Power Distribution

Routing resources
- 20-40% of all metal tracks used by Vdd, Gnd
- Increased power → denser power grid

Pins
- Vdd or Gnd pin carries 0.5-1W of power
- Pentium 4 uses 423 pins: 233 Vdd or Gnd
- More pins → higher cost package
  (+ package development, board design…)

Battery cost
- 1kg NiCd can power P4 for <1hr

Performance
- High chip temperature degrades performance
- Large across-chip temp variations induce clk skew
- High chip power limits use of high-perf circuits
- Power transients determine min supply

Courtesy: A. Kahng
Power = Package

Pentium 4 die is about 1.5g and less than 1cm³

Pentium-4 in package with interposer, heat sink, and fan can be 500g and 150cm³

Modern processor packaging is complex and adds significantly to product cost.

Courtesy: A. Kahng

Packaging

- Package functions
  - Mechanical connection of chip to board
  - Electrical connection of signals and power from chip to board with very little delay or distortion
    - Short wires with low R and L
  - Removes heat produced on chip and thermal expansion and stress
  - Protects chip from mechanical damage
  - Inexpensive to manufacture and test

- Main issues:
  - Cost
  - Thermal impedance: how effectively package removes heat from the die
    [Reading: Weste, Harris VLSI book]
  - Lead inductance

Ceramic pin grid array package – lowest
Package Types

- Through-hole vs. surface mount

Multi-Chip Modules

- Pentium Pro MCM
  - Fast connection of CPU to cache
  - Expensive, requires known good dice
Concept of a Typical Chip Package

- Chip Mounting Cavity
- Lead Frame
- Bonding Wire
- Pin

Chip-to-package Connections

- Traditionally, chip is surrounded by pad frame:
  - Metal pads on 100 – 200 μm pitch
  - Gold bond wires attach pads to package
  - Lead frame distributes signals in package
  - Metal heat spreader helps with cooling
Chip / Package Alignment

- Align top-left corners and do simple bonding

Top Left (L-shapes)

Top Left (index marker)

die

package

Example Package Data Sheet

SSM P/N CPG12028

www.spectrum-semi.com
**Advanced Packages**

- Bond wires contribute parasitic inductance
- Fancy packages have many signal, power layers
  - Like tiny printed circuit boards
- **Flip-chip** places connections across surface of die rather than around periphery
  - Top level metal pads covered with solder balls
  - Chip flips upside down
  - Carefully aligned to package (done blind!)
  - Heated to melt balls
  - Also called **C4** (Controlled Collapse Chip Connection)
Package Parasitics

- Use many $V_{DD}$, Gnd in parallel
  - Inductance, $I_{DD}$

Heat Dissipation

- Comparison
  - Light bulb: 60W, surface area ~120cm$^2$ (too hot to touch)
  - Itanium 2: 130W, die area ~4cm$^2$ (60x higher power density)

- The heat flows from the transistor junctions through the substrate and package
  - Can be spread across a heat sink,
  - Then carried away through the air by convection
  - Liquid cooling used in extreme cases ($$$)

- Analogy
  - Current flow: voltage difference / resistance
  - Heat flow: temperature difference / thermal resistance
**Thermal Resistance**

- $\Delta T = \theta_{ja} P$
  - $\Delta T$: temperature rise on chip
  - $\theta_{ja}$: thermal resistance of chip junction to ambient
  - $P$: power dissipation on chip

- Thermal resistances combine like resistors
  - Series and parallel

- $\theta_{ja} = \theta_{jp} + \theta_{pa}$
  - Series combination

**Thermal Impedance**

- Ceramic pin-grid arrays – 15 to 30 °C/Watt

- Plastic Quad Flat Packs – 40 to 50 °C/Watt

- Heat dissipation:
  - Finned heat sinks
  - Embedded metal slugs

- High-cost packages:
  - Forced air or liquid cooling through package ducts
  - Example: IBM Thermal Conduction Module
Example: Thermal Resistance and Power

- Your chip has a heat sink with a thermal resistance to the package of 4.0° C/W
- The resistance from chip to package is 1° C/W
- The system box ambient temperature may reach 55° C
- The chip temperature must not exceed 100° C
- What is the maximum chip power dissipation?

