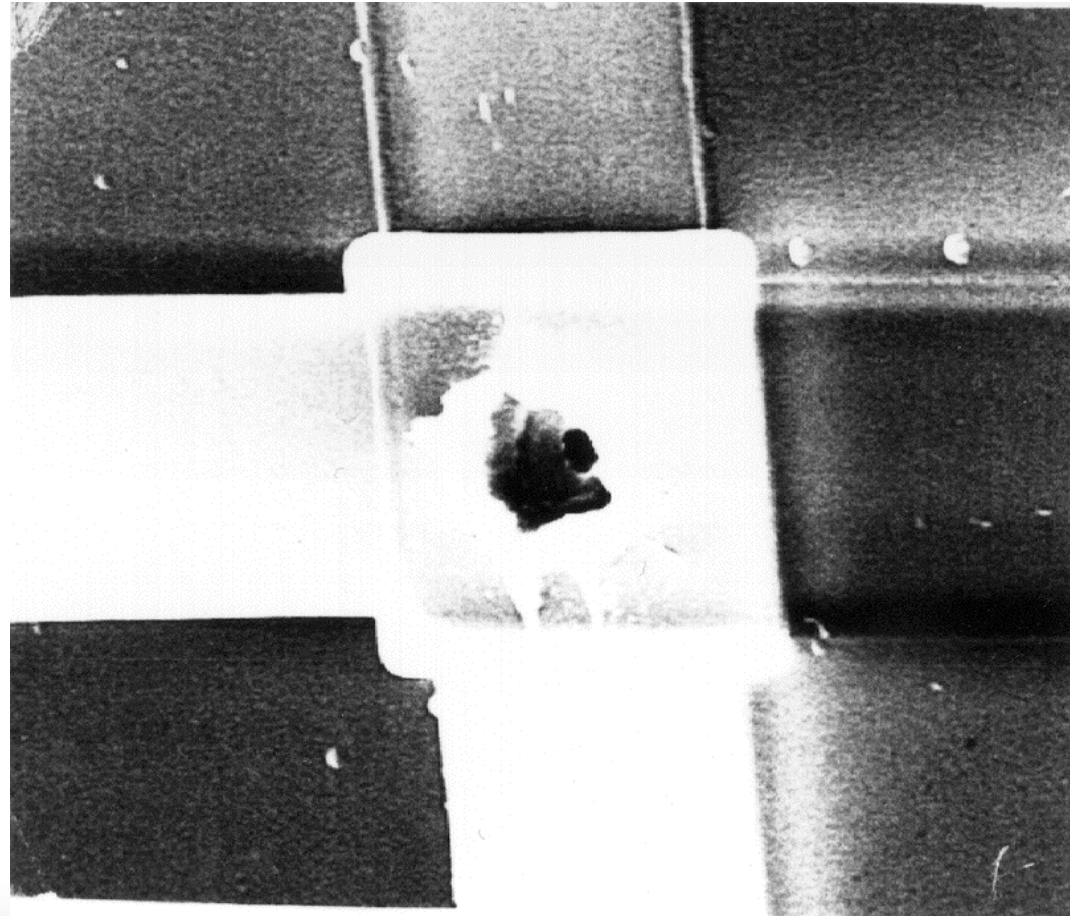
Metal 1

Electromigration (1)

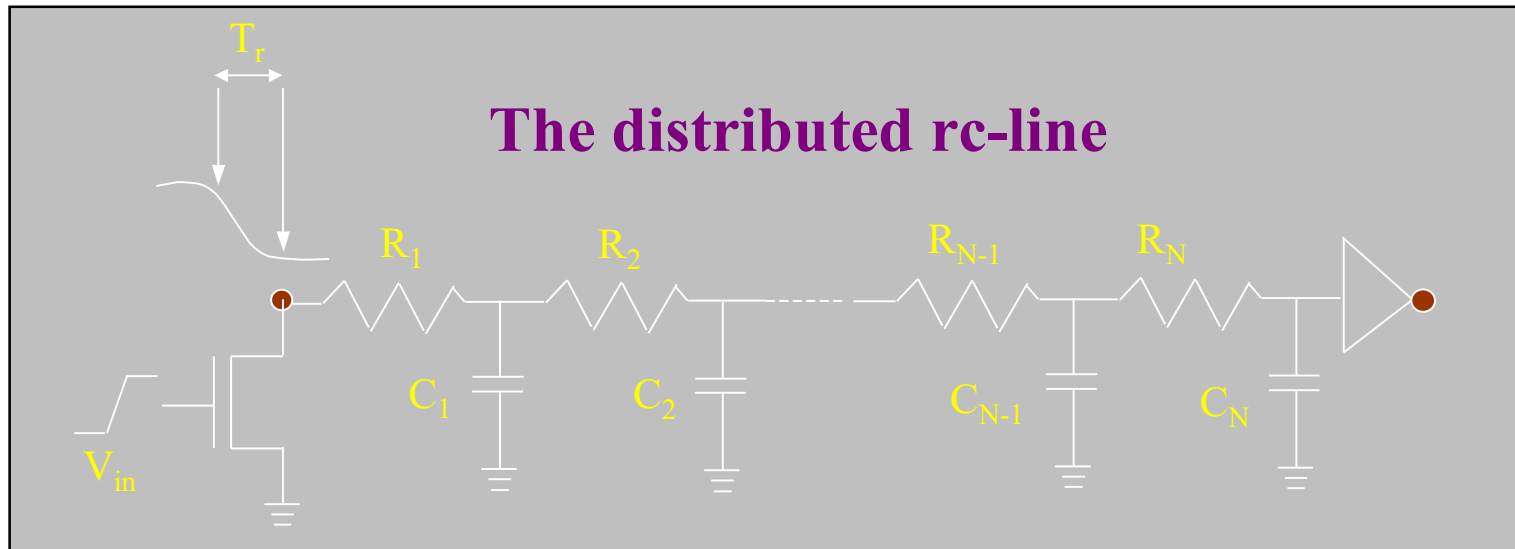


Limits dc-current to $1 \text{ mA}/\mu\text{m}$

Electromigration (2)

