Sad fact: Computers turn electrical energy into heat. Computation is a byproduct.

Energy and Performance

Air or water carries heat away, or chip melts.
The Joule: Unit of energy. Can also be expressed as Watt-Seconds. Burning 1 Watt for 100 seconds uses 100 Watt-Seconds of energy.

This is how electric tea pots work...

1 Joule heats 1 gram of water 0.24 degree C

The Watt: Unit of power. The amount of energy burned in the resistor in 1 second.

20 W rating: Maximum power the package is able to transfer to the air. Exceed rating and resistor burns.

The Joule of Heat Energy per Second

Cooling an iPod nano ...

Like resistor on last slide, iPod relies on passive transfer of heat from case to the air.

Why? Users don’t want fans in their pocket ...

To stay “cool to the touch” via passive cooling, power budget of 5 W.

If iPod nano used 5W all the time, its battery would last 15 minutes ...
Powering an iPod nano (2005 edition)

1.2 W-hour battery:
Can supply 1.2 watts of power for 1 hour.

$$\frac{1.2 \text{ W}}{5 \text{ W}} = 15 \text{ minutes}.$$ 

More W-hours require bigger battery and thus bigger “form factor” -- it wouldn’t be “nano” anymore :-).

Real specs for iPod nano:
14 hours for music,
4 hours for slide shows.

- 85 mW for music.
- 300 mW for slides.

Finding the (2005) iPod nano CPU ...

A close relative ...

Two 80 MHz CPUs.
One CPU used for audio, one for slides.

Low-power ARM roughly 1 mW per MHz ... variable clock, sleep modes.

85 mW system power realistic ...
Year-to-year: continuous improvements

iPod nano 2005
14 hours battery life (audio playback)

iPod nano 2006
24 hours battery life (audio playback)

What changed inside?

Source: ifixit.com
iPod nano 2005 - a C-shaped PC board, with a battery in the "C" opening.

iPod nano 2006 - battery lies on top of PC board.
How? Small IC packages, fewer parts

iPod nano 2006

iPod nano 2005

Source: arstechnica.com

Aluminum permits thinner case ...

What's happened since 2006?

Source: ilounge.com
2010 Nano:
“up to” 24 hours audio playback

2010 Shuffle:
“up to” 15 hours audio playback

0.74 ounces

0.44 ounces

0.39 W Hr
(33% of 2005 Nano)

0.19 W Hr

Desired screen size sets smartphone W x L

Depth? : Thin body vs. Battery life
22% gain in battery energy over 5 iterations

<table>
<thead>
<tr>
<th>Year</th>
<th>Device</th>
<th>Battery</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>iPhone</td>
<td>Li-ion Polymer, 3.7V, 1170mmAh</td>
</tr>
<tr>
<td>2008</td>
<td>iPhone 3G</td>
<td>Li-ion Polymer, 3.7V, 1150mmAh</td>
</tr>
<tr>
<td>2009</td>
<td>iPhone 3GS</td>
<td>Li-ion Polymer, 3.7V, 1220mmAh</td>
</tr>
<tr>
<td>2010</td>
<td>iPhone 4</td>
<td>Li-ion Polymer, 3.7V, 1420mmAh</td>
</tr>
<tr>
<td>2011</td>
<td>iPhone 4 (CDMA)</td>
<td>Li-ion Polymer, 3.7V, 1430mmAh</td>
</tr>
<tr>
<td>Today</td>
<td>iPhone 4S</td>
<td>Li-ion Polymer, 3.7V, 1430mmAh</td>
</tr>
</tbody>
</table>

iPhone (2007)

- Mainboard
- Antennas
- Battery
iPhone 4S

- Battery
- L-shape Main Board
- Metal frame acts as antenna

In 4 years:
- 6.8x increase in transistor count
- 33% max clock speed increase
- Attached DRAM: 128 MB -> 512 MB
- 6.8x transistors: Dual CPU and GPU, and to save energy.
Notebooks ... as designed in 2006 ...

2006 Apple MacBook -- 5.2 lbs

Performance: Must be “close enough” to desktop performance ... most people no longer used a desktop (even in 2006).

Size and Weight. Ideal: paper notebook.

Heat: No longer “laptops” -- top may get “warm”, bottom “hot”. Quiet fans OK.

Battery: Set by size and weight limits ...

Battery rating: 55 W-hour.

At 2.3 GHz, Intel Core Duo CPU consumes 31 W running a heavy load - under 2 hours battery life! And, just for CPU!

46x more energy than iPod nano battery. And iPod lets you listen to music for 14 hours!

Almost full 1 inch depth. Width and height set by available space, weight.

At 1 GHz, CPU consumes 13 Watts. “Energy saver” option uses this mode ...
55 W-hour battery stores the energy of 1/2 a stick of dynamite.

If battery short-circuits, catastrophe is possible ...
MacBook Air ... design the laptop like an iPod

2011 Air: 11.8 in x 7.56 in x 0.68 in; 2.38 lbs

2006 Macbook: 12.8 in x 8.9 in x 1 in; 5.2 lbs
Mainboard: fills about 25% of the laptop

35 W-h battery: 63% of 2006 MacBook’s 55 W-h
2011 Air: 35 W-h battery, 5 hour battery life*
iPad 2: 25 W-h battery, 10 hour battery life*

*For a content-consumption workload.

Battery-Life-Hour/W-h: 2.8x iPad advantage

“Content Creation vs. Content Consumption”

2011 Air: $999 -- 64 GB SSD, 2 GB RAM, x86
iPad 2: $699 -- 64 GB SSD, 512 MB RAM, ARM
iPhone 4S and iPad 2: Identical CPU/RAM stack
The CPU is only part of power budget!

