NERSC Power Efficiency Analysis

John Shalf
SDSA Team Leader
jshalf@lbl.gov

NERSC User Group Meeting
September 17, 2007
Microprocessors: Up Against the Wall(s)

From Joe Gebis

- Microprocessors are hitting a power wall
  - Higher clock rates and greater leakage increasing power consumption
- Reaching the limits of what non-heroic heat solutions can handle
- Newer technology becoming more difficult to produce, removing the previous trend of “free” power improvement

Intel Desktop Processor Max Power Consumption, Pentium through P4

Source: sandpile.org

NERSC User Group Meeting, September 17, 2007
Growth in Power Consumption (Top50)
**Excluding Cooling**

- System Power (kW)
- Avg. Power Top5
- Avg. Power Top50
• Immediate need to add 8 MW to prepare for 2007 installs of new systems
• NLCF petascale system could require an additional 10 MW by 2008
• Need total of 40-50 MW for projected systems by 2011
• Numbers just for computers: add 75% for cooling
• Cooling will require 12,000 – 15,000 tons of chiller capacity

Cost estimates based on $0.05 kW/hr

Data taken from Energy Management System-4 (EMS4). EMS4 is the DOE corporate system for collecting energy information from the sites. EMS4 is a web-based system that collects energy consumption and cost information for all energy sources used at each DOE site. Information is entered into EMS4 by the site and reviewed at Headquarters for accuracy.
Need Power Efficiency Metrics based on Effective Performance

• We want to push industry in the *right* direction
• Leverage *established* performance benchmarks to serve as numerator for “power efficiency” ratio
• Segregate by workload
  – Small/Workstation: *Spec2006/Watt*
  – Midrange Cluster: *NAS Parallel Benchmarks MOPS/Watt*
  – HEC/Top500: *LINPACK/Watt? HPCC/Watt? SSP/Watt?*
• Role of Top500
  – Collected history of largest HEC investments in the world
  – Archive of system metrics plays important role in analyzing industry trends
  – Can play an important role in collecting data necessary to understand power efficiency trends
  – Feed data to studies involving benchmarks other than LINPACK as well
Broad Objective for Top500

• Use Top500 List to track power efficiency trends
• Raise Community Awareness of HPC System Power Efficiency
• Push vendors toward more power efficient solutions by providing a venue to compare their power consumption
Single Node Tests: AMD Opteron

- Highest power usage is 2x NAS FT and LU
Similar Results when Testing Other CPU Architectures

- Power consumption far less than manufacturer’s estimated “nameplate power”
- Idle power much lower than active power
- Power consumption when running LINPACK is very close to power consumed when running other compute intensive applications
Full System Test

Tests run across all 19,353 compute cores
Throughput: NERSC “realistic” workload composed of full applications
idle() loop allows powersave on unused processors; (generally more efficient)
Single Rack Tests

- Administrative utility gives rack DC amps & voltage
- HPL & Paratec are highest power usage
Modeling the Entire System: Disks

- Must take into account disk subsystem
- Drive model *matters*
  - Deskstar 9.6W idle, 13.6W under load
  - Tonka 7.4W idle, 12.6W under load
- Using DDN-provided numbers, estimated power draw for model disk subsystem is 50KW idle, 60KW active
- Observed using PDU panel: ~48KW idle
Modeling the Entire System (projecting from single cabinet)

- Error factor is 0.05 if we assume 90% efficiency
Conclusions

- Power utilization under an HPL/Linpack load is a good estimator for power usage under mixed workloads for single nodes, cabinets / clusters, and large scale systems
  - Idle power is not
  - Nameplate and CPU power are not
- LINPACK running on one node or rack consumes approximately same power as the node would consume if it were part of full-sys parallel LINPACK job
- We can estimate overall power usage using a subset of the entire HPC system and extrapolating to total number of nodes using a variety of power measurement techniques
  - And the estimates mostly agree with one-another!
- Disk subsystem is a small fraction of overall power (50-60KW vs 1,200 KW)
  - Disk power dominated by spindles and power supplies
  - Idle power for disks not significantly different from active power
Some Food For Thought
New Design Constraint: POWER!

