Towards Efficient Supercomputing: A Quest for the Right Metric

Chung H. Hsu, W. Feng, J. Archuleta

Computer & Computational Sciences Division
Los Alamos National Laboratory
LA-UR 05-0936

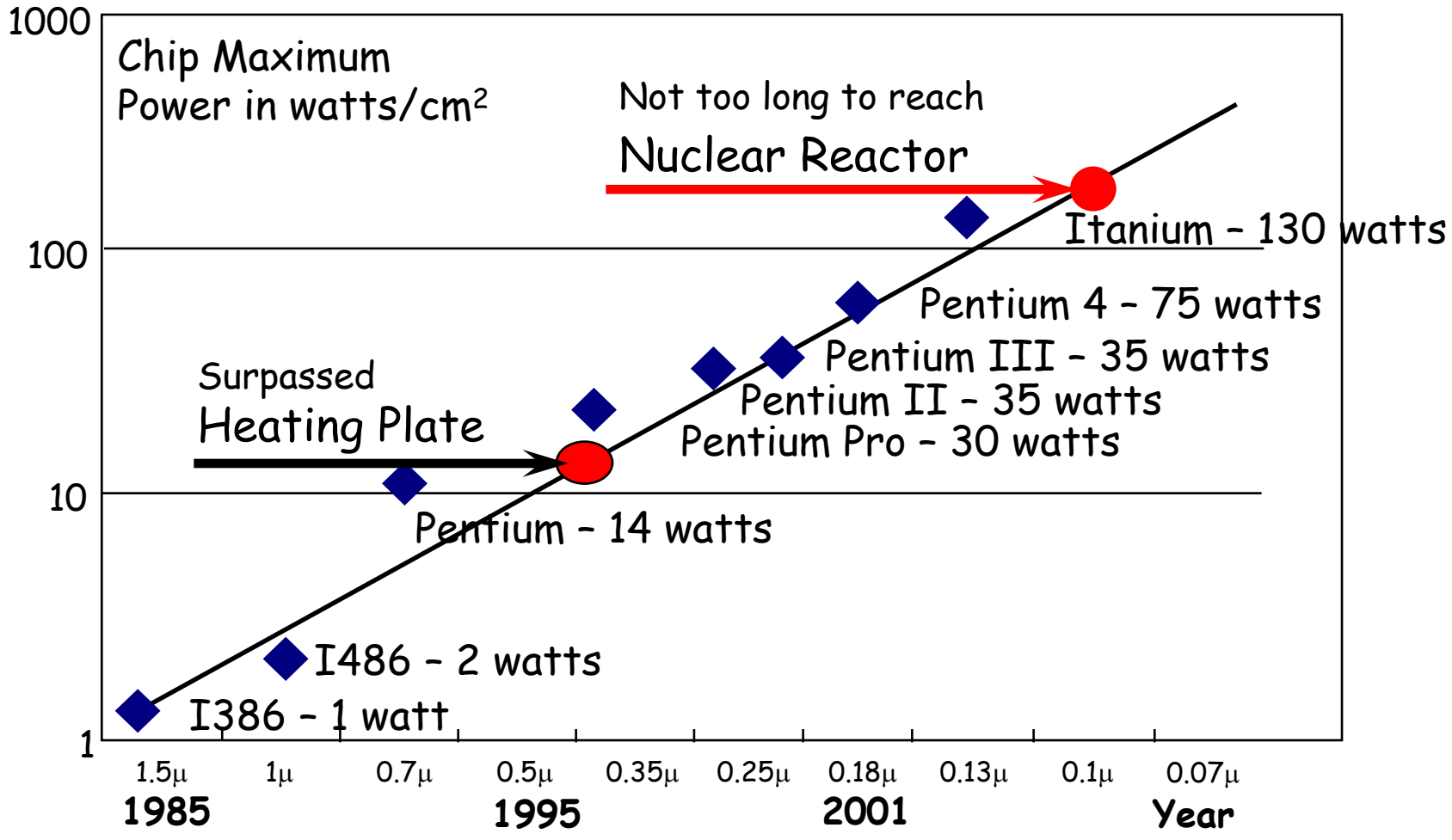
Workshop on High-Performance Power-Aware Computing; Denver, CO; April 4, 2005.