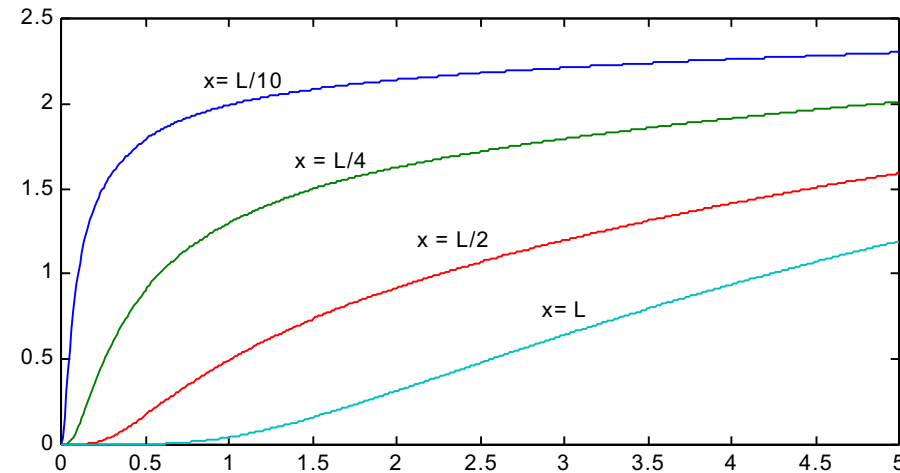


Resistivity and Performance



Diffused signal propagation

$$\text{Delay} \sim L^2$$



The Global Wire Problem

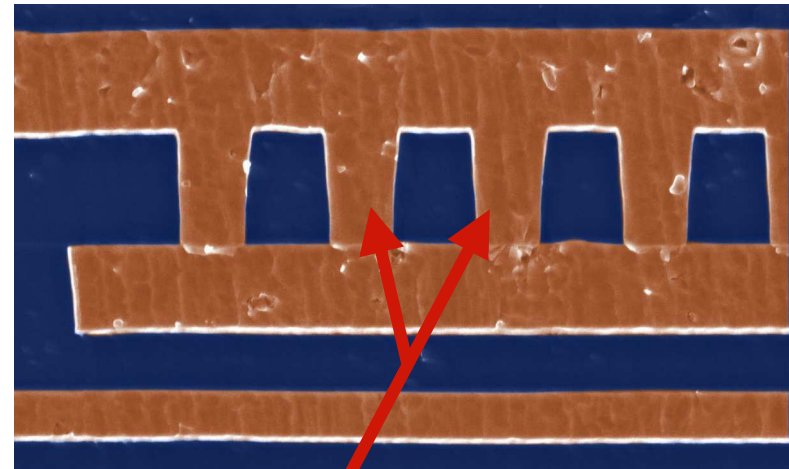
$$T_d = 0.377R_wC_w + 0.693(R_dC_{out} + R_dC_w + R_wC_{out})$$

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?

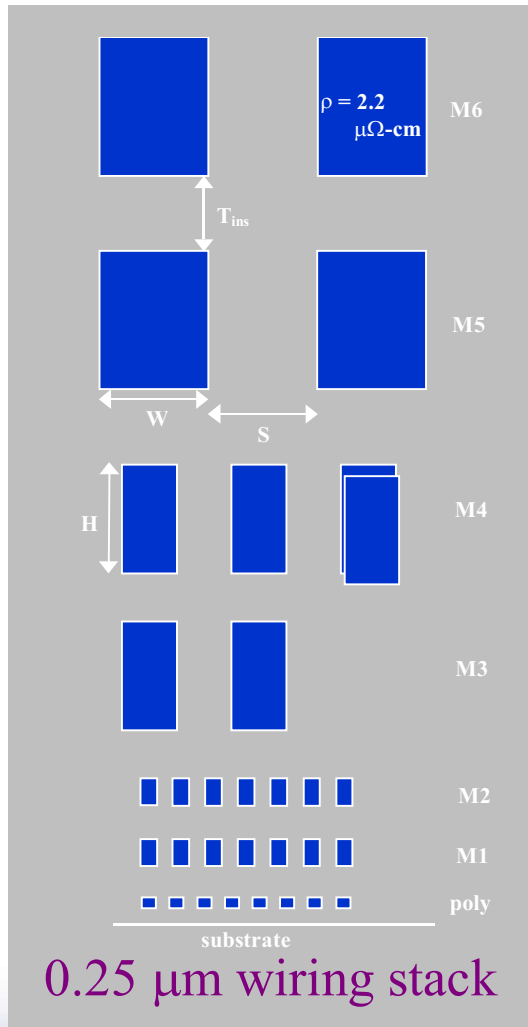
Interconnect Projections: Copper

- ❑ Copper is planned in full sub-0.25 μm process flows and large-scale designs (IBM, Motorola, IEDM97)
- ❑ With cladding and other effects, Cu $\sim 2.2 \mu\Omega\text{-cm}$ vs. 3.5 for Al(Cu) \Rightarrow 40% reduction in resistance
- ❑ Electromigration improvement; 100X longer lifetime (IBM, IEDM97)
 - Electromigration is a limiting factor beyond 0.18 μm if Al is used (HP, IEDM95)



Vias

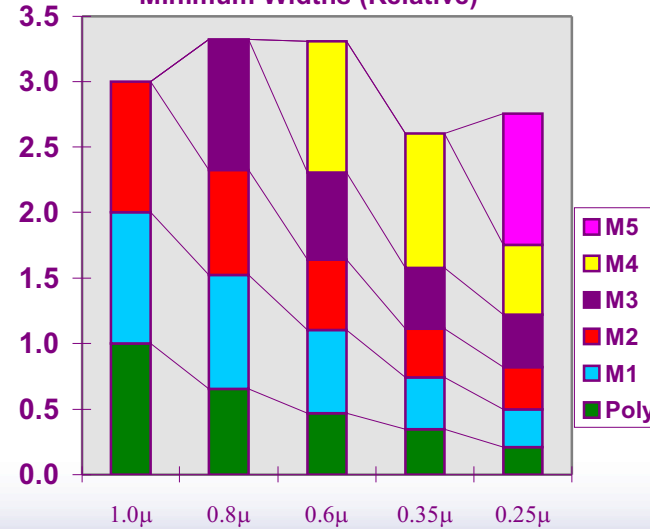
Interconnect: # of Wiring Layers



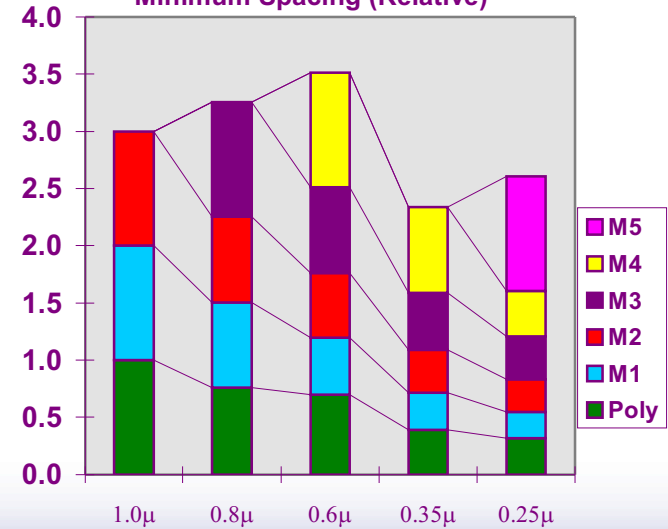
of metal layers is steadily increasing due to:

- Increasing die size and device count: we need more wires and longer wires to connect everything
- Rising need for a hierarchical wiring network; local wires with high density and global wires with low RC

