
EE273 Lecture 15

On-Chip Power Distribution

March 7, 2001

William J. Dally
Computer Systems Laboratory
Stanford University
billd@csl.stanford.edu

Logistics

- Project
 - Due Today!
 - Automatic extension until 3/12
- Final Exam
 - Friday 3/23, 8:30AM to 10:30AM
 - Location TBD
- Upcoming Lecture Schedule
 - 3/12 – Wrapup – and project discussion
 - 3/14 – Guest lecture – Jeff Cain (Cisco)

A Quick Overview

- Power is distributed on-chip from power pads to point of use
 - from edges for peripheral-bonded chips
- Load profile is determined by characteristics of logic
 - fan-out and select
- The distribution network is sized to handle the worst-case current while keeping the IR drops within margins
- On-chip bypass capacitors can reduce the amount of metal needed for distribution
 - change *peak* requirement to *average* requirement
- Symbiotic bypass capacitors are always present
 - this is why most circuits work!
- Isolate delicate circuits from noisy circuits

The On-Chip Power Distribution Problem

- Supply current is brought on chip at specific locations
 - on the edge for most chips which are *peripherally bonded*
 - distributed over the area of the chip for *area bonded* (C4, solder ball) chips
- Loads consume this current at different locations on the chip at different times
- There is often a large parasitic inductance associated with each bond-wire or solder-ball (0.1-10nH)
- Current is distributed from the bond pads to the loads on thin metal wires
 - $0.04\Omega/\bullet$ typical
- Load currents may be very high
 - average current may be as large as 20A for *very hot* chips (50W at 2.5V)
 - peak current may be 4-5x this amount (100A!)
- L di/dt of bond wire and IR drop across on-chip wires are often a major source of supply noise

Power Distribution Trends

- The on-chip power distribution problem is getting *much* harder as technology evolves
- Combination of
 - lower voltages
 - higher current density
 - thinner metal layers
 - larger chips
- We are quickly approaching the point where peripheral bonding will not be adequate for high-performance chips

Logic Current Profile

- Why does on-chip logic produce a 'spikey' current profile?
- Consider the logic that generates the current
- Current is drawn to charge gate and wire capacitance $Q=CV$, $E=CV^2$
- Typical behavior includes
 - circuit idle before clock edge
 - very little current
 - exponential clock amplification just before clock edge
 - exponential ramp up in current
 - flip-flops are clocked
 - current depends on activity factor
 - Fanout in a logic circuit
 - exponential ramp up in current
 - Fan-in or selection in a logic circuit
 - drop in current

Logic Current Profile Examples (cont.)

- a memory or register array
 - fanout in decoder
 - selection of word-line
 - amplification in S/A
 - selection in multiplexer
- A carry-lookahead adder
 - modest fan-out in PGK blocks
 - fan-out in carry tree
 - fan-in in carry tree, but with amplification
 - amplification to drive output
- Control logic
 - much like the decoder
 - fan-out in the first few stages
 - then tree fanning in to flip-flops
 - often a long, serial tail
- Overall current profile is the superposition of all of these profiles

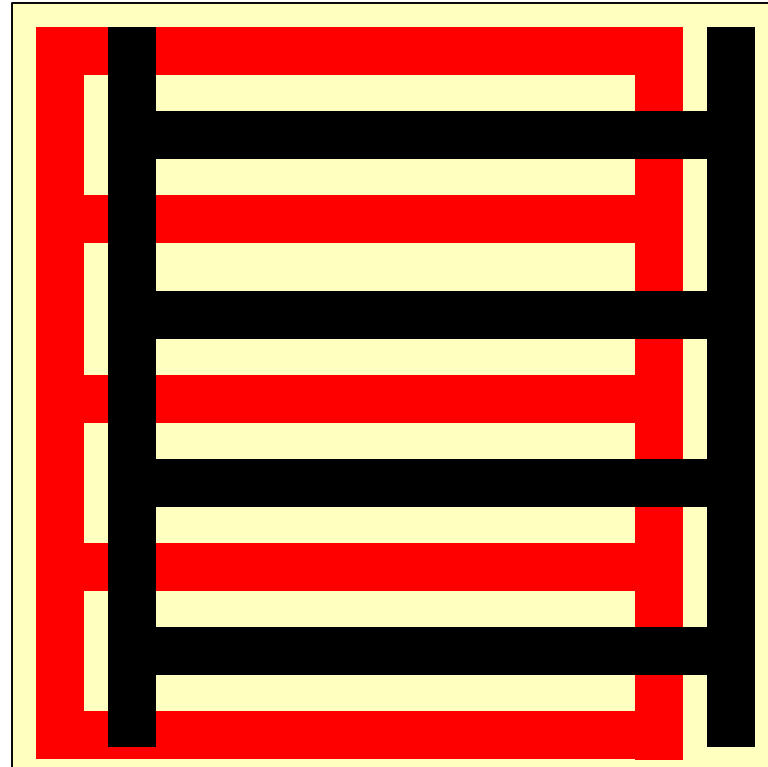
Worst-Case vs. Average Logic Current Profile