Efficiency

The ratio of the useful output to the input in any system. [The American Heritage Dictionary of the English Language, 4th Edition]

- Performance/1
- Performance/$
- Performance/watt
- Performance/ft^2
- GB/ft^2

Moore's Law for Power



Source: Fred Pollack, Intel. New Microprocessor Challenges in the Coming Generations of CMOS Technologies, MICRO32 and Transmeta

Efficient Supercomputing

What metric should one use to evaluate how efficiently a given system delivers useful output?

Machine	Green Destiny+	Earth Simulator
Year	2002	2002
LINPACK Performance (Gflops)	*101*	35,860
Cost efficiency ($/Mflop)	3.35	11.15
Space efficiency (Mflops/ft^2)	16,833	1,041
Power efficiency (Mflops/watt)	20.20	5.13

Efficient Supercomputing

What metric should one use to evaluate how efficiently a given system delivers useful output?

Disclaimer: This presentation raises more questions than it answers.

Outline

- Optimize performance and power simultaneously
- Use existing metrics
- Devise new metrics
- Summarize

Performance * Power

- Optimize performance alone
  - may consume too much power
- Optimize power alone
  - can be done thru lower performance
- Optimize both simultaneously

Performance * Power

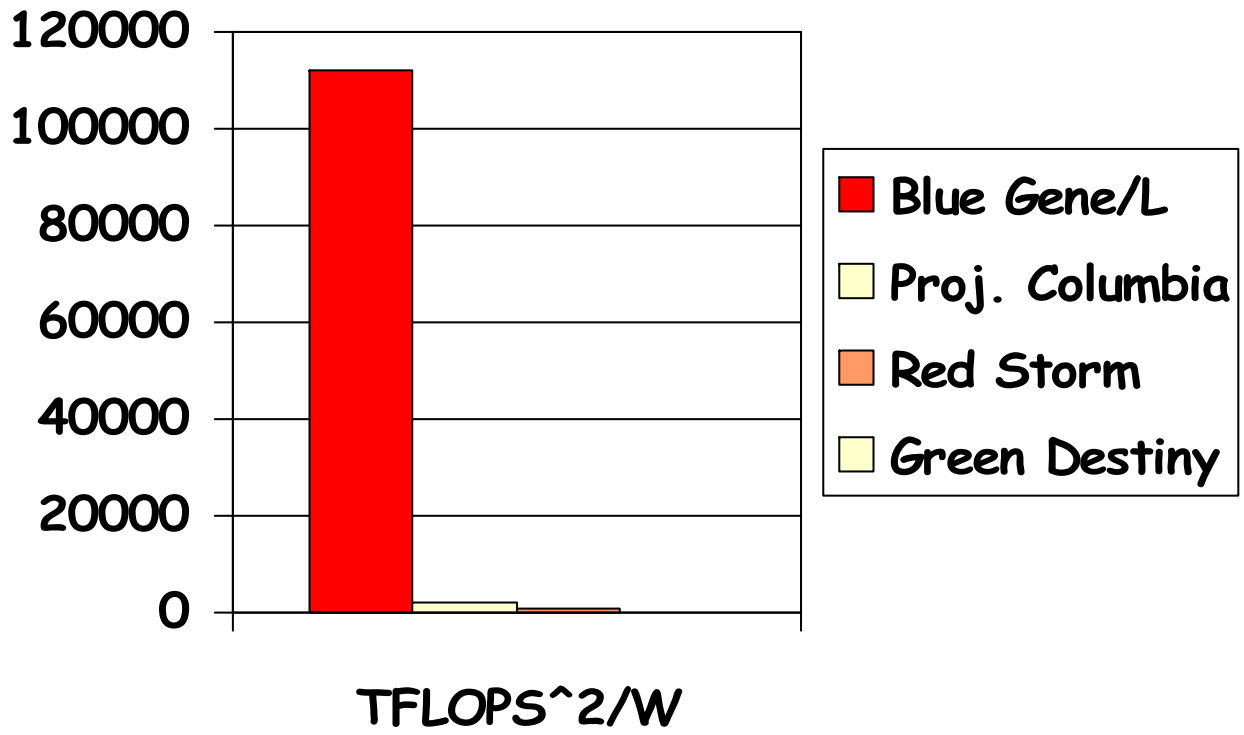
- Product form:
  - $Performance^n/Power$
  - $SPEC^2/W$, $MIPS^2/W$, $FLOPS^2/W$
  - processor design
- Sum form:
  - c1 * AccessTime + c2 * Power
  - CACTI (Cache Timing, Power, & Area Ratio)