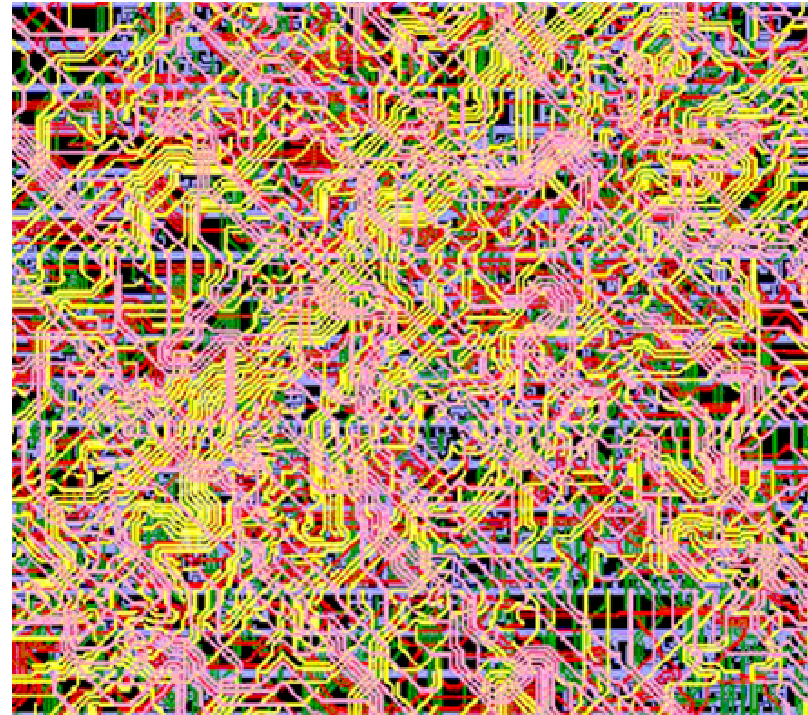
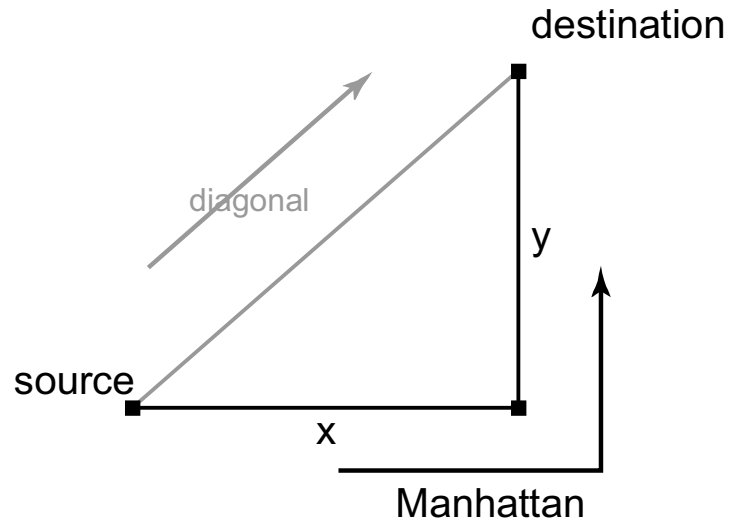
Minimum Widths (Relative)



Minimum Spacing (Relative)

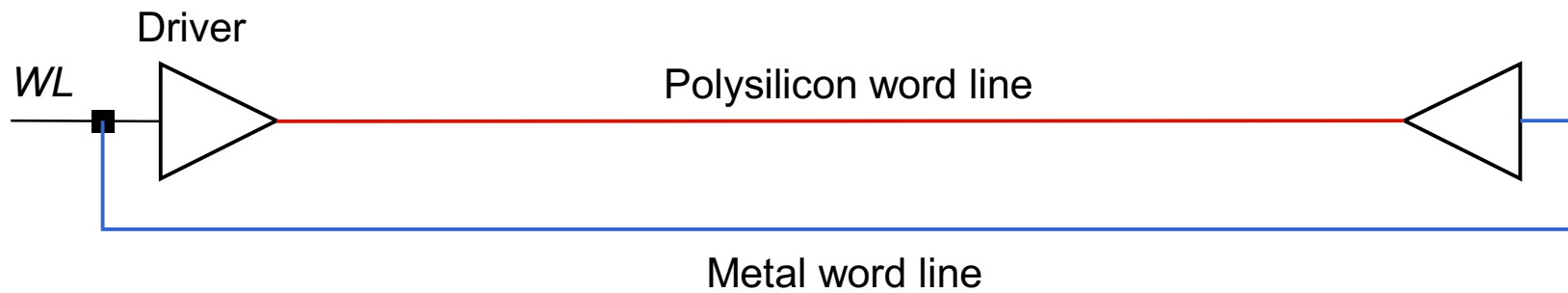


Diagonal Wiring

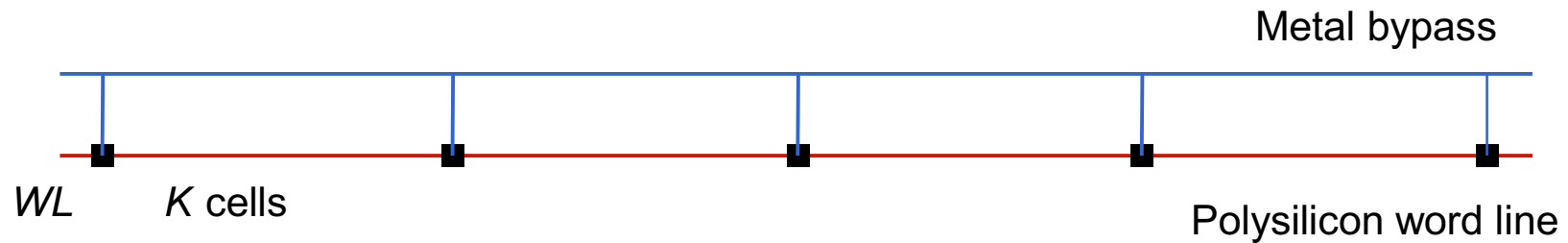


- 20+% Interconnect length reduction
- Clock speed
- Signal integrity
- Power integrity
- 15+% Smaller chips
plus 30+% via reduction

