The Intel® Core™ Duo Power and thermal aware processor

Efi Rotem
Mobile Platform Group – Intel® Corp.

Key Contributors
Avi Medelson
Alon Naveh
Mobile Platform Group Intel® Corp.

June, 2006
Moore’s Law

• “Doubling the number of transistors on a die every 2 years”
• Providing compute density

*Gordon Moore, Intel Corporation
Moore’s law for power performance

Power = C * V^2 * F + Leakage

Theoretical scaling of new Process technology:

- Linear Dimensions: Shrinks by 0.7
- Area: Shrinks by 0.5
- Capacitance: Shrinks by 0.7
- Voltage: Scale down by 0.7
- Frequency: Scale up by 1/0.7
- Power: Scale down by 0.5

Half the area
Half the power

There’s additional transistor and power budget for:

- New features
- Architectural extensions
- Performance improvement → 2X transistors ~30%-40% performance

Sustainable Performance improvement at same power consumption
**Recent Reality**

Practical scaling factors:
- Linear Dimensions and active cap. continue to shrink
- Voltage: Roughly the same!!!
- Power = \( C \times V^2 \times F \) Roughly the same!!!
- To get increased frequency significant leakage growth

The Power wall:
- Higher power density for frequency speedup - Harder to cool
- Any architectural additions come at cost on higher power
- Requires real architectural tradeoffs
  - Wide vs. Fast, Power efficient performance etc.
- Requires power efficient architectures

Quickly getting to the power wall
**Power & Density increases**

Think Watts/Cm²

- Complex architecture at faster and smaller technology
- Denser power is harder to cool

---


---

Intel® Core™ Duo power architecture
The Intel® Core™ Microarchitecture Coming in 2006

Examples:
- EPIC* (Itanium®)
- x86
- IXA* (xScale)

Examples:
- P5
- P6
- Intel NetBurst®
- Banias
- Intel® Core™

- Pentium®
- Pentium® Pro
- Pentium® II/III
- Pentium® 4
- Pentium® D
- Xeon®
- Pentium® M
- Core™ Duo

Architected to deliver performance and power efficiency

* IXA – Intel Internet Exchange Architecture
* EPIC – Explicitly Parallel Instruction Computing

Intel® Core™ Duo power architecture
June 2006
It is all about power and energy

- Power
  - Cooling, ergonomics and power delivery

- Active Energy
  - Electricity bills, cooling, form factors

- Idle and low activity energy
  - Battery life and ergonomics
Outline

• Power efficient architecture
• Thermal management
• Average power monument
Outline

• Power efficient architecture
• Thermal management
• Average power monument
Delivered Performance = Frequency * Instructions Per Cycle (IPC)

Goal is higher performance and lower power

Power $\propto C_{\text{dynamic}} \times V^2 \times$ Frequency

- $C_{\text{dynamic}}$ is roughly a product of area and activity (AF)
  “how many bits” * “how much do they toggle”
- Activity Factor is a function of IPC
  Same energy/instruction at a shorter time
Delivered Performance = Frequency * Instructions Per Cycle (IPC)

Power α $C_{dynamic}$ * $V^2$ * Frequency

- $C_{dynamic}$ is roughly a product of area and activity (AF)
  “how many bits” * “how much do they toggle”
- Activity Factor is a function of IPC
  Same energy/instruction at a shorter time
Delivered Performance = Frequency * Instructions Per Cycle (IPC)

\[
\frac{\text{Perf}}{\text{Power}} = \frac{\text{Freq} \times \text{IPC}}{C_{\text{dynamic}} \times \text{AF} \times K \times \text{IPC} \times V^2 \times \text{Freq}}
\]

\[
\frac{\text{Perf}}{\text{Power}} \approx \frac{1}{C_{\text{dynamic}} \times \text{AF} \times K \times V^2}
\]

Grow performance by X

Grow activity by less than X

Power efficiency improvements need to come from architecture
Intel® Core™ Microarchitecture

Block Diagram and Energy efficiency Walkthrough
**In Order Fetch and Decode**

- Wide dynamic execution
  - “Smart” vs. Fast
  - 4 Wide - 33% wider
    - \( P = C_{dyn} \cdot V^2 \cdot F \)
    - Higher F, \( \sim \) same \( C_{dyn} \)
    - Higher power
- Macro and micro uOp fusion
  - Lower activity – Same work
- Decoded loop storage
  - Reduced decode energy
Out of Order Execution

- uOP Fusion
- Packed 128 bit FP execution
  - More data – Same control
- Split busses
  - Only required data toggle
Out of Order Memory

- Memory Disambiguation
  - Less time on the fly
  - Lower energy
- Smart Cache
  - Effectively larger cache
  - "Cold" Transistors
  - Well power managed
Advanced Smart Cache
Dynamic Cache Allocation

Advanced Smart Cache

Independent Cache (today)

Shared Cache adapts to mismatched loads. Independent Cache can thrash heavy app even when other cache is under-utilized
Advanced Smart Cache
Efficient Data Sharing

Advanced Smart Cache

Core1

Core2

L2 Cache

FSB

Chipset MCH

Independent Cache (today)

