Making a Case for a Green500 List

S. Sharma[†], C. Hsu[†], and W. Feng[‡]

[†] Los Alamos National Laboratory

[‡] Virginia Tech

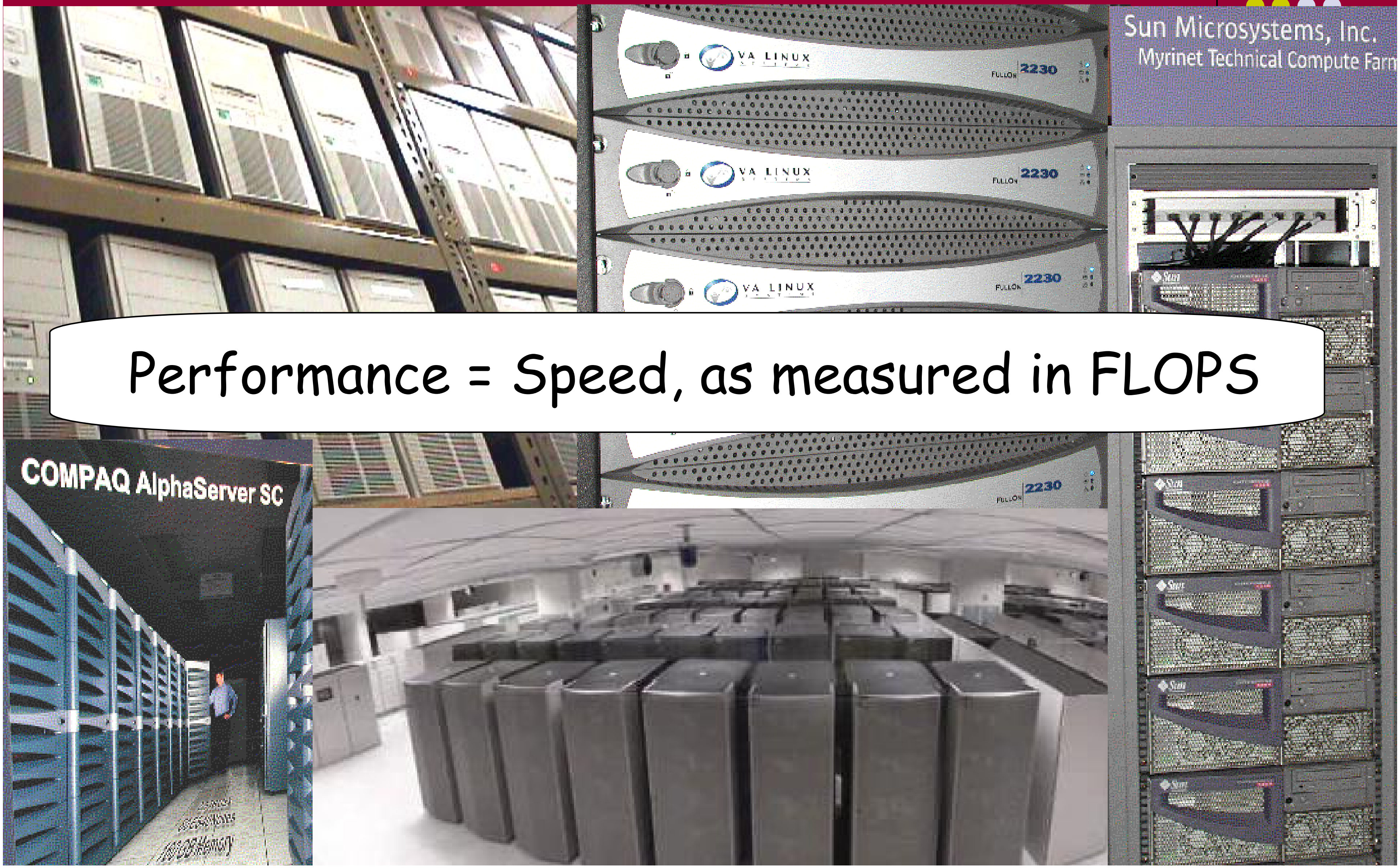
Los Alamos NATIONAL LABORATORY

VirginiaTech

Outline

- Introduction
  - What Is Performance?
  - Motivation: The Need for a Green500 List
- Challenges
  - What Metric To Choose?
  - Comparison of Available Metrics
- TOP500 as Green500
- Conclusion

Where Is Performance?



Performance = Speed, as measured in FLOPS

What Is Performance?