Using Bypasses

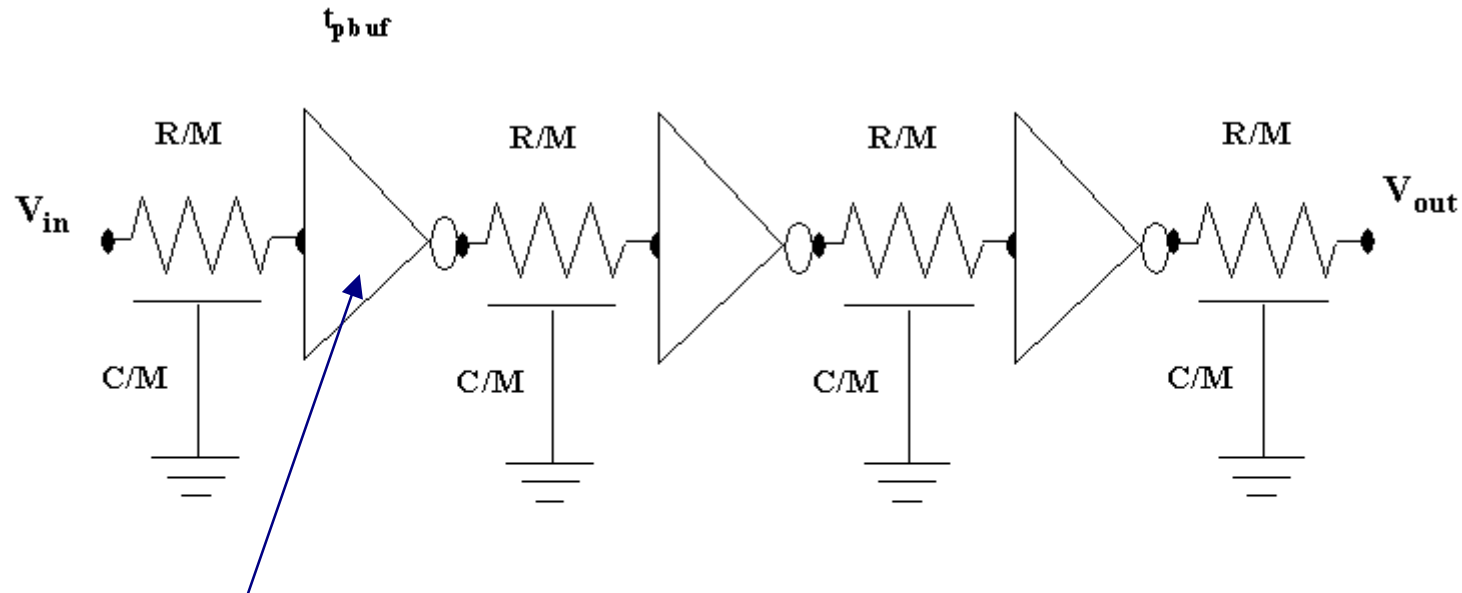


Driving a word line from both sides



Using a metal bypass

Reducing RC-delay



Repeater

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

Repeater Insertion (Revisited)

Taking the repeater loading into account

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

$$s_{opt} = \sqrt{\frac{R_dc}{rC_d}}$$

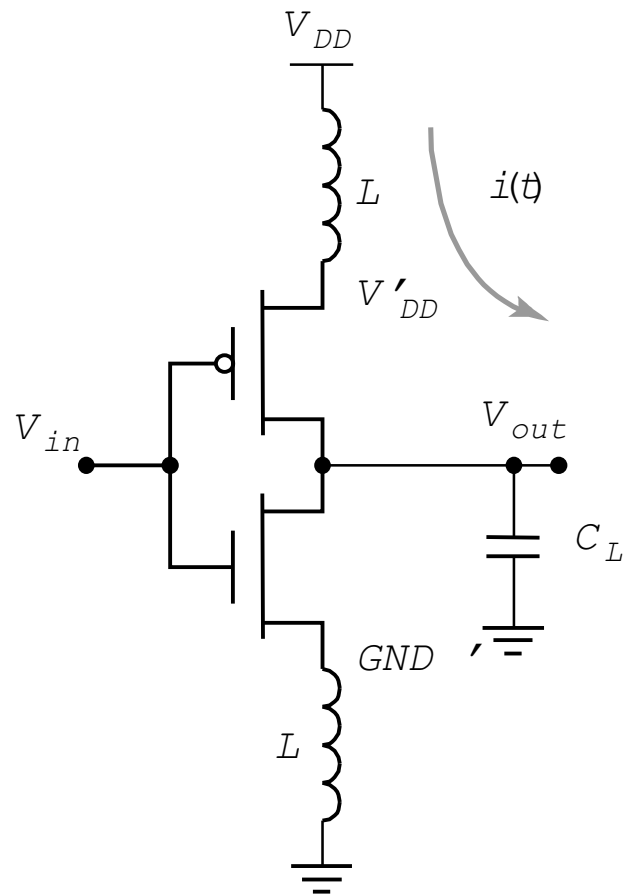
For a given technology and a given interconnect layer, there exists an optimal length of the wire segments between repeaters. The delay of these wire segments is **independent of the routing layer!**

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

INTERCONNECT

Dealing with Inductance

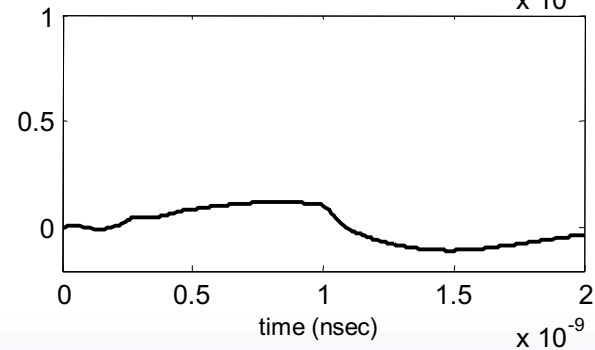
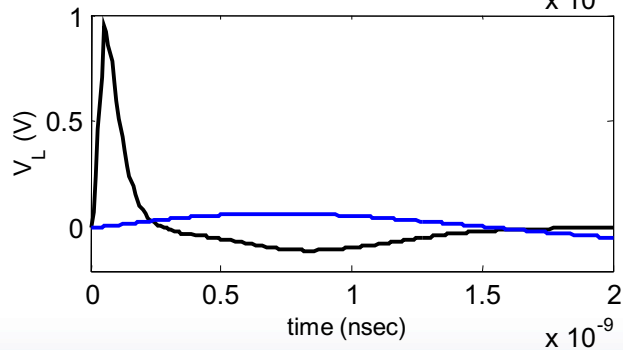
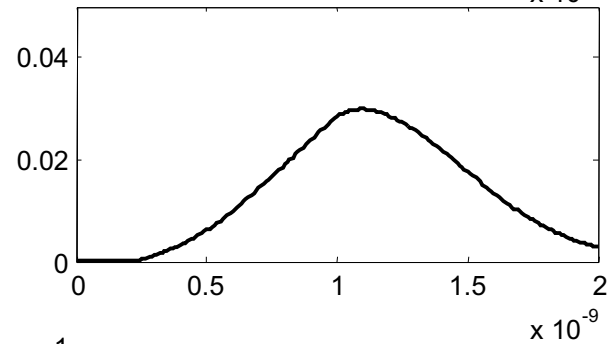
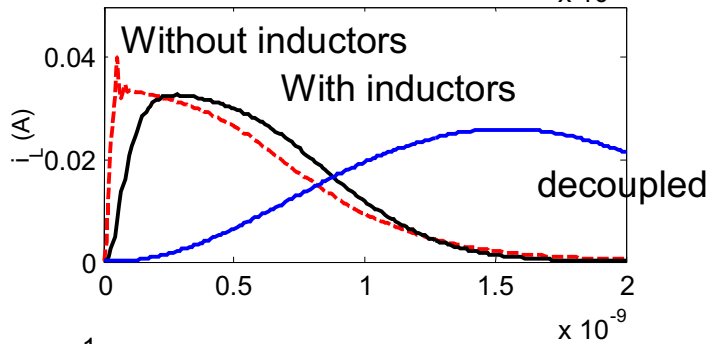
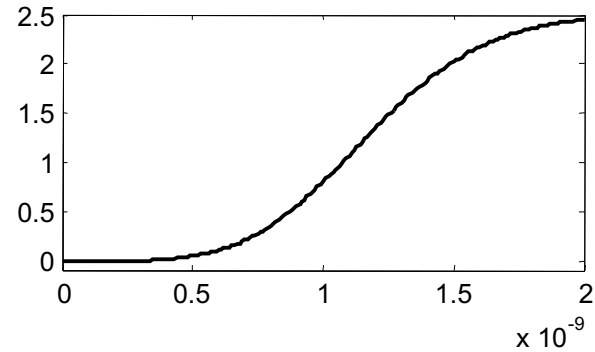
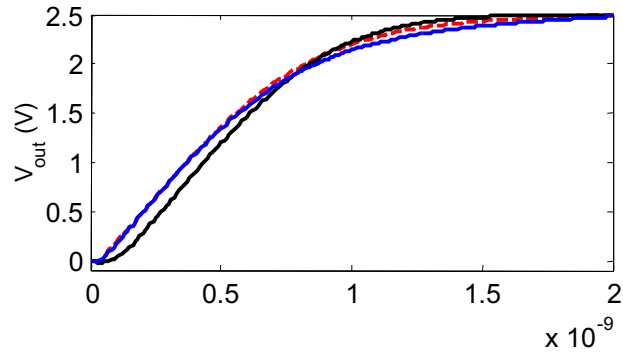
$L \frac{di}{dt}$



