EECS 427
VLSI Design 1

Lecture 12: Power Supply

(This lecture is mostly derived from EECS 627 slides)

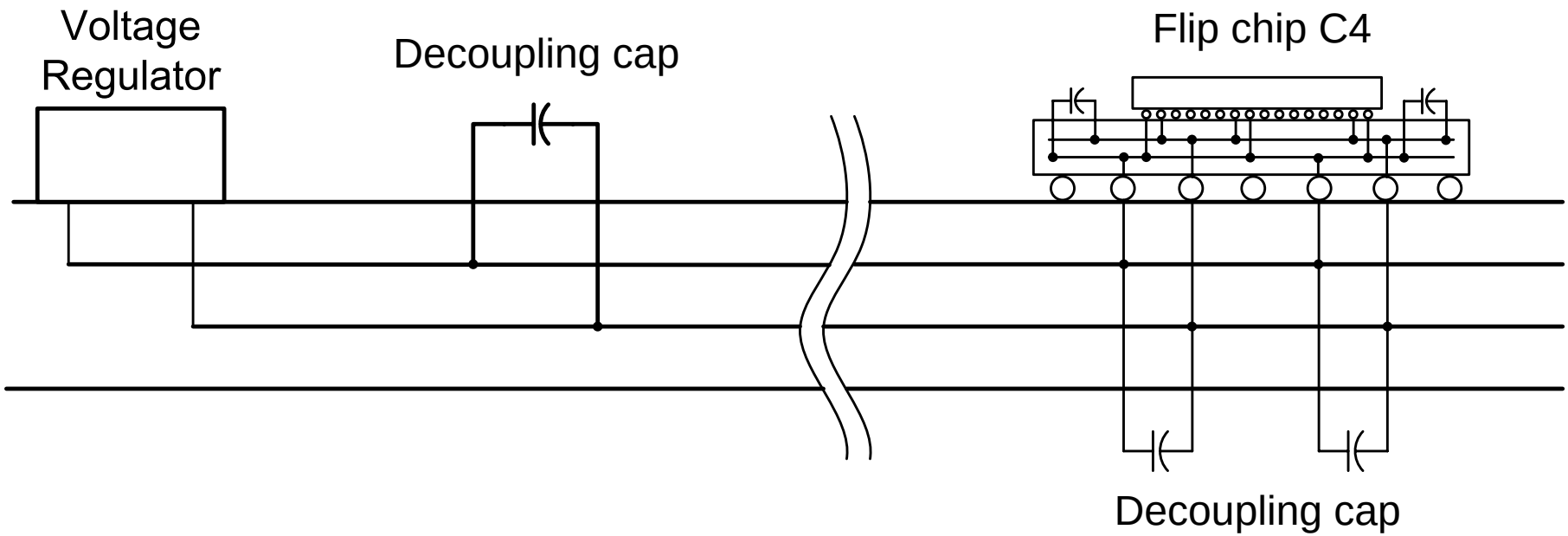
Outline

- Power supply
- Packaging
- Power grid
- Package
- On-chip decoupling capacitors
- Transient analysis
- Movies

Power supply

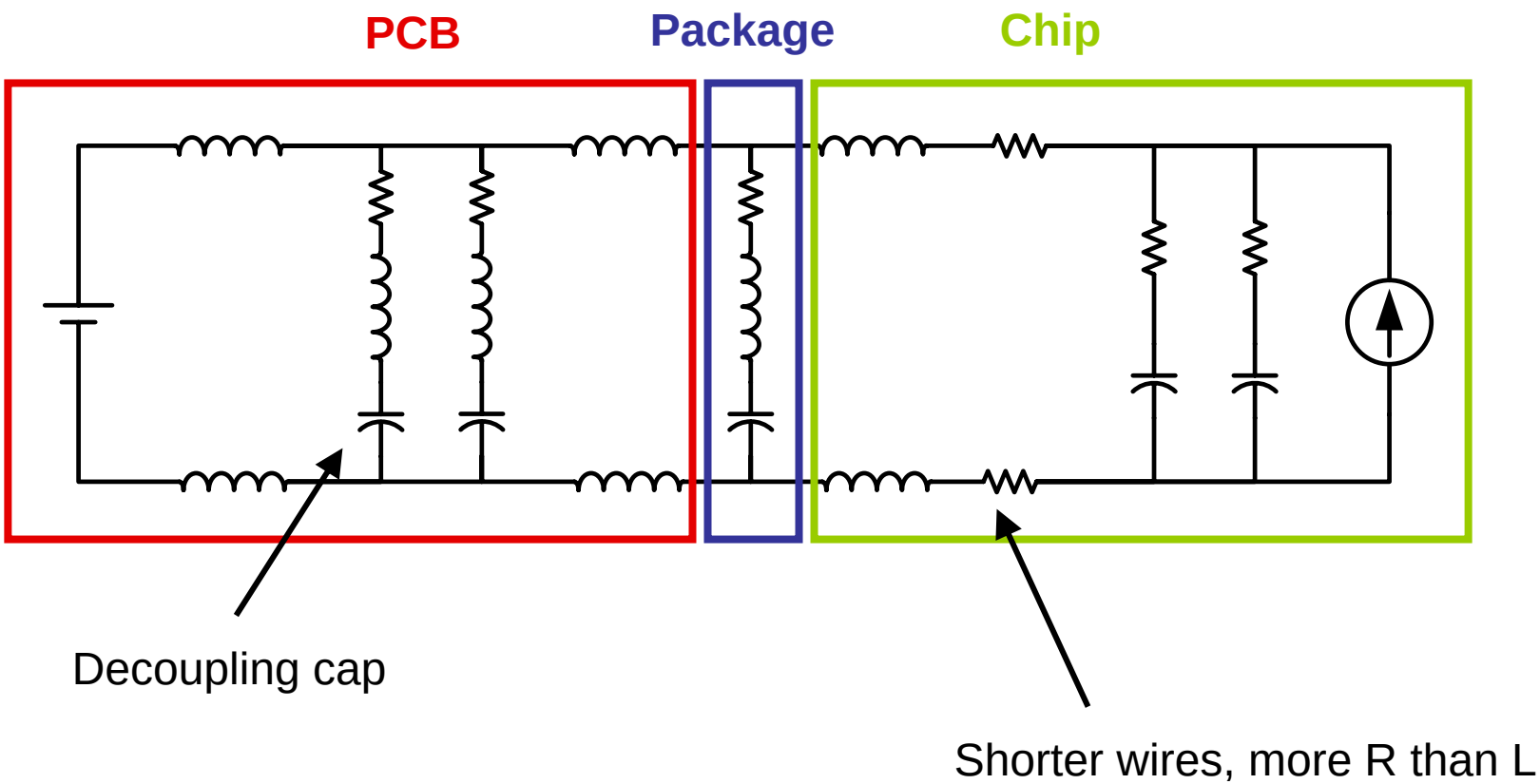
- Goal:
  - Supply a constant voltage (temporal/spatial) to all devices on chip
- Problem:
  - Must get current from voltage regulator to chip
- Supply system does not start from chip pins but includes board
- Supply network design is very difficult, even though small number of transistors

