NERSC Workload Analysis and Benchmark Approach

Harvey Wasserman
NERSC Science Driven System Architecture Group
www.nersc.gov/projects/SDSA
August 27, 2008
Acknowledgments

• Contributions to this talk by many people:

Bill Kramer  
NERSC-6 Project Manager

Jim Craw  
NERSC-6 Deputy Project Manager

Lynn Rippe  
Procurement Lead

Katie Antypas

John Shalf

Tina Butler

Rei Lee

R.K. Owen
Science Driven System Architecture Group

• Analyze requirements of broad scientific workload
  – Benchmarking
  – Algorithm tracking

• Track future trends in supercomputing architecture
  – Assess emerging system technologies

• Understand bottlenecks in current computing architecture
  – Use the NERSC workload to drive changes in computing architecture.

http://www.nersc.gov/projects/SDSA
• Increase user scientific productivity via a timely introduction of the best new technologies designed to benefit the broadest subset of the NERSC workload.*

• Question: what is that workload? How do we characterize it?

Science Driven Evaluation

• Translate scientific requirements into computational needs and then to a set of hardware and software attributes required to support them.

• Question: how do we represent these needs so that we can communicate them to others?
  – Answer: a set of carefully chosen benchmark programs.
Thoughts

• “For better or for worse, benchmarks shape a field.”
  – Prof. David Patterson, CS252 Lecture Notes, University of California, Berkeley, Spring, 1998.

• “Benchmarks are only useful insofar as they model the intended computational workload.”
  – Bucher & Martin, LANL, 1982
NERSC Benchmarks Serve 3 Critical Roles

• Carefully chosen to represent characteristics of the expected NERSC-6 workload.

• Give vendors opportunity to provide NERSC with concrete performance and scalability data;
  – Measured or projected.

• Part of the acceptance test and a measure of performance throughout the operational lifetime of NERSC-6.
Why Measure Performance?

Overarching goal: improve the state of computer architecture
Workload Analysis

• Understand D.O.E. Office of Science computational requirements
  – Augment with anticipated algorithm / science / technology trends
  – Workshop reports, individual discussions, etc.

• NERSC Workload overview
  – ~3000 users
  – 300 - 400 projects representing a broad range of science and algorithms
  – ~700 codes (>2 codes per project on average)
  – 15 science areas for 6 D.O.E Office of Science divisions.
NERSC 2008 Allocations
By DOE Office

<table>
<thead>
<tr>
<th>Office</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCR</td>
<td>Advanced Scientific Computing Research</td>
</tr>
<tr>
<td>BER</td>
<td>Biological &amp; Environmental Research</td>
</tr>
<tr>
<td>BES</td>
<td>Basic Energy Sciences</td>
</tr>
<tr>
<td>FES</td>
<td>Fusion Energy Sciences</td>
</tr>
<tr>
<td>HEP</td>
<td>High Energy Physics</td>
</tr>
<tr>
<td>NP</td>
<td>Nuclear Physics</td>
</tr>
</tbody>
</table>

- ASCR: 8%
- BER: 22%
- BES: 31%
- FES: 14%
- HEP: 14%
- NP: 11%
NERSC Allocations 2008
By Science Area

- Accelerator Physics; 7%
- Applied Math
- Astrophysics
- Chemistry
  - Climate Research
- Combustion
- Computer Sciences
- Engineering
- Environmental Sciences
- Fusion Energy
- Geosciences
- High Energy Physics
- Lattice Gauge Theory
- Life Sciences
- Materials Sciences

- Materials Sciences; 17%
- Climate Research; 16%
- Fusion Energy; 14%
- Combustion; 7%
- Lattice Gauge Theory; 13%
- Chemistry; 11%
- Astrophysics
NERSC workload is diverse and time varying.
Now we “drill down” to the code level.

What follows are some case studies:

- Science area
- Science driver
- Code distribution
- Best benchmark choice
- Some characteristics of the benchmarks.
Example: Climate Modeling

- CAM dominates CCSM computational requirements.

- FV-CAM increasingly replacing Spectral-CAM in future CCSM calculations.