Impact of inductance on supply voltages:

- Change in current induces a change in voltage
- Longer supply lines have larger L

L di/dt: Simulation



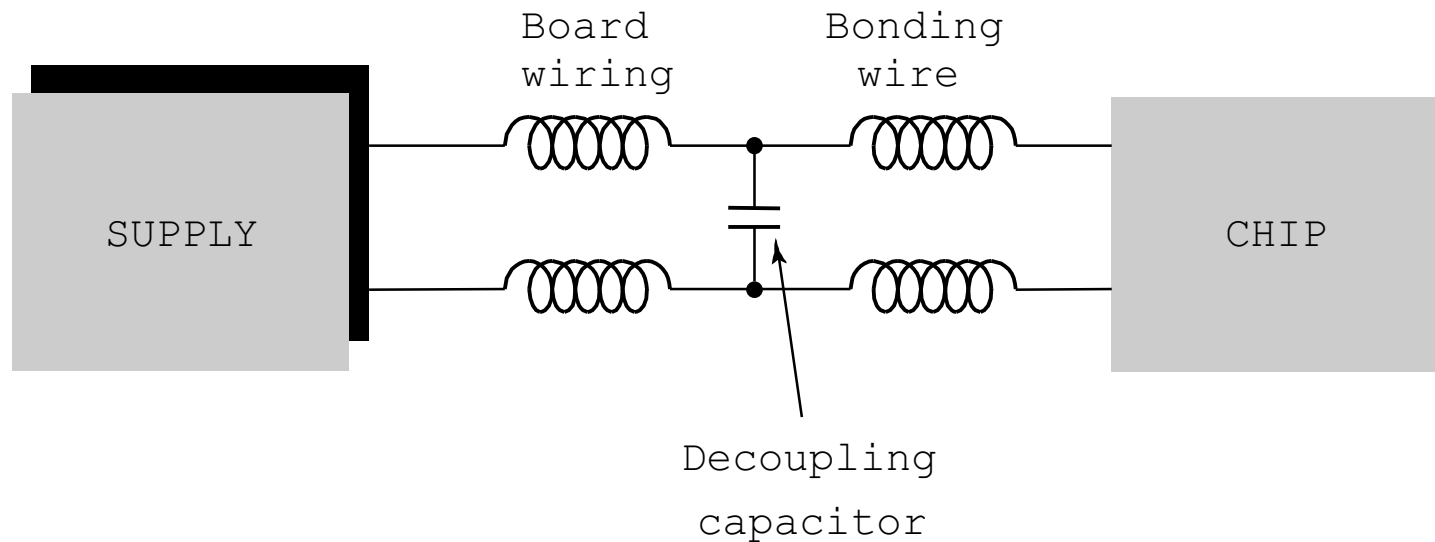
Input rise/fall time: 50 psec

Input rise/fall time: 800 psec

Dealing with Ldi/dt

- ❑ **Separate power pins for I/O pads and chip core.**
- ❑ **Multiple power and ground pins.**
- ❑ **Careful selection of the positions of the power and ground pins on the package.**
- ❑ **Increase the rise and fall times of the off-chip signals to the maximum extent allowable.**
- ❑ **Schedule current-consuming transitions.**
- ❑ **Use advanced packaging technologies.**
- ❑ **Add decoupling capacitances on the board.**
- ❑ **Add decoupling capacitances on the chip.**

Decoupling Capacitors



Decoupling capacitors are added:

- on the board (right under the supply pins)
- on the chip (under the supply straps, near large buffers)

De-coupling Capacitor Ratios

□ EV4

- total effective switching capacitance = 12.5nF
- 128nF of de-coupling capacitance
- de-coupling/switching capacitance ~ 10x

□ EV5

- 13.9nF of switching capacitance
- 160nF of de-coupling capacitance

□ EV6

- 34nF of effective switching capacitance
- 320nF of de-coupling capacitance -- not enough!