From the regulator to the chip



C4: Controlled Collapsed Chip Connects

Supply network model



Voltage drop (1/2)

- IR drop $\Delta V_{IR} = IR$
  - Max at peak current
  - Scaling of I_{dc} P ↑ V ↓ I ↑↑
    - P=15W V_{dd}=2.5V
    - P=150W (10x) V_{dd}=1.25V (0.5x)

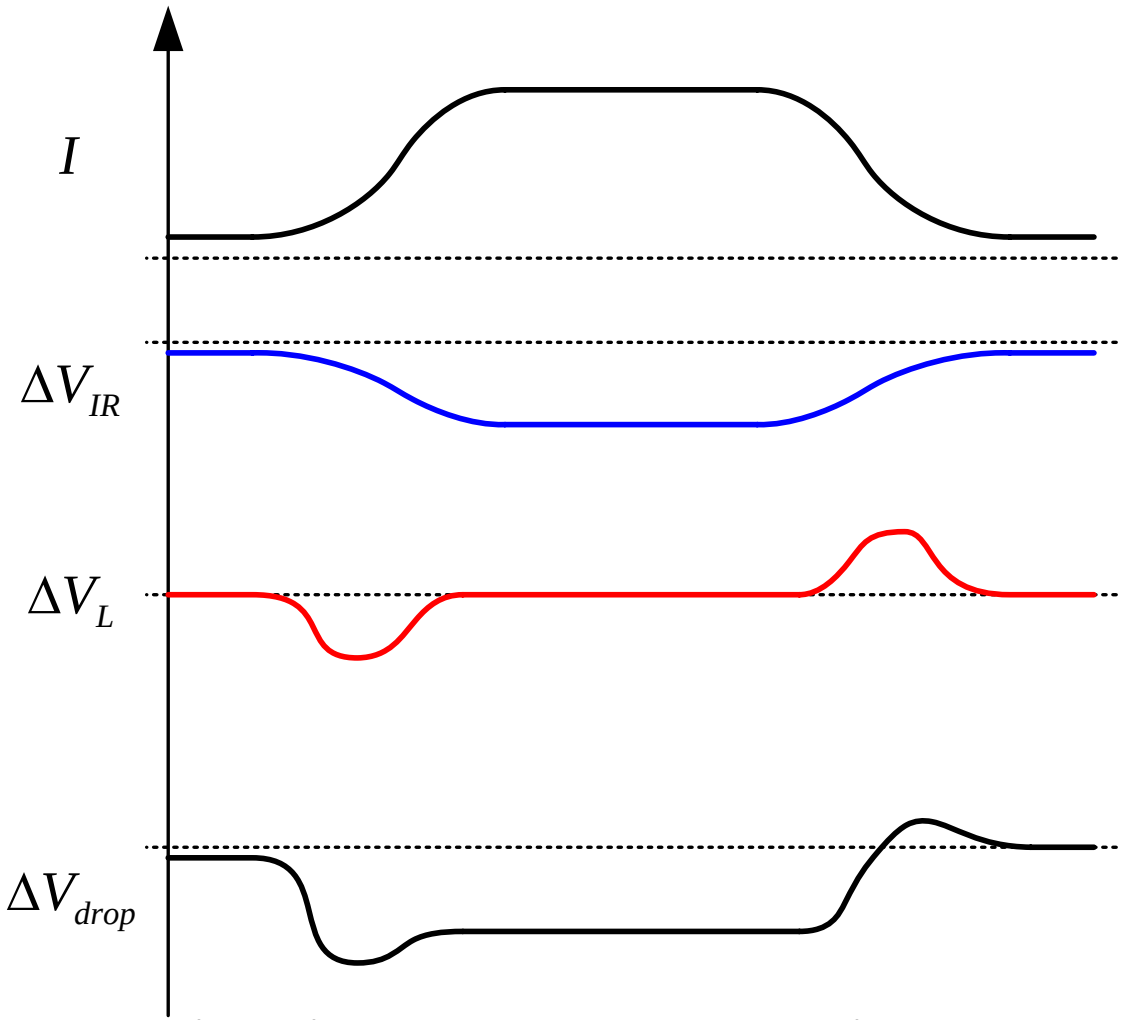
$$\Delta I = \frac{\Delta P}{\Delta V} = 20x$$

- Ldi/dt drop $\Delta V_L = L\frac{dI}{dt}$
  - Due to change in current
  - Scaling of V_{ac} dI ↑ dt ↓ dI/dt ↑↑↑
    - f=200MHz
    - f=1GHz (5x)

$$\Delta I = 7x \qquad \Delta V_L = 35x$$

Voltage drop (2/2)

- Time of drop is different. Do not strictly add



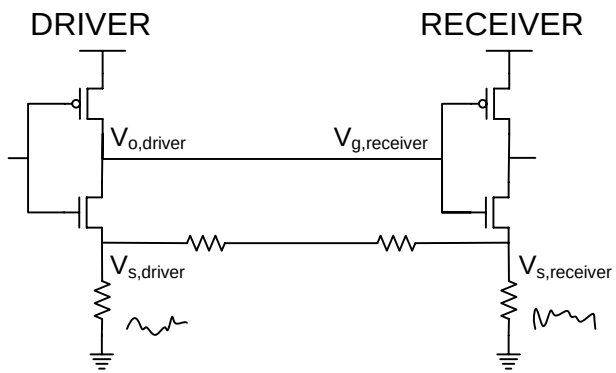
- Pulsed noise
- DC noise (sustained)
- under/over shoot
- Supply variation both:
  - Time (temporal)
  - Area (spatial)

Problems due to power grid noise (1/3)

- Sustained drop: Performance degradation

$$t_d \propto \frac{CV}{I} \propto \frac{1}{V_{gs} - V_t} \propto \frac{1}{V_{dd} - V_{ss} - V_t} \qquad \Delta V \uparrow \rightarrow t_d \uparrow$$

- AC Noise: Functional/delay noise
  - Acts as standard noise pulse
  - Difference not absolute voltage drop
  - Uniform drop is OK