- Drivers:
  - Critical support of U.S. submission to the Intergovernmental Panel on Climate Change (IPCC).
    - Schedule coincident with arrival of NERSC-6 system.
  - V & V for CCSM-4

- Focus on ensemble runs - 10 simulations per ensemble, 5-25 ensembles per scenario, relatively small concurrencies.
CAM Characteristics

- Unusual interprocessor communication topology – stresses interconnect.
- Relatively low computational intensity – stresses memory subsystem.
- MPI messages in bandwidth-limited regime.
- Limited parallelism requires faster processors (counter to current microprocessor architectural trends).

*Computational intensity is the ratio of # of Floating Point Operations to # of memory operations.
Material Science by Code

• 7,385,000 MPP hours awarded
• 62 codes, 65 users
• Typical code used in 2.15 allocation requests

<table>
<thead>
<tr>
<th>Code</th>
<th>MPP Hours</th>
<th>Percent</th>
<th>Cumulative%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 VASP</td>
<td>1,992,110</td>
<td>26%</td>
<td>26%</td>
</tr>
<tr>
<td>2 LSMS</td>
<td>600,000</td>
<td>8%</td>
<td>34%</td>
</tr>
<tr>
<td>3 FLAPW, DMoI3</td>
<td>350,000</td>
<td>5%</td>
<td>39%</td>
</tr>
<tr>
<td>4 CASINO</td>
<td>312,500</td>
<td>4%</td>
<td>43%</td>
</tr>
<tr>
<td>5 QBox</td>
<td>262,500</td>
<td>3%</td>
<td>46%</td>
</tr>
<tr>
<td>6 SIESTA</td>
<td>346,500</td>
<td>5%</td>
<td>51%</td>
</tr>
<tr>
<td>7 RGWBS</td>
<td>232,500</td>
<td>3%</td>
<td>54%</td>
</tr>
<tr>
<td>8 PEScan</td>
<td>220,000</td>
<td>3%</td>
<td>57%</td>
</tr>
<tr>
<td>9 PARATEC</td>
<td>337,500</td>
<td>4%</td>
<td>61%</td>
</tr>
<tr>
<td>10 PARSEC</td>
<td>182,500</td>
<td>2%</td>
<td>64%</td>
</tr>
<tr>
<td>Other</td>
<td>167,300</td>
<td>34%</td>
<td>66%</td>
</tr>
</tbody>
</table>
Materials Science by Algorithm

- **Density Functional Theory** codes
  - >70% of the MatSci. workload!
  - Majority are planewave DFT.

- Common requirements for DFT:
  - 3D global FFT
  - Dense Linear Algebra for orthogonalization of wave basis functions and calculating pseudopotential

- Dominant Code: VASP

- Science driver: nanoscience, ceramic crystals, novel materials, quantum dots, ...

- Similar Codes (planewave DFT)
  - Qbox, PARATEC
  - PETOT/PESCAN
PARATEC: Parallel Total Energy Code

- **Authors**: LBNL + UC Berkeley.
- **Relation to NERSC Workload**
  - Represents / captures the performance of a wide range of codes (VASP, CPMD, PETOT, QBox).
  - 70% of NERSC MatSci computation done via Plane Wave DFT codes.
- **Description**: Planewave DFT; calculation in both Fourier and real space; has custom 3-D FFT to transform between.
- **Coding**: 50,000 lines of Fortran90; uses SCALAPACK / FFTW / BLAS3; vectorizable version;
- **Parallelism**: fine-grain parallelism over DF grid points via MPI.
- **NERSC-6 tests**: strong scaling on 256 and 1024 cores.
- **Profile**: all-to-all data transpositions dominate communications time; Good differentiation between systems.
- **Special**: Also used for NSF Trac-I/II benchmarking.
Paratec Characteristics

- All-to-all communications
- Strong scaling emphasizes small MPI messages.
- Overall rate dominated by FFT speed and BLAS.
- Achieves high per-core efficiency on most systems.
- Good system discrimination.