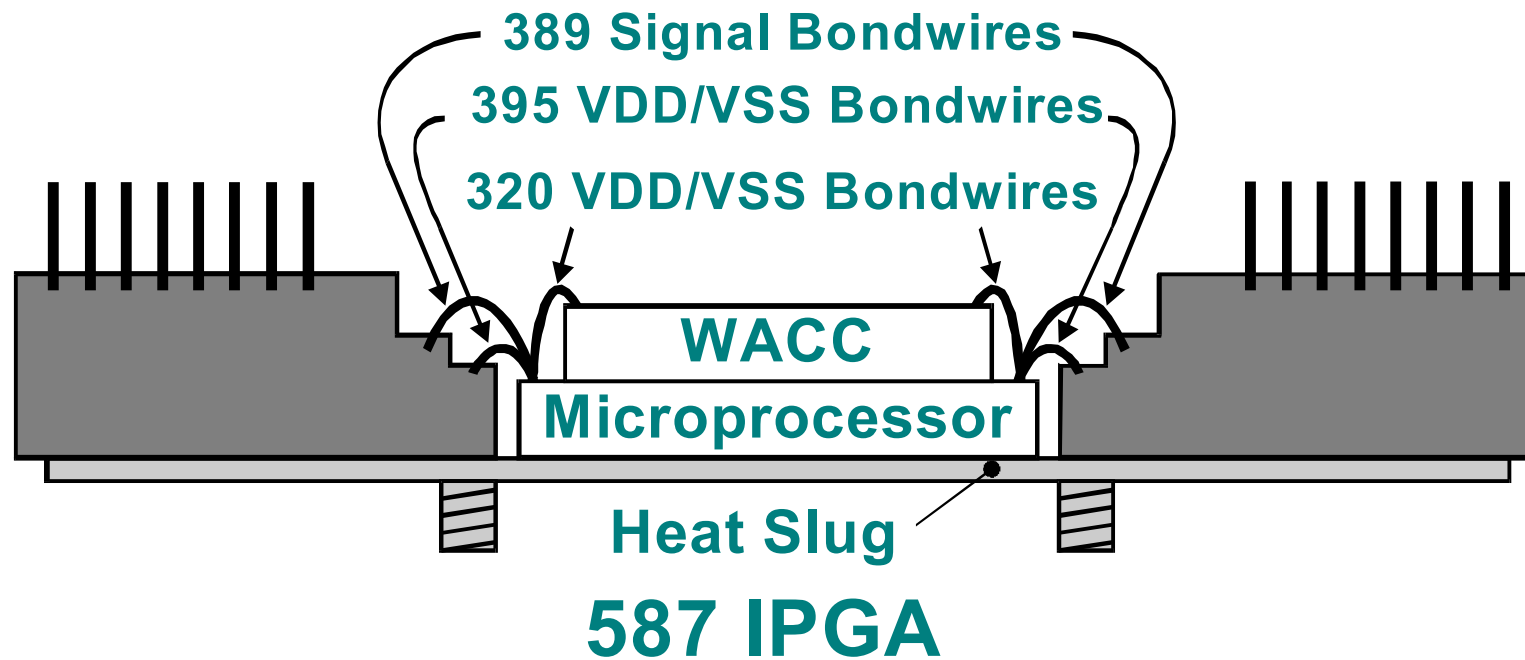
EV6 De-coupling Capacitance

Design for $\Delta I_{dd} = 25 \text{ A}$ @ $V_{dd} = 2.2 \text{ V}$, $f = 600 \text{ MHz}$

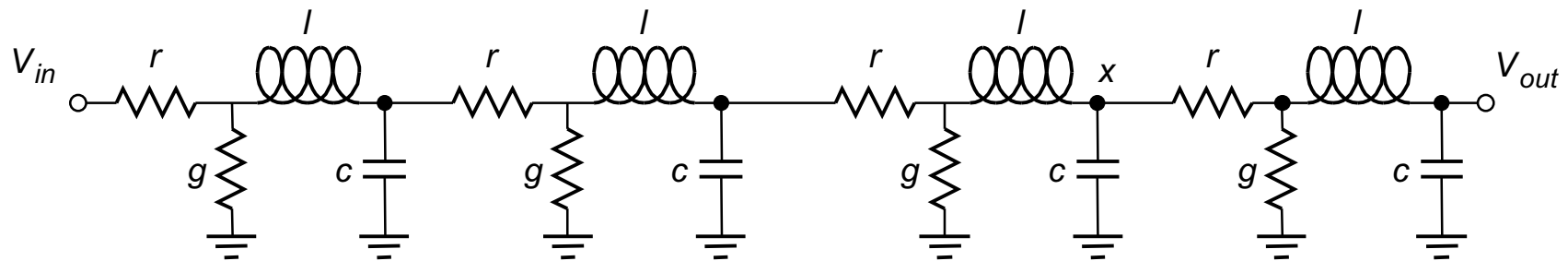
- 0.32- μF of on-chip de-coupling capacitance was added
 - Under major busses and around major gridded clock drivers
 - Occupies 15-20% of die area
- 1- μF 2- cm^2 Wirebond Attached Chip Capacitor (WACC) significantly increases “Near-Chip” de-coupling
 - 160 V_{dd}/V_{ss} bondwire pairs on the WACC minimize inductance

EV6 WACC

389 Signal - 198 VDD/VSS Pins



The Transmission Line



$$\frac{\partial^2 v}{\partial x^2} = rc \frac{\partial v}{\partial t} + lc \frac{\partial^2 v}{\partial t^2}$$

The Wave Equation

Design Rules of Thumb

- Transmission line effects should be considered when the rise or fall time of the input signal (t_r , t_f) is smaller than the time-of-flight of the transmission line (t_{flight}).

$$t_r (t_f) \ll 2.5 t_{flight}$$

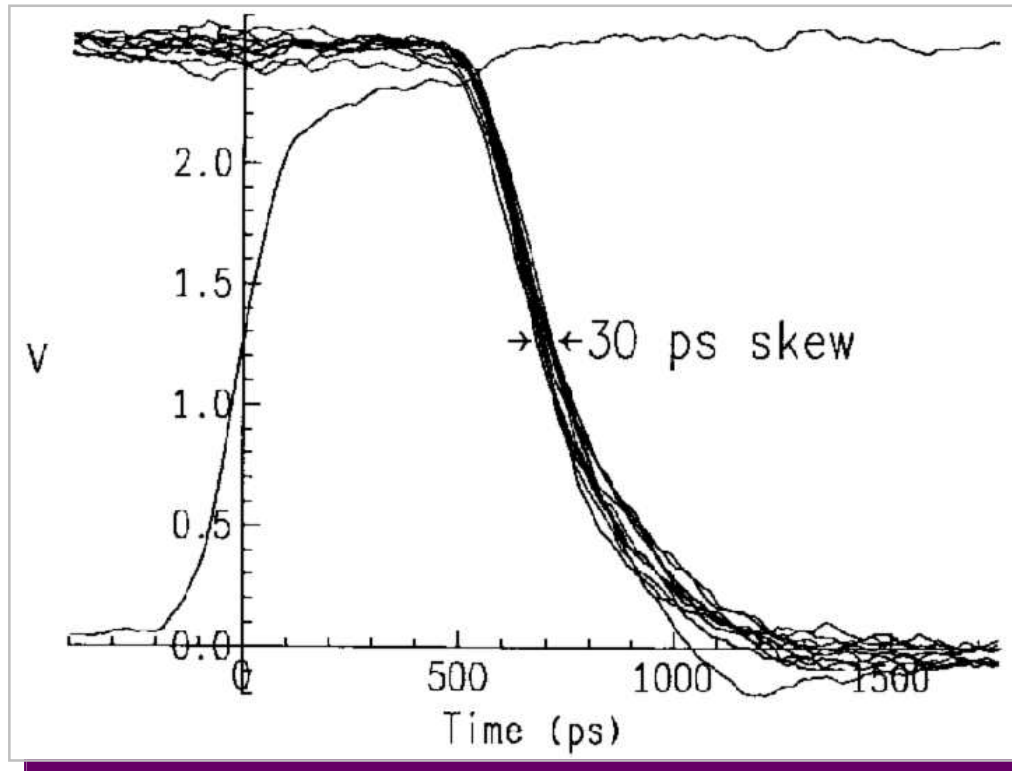
- Transmission line effects should only be considered when the total resistance of the wire is limited:

$$R < 5 Z_0$$

- The transmission line is considered lossless when the total resistance is substantially smaller than the characteristic impedance,

