Why is Power Important?

• My top 5 reasons
  1. How many Hoover dams does it take?
  2. The true (economic) cost of power
  3. Anytime, Anywhere (or more for less)
  4. The end of (classical) scaling
  5. Power impacts correct IC operation
How many Hoover dams?
Then ...

- Name: ENIAC
  - First all-electronic, Turing-complete computer
- Built: Feb. 1946
- Speed: 5000 additions/sec
- Space: 1800 sq. ft.
- Weight: ~30 tons

Power Consumed: 174,000 Watts

Avg. power consumption of a US home: 1020W
and Now….

- **Name:** Earth Simulator
  - World’s fastest supercomputer (June 2002 – Nov 2004)
- **Built:** Feb 2002
- **Speed:** $35.61 \times 10^{12}$ Flops
- **Space:** ~35000 sq. ft.

Power consumed: 
~$13,000,000$ Watts
Earth Simulator Facilities

- Users
- Computer
- Cooling

Earth Simulator facilities include a research building, a simulator building, and a cooling system. The facility is equipped with large-scale computing resources and cooling infrastructure to support high-performance simulations.
What if you design for power?

- **Name:** MP211
  - First multi-core processor for mobile phones
- **Built:** Sep 2004
- **Speed:** > $10^9$ Inst/sec
- **Size:** $\sim 1$ sq. cm.

**Power:** 0.12-0.25 Watts (active), 0.00005 Watts (idle)
Societal Impact

- **Environmental burden**
  - Total power consumption of CPUs in world’s computers
    - 2001: 9 Billion Watts (> 4 Hoover Dams!)
    - 2005: ~20 Hoover Dams
  - Servers and associated infrastructure
    - 123 Billion kWh worldwide in 2005 = over 170 Billion tons of CO₂
      (over 250 Billion kWh by 2010)
    - Fluorescent lighting: 200 Billion kWh
    - 2010: Data centers use 1.3 percent of worldwide electricity

http://www.greenpeace.org/
Green is IN!

Earth Day – April 22nd, 2012

New research funding agency (ARPA-E) formed - $400 Million For Off the Wall Energy Ideas
The True (Economic) Cost of Power
The True Cost of Power

- **Product Cost**
  - Packaging / cooling has increased impact on IC and system cost

- **Operation Cost**
  - $180/yr electricity bill for an always-on PC
  - $4-$8/Watt of load for data centers

- **Replacement Cost**
  - Mean Time to Failure is halved for every 10 deg. C increase in temperature
Anytime, Anywhere (More for Less)

“By having the phone with you, you should not need anything else but your clothes!”

Takashi Natsuno
Managing Director
i-mode strategy
NTT DoCoMo

[Technology Review, July ’04]

i-mode: Japan’s most successful wireless internet service (46 M subscribers)
Battery-powered Mobile Appliances

- Large portion of the electronics and semiconductor industries!
  - Largest by volume
    - Wireless handset sales: 990 million units in 2006
  - Significant by revenue
    - Semiconductors for handsets: $33 Billion in 2007
- Battery life is a primary concern
Evolution of Mobile Applications & Services

Service complexity

- Phone
- Voice

+ Organizer + IM

IM
SMS
Calendar

... + Internet + multimedia

Email
Mobile music
Digital TV
Mobile games
MMS
Location-based services
Ringtones, ringback
Browser

... + wallet/purse + sensor

Web services
Traffic surveillance
Smart navigation
Service monitoring
Billing & Payment
E-wallet
Image recognition
Image search
Mobile commerce
The Battery Gap

- Diverging gap between actual battery capacities and energy needs (e.g., fuel cells) may help, but not enough!

**Energy Required (1 day) = 33,000J**

**Req. Battery Capacity ~ 2500 mAh**

**Available Battery Capacity ~1000 mAh**
The End of Classical Scaling
Historical Trends: IC Scaling

No exponential is forever, but you can delay forever…
–Gordon Moore
The End of (Classical) Scaling

- **Classical scaling**: All device dimensions and Voltage are scaled equally
  - Ended due to unacceptable gate leakage, reliability, ...
  - Causes dramatic rise in power density with scaling

Source: Bernard Meyerson, “How Does One Define "Technology" Now That Classical Scaling Is Dead (and Has Been for Years)?”, DAC 2005 Keynote
• This clearly unsustainable trend led to a transformation in the computing industry that we now call “multi-core”
The End of (Classical) Scaling – Power Cliffs

• The switch from Bipolar to CMOS occurred when power became a showstopper
  – No viable alternative in sight this time!