2004-era notebook running a full workload.

"Amdahl’s Law for Power"

If our CPU took no power at all to run, that would only double battery life!
Servers: Total Cost of Ownership (TCO)

Machine rooms are expensive. Removing heat dictates how many servers to put in a machine room.

Electric bill adds up! Powering the servers + powering the air conditioners is a big part of TCO.

Reliability: running computers hot makes them fail more often.

Computations per W-h doubles every 1.6 years, going back to the first computer.

(Jonathan Koomey, Stanford).
Processors and Energy

Device engineers trade speed and power

- We can reduce $CV^2(P_{\text{active}})$ by lowering $V_{dd}$.
- We can increase speed by raising $V_{dd}$ and lowering $V_t$.
- We can reduce leakage ($P_{\text{standby}}$) by raising $V_t$.

From: Silicon Device Scaling to the Sub-10-nm Regime
Meikei Ieong, Bruce Doris, Jakub Kedzierski, Ken Rim, Min Yang
Customize processes for product types...


Five low-power design techniques

- Parallelism and pipelining
- Power-down idle transistors
- Slow down non-critical paths
- Clock gating
- Thermal management
Trading Hardware for Power

via Parallelism and Pipelining ...

Gate delay roughly linear with Vdd

And so, we can transform this:

Block processes stereo audio. 1/2 of clocks for “left”, 1/2 for “right”.

Into this:

CV^2 power only

Ex: Top block processes audio channel 1, bottom block processes audio channel 2.

THIS MAGIC TRICK BROUGHT TO YOU BY CORY HALL ...
### Multiple Cores for Low Power

Trade hardware for power, on a large scale ...
Cell: The PS3 chip

Cell (PS3 Chip): 1 CPU + 8 “SPUs”

- L2 Cache: 512 KB
- 8 Synergistic Processing Units (SPUs)
- PowerPC
One Synergistic Processing Unit (SPU)

SPU issues 2 inst/cycle (in order) to 7 execution units
256 KB Local Store, 128 128-bit Registers
SPU fills Local Store using DMA to DRAM and network

A “Schmoo” plot for a Cell SPU ...

The lower Vdd, the less dynamic energy consumption.

\[ E_{0\rightarrow1} = \frac{1}{2} C V_{dd}^2 \]

The lower Vdd, the longer the maximum clock period, the slower the clock frequency.

\[ E_{1\rightarrow0} = \frac{1}{2} C V_{dd}^2 \]
Clock speed alone doesn’t help E/op ...

But, lowering clock frequency while keeping voltage constant spreads the same amount of work over a longer time, so chip stays cooler ...

\[ E_{0\to1} = \frac{1}{2} CV_d^2 \]

scaling V and f does lower energy/op

1 W to get 2.2 GHz performance, 26 C die temp. 7W to reliably get 4.4 GHz performance, 47C die temp.

If a program that needs a 4.4 Ghz CPU can be recoded to use two 2.2 Ghz CPUs ... big win.
How iPod nano 2005 puts its 2 cores to use...

Two 80 MHz CPUs. Was used in several nano generations, with one CPU doing audio decoding, the other doing photos, etc.

Design Technique #2 (of 5)

Powering down idle circuits
Add “sleep” transistors to logic ...

- Example: Floating point unit logic.
- When running fixed-point instructions, put logic “to sleep”.
  
  +++ When “asleep”, leakage power is dramatically reduced.

--- Presence of sleep transistors slows down the clock rate when the logic block is in use.

Intel example: Sleeping cache blocks

>3x SRAM leakage reduction on inactive blocks

Slow down “slack paths”

Fact: Most logic on a chip is “too fast”

The critical path
Most wires have hundreds of picoseconds to spare.

From “The circuit and physical design of the POWER4 microprocessor”, IBM J
Use several supply voltages on a chip ...

Why use multi-Vdd? We can reduce dynamic power by using low-power Vdd for logic off the critical path.

What if we can’t do a multi-Vdd design?
In a multi-Vt process, we can reduce leakage power on the slow logic by using high-Vth transistors.


LOW POWER ARM 1136JF-S™ DESIGN

George Kuo, Anand Iyer
Cadence Design Systems, Inc.
San Jose, CA 95134, USA

Logical partition into 0.8V and 1.0V nets done manually to meet 350 MHz spec (90nm).

Level-shifter insertion and placement done automatically.

Dynamic power in 0.8V section cut 50% below baseline.

Leakage power in 1.0V section cut 70% below baseline.

From a chapter from new book on ASIC design by Chinnery and Keutzer (UCB).
Gating clocks to save power

On a CPU, where does the power go?

Half of the power goes to latches (Flip-Flops).

Most of the time, the latches don’t change state.

So (gasp) gated clocks are a big win.
But, done with CAD tools in a disciplined way.

From: Bose, Martonosi, Brooks: Sigmetrics-2001 Tuto
Synopsis Power Compiler can do this ...

"Up to 70% power savings at the block level, for applicable circuits"

Synopsis Data Sheet

Design Technique #5 (of 5)

Thermal Management
Keep chip cool to minimize leakage power

A recipe for thermal runaway

Figure 3: \(I_{CCINTQ} \) vs. Junction Temperature with Increase Relative to 25°C

IBM Power 4: How does die heat up?

4 dies on a multi-chip module

2 CPUs per die
115 Watts: Concentrated in “hot spots”

- Fixed point units
- Cache logic

66.8 °C == 152 °F  
82 °C == 179.6 °F

Idea: Monitor temperature, servo clock speed