TOP500 Supercomputer List

- Benchmark
  - LINPACK: Solves a (random) dense system of linear equations in double-precision (64 bits) arithmetic.
    - Introduced by Prof. Jack Dongarra, U. Tennessee
- Evaluation Metric
  - Performance (i.e., Speec
    - Floating-Operations Per
- Web Site
  - http://www.top500.org
- Next-Generation Bench k: HPC Challenge
  - http://icl.cs.utk.edu/h

Performance, as defined by speed, is an important metric, but…

W. Feng, feng@cs.vt.edu, (540) 231-1192

Reliability & Availability of HPC

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 (as of 2003)	~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, RENCI, 2004

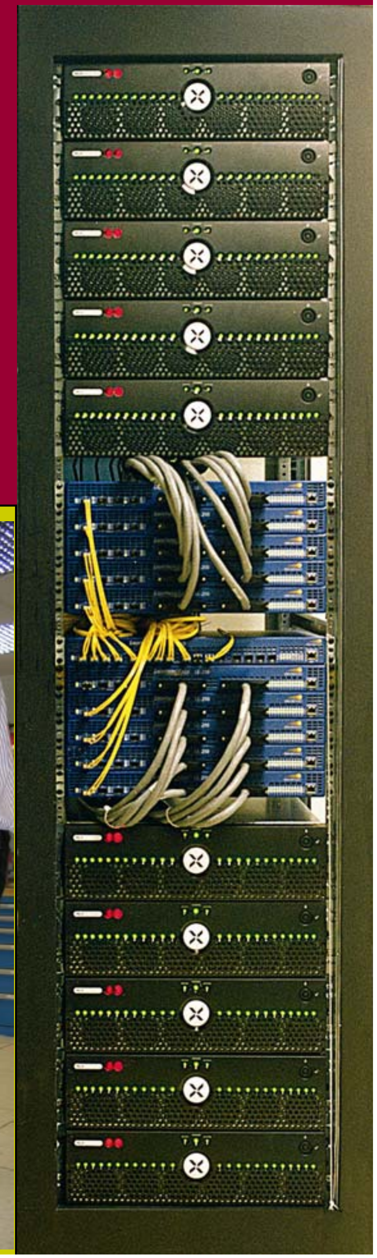
Costs Associated with HPC

- Infrastructure
  - Sizable costs associated with system administration and maintenance. (People resources are $$$.)
  - Massive construction and operational costs associated with powering and cooling.
    - Google
      - $2M to buy 30 acres of land by The Dalles Dam (Columbia River)
        - Inexpensive power to satisfy their high electrical demand.
        - Water can be used to cool its massive server-filled facility directly rather than relying on more expensive A/C.
    - Lawrence Livermore National Laboratory
      - Building for Terascale Simulation Facility: $55M
      - Electrical Costs: $14M/year to power and cool.
- Productivity
  - Downtime means no compute time, i.e., lost productivity.

Recent Trends in HPC

- Low(er)-Power Multi-Core Chipsets
  - AMD: Athlon64 X2 (2) and Opteron (2)
  - ARM: MPCore (4)
  - IBM: PowerPC 970 (2)
  - Intel: Smithfield (2) and Montecito (2)
  - PA Semi: PWRficient (2)
- Low-Power Supercomputing
  - Green Destiny (2002)
  - Orion Multisystems (2004)
  - BlueGene/L (2004)
  - MegaProto (2004)

October 2003
BG/L half rack prototype
500 Mhz
512 nodes/1024 proc.
2 TFlop/s peak
1.4 Tflop/s sustained



Perspective

- FLOPS Metric of the TOP500
  - Performance = Speed (as measured in FLOPS with Linpack)
  - May not be "fair" metric in light of recent low-power trends to help address reliability, availability, and total cost of ownership.
- The Need for a Different Performance Metric
  - Performance = *f* (speed, "time to answer", power consumption, "up time", total cost of ownership, usability, …)
  - Easier said than done …
    - Many of the above dependent variables are difficult, if not impossible, to quantify, e.g., "time to answer", TCO, usability, etc.
- The Need for a Green500 List
  - Performance = *f* (speed, power consumption) as speed and power consumption can be quantified.

W. Feng, feng@cs.vt.edu, (540) 231-1192

Outline

- Introduction
  - What Is Performance?
  - Motivation: The Need for a Green500 List
- Challenges
  - What Metric To Choose?
  - Comparison of Available Metrics
- TOP500 as Green500
- Conclusion

Challenges for a Green500 List

- What Metric To Choose?
  - *ED^n*: Energy-Delay Products, where *n* is a non-negative int.
    (borrowed from the circuit-design domain)
  - *Variants of ED^n*
  - *Speed / Power Consumed*
    - FLOPS / Watt, MIPS / Watt, and so on
- What To Measure? Obviously, energy or power … but
  - Energy (Power) consumed by the computing system?
  - Energy (Power) consumed by the processor?
  - Temperature at specific points on the processor die?
- How To Measure Chosen Metric?
  - Power meter? But attached to what? At what time granularity should the measurement be made?

ED^n : Energy-Delay Products

- Original Application
  - Circuit Design
- Problem
  - For $n \geq 1$, the metric is biased towards systems with a larger number of processors as the "delay component" (i.e., aggregate speed) dominates.
  - As *n* increases, the bias towards aggregate speed, and hence, HPC systems with larger numbers of processors, increases dramatically.