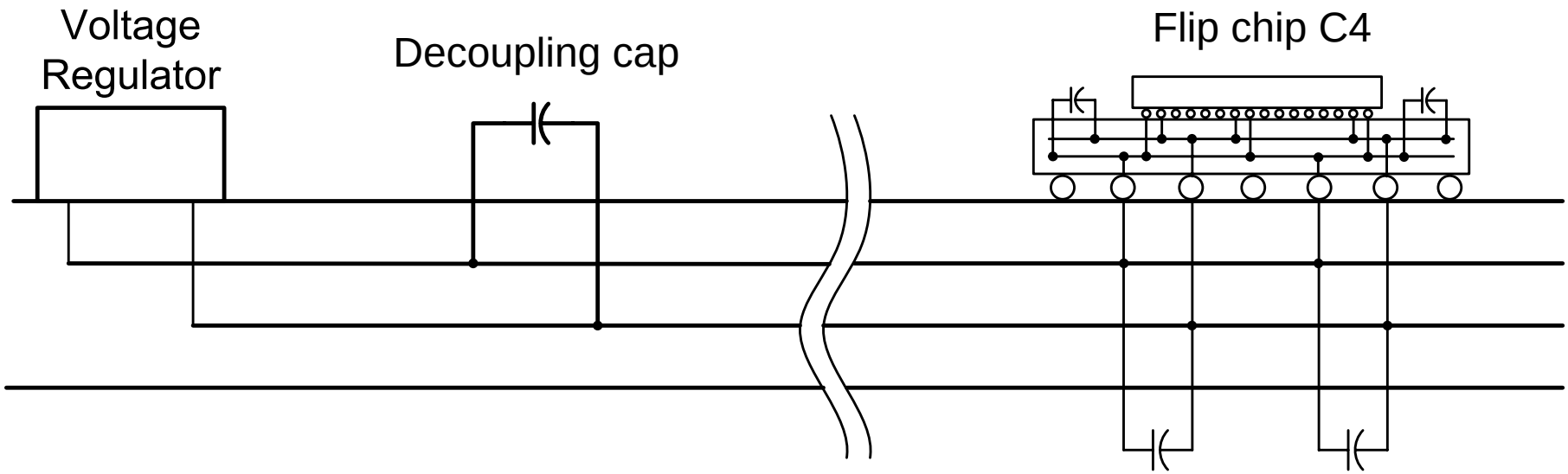
Problems due to power grid noise (2/3)

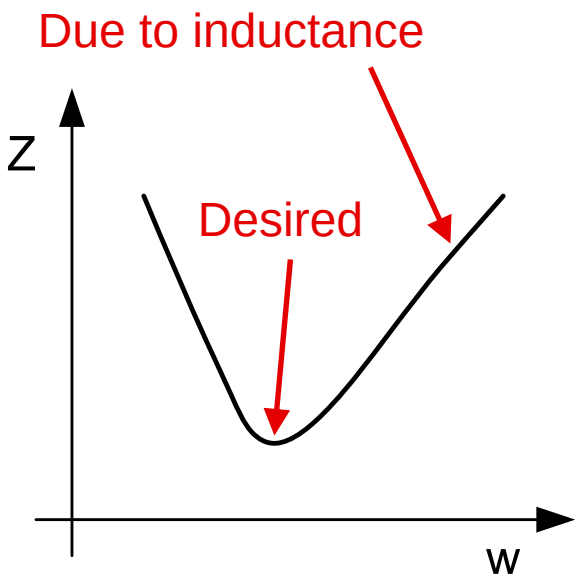
- TDDB: Time Dependent Dielectric Breakdown
  - Due to overshoots (inductance)
  - Hot electrons cause injection into gate oxide, may get trapped → degrade the performance of the transistor
  - Limit on Vdd
- Electromigration
  - High current densities displace metal atoms out of place due to momentum transfer.

Elements of power supply system - Outline

- Board
- Package
- Chip
  - Power grid
  - Decoupling cap

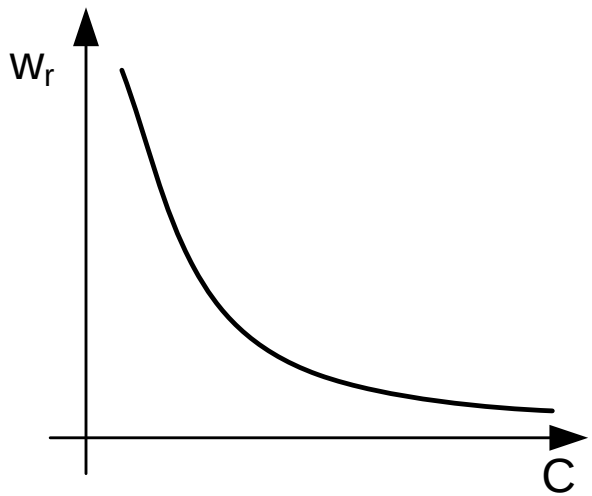
Board design (1/3)



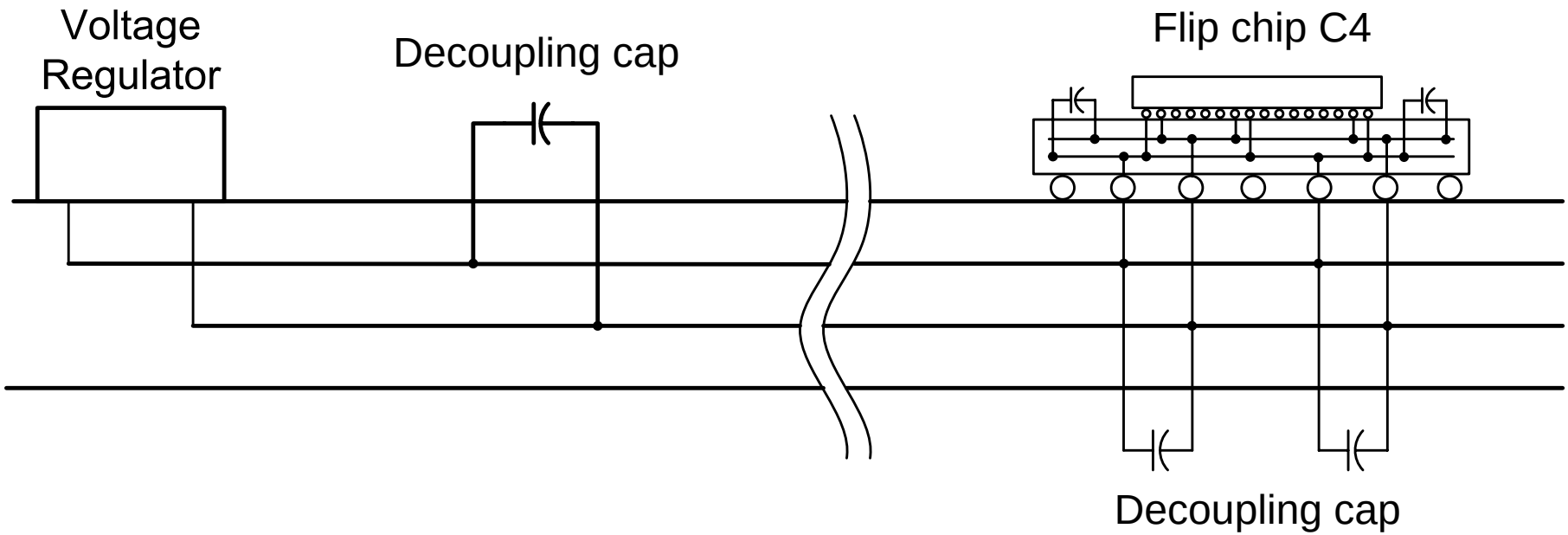


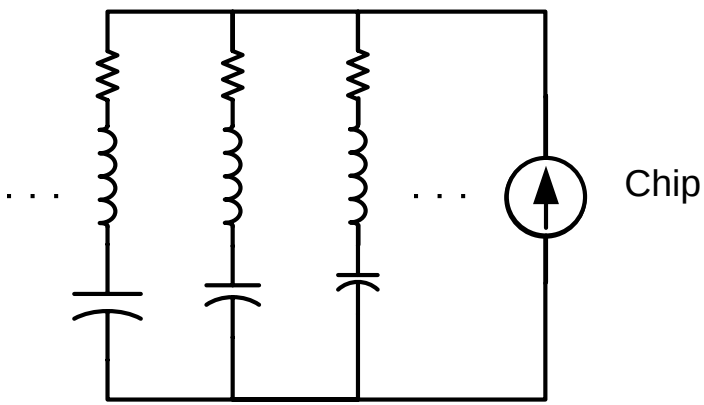
$$Z = R + j\omega L + \frac{1}{j\omega C}$$

