PWRficient Architecture in Critical Embedded Systems

Breakthrough Performance/Watt Multicore Processor

Peter Bannon
VP of Architecture and Verification

Bus & Boards, 2007
January 15th, 2007
Agenda

- The Power Problem
- PWRficient™ Platform Processors
  - PWRficient 1682M platform processor
    - PA6T core
    - CONEXIUM coherent interconnect architecture
    - ENVOI I/O System
  - Frequency & Power Summary
- Example System Designs
- Low Power Design
- Conclusion
The Power Problem
Energy Efficiency is a Dominant Industry Theme

“"We are at the point of new wealth creation when it comes to green technology”"

Bill Joy, co-founder, Sun Microsystems

“"Customers continue to demand solutions that focus on low-power consumption and quieter operation.”"

Bob Brewer, corporate vice president, Desktop Division, AMD; from Dec 2006 press release

“The industry is going through the most profound shift in decades, moving to an era where performance and energy efficiency are critical in all market segments and all aspects of computing”

Paul Otellini, President & CEO Intel
IDF, September 2006

P.A. Semi’s green contribution: the world’s most power-efficient high-performance processor
Useful Power Is Decreasing

Useful power peaked in 2003
The Escalating Power Problem

Shrinking device geometries provides
- Faster gates
- Increased density

BUT

Moore’s Law means more power

EXCESSIVE POWER DISSIPATION LIMITS USABLE GATE CAPACITY
3-Year Operating Cost vs. Server Cost

- 3-Yr. Operating Cost
- Server Cost
Choosing Power Over Ultimate Performance

- Look for the exponential opportunities in power/performance
- Give up some performance for substantial power decrease
Design Choices for Low Power

Design choices at several levels favor low power
- CMOS process target
- Circuit design style and sizing
- Micro-architecture features

Integration
- Saves interface power

Management
- Voltage/frequency scaling
- Multiple power planes for optimal voltage selection per region
- Clock gating to reduce power of idle circuits
- Active and pre-charge standby modes in DRAM array
- PCIe power saving modes
- Nap and Sleep modes for CPU
PWRficient Family Overview

Platform Processor
P.A. Semi Introduction

Santa Clara-based fabless processor company
- Developing Power Architecture™ microprocessors under license from IBM
- Industry veterans lead 150-strong team
  - Pioneering methodology for high-performance, low-power design

Founded in 2003, broke out of stealth mode in October 2005
- Started design in September 2004
- Unveiled PWRficient™ Platform Processors at MPF 2005
- Now sampling family of high-performance, 64-bit, multicore SoC devices
  - High-speed I/O integration
- Fabricated on 65nm technology
- Target markets
  - Telecom, datacom, storage
  - Military/aerospace embedded computing and control

Vision to dramatically reposition the ‘performance per watt’ experience for high-performance embedded computing.
Power Architecture Processors (PowerPC)

- Will be in 50% of the world’s automobiles by 2007
- Runs 4 of the world’s top 5 fastest supercomputers and 44 of the top 100
- Dominant in major embedded markets
  - automotive control
  - networking & communications
  - gaming consoles
- 64% of VME and CompactPCI SBC’s
  - North America and Europe (Source: VDC)
PWRficient Family

▶ **PA6T core—the CPU**
  ▶ Power Architecture compliant, 64-bit, high-performance FPU & VMX
  ▶ 7W @ 2GHz worst-case power dissipation
  ▶ Provides a high-performance alternative to leverage existing Power Architecture software (from IBM, Freescale, AMCC)

▶ **CONEXIUM™—the on-chip coherent interconnect**
  ▶ Scalable cross-bar interconnect
  ▶ 1–8 SMP cores
  ▶ 1 or 2 L2 caches, sized 512KB–8MB
  ▶ 1–4 64-bit DDR2 memory controllers

▶ **ENVOI™—the I/O system**
  ▶ SERDES I/O—PCI Express™, XAUI, SGMII, SATA
  ▶ Offload engines—TCP/IP, iSCSI, cryptography, and RAID
  ▶ Support I/O—Boot bus, UARTs, SMBus, GPIOs
PWRficient PA6T-1682M Block Diagram

*Transaction trace memory †Peripheral trace memory
PWRficient PA6T Core

Power Architecture rev 2.04
- 2.0GHz 64-bit core with FP and VMX
- Hypervisor and virtualization support
- 64KB L1 I Cache & 64KB D Cache
- Idle and sleep modes
- 11M logic transistors

Super-scalar, out-of-order design
- Quad fetch into 64-entry scheduler
- Issue up to 3 per cycle
- Strongly ordered memory model
  - Issue out of order, retire in order
- 15 transactions in flight

Max Power
- 7W @ 2.0GHz (“thermal virus”)

Performance
- Estimated 4,400 Dhrystone MIPS per core
Processed Pipeline
Multi-Issue, Out-of-Order, Superpipelined

Reduced instruction issue width to save power ~ $N^2$ to $N^3$

Longer pipeline to accommodate slower wire delays

Increased branch prediction resources to reduce false speculation
DDR2 Memory Controller

- ~5 Watts for active DRAM rank
- ~1 Watt for standby DRAM rank
- DRAM is a significant percentage of system power
  - SOC 13 Watt typical
  - 2 x 2 DIMM memory is ~12 Watts
- DRAM controller manages power
  - Extensive clock gating within the controller
  - Use of DDR2 power saving modes
    - Precharge standby
    - Active standby
  - Sorts requests to maximize power saving
ENVOI — Intelligent I/O

- Interoperability between PCI Express, packets, DMA, and memory
  - Centralized DMA model
  - Shared resources saves area and power, which enables integration
  - Idle functions are clock gated or powered off