<table>
<thead>
<tr>
<th>Total Message Count</th>
<th>256 cores</th>
<th>1024 cores</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 &lt;= MsgSz &lt; 256</td>
<td>114,432</td>
<td></td>
</tr>
<tr>
<td>256 &lt;= MsgSz &lt; 4KB</td>
<td>20,337</td>
<td>1,799,211</td>
</tr>
<tr>
<td>4KB &lt;= MsgSz &lt; 64KB</td>
<td>403,917</td>
<td>4,611</td>
</tr>
<tr>
<td>64KB &lt;= MsgSz &lt; 1MB</td>
<td>1,256</td>
<td>22,412</td>
</tr>
<tr>
<td>1 MB &lt;= MsgSz &lt; 16MB</td>
<td>2,808</td>
<td></td>
</tr>
</tbody>
</table>
Other Application Areas

- **Fusion**: 76 codes
  - 5 codes account for >50% of workload: OSIRIS, GEM, NIMROD, M3D, GTC
  - Further subdivide to PIC (OSIRIS, GEM, GTC) and MHD (NIMROD, M3D) code categories

- **Chemistry**: 56 codes for 48 allocations
  - Planewave DFT: VASP, CPMD, DACAPO
  - Quantum Monte Carlo: ZORI
  - Ab-initio Quantum Chemistry: Molpro, Gaussian, GAMESS
  - Planewave DFT dominates (but already covered in MatSci workload)
  - Small allocations Q-Chem category add up to dominant workload component

- **Accelerator Modeling**
  - 50% of workload consumed by 3 codes: VORPAL, OSIRIS, QuickPIC
  - Dominated by PIC codes

---

<table>
<thead>
<tr>
<th>Code</th>
<th>Award</th>
<th>Percent</th>
<th>Cumulative%</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSIRIS</td>
<td>2,112,500</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>GEM</td>
<td>2,058,333</td>
<td>11%</td>
<td>22%</td>
</tr>
<tr>
<td>NIMROD</td>
<td>2,229,167</td>
<td>12%</td>
<td>34%</td>
</tr>
<tr>
<td>M3D</td>
<td>1,921,667</td>
<td>10%</td>
<td>45%</td>
</tr>
<tr>
<td>GTC</td>
<td>1,783,333</td>
<td>10%</td>
<td>54%</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Code</th>
<th>Award</th>
<th>Percent</th>
<th>Cumulative%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZORI</td>
<td>695,000</td>
<td>12%</td>
<td>12%</td>
</tr>
<tr>
<td>MOLPRO</td>
<td>519,024</td>
<td>9%</td>
<td>21%</td>
</tr>
<tr>
<td>DACAPO</td>
<td>500,000</td>
<td>9%</td>
<td>29%</td>
</tr>
<tr>
<td>GAUSSIAN</td>
<td>408,701</td>
<td>7%</td>
<td>36%</td>
</tr>
<tr>
<td>CPMD</td>
<td>396,607</td>
<td>7%</td>
<td>43%</td>
</tr>
<tr>
<td>VASP</td>
<td>371,667</td>
<td>6%</td>
<td>49%</td>
</tr>
<tr>
<td>GAMESS</td>
<td>364,048</td>
<td>6%</td>
<td>56%</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Code</th>
<th>MPP Award</th>
<th>Percent</th>
<th>Cumulative%</th>
</tr>
</thead>
<tbody>
<tr>
<td>VORPAL</td>
<td>1,529,786</td>
<td>33%</td>
<td>33%</td>
</tr>
<tr>
<td>OSIRIS</td>
<td>784,286</td>
<td>16%</td>
<td>49%</td>
</tr>
<tr>
<td>QuickPIC</td>
<td>610,000</td>
<td>13%</td>
<td>62%</td>
</tr>
<tr>
<td>Omega3p</td>
<td>210,536</td>
<td>4%</td>
<td>66%</td>
</tr>
<tr>
<td>Track3p</td>
<td>210,536</td>
<td>4%</td>
<td>70%</td>
</tr>
</tbody>
</table>
Benchmark Selection Criteria

