Facing the Hot Chip Challenge (Again)

Bill Holt
General Manager
Technology and Manufacturing Group
Intel Corporation

Hot Chips 17
August, 2005
Key Messages

• The driving force behind Moore’s Law

• Power has always been a consideration

• Process and design collaboration required to address power challenges

• Technology Advances provide a transistor budget to support innovation

• Efficient Design utilizes transistor budget to deliver product performance
“Reduced cost is one of the big attractions of integrated electronics, and the cost advantage continues to increase as the technology evolves toward the production of larger and larger circuit functions on a single semiconductor substrate.”

Electronics, Volume 38, Number 8, April 19, 1965

Source: Intel
Moore’s Law - 2005

Transistors Per Die

10^10
10^9
10^8
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0


Source: Intel
As the number of transistors goes UP

Cost per transistor goes DOWN
Scaling: The Fundamental Cost Driver

- Twice the circuitry in the same space (architectural innovation)
- The same circuitry in half the space (cost reduction)
- Half the die size for the same capability than in the prior process

Die sizes:
- 350nm 200mm
- 250nm 200mm
- 180nm 200mm
- 130nm 200mm
- 90nm 300mm
- 65nm 300mm

Dual Core
Wafer Size: Enables Cost Efficiency

Year That Industry Exceeds 3 Million wafers/year
Processed Wafer Cost

Wafer size conversions offset trend of increasing wafer processing cost

Source: Intel
Moore’s Law + Bigger Wafers = Lower Cost/function

100 nanodollars per transistor
(1 cent = 100,000 transistors)

Source: WSTS/Intel, 5/05
Key Messages

- The driving force behind Moore’s Law
- Power has always been a consideration
- Process and design collaboration required to address power challenges
- Technology Advances provide a transistor budget to support innovation
- Efficient Design utilizes transistor budget to deliver product performance
Power challenges are neither new nor fundamental

“Will it be possible to remove the heat generated by 10’s of thousands of components?”

G. Moore, *Cramming more components onto integrated circuits*, Electronics, Volume 38, Number 8, April 19, 1965
“The **power barriers** now facing alternative semiconductor processes indicate that only CMOS will allow chip makers to capitalize on the density that can be achieved with gate arrays and standard cells.”

“Once, maybe twice a decade the electronics industry encounters a force that affects not only the way circuits are physically designed but also the way the industry thinks. **CMOS is just such a force.**”

*Source: Electronic Design, October 1984*
Silicon Technology has Changed to Increase Power Efficiency

1960’s: Bipolar
1970’s: PMOS, NMOS
1980’s: CMOS
1990’s: Voltage scaling (P = CV^2f)
2000’s: Power efficient scaling/design
Moore’s Law Will Outlive CMOS

Through Innovation


1965 Data (Moore)

10^0 10^1 10^2 10^3 10^4 10^5 10^6 10^7 10^8 10^9 10^10 10^11 10^12 10^13

Transistors/Die

Bipolar PMOS NMOS CMOS Voltage Scaling Pwr Eff Scaling New Nanostructures

Beyond CMOS? Spin based? Molecular? Other?

Source: Intel
Key Messages

- The driving force behind Moore’s Law
- Power has always been a consideration
- Process and design collaboration required to address power challenges
- Technology Advances provide a transistor budget to support innovation
- Efficient Design utilizes transistor budget to deliver product performance
Process Advances Still Scale Power

but the rate has slowed and collaboration is required
Leakage becomes Significant

Power scaling vs. process for the last 10 years
(includes frequency increasing with process speed)
However, Power Density Has Leveled Off

Source: Intel
Effective Process and Design Collaboration Succeeds in Power Improvements

- 130nm
- Next generation product if 130nm
- 90nm process only
- Process plus design efficiency
- Process, design efficiency, and frequency increase

Relative Power
Key Messages

• The driving force behind Moore’s Law
• Power has always been a consideration
• Process and design collaboration required to address power challenges
• Technology Advances provide a transistor budget to support innovation
• Efficient Design utilizes transistor budget to deliver product performance
Silicon Technology Advances

Feature Size Scaling

- 0.7x every 3 years
- 0.7x every 2 years

New technology generation every 2 years
Transistor gate length ~60% of other minimum features
Key Density Indicator Continues to Scale

Gate pitch continues to scale 0.7x per generation, providing ~2x transistor density improvements
Strained Silicon Improves Transistor Performance and Leakage Today

Transistor Drive Current (mA/μm)

Transistor Leakage Current (nA/μm)

PMOS
NMOS

0.04x $I_{OFF}$
0.20x $I_{OFF}$

+25% $I_{ON}$
+10% $I_{ON}$

Source: Intel
High-k Dielectric Can Reduce Gate Leakage Tomorrow

<table>
<thead>
<tr>
<th></th>
<th>High-k vs. SiO₂</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gate capacitance</td>
<td>60% greater</td>
<td>Faster transistors</td>
</tr>
<tr>
<td>Gate dielectric leakage</td>
<td>&gt; 100x reduction</td>
<td>Lower power</td>
</tr>
</tbody>
</table>

Process integration is the key challenge

Source: Intel
Transistors Require Optimization to the Application

Performance vs. Leakage

Optimized transistors can provide ~1000x lower leakage
65 nm Generation Transistors Today

- 35 nm gate length
- 1.2 nm gate oxide
- 220 nm gate pitch
- NiSi for low resistance
- 2\textsuperscript{ND} generation strained silicon for enhanced performance/power
Innovations Required to Reduce Interconnect RC Challenges

- Clock Period
- RC delay of 1mm interconnect
- Low K
- Copper Interconnect

Source: IDF
3D Silicon Stacking

Wafer

Die

Top
Thin
Wafer

Bottom
Wafer

Logic, Memory, or ?