Power Distribution

- Power Distribution Network functions
  - Carry current from pads to transistors on chip
  - Maintain stable voltage with low noise
  - Provide average and peak power demands
  - Provide current return paths for signals
  - Avoid electromigration & self-heating wearout
  - Consume little chip area and wire
  - Easy to lay out
Power Supply Drop/Noise

- Supply noise is variations in power supply voltage that manifest as noise onto the logic gates.
  - Power supply wiring resistance creates voltage variations with current surges.
  - The current surge for static CMOS depend on dynamic behavior of circuit.
- Tackling the drop
  1. \( V_{DD} \)-GND capacitance
     - Based on total maximum switched capacitance (10x)
  2. Redesign power/ground network to reduce resistance.
     - Based on maximum current required by each block
  3. Adjust activity to another clock cycle to reduce peak current.
     - Scheduling

Power Requirements

- \( V_{DD} = V_{DD\text{nominal}} - V_{\text{droop}} \)
- Want \( V_{\text{droop}} < +/- 10\% \) of \( V_{DD} \)
- Sources of \( V_{\text{droop}} \)
  - IR drops
  - L di/dt noise
- \( I_{DD} \) changes on many time scales
**Issue #1: RI Introduced Noise**

![Diagram of noise introduction](image)

- $V_{DD} - \Delta V$
- $I$
- $R$
- $M_1$
- $\Delta V$

**Power IR Drop Example**

- Drive a 32-bit bus, total load of each wire: 2pF, $R = 0.125\Omega$/square, want delay ~0.5ns, < 10% drop
  - $R$ for each transistor needs to be < 0.25kΩ
    - To meet $RC = 0.5ns$
  - Effective $R$ of bits together is $250/32 = 7.5\Omega$
  - For < 10% drop, Power $R$ must be < 1Ω
  - That is only 8 squares

- Must support Total Power
  - Chips today dissipate 5-50W
  - Implies total current is 4-40A (Power = IV)
    - Supply is now as low as 1.2V!
  - Very big problem currently
  - Use many supply pins (@0.2mA each), and wide wires for low $R$
  - Grids of higher level metal for power is a must!
    - Thicker metal... lower $R$
Resistance and Power Distribution Problem

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

Source: Cadence

Issue #2: L di/dt

Impact of inductance on supply voltage

- Change in current induces the change in voltage
- Longer supply lines have larger \( L \)
L di/dt: Example

- (12.3.3 W&H) A 1GHz chip transitions from idle (20 A) to full power (60 A) operation in a single cycle. If the power supply has 20 pH of series inductance, estimate the power supply noise caused by this transition if the chip has no internal bypass capacitance.

Design Techniques to Address L di/dt

- Separate power pins for I/O pads and chip core
- Multiple power and ground pins
- Position of power and ground pins on package
- Increase $t_r$ and $t_f$
- Advanced packaging technologies
- Decoupling capacitances on chip and on board
Bypass Capacitors

- Need low supply impedance at all frequencies
- Ideal capacitors have impedance decreasing with $\omega$
- Real capacitors have parasitic $R$ and $L$
  - Leads to resonant frequency of capacitor

Frequency Response

- Use multiple capacitors in parallel
  - Large capacitor near regulator has low impedance at low frequencies
  - But also has a low self-resonant frequency
  - Small capacitors near chip and on chip have low impedance at high frequencies
- Choose caps to get low impedance at all frequencies
Decoupling Capacitors

Decoupling capacitors are added:
- On the board (right under the supply pins)
- On the chip (under the supply straps, near large buffers)

On-chip Decoupling Capacitance

- Static CMOS logic dynamically switches (no dc current).
  - Supply just need to provide the average current.
  - Peak current needs to come from nearby capacitance.
- Basically the same as charge sharing
  - Use $C_{\text{decoup}} > 10 \times C_{\text{switched}}$ to guarantee $<10\%$ $V_{\text{DD}}$ drop.
  - Put capacitance near load with little resistance.
    - Disseminate the capacitance throughout the standard cell.
    - The more the merrier.
    - Part of the P&R tool.
Bypass Capacitances in Real Life

- Left: package bypass; Right: PCB bypass

Courtesy: M. Horowitz, K. Mai

Power Distribution Network

- AC/DC converter
  - Usually 110VAC to 12 or 5VDC in desktop PCs
- Voltage Regulator Module
  - Converts one DC level to another (5V to 1.2V)
- Printed circuit board
  - Planes send current from VRM to the package
  - Planes have capacitance for bypass; use discretes too
- Package
  - Deliver current to the chip itself using balls or bonds
  - Can use bypass caps on the package as well
- Chip power grid
  - Use device bypass capacitors

Courtesy: M. Horowitz
Power System: Lumped Model

- Power comes from regulator on system board
  - Board and package add parasitic R and L
  - Bypass capacitors help stabilize supply voltage
  - But capacitors also have parasitic R and L
- Simulate system for time and frequency responses

---

On-chip Power Network: Distributed Model

[Diagram of on-chip power network]

- Base grid
- Core grid
- Global grid

*Courtesy: A. Todri*
**Power Distribution Strategy**

- 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
Top-Level Power and Gnd Routing

- Usually on the top layer of metal, and then distributed to the lower levels

Power and Ground Need Wider Wires

- The power supply needs to be distributed to all the cells in the circuit.
  - A tree
    - Trunk of the tree must supply current to all branches.
    - Resistance in these lines must be very small, since when a gate switches, its current flows through the supply lines.
      - If the resistance of the supply lines is too large, the voltage supplied to gates will drop, which can cause the gate to malfunction.
      - Usually you don’t want the supply to change more than 5-10% due to supply resistance.
- So power supply must be on the metal layer
- Is that enough? Usually, they have to be wider too.
  - $R_{\text{trans}}$ is much greater (by $10^5$) than $R_{\text{metal}}$
  - But one builds wide devices, and long wires
  - And in a chip there are many devices connected in parallel to the supplies. So you need still need to be careful even with metal layers, and make the special wires wide enough.
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