- **Coverage**
  - Cover science areas
  - Cover algorithm space

- **Portability**
  - Robust ‘build’ systems
  - Not an architecture specific implementation

- **Scalability**
  - Do not want to emphasize applications that do not justify scalable HPC resources

- **Open Distribution**
  - No proprietary or export-controlled code

- **Availability of Developer for Assistance/Support**
## NERSC-6 Application Benchmarks

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Science Area</th>
<th>Algorithm Space</th>
<th>Base Case Concurrency</th>
<th>Problem Description</th>
<th>Lang</th>
<th>Libraries</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAM</td>
<td>Climate (BER)</td>
<td>Navier Stokes CFD</td>
<td>56, 240 Strong scaling</td>
<td>D Grid, (~5° resolution); 240 timesteps</td>
<td>F90</td>
<td>netCDF</td>
</tr>
<tr>
<td>GAMESS</td>
<td>Quantum Chem (BES)</td>
<td>Dense linear algebra</td>
<td>384, 1024 (Same as Ti-09)</td>
<td>DFT gradient, MP2 gradient</td>
<td>F77</td>
<td>DDI, BLAS</td>
</tr>
<tr>
<td>GTC</td>
<td>Fusion (FES)</td>
<td>PIC, finite difference</td>
<td>512, 2048 Weak scaling</td>
<td>100 particles per cell</td>
<td>F90</td>
<td></td>
</tr>
<tr>
<td>IMPACT-T</td>
<td>Accelerator Physics (HEP)</td>
<td>PIC, FFT</td>
<td>256, 1024 Strong scaling</td>
<td>50 particles per cell</td>
<td>F90</td>
<td></td>
</tr>
<tr>
<td>MAESTRO</td>
<td>Astrophysics (HEP)</td>
<td>Low Mach Hydro; block structured -grid multiphysics</td>
<td>512, 2048 Weak scaling</td>
<td>16 32^3 boxes per proc; 10 timesteps</td>
<td>F90</td>
<td>Boxlib</td>
</tr>
<tr>
<td>MILC</td>
<td>Lattice Gauge Physics (NP)</td>
<td>Conjugate gradient, sparse matrix; FFT</td>
<td>256, 1024, 8192 Weak scaling</td>
<td>8x8x8x9 Local Grid, ~70,000 iters</td>
<td>C, assemb.</td>
<td></td>
</tr>
<tr>
<td>PARATEC</td>
<td>Material Science (BES)</td>
<td>DFT; FFT, BLAS3</td>
<td>256, 1024 Strong scaling</td>
<td>686 Atoms, 1372 bands, 20 iters</td>
<td>F90</td>
<td>Scalapack, FFTW</td>
</tr>
</tbody>
</table>
## Algorithm Diversity

<table>
<thead>
<tr>
<th>Science areas</th>
<th>Dense linear algebra</th>
<th>Sparse linear algebra</th>
<th>Spectral Methods (FFT)s</th>
<th>Particle Methods</th>
<th>Structured Grids</th>
<th>Unstructured or AMR Grids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accelerator Science</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Astrophysics</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Chemistry</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Climate</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Combustion</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Fusion</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Lattice Gauge</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Material Science</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

NERSC users require a system which performs adequately in all areas.
# N6 Benchmarks Coverage

<table>
<thead>
<tr>
<th>Science areas</th>
<th>Dense linear algebra</th>
<th>Sparse linear algebra</th>
<th>Spectral Methods (FFT)s</th>
<th>Particle Methods</th>
<th>Structured Grids</th>
<th>Unstructured or AMR Grids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accelerator Science</td>
<td>X</td>
<td></td>
<td>X IMPACT-T</td>
<td>X IMPACT-T</td>
<td>X IMPACT-T</td>
<td>X</td>
</tr>
<tr>
<td>Astrophysics</td>
<td>X MAESTRO</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X MAESTRO</td>
<td>X MAESTRO</td>
</tr>
<tr>
<td>Chemistry</td>
<td>X GAMESS</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Climate</td>
<td></td>
<td></td>
<td>X CAM</td>
<td></td>
<td>X CAM</td>
<td></td>
</tr>
<tr>
<td>Combustion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X MAESTRO</td>
<td>X AMR Elliptic</td>
</tr>
<tr>
<td>Fusion</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X GTC</td>
<td>X GTC</td>
<td>X</td>
</tr>
<tr>
<td>Lattice Gauge</td>
<td>X MILC</td>
<td>X MILC</td>
<td>X MILC</td>
<td></td>
<td>X MILC</td>
<td></td>
</tr>
<tr>
<td>Material Science</td>
<td>X PARATEC</td>
<td>X PARATEC</td>
<td></td>
<td></td>
<td>X PARATEC</td>
<td></td>
</tr>
</tbody>
</table>
**Summary: CI & %MPI**