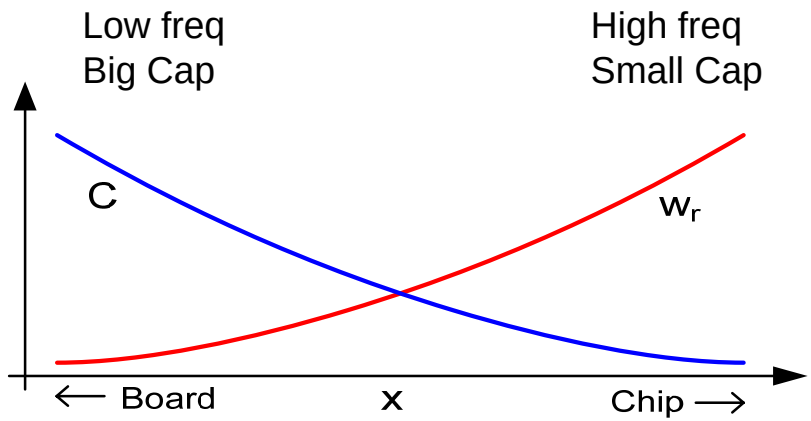
$$\omega_r = \sqrt{\frac{1}{LC}}$$



Board design (2/3)







Board design (3/3)

- Make sure resistance is small
- Decap design: act to filter high frequency current
  - Small cap w/low L → near chip to suppress high frequency current
    - Try to get as close to chip as possible
  - Large cap w/higher L → further away from chip

Package types (1/2)

- Wire bonding
  - Advantages:
    - Cheap
    - Allows for thermal expansion
  - Disadvantages:
    - High inductance (typical 5nH)
    - Limited pads (~200)
  - To reduce R → More metal layers
  - To reduce L → Lower distance from connections, bring cap closer

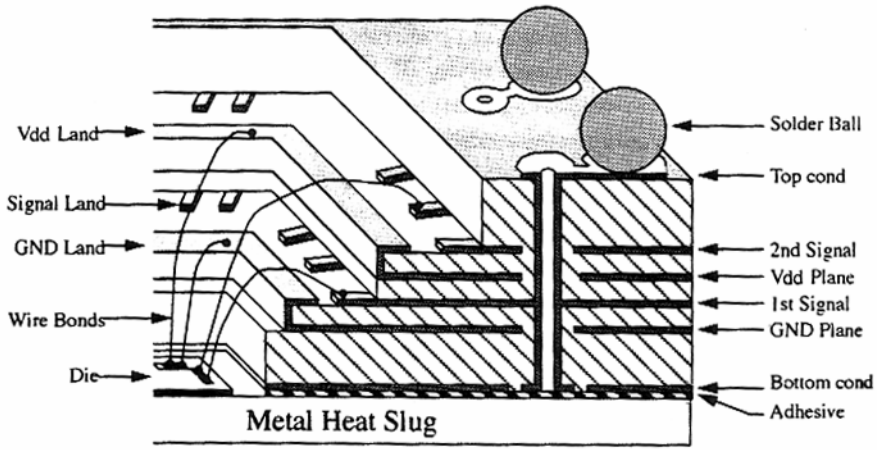


FIGURE 2-8 Typical Type-II BGA Construction

Package types (2/2)

- Flip-chip
  - Advantages:
    - Lower inductance (0.1nH) → lower distance
    - High number of pads (~1000)

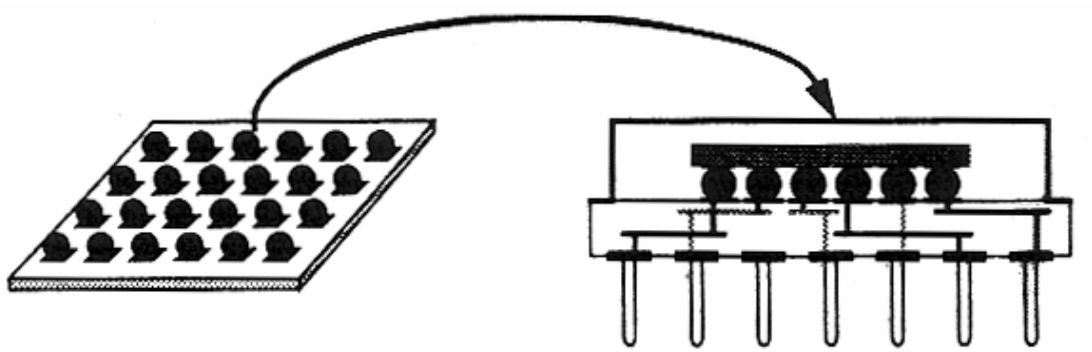
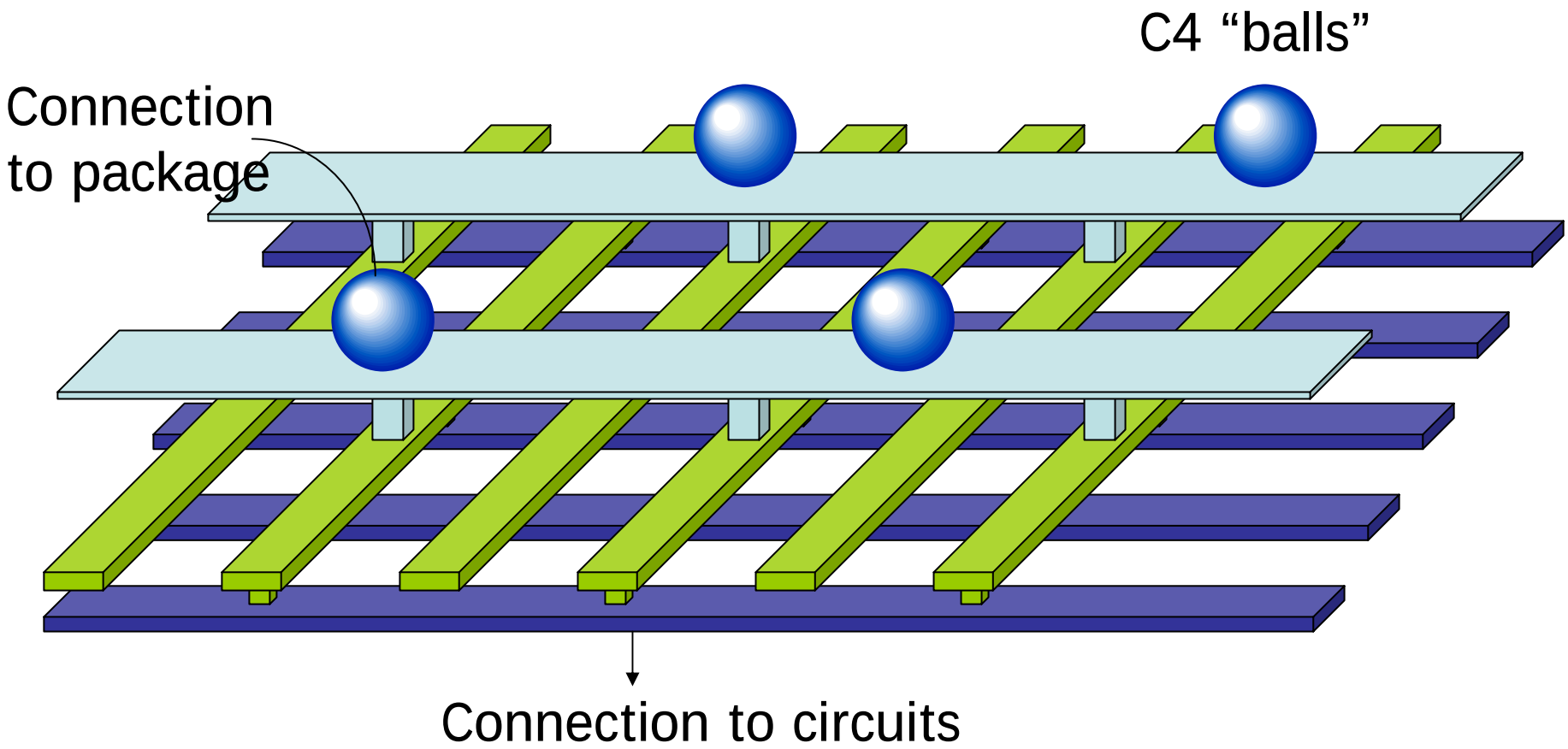