$$R < Z_0/2$$

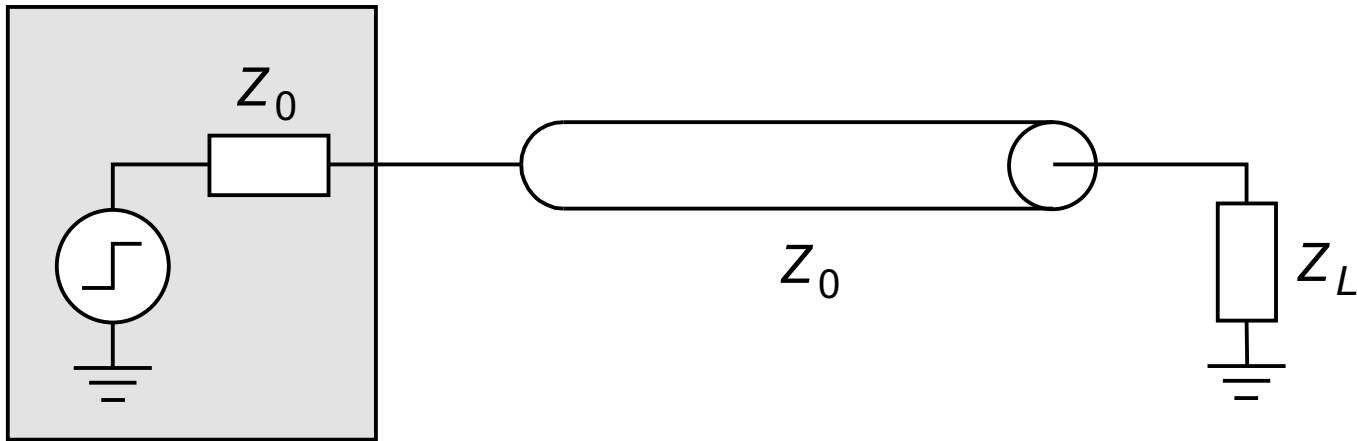
Should we be worried?



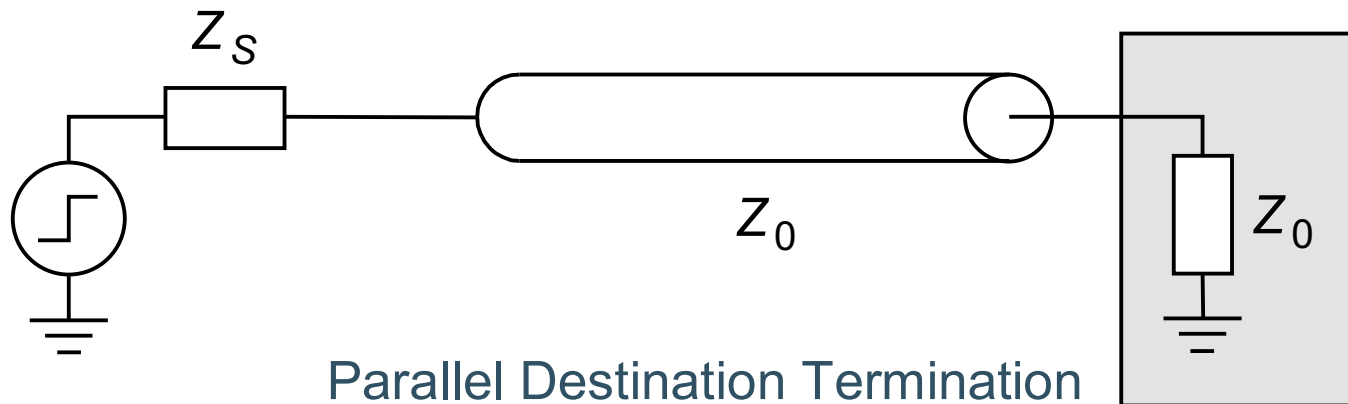
Transmission line effects cause overshooting and non-monotonic behavior