Logic

Source: Intel
65 nm Generation Interconnects

- Cu Line
- Cu Via
- CDO
- Low-k
- Dielectric

M8
M7
M6
M5
M4
M3
M2
M1
Key Messages

• The driving force behind Moore’s Law

• Power has always been a consideration

• Process and design collaboration required to address power challenges

• Technology Advances provide a transistor budget to support innovation

• Efficient Design utilizes transistor budget to deliver product performance
Power Reduction Techniques

- Optimum Design
- Leakage Control
- Active Power Reduction
- Increase On-die Memory
- Multi-threading
- Dual Core and Multi-core
- Special Purpose Hardware
- Function Integration through SOC/SIP
Circuit Techniques Reduce Source Drain Leakage

Body Bias

\[ V_{dd} \quad V_{bp} \quad \text{+ Ve} \quad \text{- Ve} \quad V_{bn} \]

Leakage Reduction 2 - 10X

Stack Effect

Equal Loading

Sleep Transistor

\[ \text{Logic Block} \]

Leakage Reduction 5 - 10X

2 - 1000X
Sleep Transistor Reduces SRAM Leakage Power

>3x SRAM leakage reduction on inactive blocks

Source: Intel
Active Power Reduction

Slow $\rightarrow$ Fast $\rightarrow$ Slow

Low Supply Voltage $\rightarrow$ High Supply Voltage

Multiple Supply Voltages

Replicated Designs

- **Logic Block**
  - Freq = 1
  - Vdd = 1
  - Throughput = 1
  - Power = 1
  - Area = 1
  - Pwr Den = 1

- **Logic Block**
  - Freq = 0.5
  - Vdd = 0.5
  - Throughput = 1
  - Power = 0.25
  - Area = 2
  - Pwr Den = 0.125

- **Logic Block**
  - Vdd/2
Dual Core

Rule of thumb

<table>
<thead>
<tr>
<th>Voltage</th>
<th>Frequency</th>
<th>Power</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>1%</td>
<td>3%</td>
<td>0.66%</td>
</tr>
</tbody>
</table>

In the same process technology...

```
Voltage = 1
Freq    = 1
Area    = 1
Power   = 1
Perf    = 1
```

```
Voltage = -15%
Freq    = -15%
Area    = 2
Power   = 1
Perf    = ~1.8
```
Special Purpose Hardware

Special Purpose HW Engine

2.23 mm x 3.54 mm, 260K transistors

Opportunities:
MPEG Encode/Decode
Speech recognition
Graphics

Special purpose HW—Best Mips/Watt
Value of Integration

- Special-purpose hardware → more MIPS/mm²
- SIMD integer and FP instructions in several ISAs

<table>
<thead>
<tr>
<th></th>
<th>Die Area</th>
<th>Power</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Purpose</td>
<td>2X</td>
<td>2X</td>
<td>~1.4X</td>
</tr>
<tr>
<td>Multimedia Kernels</td>
<td>&lt;10%</td>
<td>&lt;10%</td>
<td>1.5 - 4X</td>
</tr>
</tbody>
</table>

Si Monolithic

- Special HW
- CMOS RF
- CPU
- Memory

Polyolithic

- Wireline
- Opto-Electronics
- RF
- Dense Memory

Heterogeneous Si, SiGe, GaAs
## Joint Power Reduction Technology Roadmap

<table>
<thead>
<tr>
<th>Process Technology</th>
<th>Dual Vt Copper</th>
<th>Dual Vt (Le) Strain engineering Low K ILD</th>
<th>FinFET (Tri-Gate) Metal Gate 3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circuits and Design</td>
<td>Sizing Clock gating</td>
<td>Sleep transistors Stack effect Multiple supply voltages Body Bias</td>
<td></td>
</tr>
<tr>
<td>Micro-Architecture</td>
<td>Shallower pipelines Large caches Multi-threading</td>
<td>Multi-threaded Multi-core Increasing multi-processing Special purpose HW</td>
<td></td>
</tr>
</tbody>
</table>

Future options subject to change
Effective Process and Design Collaboration Succeeds in Power Improvements

- 130nm
- Next generation product if 130nm
- 90nm process only
- Process plus design efficiency
- Process, design efficiency, and frequency increase

Relative Power
... and it works

<table>
<thead>
<tr>
<th></th>
<th>Madison</th>
<th>Montecito</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cores/Threads</td>
<td>1/1</td>
<td>2/4</td>
</tr>
<tr>
<td>Transistors</td>
<td>0.41</td>
<td>1.72 Billion</td>
</tr>
<tr>
<td>L3 Cache</td>
<td>6</td>
<td>24 MByte</td>
</tr>
<tr>
<td>Frequency</td>
<td>1.5</td>
<td>&gt;1.7 GHz</td>
</tr>
<tr>
<td>Relative Performance</td>
<td>1</td>
<td>&gt;1.5x</td>
</tr>
<tr>
<td>Thermal Design Power</td>
<td>130</td>
<td>~100 Watt</td>
</tr>
</tbody>
</table>

Source: Intel
Conclusions

- Economics is driving force behind Moore’s Law
- Power has always been a consideration
- Process and design collaboration required to address power challenges
- Technology Advances provide a transistor budget to support innovation
- Efficient Design utilizes transistor budget to deliver product performance
Thank you