Performance^n/Power

- Besides SPEC^n/W, MIPS^n/W, FLOPS^n/W
  - ED^n form: PDP, EDP, ED^{2}P
  - Narrower gap between processor designs when using SPEC^2/W or SPEC^2/Wλ^2 [JSSC'96]
  - SPEC^2/W or SPEC^3/W for high-end and SPEC/W for low-end [Micro 2000]
  - Energy complexity of computation [IPL 2001]
  - An elusive metric [WCED 2003]
- For HPC
  - Biased towards massively parallelism
  - Loose control in performance loss

Performance^n/Power

- Biased towards massive parallelism



Performance^n/Power

- Biased towards massive parallelism
  - Per processor, F MIPS @ P watts
  - Total s processors
  - $MIPS^n/W = (s * F)^n/(s * P) = s^{n-1} * F^n/P$
  - Performance^n/Power/processor will not work
- Loose control in performance loss

Low Power = Low Performance

Not always true.

Machine	Power	HPL Perf.
2 x 2P 1-GHz Itanium-2	601	11
1 x 4P 2-GHz Opteron	403	12
12 x 1P 1.2-GHz Efficeon	185	14

Yokogawa WT210
5mA – 26A
Sample rate: 20us

Performance?

- Throughput (MIPS, MFLOPS)
- Execution Time (seconds, hours)
- Which system resources are stressed?
- The war of means (arithmetic, geometric, …)
  - HPCC lets you decide it yourself

- HPL is a MICRO-benchmark
  - TOP500 list, November 2004
  - 400 systems ran shorter than 2 hours
  - The longest ran 18 hours.

Power?

- How to measure Power in HPC ?
- Component or System Power?
  - Monotonically increasing if perf. vs. power
  - U-shape if perf. vs. energy

- Power, Energy, or Temperature?
  - temperature relates to reliability [Queue'03]
  - temperature ?= average power [Micro 2003]
  - temperature ?= instantaneous power [Micro 2003]

Reliability & Availability

Systems	CPUs	Reliability & Availability
ASCI Q	**8,192**	**MTBI: 6.5 hrs.** 114 unplanned outages/month. ◆ HW outage sources: storage, CPU, memory.
ASCI White	**8,192**	**MTBF: 5 hrs. (2001) and 40 hrs. (2003).** ◆ HW outage sources: storage, CPU, 3rd-party HW.
NERSC Seaborg	**6,656**	**MTBI: 14 days. MTTR: 3.3 hrs.** ◆ SW is the main outage source. **Availability: 98.74%.**
PSC Lemieux	**3,016**	**MTBI: 9.7 hrs.** **Availability: 98.33%.**
Google	**~15,000**	**20 reboots/day; 2-3% machines replaced/year.** ◆ HW outage sources: storage, memory. **Availability: ~100%.**

MTBI: mean time between interrupts; MTBF: mean time between failures; MTTR: mean time to restore

Source: Daniel A. Reed, UNC

An Initial Attempt

- Efficiency, Reliability, Availability
- 5-year Total Cost of Ownership (TCO)

- TCO = A + E + D
- E = $0.1 kWh * Power * 5 years * (1 + 0.5)
- D = $1K * FailureRate * 5 years
- FailureRate = exp(26.6 – 1/(8.6*10^{-5}*Tcpu))
- Tcpu = 273 °K + 45 °C + 0.2 °C/W * Power
  - 1% failure rate at 100 °C

Summary

- Current HPC systems are less efficient because of the Moore's law for power consumption.
- Existing ("borrowed") efficiency metrics may not fit power-aware HPC use.
- New factors such as reliability, availability, & productivity need to be factored in.
- We do not know how to relate these factors with H/W characteristics.

SUPERCOMPUTING
in SMALL SPACES

http://sss.lanl.gov