**Clock signals in 400 MHz IBM Microprocessor
(measured using e-beam prober) [Restle98]**

Matched Termination

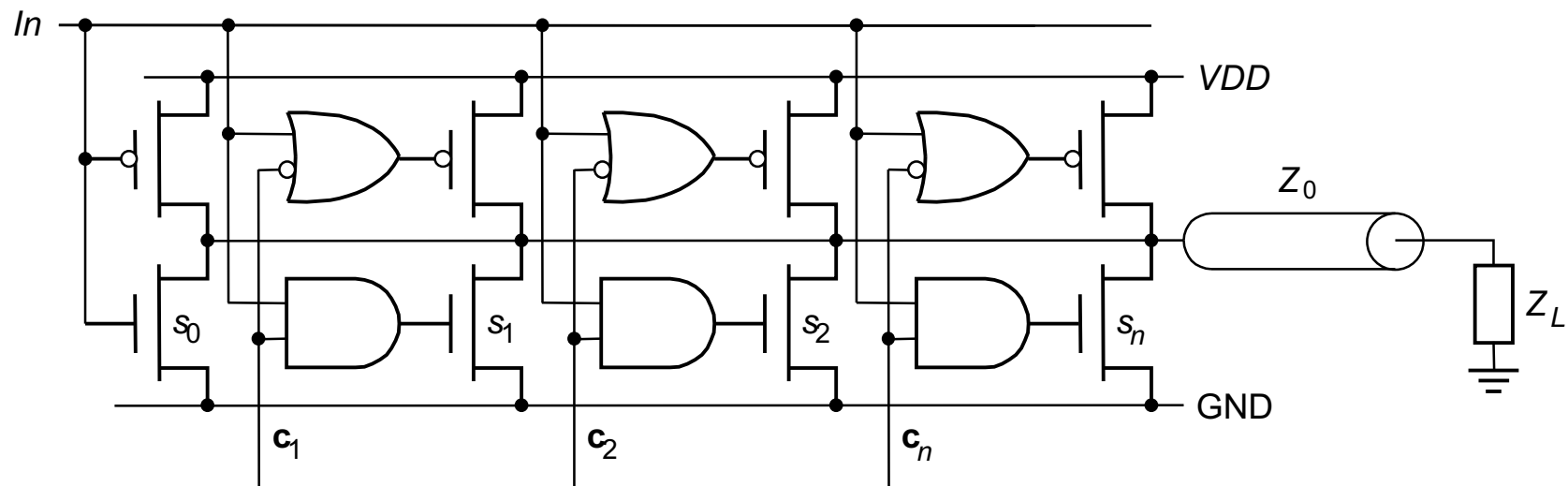


Series Source Termination

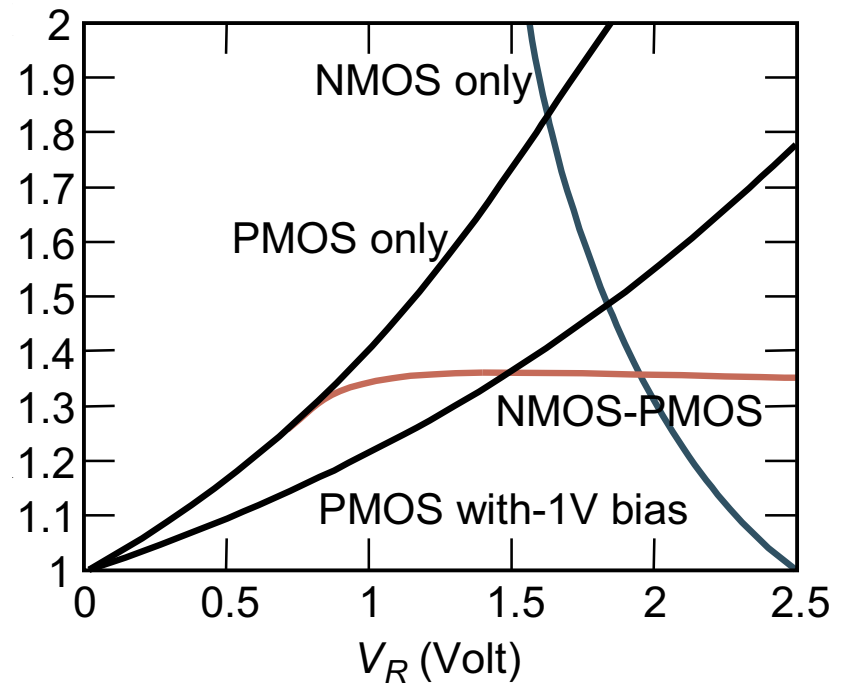
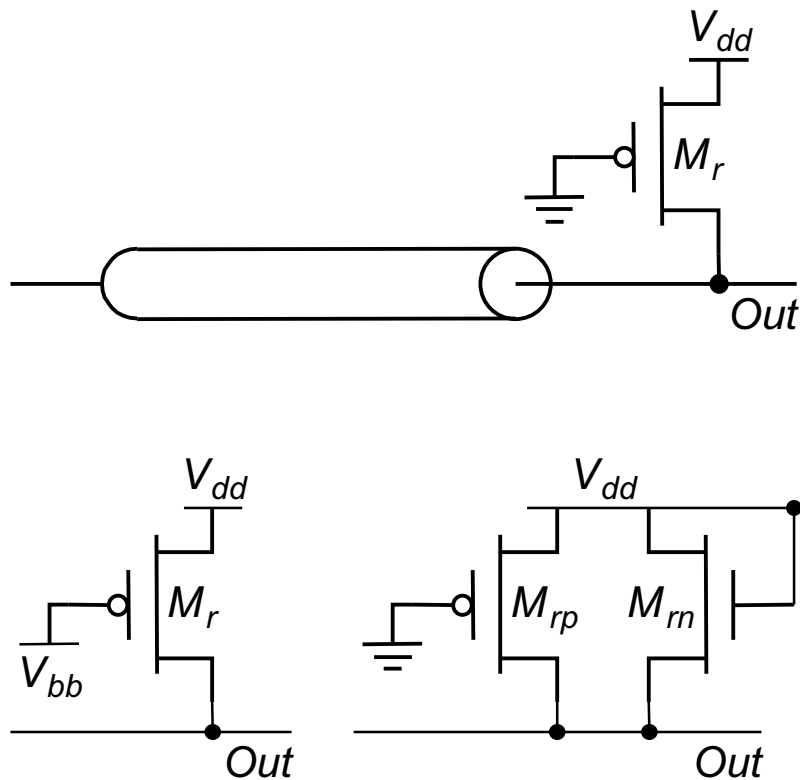


Parallel Destination Termination

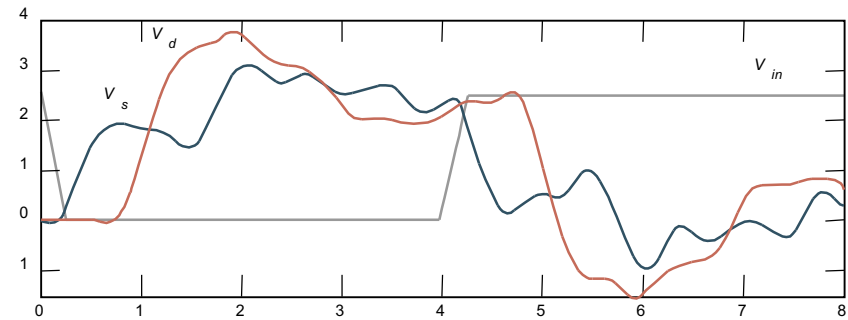
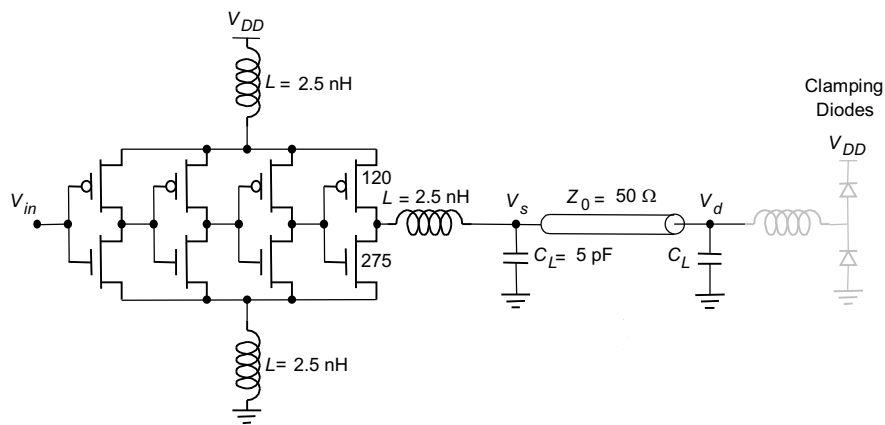
Segmented Matched Line Driver



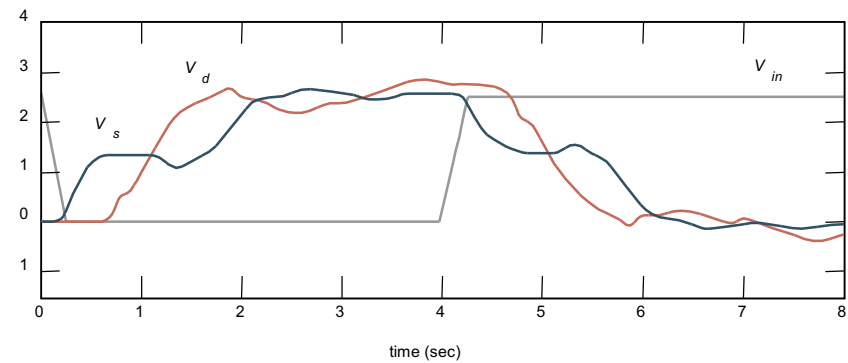
Parallel Termination— Transistors as Resistors



Output Driver with Varying Terminations



Initial design



Revised design with matched driver impedance

The “Network-on-a-Chip”