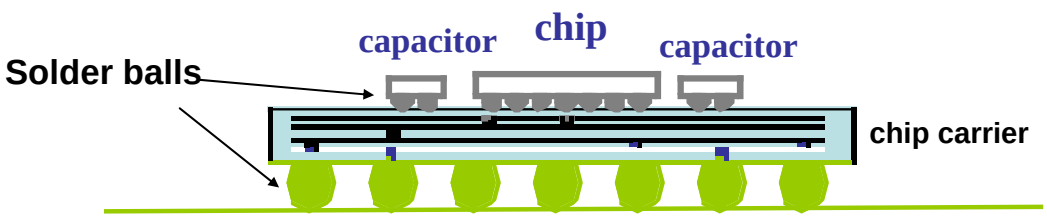
FIGURE 2-3 Flip-Chip Bonding

On-Chip Power Delivery



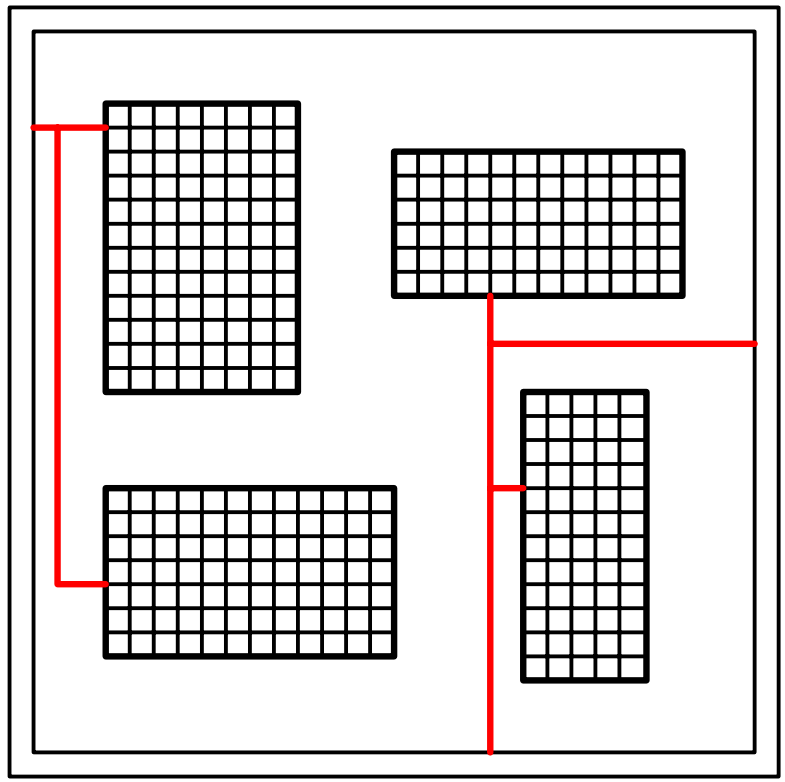
Power Integrity & Package

- Package design becoming increasingly complicated
- Large number of power/ground layers in current packages
- Significant Power supply variations caused by the package



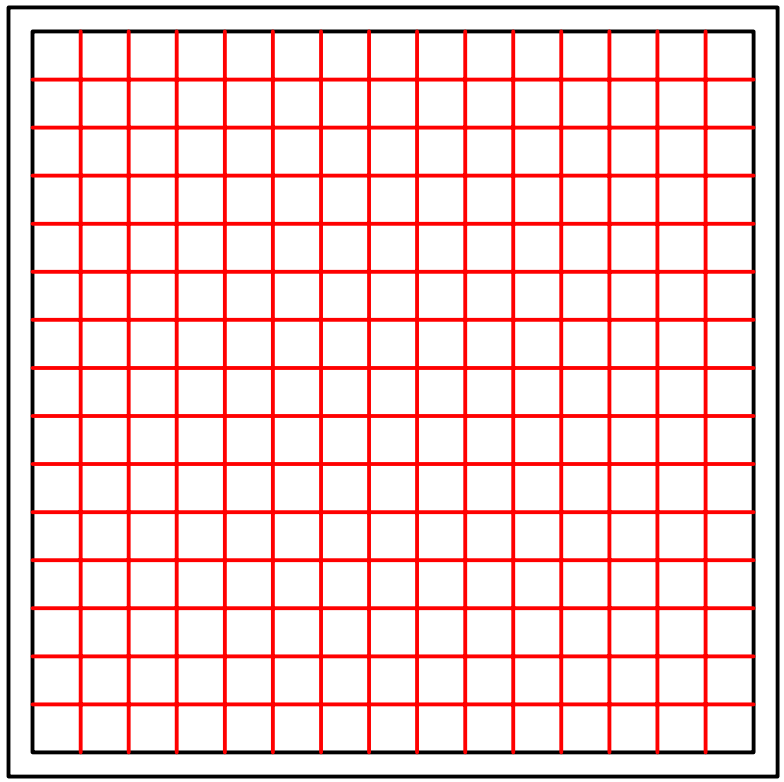
Power Grid design (1/2)