• Transistors still getting smaller
  – Moore’s Law is alive and well
• But Denard scaling is dead!
  – No power efficiency improvements with smaller transistors
  – No clock frequency scaling with smaller transistors
  – All “magical improvement of silicon goodness” has ended
• Traditional methods for extracting more performance are well-mined
  – Cannot expect exotic architectures to save us from the “power wall”
Path to Power Efficiency

• Mark Horowitz: Years of Low Power Research . . .
  – Have shown only one design technique to reduce power: Reduce 
    waste
  – Sources of Waste
    • Wasted transistors (surface area)
    • Wasted computation (useless work/speculation)
    • Wasted bandwidth (data movement)
    • Energy (clock gating, leakage control, etc)
    • Performance: Adding additional constraints to operation flow

• If technology scaling has stalled, need to focus on reducing waste in our systems!
  – Exploit Specialization!
  – Optimize execution units for specific applications
  – Reformulate the hardware to reduce needed work
  – Can improve energy efficiency for a class of applications
Learning from Embedded Market

• Desktop CPU market motivated to provide max performance at any cost.
  – Maximizing clock frequency
  – Long pipelines, complex o-o-o execution = extra power
  – Add features to cover virtually every conceivable application
  – Power consumption limited only by ability to dissipate heat
  – Cost around $1K for high-end chips

• Embedded market motivated to maximize performance at min cost and power
  – Want cell phones that last forever on tiny battery and cost ~$0
  – Specialized: remove unused features
  – Effective performance per watt is critical metric

• The world has changed
  – Clock frequency scaling has ended
  – At limited for cost effective air-cooled systems
  – Price point for desktops/portables dropping (portables dominate market)
  – For HPC, cost of power is exceeding procurement costs!
  – Technology from embedded market is now trickling up into server designs
    • Rather than traditional trickle down flow of innovations

• What will HPC learn from the embedded market?
  – Simpler, smaller cores
  – Many cores on chip (100’s of cores, not 2,4,8)
  – Lower clock rates
  – More specialization to applications
Examples: Power Efficiency Benefits of Tailoring Hardware to Application

**Consumer Electronics**

- Optimized ConsumerMarks/MHz
  - 2.0
  - 0.087, 0.080, 0.059, 0.058, 0.039

**DSP**

- Optimized TeleMarks/MHz
  - 0.473
  - 0.03, 0.016, 0.013, 0.011

**Networking**

- Optimized NetMarks/MHz
  - 0.123
  - Extensible optimized, Extensible out-of-box, MIPS64 20Kc, ARM1020E, MIPS64b (NEC VR5000), MIPS32b (NEC VR4122)

Source: EEMBC Certified Benchmarks

Chris Rowen, Tensilica
How Small is “Small”

- **Power5 (Server)**
  - 389mm^2
  - 120W@1900MHz
- **Intel Core2 sc (laptop)**
  - 130mm^2
  - 15W@1000MHz
- **ARM Cortex A8 (automobiles)**
  - 5mm^2
  - 0.8W@800MHz
- **Tensilica DP (cell phones / printers)**
  - 0.8mm^2
  - 0.09W@600MHz
- **Tensilica Xtensa (Cisco router)**
  - 0.32mm^2 for 3!
  - 0.05W@600MHz

Each core operates at 1/3 to 1/10th efficiency of largest chip, but you can pack 100x more cores onto a chip and consume 1/20 the power.
Consumer Electronics Convergence

Market in Japan (B$)

- PC
- TV
- TV+DVD+DSC
- DSC
- DVD
- Analog
- Digital

Years: 2001 to 2005
Consumer Electronics has Replaced PCs as the Dominant Market Force in CPU Design!!
Consumer Electronics has Replaced PCs as the Dominant Market Force in CPU Design!!

Apple Introduces iPod

IPod+ITunes exceeds 50% of Apple’s Net Profit

Apple Introduces Cell Phone (iPhone)

Brief History of PC
1975 Altair/ MITS
1978 Apple II
1981 IBM PC(MSDOS)
1985 Windows 1.0

Source: IDC

IBM Started PC Business

IBM Sold PC Business to Lenovo

1981

Revenue

100B

100M

10B

1B


Revenue($)
Rise of the Embedded Processor?

• 1990’s: Rise of the desktop PC CPUs led to sea-change in HPC system architecture
  – Revenues and volumes of PC CPUs made it difficult for custom designs to compete due to technology investments
  – Rise of the commodity cluster

• 2005+: Processor cores from consumer electronics space now dominating revenues and volumes
  – Correct design point for power efficiency
  – BG/L, BG/P, SiCortex, based on embedded/consumer-electronics CPU designs
  – But can we program them effectively? (massively concurrent)