Variants of ED^n : $V_\partial = E^{(1-\partial)} D^{2(1+\partial)}$

- Negative values of ∂ (particularly more negative values) marginally offset the bias of the ED^n towards speed.
  - In our benchmarking, they produced identical rankings to the ED^n metric.
- Positive values of ∂ place greater emphasis on performance.
  - As ∂ increases towards one, the metric approaches the limit E^0D^4 and behaves more like the standard FLOPS metric, which is used for TOP500 List.

Metric of Choice: FLOPS / Watt (again)

Outline

- Introduction
  - What Is Performance?
  - Motivation: The Need for a Green500 List
- Challenges
  - What Metric To Choose?
  - Comparison of Available Metrics
- TOP500 as Green500
- Conclusion

Efficiency of Four-CPU Clusters

Name	CPU	LINPACK (Gflops)	Avg Pwr (Watts)	Time (s)	ED ($*10^6$)	ED2 ($*10^9$)	Flops/W	$V_{\partial=-0.5}$
C1	3.6G P4	19.55	713.2	315.8	71.1	22.5	27.4	33.9
C2	2.0G Opt	12.37	415.9	499.4	103.7	51.8	29.7	47.2
C3	2.4G Ath64	14.31	668.5	431.6	124.5	53.7	21.4	66.9
C4	2.2G Ath64	13.40	608.5	460.9	129.3	59.6	22.0	68.5
C5	2.0G Ath64	12.35	560.5	499.8	140.0	70.0	22.0	74.1
C6	2.0G Opt	12.84	615.3	481.0	142.4	64.5	20.9	77.4
C7	1.8G Ath64	11.23	520.9	549.9	157.5	86.6	21.6	84.3

W. Feng, feng@cs.vt.edu, (540) 231-1192

Green500 Ranking of Four-CPU Clusters

Green500 Ranking							TOP 500	Power 500
Rank	ED	ED^2	ED^3	$V_{\partial=-0.5}$	$V_{\partial=0.5}$	FLOPS /Watt	FLOPS	Watts
1	C1	C1	C1	C1	C1	C2	C1	C2
2	C2	C2	C2	C2	C3	C1	C3	C7
3	C3	C3	C3	C3	C4	C5	C4	C5
4	C4	C4	C4	C4	C2	C4	C6	C4
5	C5	C5	C5	C5	C5	C7	C2	C6
6	C6	C6	C6	C6	C6	C3	C5	C3
7	C7	C7	C7	C7	C7	C6	C7	C1

W. Feng, feng@cs.vt.edu, (540) 231-1192

Outline

- Introduction
  - What Is Performance?
  - Motivation: The Need for a Green500 List
- Challenges
  - What Metric To Choose?
  - Comparison of Available Metrics
- TOP500 as Green500
- Conclusion

W. Feng, feng@cs.vt.edu, (540) 231-1192

TOP500 Power Usage

Name	Linpack	Peak Power	MFLOPS/W	TOP500 Rank
BlueGene/L	367,000	2,500	146.80	1
ASC Purple	77,824	7,600	10.24	3
Columbia	60,960	3,400	17.93	4
Earth Simulator	40,960	11,900	3.44	7
MareNostrum	42,144	1,071	39.35	8
Jaguar-Cray XT3	24,960	1,331	18.75	10
ASC Q	20,480	10,200	2.01	18
ASC White	12,288	2,040	6.02	47

TOP500 as Green500

Relative Rank	TOP500	Green500
1	BlueGene/L (IBM)	BlueGene/L (IBM)
2	ASC Purple (IBM)	MareNostrum (IBM)
3	Columbia (SGI)	Jaguar-Cray XT3 (Cray)
4	Earth Simulator (NEC)	Columbia (SGI)
5	MareNostrum (IBM)	ASC Purple (IBM)
6	Jaguar-Cray XT3 (Cray)	ASC White (IBM)
7	ASC Q (HP)	Earth Simulator (NEC)
8	ASC White (IBM)	ASC Q (HP)

Conclusion

- Metrics for a Green500 Supercomputer List
  - Still no definitive metric to use
    - By process of elimination, we converged on FLOPS/watt, which is relatively easy to derive from the TOP500 Supercomputer List.
  - Insight with respect to current metrics
  - Insight with respect to when to use processor energy (or power) versus system energy (or power)
- TOP500 as Green500
  - From the data presented, IBM and Cray make the most energy-efficient HPC systems today.

For More Information

- Visit "Supercomputing in Small Spaces" at http://sss.lanl.gov
  - Soon to be re-located to Virginia Tech
- Contact Wu-chun Feng (a.k.a. "Wu")
  - E-mail: feng@cs.vt.edu
  - Phone: (540) 231-1192
  - Mailing Address
    - 2200 Kraft Drive
      Department of Computer Science
      Virginia Tech
      Blacksburg, VA 24060