ENVOI™ Intelligent I/O

CONEXIUM™ Interchange

- I/O Cache
- Bridge
- DMA, Offload
- 8 x PCI Express
- 2 x 10GbE XAUI
- 4 x GbE SGMII

ENVOI™ Intelligent I/O

24 multi-mode SERDES
Boot time configurable
Flexible mapping to MACs & PCI Express
Server-Class Reliability Features

Guiding principles
- Arrays protected based on predicted soft error rates
- Machine check architecture logs all detected errors
  - Interrupt notification based on programmable thresholds
  - Error propagation ensures that erroneous data is never used
- I/O protocol engines provide in-band and/or out-of-band error notification where possible per protocol definition

ECC protection
- L1 data cache data
- L2 tags and data
- I/O cache and buffers
- DDR memory controller supports ECC memories
  - Includes scrubbing engine
  - Combined CRC/ECC for single bit correct, double detect, chip fail detect

Parity protection
- CONEXIUM bus
- L1 instruction cache tags and data
- L1 data cache tags
Example System Block Diagrams
Advanced Mezzanine Card (AMC)

- **High-performance, low power**
  - Dual-core 2GHz under 40W in a single-width mid-size module

- **Flexible I/O options**
  - Support varying widths of PCIe to card edge
  - Support XAUI or SGMII to the card edge
  - Support GbE or XAUI to the front panel

- **Multiple boot options**
  - LPC flash
  - SPI flash
ATX form-factor Motherboard

- **Extensive I/O**
  - Four PCI slots
    - One x16 PCI Express
    - One x4 PCI Express
    - One x1 PCI Express
    - PCI 32/33MHz
  - Two GbE ports
    - Separate XAUI option card for x4 PCIe slot
  - IDE port

- **Memory sockets**
  - Four DIMM sockets
  - CompactFlash socket
  - Boot ROM emulator socket

- **Standard JTAG Emulators**
  - Corelis JTAG controller
  - WindRiver ICE and Probe
Higher-end fileserver/NAS

- **Full Redundancy**
  - Requires use of SAS discs
  - No single point of failure (for RAID discs)

- **PCIe Links to SAS controller**

- **PCIe/XAUI Link between CPUs**
  - Allows metadata log from one to other

- **Memory controller supports backup battery**
  - Can use memory for file log
  - Care taken for reset and power fail

- **Assist engines**
  - TCP assist
  - CRC engine or signatures for digest
  - RAID XOR option
Low-Power Design
Power Efficient

POWER IS FIRST-ORDER DESIGN PRINCIPLE

> 25,000 gated clocks

Separate power rails for cores, I/O pads

Turn off unused I/Os

MC optimizes DRAM power

L2 cache and I/O are coherent with core off

Dynamic power-control hardware and software voltage/frequency management
Fine-Grained Clock Gating Reduces Dynamic Power

% of Flops Clocked

80
70
60
50
40
30
20
10

Time

Coarse-Grained Clock Gating

Normal Operation

Thermal Virus
**Device-Specific $V_{dd}$ Reduces Static and Dynamic Power**

**Conventional approach**
- Operate at 1.1V across entire process range
- Fast parts tend to be very leaky

**P.A. Semi approach**
- Operate at device-specific optimal $V_{dd}$
- Partition power plane for optimal voltage selection per region
- Enables full process range for power yield
Voltage/Frequency Scaling

- Multiple power planes for maximum control
  - VID regulators for cores, SoC
  - Support for either or both cores powered down
    - Low power I/O coherent nap mode

<table>
<thead>
<tr>
<th></th>
<th>Max Freq</th>
<th>Typ</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA6T-1682M-FCN</td>
<td>2.0GHz</td>
<td>13W</td>
<td>25W</td>
</tr>
<tr>
<td>PA6T-1682M-FCG</td>
<td>1.5GHz</td>
<td>8W</td>
<td>15W</td>
</tr>
<tr>
<td>PA6T-1682M-FCD</td>
<td>1.0GHz</td>
<td>6W</td>
<td>10W</td>
</tr>
<tr>
<td>I/O coherent nap</td>
<td></td>
<td></td>
<td>2W*</td>
</tr>
</tbody>
</table>

*PA6T-1682M-FCN nap power may be higher
Comparison with 65nm 2GHz processors

- PWRficient 1682M takes less total power than competing support chipsets alone
- 3x - 4x advantage in total system power* compared to contemporary 65nm dual-core processors
- Reduced system cooling and chip count result in improved system reliability

![Power Comparison Chart]

- CPUs @ 2GHz
- I/O (Northbridge, Southbridge, 10GbE MACs)
Summary

PWRficient processors set the bar for ultimate energy conservation at full performance

- Lowest total energy for a computation or transaction

Power conservation a key initiative from architectural concept though design implementation

High levels of integration enable low latency, high bandwidth interfaces to cache, memory, & IO
Contact P.A. Semi

For further information, please visit P.A. Semi web site at:

www.pasemi.com

Kindly direct sales inquiries to:

pasales@pasemi.com

Full contact information:

P.A. Semi, Inc.
3965 Freedom Circle, Floor 8
Santa Clara
CA 95054-1203 USA
Main: 408.200.4500
Fax: 408.200.4501
Thank You

The P.A. Semi name and the P.A. Semi logo and combinations thereof are trademarks of P.A. Semi, Inc. The Power name is a trademark of International Business Machines Corporation, used under license therefrom. SPECint and SPECfp are registered trademarks of the Standard Performance Evaluation Corporation (SPEC). All other trademarks are the property of their respective owners.