- Tree
  - Local grids
  - Simple
  - Metal utilization is efficient
  - Single wire failure

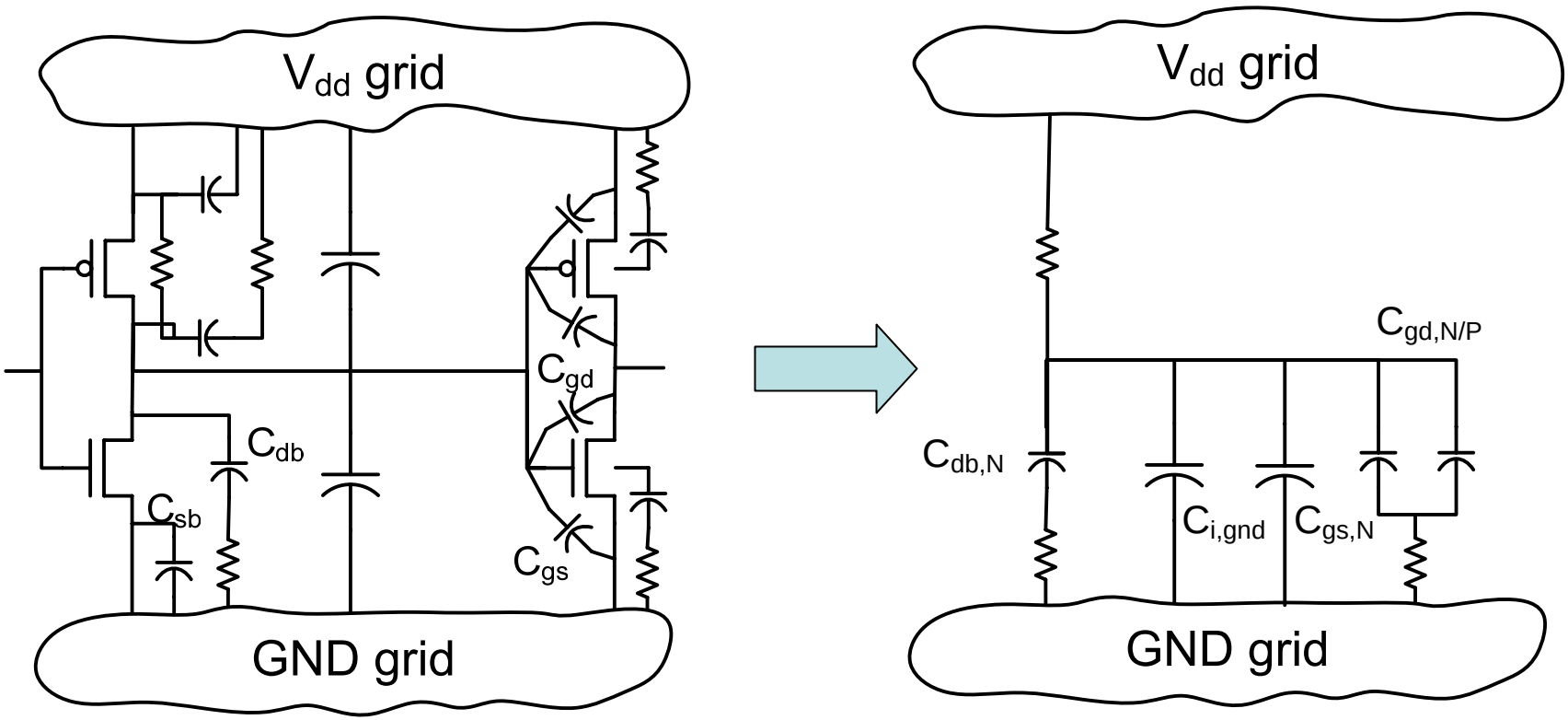


Power Grid design (2/2)

- Grid
  - High end processor
  - Lots of metal
  - Lower resistance
  - Lines up with C4
  - Less J, better electromigration
  - Spatial variation is lower
  - Highly redundant



Decoupling cap – Implicit (2/3)



$$C_{eff} = \frac{1}{2}\left(C_{db,N} + C_{gs,N} + C_{db,P} + C_{gs,P}\right) + \frac{1}{2}\left(C_{i,gnd} + C_{i,Vdd}\right) + C_{gd,N} + C_{gd,P}$$

- Depends on state of circuit
- Current injected/extracted at different locations
- Includes both device and interconnect parasitics

Decoupling cap – Back of the envelope (3/3)

- Implicit decoupling capacitance is the capacitance that is not switching
- Assuming a switching factor of s

$$P = fC_{switch}V_{dd}^{2}$$

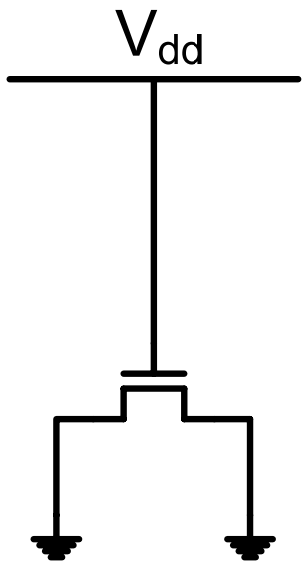
$$C_{switch} = \frac{P}{fV_{dd}^{2}}$$

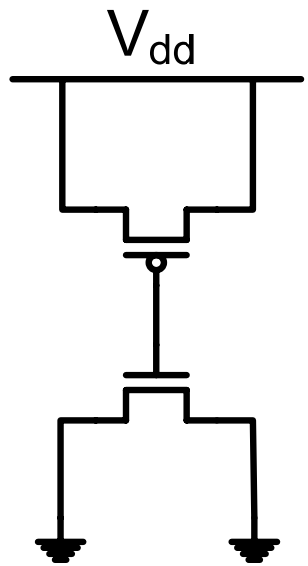
$$C_{switch} = sC_{total}$$

$$C_{decap} = (1-s)C_{total}$$

$$C_{decap} = \frac{1-s}{s}\frac{P}{fV_{dd}^{2}}$$

Decoupling cap – explicit (1/3)