- Current drawn is often very data dependent
 - e.g., a data path may switch 64-bits from all 0s to all 1s
 - on average only 1/4 of the bits will have this transition when there is a transition at all
- For noise analysis we must consider *worst-case* power
 - cannot allow possible, but unlikely events to cause system failure
- For battery life we may consider *average* power

IR-Drops

- Power distribution network is designed to keep IR drop on V_{DD} and GND networks within limits
 - e.g., for 10% supply variation, can drop at most 5% on each supply
- Networks are usually designed specifically for the loads of a given chip.
- However, we can gain insight into the process by considering a uniform load
- For example, suppose current densities are
 - $J_{peak} = 0.3A/mm^2$
 - $J_{avg} = 0.05A/mm^2$
- For a peripheral bonded chip, V_{DD} and GND are usually distributed by *combs* with interdigitated fingers
 - a hierarchy of such combs is often used
- How much of a metal layer (or how many layers) do we need to distribute this power

IR Drops, The Picture



IR Drops, the Calculation

- Consider a thin *slice* of the chip, say 1mm high
- Suppose a fraction, k_P , of a metal layer is devoted to each supply
- In a slice of length L , say 16mm, current is drawn from both sides over a 8mm path

$$I_{\max} = \frac{LJ_{\text{peak}}}{2}$$

This current ramps linearly, so current at point x is

$$I(x) = 2xI_{\max} / L$$

- The resistance of a 1mm length of the supply net is

$$R = \frac{R_{\text{sq}}}{k_P} = \frac{0.04}{k_P} \Omega$$

- Voltage integrates along the strip, so

$$V(x) = \frac{x^2 r_W I_{\max}}{k_P L} = \frac{0.04 x^2 J_{\text{peak}}}{2k_P}$$

$$V\left(\frac{L}{2}\right) = \frac{r_W L^2 J_{\text{peak}}}{8k_P} = \frac{0.005 L^2 J_{\text{peak}}}{k_P}$$

IR Calculation, Cont.

- So, to get k_p , for a given voltage drop, V , say 125mV we have
- Clearly this is not acceptable.

$$\begin{aligned}k_P &= \frac{r_W L^2 J_{\text{peak}}}{8V} \\ &= \frac{(0.04)(15)^2 (0.3)}{8(.125)} \\ &= 2.7\end{aligned}$$

- This is almost 3 metal layers for each of power and GND!

Second Example

- 15mm x 15mm Chip draws 10A from a 1.2V supply
- Can tolerate 0.1V IR drop
 - 0.05 on each of Vdd and GND
- Peak to average ratio is 4:1
 - $J_{\text{peak}} = 0.178 \text{ A/mm}^2$
 - $J_{\text{avg}} = 0.044 \text{ A/mm}^2$
- Consider 1mm strip of chip
 - Integrate current to calculate voltage drop
 - $i(x) = Jx$
- 0.4V drop (on Vdd and GND) if one metal layer devoted to power and gnd
 - Only 0.4V left!

$$\begin{aligned}\Delta V &= \frac{r_w L^2 J_{\text{peak}}}{8W} \\ &= \frac{(0.04)(15)^2(0.178)}{8(.5)} \\ &= 0.4V\end{aligned}$$

Metal Migration

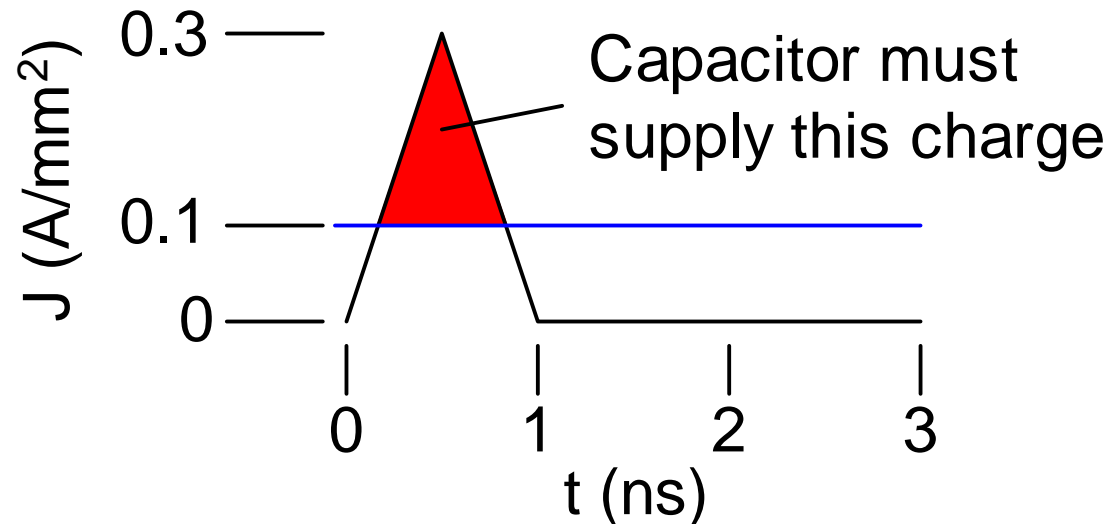
- Over long distances, current density must be kept low to avoid large IR drops
- Over short distances, current density must *still* be kept low to avoid *metal migration*
- Over time, wires that carry high current densities will fail as the metal is eroded away
 - think of your wire as a fuse
- Migration threshold varies depending on process, temperature, and lifetime
- A typical number is
 - $J_{MM} = 1\text{mA}/\mu\text{m}^2$
- This is often a factor on short power buses that connect from the main bus to a point of high current use
- It can also be a factor on the output of high-current drivers
- Migration applies to vias as well as wires

