Lecture 3: Benchmarks

William Gropp
www.cs.illinois.edu/~wgropp

How Are Systems Measured?

- Why do you want to measure or rank a system?
  - What is the purpose of the extreme scale system?
  - If it is to do science, that should be the measurement
- There is value is having a way to compare systems before making a purchase or request
  - Will this system be able to solve my problem?
  - Is the system the right “size”?

Benchmarks

- Benchmarks are methods used to provide a measurement that allows similar things to be compared.
- High performance computing benchmarks are typically one or more program and defined input that must be run correctly; the measurement is usually either the time or the rate (operations/second)

Some Popular Benchmarks

- HPLinpack http://top500.org
- STREAM http://www.cs.virginia.edu/stream/
- HPC Challenge http://icl.cs.utk.edu/hpcc
- Graph 500 http://graph500.org
- NAS Parallel Benchmarks http://www.nas.nasa.gov/publications/npb.html
- Many others

HP Linpack

- The most famous HPC benchmark – used for the “Top500” ranking
- Solve a system of n linear equations using Gaussian elimination (matrix is dense)
- Time is roughly $2n^3/3$ (floating point operations only)
- Memory is roughly n^2
- Results are updated twice a year
- Solving a system of linear equation is at the heart of many computational science problems
    - ♦ But almost all large systems are sparse and are not solved with Gaussian elimination (or at best with a *sparse* Gaussian elimination algorithm)

STREAM

- Measure "Sustainable Memory Bandwidth"
  - For four operations:
    - COPY (x(i) = y(i))
    - SCALE (x(i) = a * y(i))
    - ADD (x(i) = y(i) + z(i))
    - TRIAD (x(i) = y(i) + a * z(i))
- Very large arrays x, y, and z
  - We'll explain why "very large" later in this course

HPC Challenge

- Attempt to broaden the HPLinpack benchmark to a suite of benchmarks
  - HPLinpack
  - DGEMM – dense matrix-matrix multiply
  - STREAM – memory bandwidth
  - PTRANS – parallel matrix transpose
  - RandomAccess – integer accumulates anywhere (race conditions allowed)
  - FFT – 1d FFT
  - Communication (from beff); bandwidth and latency
- Characteristics are not distinct
  - E.g., DGEMM a major part of HPL
  - Infrequently used today

Graph 500

- An attempt to provide an alternative to HP Linpack for graph problems
  - Results available since Nov 2010
  - Breadth First Search (BFS) is the first benchmark
  - Reference implementations for OpenMP and MPI
- Significant early progress through better algorithms
  - From the description: "However, we do not constrain the choice of BFS algorithm itself, as long as it produces a correct BFS tree as output"
- Additional graph benchmarks under development

NAS Parallel Benchmarks

- Derived from applications important to NASA
- Original version described problems to solve but left implementation to the user
  - Defined before there was a standard for programming parallel systems
- Most uses today based on the MPI or MPI +OpenMP hybrid versions available from NASA
- Benchmarks include
  - Integer sort
  - Conjugate gradient
  - Multigrid
  - 3D FFT
  - 3 "pseudo applications" (solvers)

More Recent Collections

- Sustained Petascale Performance
  - ♦ Measures full applications, weighted to represent workload
  - ♦ Used for Blue Waters
  - ♦ More accurate for specific system but hard to compare over time
- Coral
  - ♦ Used for most recent big DOE procurement
  - ♦ Large set ranging from applications to microbenchmarks
  - ♦ https://asc.llnl.gov/CORAL-benchmarks/

The Top 5 systems in Nov 2014 (by HPLinpack)

1. Tianhe-2 (China), 3,120,000 cores
2. Titan Cray XK (US), 560,640 cores
3. Sequoia Blue Gene/Q (US), 1,572,864 cores
4. Fujitsu K Computer (Japan), 705,024 cores
5. Mira Blue Gene/Q (US), 786,432 cores

- #1 has Intel Phi; #2 NVIDIA Kepler
- Blue Waters (with 792,064 cores) would be around #4 if we bothered to run this benchmark