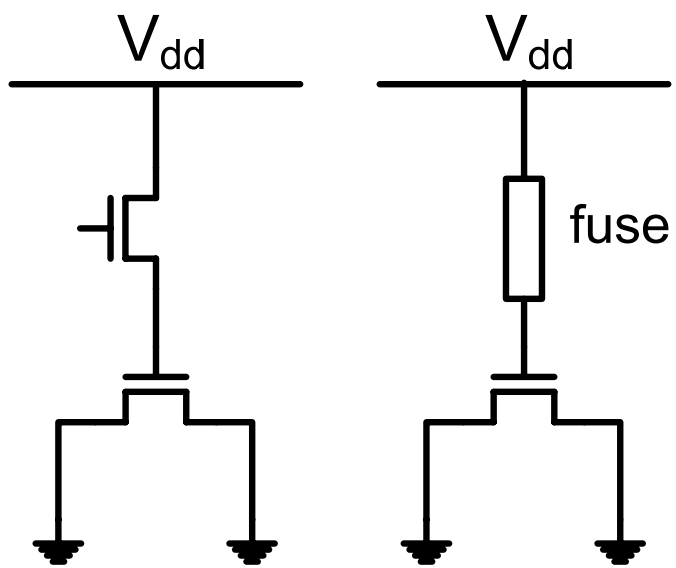




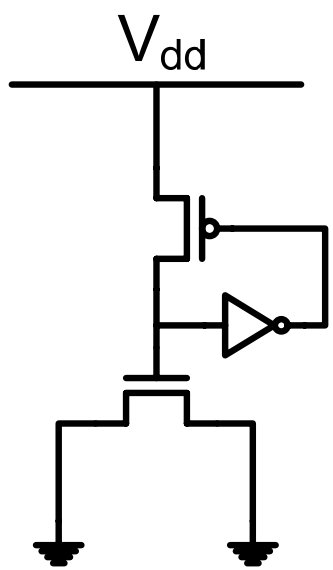
- Reliability: Thin oxide gate can short. Yield problem
- High gate leakage

- Short only if double failure
- Gate voltage is half, less sensitive to gate failure
- Less cap

Decoupling cap – explicit (2/3)

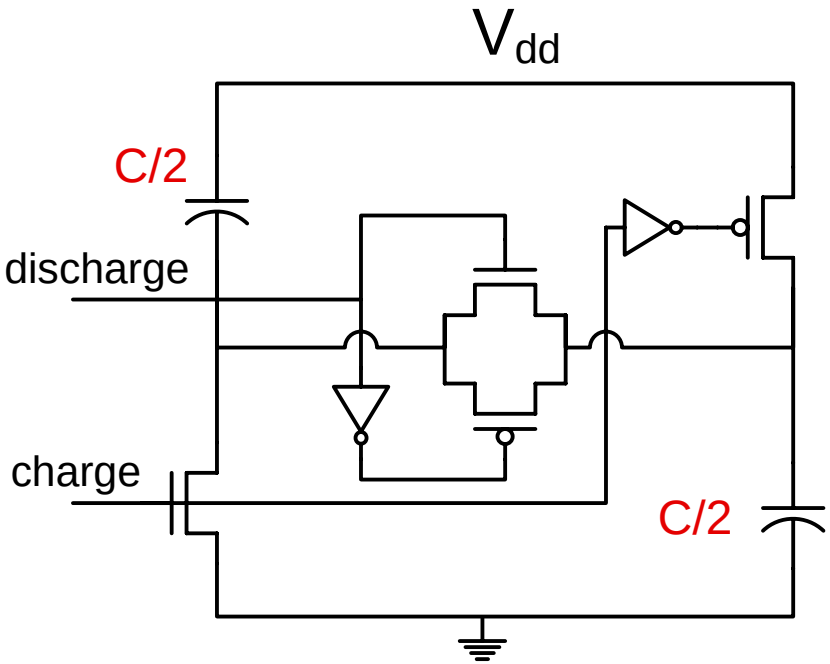


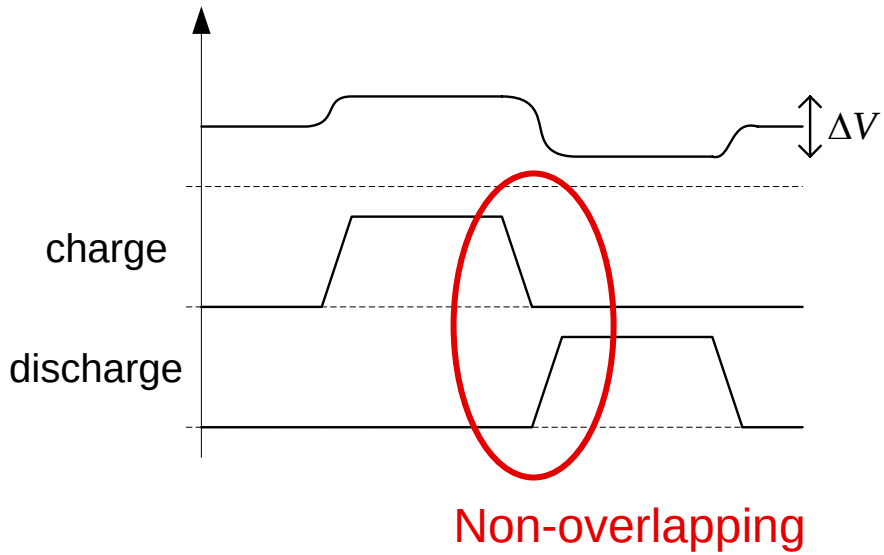
- Turns off the cap in case of oxide failure
- Less gate leakage
- Unusable if the transistor fails



- Variation of previous

Active decoupling cap – explicit (3/3)





Normal cap

$$\Delta Q = \Delta V \cdot C$$

This scheme

$$\text{Charge } Q = V_{dd} \cdot C$$

$$\text{Discharge } Q = \frac{C}{n} \cdot (V_{dd} - \Delta V)$$

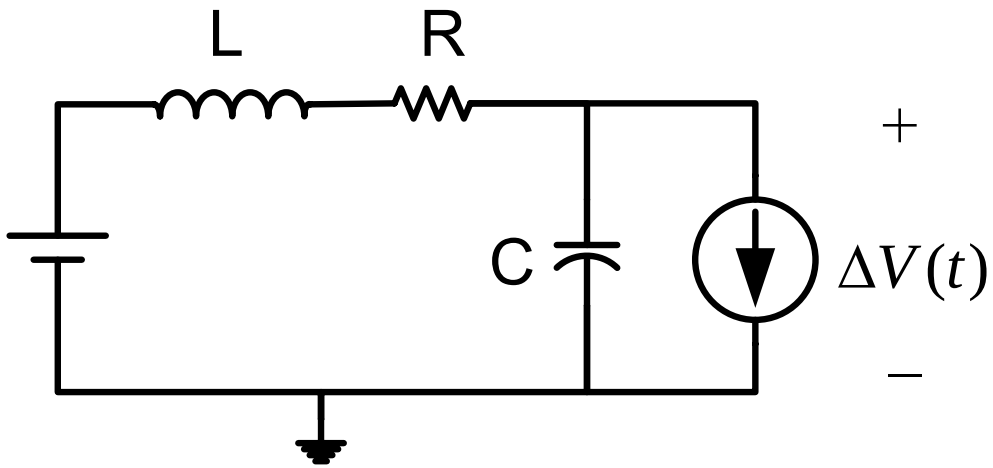
$$\Delta Q = V_{dd} \cdot C - \frac{C}{n} \cdot V_{dd} + \frac{C}{n} \cdot \Delta V$$