On-Chip Bypass Capacitors

- The good news is that the power distribution network doesn't really need to carry all of the peak current
- Much of the difference between peak and average current may be supplied by local, on-chip bypass capacitors
- Bypass capacitors are also critical in mitigating the effects of the supply bond-wire inductance
- Suppose we want to distribute power on just two metal layers ($k_p=1$) but keep the drop on each supply to 0.125V
 - Need to reduce peak current from 0.3A/mm² to 0.1A/mm²

On-Chip Bypass Capacitors

Current Profile



$$Q = \left(\frac{2}{3}\right)^2 (0.3 \text{ A/mm}^2) (1 \times 10^{-9}) (0.5) = 67 \text{ pC/mm}^2$$

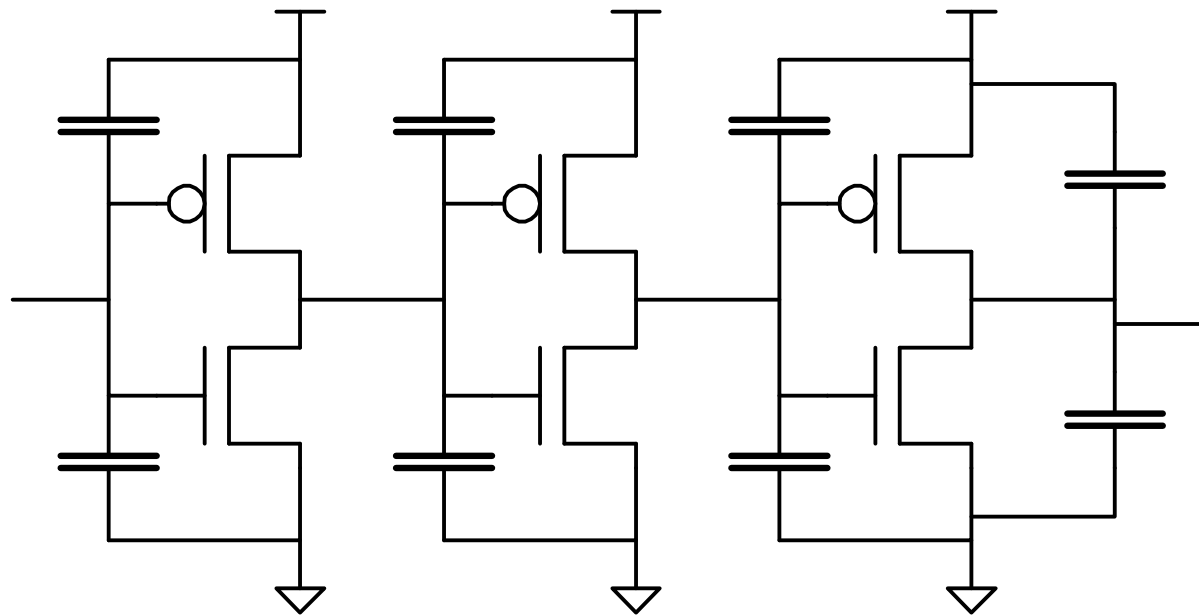
$$C = \frac{Q}{\Delta V} = \frac{67 \text{ pC/mm}^2}{0.25 \text{ V}} = 267 \text{ pF/mm}^2$$

On Chip Bypass Capacitors

- We need a bypass capacitor of about 0.25nF for each 1mm² area of the chip
- For comparison, an MOS capacitor covering a 1mm² area has a capacitance of about 5nF/mm²
- So, our bypass capacitor uses 5% of the silicon area!
- Can be made much smaller with local regulation

Symbiotic Bypass Capacitors

- Where are the bypass capacitors in this picture?
- Gates that are not switching at a given instant in time act as *symbiotic* bypass capacitors
- If only one gate in 60 switches at a given instant, the bypass capacitance is 30 times the switched capacitance



Power Supply Isolation

- Chips often contain
 - **noisy circuits**
 - pad-drivers
 - clock generators
 - large RAM arrays
 - **noise-sensitive circuits**
 - PLLs and DLLs
 - receive amplifiers
 - etc...
- We would like to isolate the noise sensitive circuits from the noise generated by the noisy circuits
- To do this we need to make sure the two circuits share as little of the power distribution network as possible
- Typically provide separate power and/or GND pins
 - **quiet GND and V_{DD}**

Next Time

- Wrap up and project discussion