Core1

Core2

L2 Cache

FSB

Chipset MCH

2X L2 to L1 Bandwidth
Advanced Smart Cache
Efficient Data Sharing

Advanced Smart Cache

Core1 | Core2
--- | ---
L2 Cache

FSB
Chipset MCH

Independent Cache (today)

Core1 | Core2
--- | ---
L2 Cache
FSB
L2 Cache
Chipset MCH

2x L2 to L1 Bandwidth
Advanced Smart Cache
Efficient Data Sharing

Advanced Smart Cache

- Core1
- Core2
- L2 Cache
- FSB
- Chipset MCH

Independent Cache (today)

- Core1
- Core2
- L2 Cache
- L2 Cache
- FSB
- Chipset MCH

2X L2 to L1 Bandwidth
In Order Retirement
Dual / Multi Core

- Efficient MT performance
  - 2X transistors ~ 2X perf.
  - Extracts more TLP
- Limited future scalability
  - On most workloads
  - Serial code, locks, sync...
Even during periods of high performance execution, many parts of the chip can be turned off.

Examples:
- Cache partially utilized
- FE turned off while IQ operating as loop cache
- Non used execution units
Resulting Power/Performance efficiency

- Efficiency is only part of the picture
  - Low performance processors are more efficient
  - Efficiency is delivered by both process technology and architecture
Resulting Power efficiency

• Same CPU can operate in different efficiency states
  - Lower Vcc $\rightarrow$ Higher efficiency
Performance – the other side of the coin

- Deliver higher performance at better efficiency
Outline

• Power efficient architecture
• Thermal management
• Average power monument
Mobile CPU Cooling mechanisms

Actually we transfer heat and not cool

• Transfers heat from hot junction to a cooler ambient

Cooling Challenges:
Junction to ambient heat transfer
Thermodynamic limits
Ergonomics constraints

• Skin temp, Exhaust temp, Fan noise

Power = Performance
Performance is limited by power and cooling
Managing power enables increased performance and improved system
Legacy Thermal Management

• Analog thermal diode measures the die temp near edge
• Control algorithm manages fans and CPU frequency to meet policy
  – Active fan control or Passive for better ergonomics and battery life
• Thermal monitor circuit to keep temp within spec at any time
• Analog diode is not at the hot spot – offset varies with workload
  – Offset increases in multi core CPU
  – Thermal monitor located at hot spot
Enabling Efficient Processor and Platform Thermal Control

**DTS – Digital Thermal Sensor**

- Several thermal sensors cover all possible hot spots
  - Internal A/D converts temperature into digital value
  - Measures the maximum temperature on the die at any given time
- Temperature reporting to S/W by interrupt and register value
- External communication via serial bus - PECI
- Accurately temperature reporting enables advanced thermal control schemes

![Diagram of DTS and PECI connections]

*Intel® Core™ Duo power architecture*
Improved thermal accuracy

- Thermal management efficiency limited by feedback accuracy
  - Need to keep guard band to meet specifications
  - Temperature offset, manufacturing accuracy, A/D accuracy, etc.
- 1°C inaccuracy $\approx 1$W (!)
- Increasingly difficult with higher power CMP processors
Adaptive thermal monitor

• Thermal monitor initiates self thermal control to meet spec.
  – Properly designed system should not accede specification
  – System designers may value other platform goodness
    • Better acoustics, improved ergonomics, longer battery life
  – Improves CPU reliability

• Legacy thermal monitor drop frequency to minimum
  – Possibly noticeable impact

• Adaptive thermal monitor introduced on Intel® Core™ Duo
  – Searches for the right operation point
  – Interrupt notifies control S/W

Enables improved ergonomics and reliability
While maintaining higher performance
Outline

• Power efficient architecture
• Thermal management
• Average power monument
CPU dynamic power

- Typical CPU usage varies over time
  - Bursts of high utilization & Long idle periods

- Mobile CPUs optimize energy consumption
  - High power when \textit{high performance} is needed
  - Low power at low activity
Legacy power management

Performance States → “P-states”

P0 is the highest frequency

• P1, P2... are lower frequency @ Vcc
• Uses the dynamic voltage scaling mechanism

Operating System controlled sleep states → “C-states”

• C0 is active state, C1-4 deeper sleep states
• Increasingly clock and voltage reduction

Each core on Intel® Core™ Duo is controlled separately by to OS

• Internal synchronization
Entering C-states

C0
Active State

C1

C2

C3/4/5

HLT

STPCLK

DPSLP
DPRSTP

Exit

Exit

Exit

increasingly lower power

additional platform savings

Intel® Core™ Duo power architecture

June 2006
Power management boundary conditions

Clock can be turned off independently for each core

Voltage reduction at C4 requires “one voice” voting for full CPU idling
Sleep state hardware synchronization

- Operating system controls each core independently
- Each core enters core C state independently
- Voltage control and package states synchronized between cores
Summary

- First Dual core architecture
- Power efficient architecture
  - High perf. at high power efficiency
  - Energy efficient IPC features
- Dual core thermal mgmt
- Average power mgmt