$$= \frac{n-1}{n} C \cdot V_{dd} + \frac{C}{n} \cdot \Delta V$$

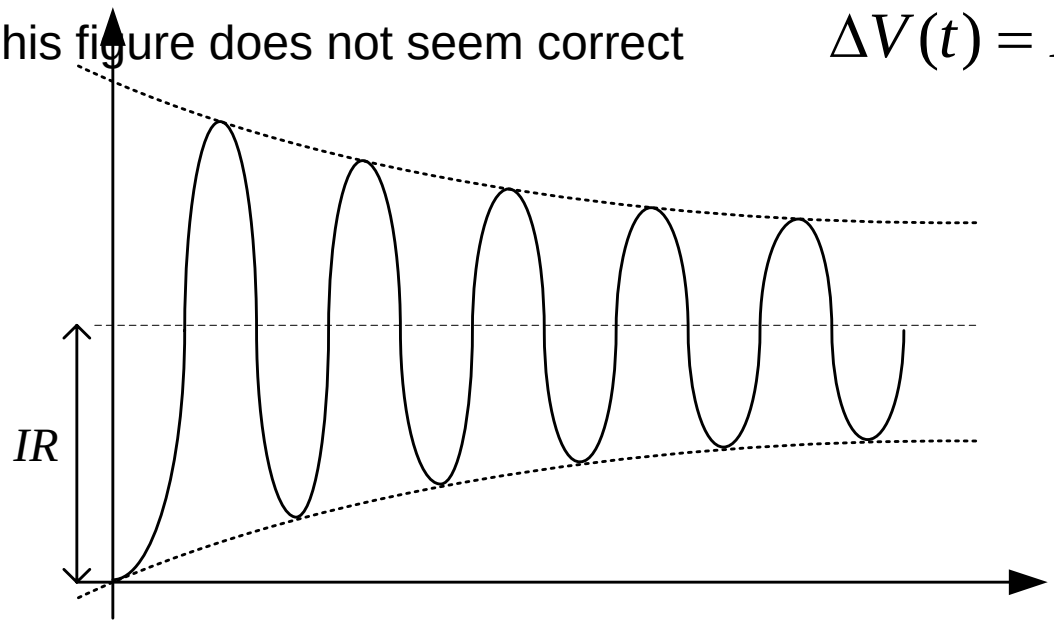
For n=2

$$\Delta Q = \frac{1}{2} C \cdot V_{dd} + \frac{C}{2} \cdot \Delta V$$

Power grid model (1/3)



This figure does not seem correct



$$\Delta V(t) = IR + I\sqrt{\frac{L}{C}}e^{-\frac{R\cdot t}{2L}}\sin(\omega_r t - \theta)$$

$$\omega_r = \sqrt{\frac{1}{LC}}$$

Power grid model (2/3)

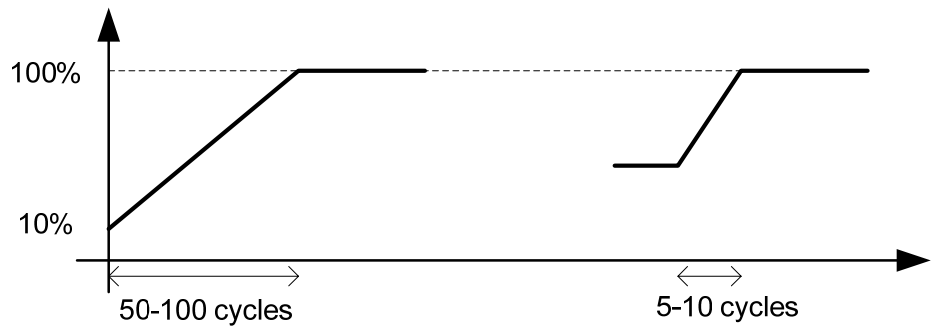
IR-drop $\Delta V_{IR} \propto I, R$

- Solved with more metal
- Spatial variation of voltage

LdI/dt-drop $\Delta V_L \propto I\sqrt{\frac{L}{C}}$

Damping $\frac{R}{2L}$

- Decoupling cap can only help with $\sqrt{1/C}$
- Current step input
  - Out of reset
  - Instructions



Power grid model (3/3)

Resonance $\omega_r = \sqrt{\frac{1}{LC}}$ Quality factor $Q = \frac{1}{R}\sqrt{\frac{L}{C}}$

We do not want the system to resonate

10-15 years ago $\omega_r > \omega_{clk}$

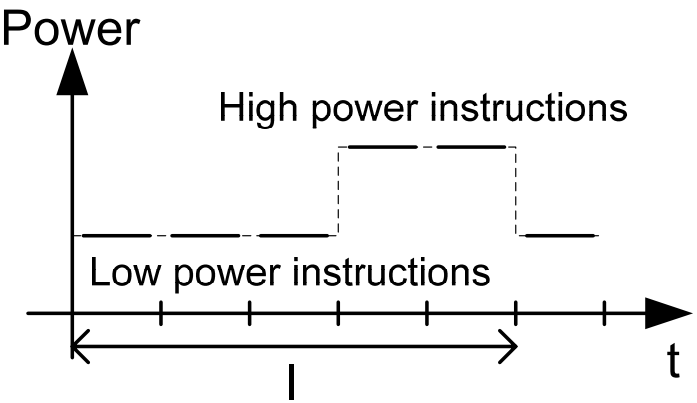
- Harmonics of the clock might hit ω_r
- Tune ω_r to avoid clock harmonic

$\omega_r \neq k\omega_{clk}$

k→harmonic

Today $\omega_r < \omega_{clk}$

- Due to circuit activity, activation frequency can be lower than ω_{clk}
- Cannot tune to avoid clock harmonic, depends on instruction stream



$\omega_r \neq \frac{k\omega_{clk}}{l}$ k→harmonic

l→loop

How to fix

Reduce R: Use more metal. Tree → Grid → Plane

ΔV_{IR} ↓ but less damping Q ↑

Reduce L: Thin package, bondwire, flip-chip. More pads

ΔV_L ↓ ω_r ↑ Q ↓ difficult to control

Increase C: Decoupling capacitance, but only $\sqrt{C}$

ΔV_{IR} ↓ ΔV_L ↓ Q ↓ Area ↑

Decrease I: Turn modules slowly, larger reset latency

ΔI ↓ ΔV_{IR} ↓ ΔV_L ↓

Transient Analysis Results

Design	Clk freq (MHz)	w_r freq (MHz)	IR-drop (mV)	Undershoot (mV)	Overshoot (mV)
P1	200	120	101	183	100
P2	300	115	67	246	159
P3	500	90	150	345	199

- No significant increase in IR-drop as speed and power increase
- However, noise due to power transience is increasing rapidly
- Decrease in resonance frequency is of critical concern.