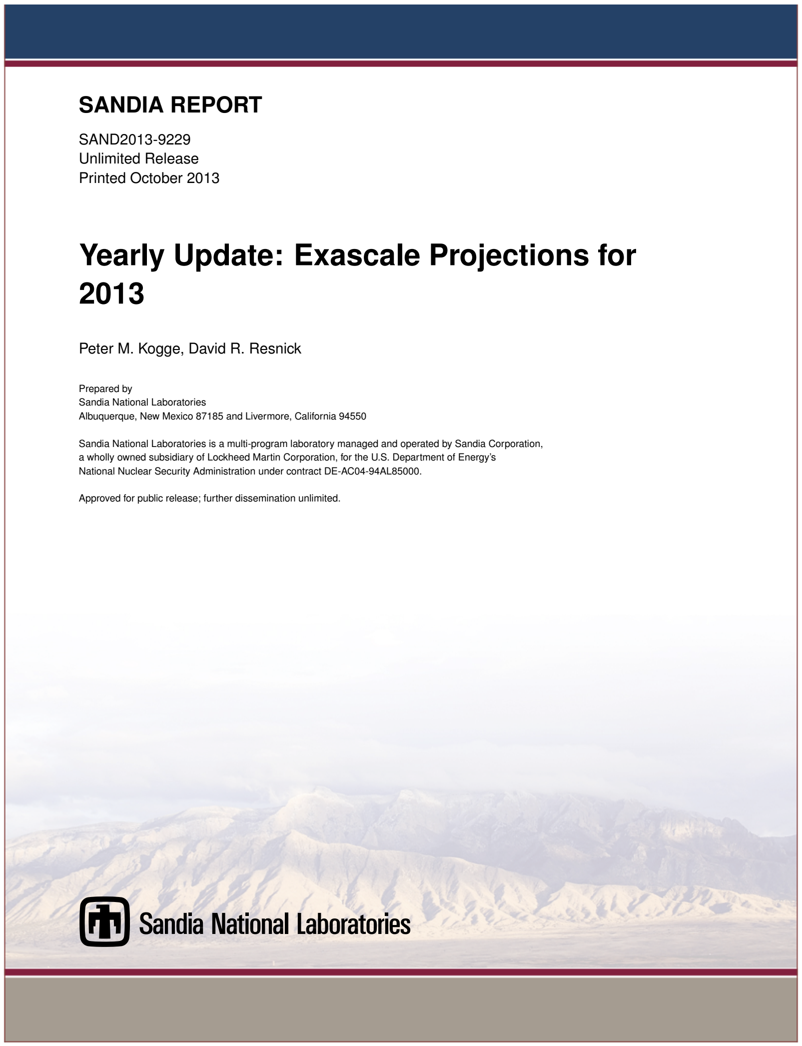
Blue Waters and Sequoia Computing Systems

System Attribute	NCSA Blue Waters	LLNL Sequoia
Vendors	Cray/AMD/NVIDIA	IBM
Processors	Interlagos/Kepler	PowerPCA2 variant
Total Peak Performance (PF)	13.34	20.1
Total Peak Performance (CPU/GPU)	7.1/6.24	20.1/0.0
Number of CPU Chips (8, 16 FPcores/chip)	49,504	98,304
Number of GPU Chips	4,224	0
Amount of CPU Memory (TB)	1,476	1,572
Interconnect	3D Torus	5D Torus
Amount of On-line Disk Storage (PB)	26	50(?)
Sustained Disk Transfer (TB/sec)	>1	0.5-1.0
Amount of Archival Storage (PB)	380	?
Sustained Tape Transfer (GB/sec)	58	?

Where Are HPC Systems Going?

- Many discussions look at the benchmarks
  - ♦ Top500 is the most common
  - ♦ STREAM and RandomAccess also common
- Increasing interest in more application-oriented benchmarks
- Short form:
  - ♦ Rates of increase in performance are slowing
  - ♦ Meeting power and performance targets leading to more specialization in hardware

Trend Data



SANDIA REPORT

SAND2013-9229
Unlimited Release
Printed October 2013

Yearly Update: Exascale Projections for 2013

Peter M. Kogge, David R. Resnick

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

- Excellent, comprehensive report on HPC architecture issues and trends
- http://www.osti.gov/scitech/biblio/1104707

Thinking about Trends

- Absolute numbers are often hard to interpret
  - Is 1usec latency good? Bad?
    - Spectacularly good for disk
    - Good for interconnect
    - Poor for main memory
    - Disastrous for register

Ratios are often better

- Rather than I/O bandwidth, memory size to I/O bandwidth (bytes / (bytes/sec)) gives seconds
  - ♦ Time to copy all of memory to disk
  - ♦ Important for "checkpoints" (see fault tolerance later in this course)

Dimensionless Quantities Often the Best

- Example: Ratio of latencies
  - L1 to L2 cache
  - L1 to Memory
  - Memory to remote memory
  - Memory latency in clock ticks
- "Best" because independent of the units chosen
- We'll use ratios and dimensionless quantities when looking at trends

Questions

- True or False:
  - ♦ There are computers using more than one million cores today
  - ♦ The Top500 benchmark predicts the performance of many applications
- What are 3 important benchmarks? What do they measure?
- Do all benchmarks specify the specific code that must be run?

Readings

- Abstract Machine Models and Proxy Architectures for Exascale Computing, J. Ang et al, http://crd.lbl.gov/Fassets/Fpubs_presos/FCALAbstractMachineModelsv1.1.pdf
- ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems, P. Kogge Editor, http://www.cse.nd.edu/Reports/2008/TR-2008-13.pdf