Source: Bernard Meyerson, “How Does One Define "Technology" Now That Classical Scaling Is Dead (and Has Been for Years)?”, DAC 2005 Keynote
Power Affects Design Complexity
The Power Delivery Challenge

Increasing “battery gap” for portable products

Power grid design a critical step (routing area vs. electromigration tradeoff)

Packaging / cooling increased impact on cost
Speed-power Inter-dependence

Timing-power inter-dependence
- Power $\Rightarrow$ I-R drop $\Rightarrow$ Reduced Vdd for embedded cell $\Rightarrow$ Increased delay
- Power $\Rightarrow$ higher temperature $\Rightarrow$ lower speed

Example: 595K gate design, 0.13um, Vdd=1.3V, Max I-R = 264mV
Our focus in 595Z

• How can we automatically estimate power consumption at the logic level?
• Synthesis (automatic) techniques to reduce power.
Outline

• Power dissipation in CMOS circuits
• Estimating power consumption
  – Power models for gates
  – Estimating value probabilities and switching activities
CMOS Power Consumption Basics

\[ \text{Power} = \text{Capacitive Switching Power} + \text{Short Circuit Power} + \text{Leakage Power} \]

- Dynamic power dominated in > 90nm technologies
- Leakage has become an equal culprit in nanoscale CMOS

Source: Shekar Borkar, “Microarchitecture and Design Challenges for Gigascale Integration”, MICRO 2004
Capacitive Switching Power

- Power required to charge and discharge capacitances

\[
\text{Energy required to charge load capacitance} = Q * V_{dd} = C_L * V_{dd} * V_{dd}
\]

\[
\text{Power} = \text{Energy/transition} * \text{transition rate} = C_L * V_{dd}^2 * f_{0\rightarrow1}
\]

\[
= C_L * V_{dd}^2 * f * P_{0\rightarrow1}
\]
Capacitive Switching Power

- Over time, at any signal in the circuit, 0 → 1 and 1 → 0 transitions are equi-probable

\[ \text{Cap. Switching Power} = \frac{1}{2} \cdot C_L \cdot V_{dd}^2 \cdot N \cdot f \]

- \( C_L \): Load Capacitance
- \( V_{dd} \): Supply Voltage
- \( N \): Switching Activity (per. clock cycle)
- \( f \): Clock Frequency

\[ C_L \cdot N \cdot f \]: Switched Capacitance
Leakage Power

- Three major components
  - Sub-threshold leakage
  - Gate leakage
  - Reverse-biased junction (Band-to-Band tunneling)
Power Models for Gates

• Dynamic Power
  – Unit capacitance
  – Lumped output capacitance
  – Pin-based model
  – Transition-based model
  – State-dependent model

• Leakage Power
  – Constant
  – Input-dependent (lookup table based on gate input values)
Dynamic Power Models for Gates

- Unit capacitance (simplest model)
  - Technology-independent
  - Power $\propto \sum_{all \ g_i} N(g_i)$
  - Extension: add per-fanout capacitance

- Lumped output capacitance
  - Specified in technology library

- Pin-based model
  - Associate a capacitance with each pin of the cell / gate

Increasing accuracy
Dynamic Power Models for Gates

• Transition-based model
  – Specify the power consumption for each possible transition at gate inputs
  – Exponential in number of gate inputs
  – Assumes that transitions on different inputs are either simultaneous or separated enough in time that they do not interact

<table>
<thead>
<tr>
<th>Transition</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>00→01</td>
<td></td>
</tr>
<tr>
<td>00→10</td>
<td></td>
</tr>
<tr>
<td>00→11</td>
<td></td>
</tr>
<tr>
<td>01→00</td>
<td></td>
</tr>
<tr>
<td>01→10</td>
<td></td>
</tr>
<tr>
<td>01→11</td>
<td></td>
</tr>
<tr>
<td>10→00</td>
<td></td>
</tr>
<tr>
<td>10→01</td>
<td></td>
</tr>
<tr>
<td>10→11</td>
<td></td>
</tr>
<tr>
<td>11→00</td>
<td></td>
</tr>
<tr>
<td>11→01</td>
<td></td>
</tr>
<tr>
<td>11→10</td>
<td></td>
</tr>
</tbody>
</table>
Dynamic Power Models for Gates

• State-dependent model
  – Accurately accounts for effect of history on internal signals within a gate
  – Requires a very accurate simulator (e.g., SPICE) to build this model
Leakage Power Models

- **Constant**
  - Depends on gate type

- **Input-dependent**
  - Look-up table based on input values
  - Considers stacking effect and loading effect
Power Estimation: What Else Do We Need to Know?

- Value probabilities and switching activities at signals inside the circuit

\[ P = \sum_{\text{all gates } g_i} P_{\text{dynamic}}(g_i) + P_{\text{leakage}}(g_i) \]

\[ P_{\text{dynamic}}(g_i) = f\left( N_{\text{transition}}(x_i), N_{\text{transition}}(y_i), N_{\text{transition}}(z_i) \right) \]

\[ P_{\text{leakage}}(g_i) = h\left( P(x_i, y_i = 00), P(x_i, y_i = 01), P(x_i, y_i = 10), P(x_i, y_i = 11) \right) \]

Note: Assume a pin-based delay model for dynamic power and an input-dependent leakage model.
Switching Activity Under Zero Delay

• Assuming zero gate delays, each gate output can switch at most once in each clock cycle
• Whether there is a transition or not depends on current and previous input vectors
Switching Activity under General Gate Delays

- Output of each gate may switch multiple times for a single input vector pair (glitching)
Glitching Example

• Data path circuits often dissipate a lot of dynamic power due to glitching

Example: 16-bit ripple carry adder