Metal 5
Metal 4
Metal 3
RP2/Vdd
RP1/Vss
Metal 2
Metal 1

Courtesy: Compaq

Decoupling Capacitor Ratios

- **EV4, 0.75um, 200MHz**
  - total effective switching capacitance = 12.5nF
  - 128nF of de-coupling capacitance
  - de-coupling/pitching capacitance \( \sim 10 \times \)

- **EV5, 0.5um, 350MHz**
  - 13.9nF of switching capacitance
  - 160nF of de-coupling capacitance

- **EV6, 0.35um, 575MHz**
  - 34nF of effective switching capacitance
  - 320nF of de-coupling capacitance -- not enough!

Source: B. Herrick (Compaq)
EV6 De-coupling Capacitance

Example:

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

- 0.32-$\mu$F of on-chip de-coupling capacitance was added
  * Under major busses and around major gridded clock drivers
  * Occupies 15-20% of die area

- 1-$\mu$F 2-cm² Wirebond Attached Chip Capacitor (WACC) significantly increases "Near-Chip" de-coupling
  * 160 Vdd/Vss bondwire pairs on the WACC minimize inductance

Source: B. Herrick (Compaq)

Electromigration (1)

Limits dc-current to 1 mA/$\mu$m
Electromigration (2)

Check for metal migration at worst power corner
Do the following checks:

<table>
<thead>
<tr>
<th>PROCESS</th>
<th>TEMP</th>
<th>VOLTAGE</th>
<th>TESTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast-n/fast-p</td>
<td>0°C</td>
<td>5.5V (3.6V)</td>
<td>Power dissipation (DC), clock races, hold time constraints</td>
</tr>
<tr>
<td>Slow-n/slow-p</td>
<td>125°C</td>
<td>4.5V (3.0V)</td>
<td>Circuit speed, setup time constraints</td>
</tr>
<tr>
<td>Slow-n/fast-p</td>
<td>0°C</td>
<td>5.5V (3.6V)</td>
<td>Pseudo-nMOS noise margin, level shifters, memory write/read, ratioed circuits</td>
</tr>
<tr>
<td>Fast-n/slow-p</td>
<td>0°C</td>
<td>5.5V (3.6V)</td>
<td>Memories, ratioed circuits, level shifters</td>
</tr>
</tbody>
</table>
Power Supply Rules

- The exact rules depend on the technology
  - Each technology file should have its rules for resistance and electromigration

- Example Rules:
  - Must have a contact for each 16\(\lambda\) of transistor width (more is better)
  - Wire must have less than 1mA/\(\mu\)m of width
  - Power/Gnd width = Length of wire \(\times\) Sum (all transistors connected to wire) / 3\(\times\)10\(^6\). (very approximate)

- For small designs, power supply design is less of an issue
  - Total power is small
  - Chip is small, so wires are short
    - Will not be an issue in this class

Working with Scaled V\(_{DD}\)
**Simple On-chip Level Converter**

![Diagram of a simple on-chip level converter](image)

**Example: $V_{DDL} = 0.4V$**

![Diagram with example values](image)
Layout (Separate N-wells)

PR boundary: 7um x 3.9um (std height)

Post-layout Simulation Results

<table>
<thead>
<tr>
<th>VddL (V)</th>
<th>tp:0-1 (ps)</th>
<th>tp:1-0 (ps)</th>
<th>tp:avg (ps)</th>
<th>tp:avg (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>1725</td>
<td>1750</td>
<td>1738</td>
<td>1889</td>
</tr>
<tr>
<td>0.5</td>
<td>435</td>
<td>880</td>
<td>658</td>
<td>792</td>
</tr>
<tr>
<td>0.6</td>
<td>214</td>
<td>754</td>
<td>484</td>
<td>573</td>
</tr>
<tr>
<td>0.7</td>
<td>145</td>
<td>717</td>
<td>431</td>
<td>501</td>
</tr>
<tr>
<td>0.8</td>
<td>115</td>
<td>701</td>
<td>408</td>
<td>470</td>
</tr>
<tr>
<td>0.9</td>
<td>98</td>
<td>695</td>
<td>396</td>
<td>452</td>
</tr>
<tr>
<td>1.0</td>
<td>86</td>
<td>693</td>
<td>389</td>
<td>442</td>
</tr>
</tbody>
</table>

Note: final implementation may have additional buffers at the output to increase drive strength before the pads

- I/O pads: see W&H, 12.4.1-3
Next Week

- Monday: Interconnect, Memory
- Wed (pre-recorded lecture): Design for Test

Happy Thanksgiving!