<table>
<thead>
<tr>
<th></th>
<th>CAM</th>
<th>GAMESS</th>
<th>GTC</th>
<th>IMPACT-T</th>
<th>MAESTRO</th>
<th>MILC</th>
<th>PARATEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI*</td>
<td>0.67</td>
<td>0.61</td>
<td>1.15</td>
<td>0.77</td>
<td>0.24</td>
<td>1.39</td>
<td>1.50</td>
</tr>
<tr>
<td>Cray XT4 %Peak per Core (largest case)</td>
<td>13%</td>
<td>12%</td>
<td>24%</td>
<td>14%</td>
<td>5%</td>
<td>14%</td>
<td>44%</td>
</tr>
<tr>
<td>Cray XT4 %MPI Medium</td>
<td>29%</td>
<td>4%</td>
<td>9%</td>
<td>20%</td>
<td>20%</td>
<td>27%</td>
<td></td>
</tr>
<tr>
<td>Cray XT4 %MPI Large</td>
<td>35%</td>
<td>6%</td>
<td>40%</td>
<td>20%</td>
<td>23%</td>
<td>64%</td>
<td></td>
</tr>
<tr>
<td>Cray XT4 %MPI ExtraL</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>30%</td>
<td>n/a</td>
</tr>
<tr>
<td>Cray XT4 Avg Msg Size Med</td>
<td>113K</td>
<td>n/a</td>
<td>1 MB</td>
<td>35KB</td>
<td>2K</td>
<td>16KB</td>
<td>34KB</td>
</tr>
</tbody>
</table>

*CI is the computational intensity, the ratio of # of Floating Point Operations to # of memory operations.*
Traditional Sources of Performance Improvement are Flat-Lining

- **New Constraints**
  - 15 years of *exponential* clock rate growth has ended

- **But Moore’s Law continues!**
  - How do we use all of those transistors to keep performance increasing at historical rates?
  - Industry Response: #cores per chip doubles every 18 months *instead* of clock frequency!
Response to Technology Trends

• Parallel computing has thrived on weak-scaling for past 15 years

• Flat CPU performance increases emphasis on strong-scaling

• Benchmarks changed accordingly
  – **Concurrency**: Increased 4x over NERSC-5 benchmarks
  – **Strong Scaling**: Input decks emphasize strong-scaled problems
  – **Implicit Methods**: Added MAESTRO application benchmark
  – **Multiscale**: Added AMR Poisson benchmark
  – **Lightweight Messaging**: Added UPC FT benchmark
Summary So Far

• Codes represent important science and/or algorithms and architectural stress points such as CI*, message type/size/topology.
• Codes provide a good means of system differentiation during acquisition and validation during acceptance.
• Strong suite of scalable benchmarks (256-8192+ cores).

*CI = Computational Intensity, # FLOPs / Memory references
Other NERSC Benchmark Tests

Validation Efforts
Use a Hierarchy of Tests

Integration (reality) increases

Understanding Increases

Full Workload

composite tests

full application

stripped-down app

kernels

system

component tests
# Lower-Level Benchmarks

<table>
<thead>
<tr>
<th>CODE</th>
<th>PURPOSE / DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>STREAM</td>
<td>Single- and multi-core memory bandwidth.</td>
</tr>
<tr>
<td>FCT</td>
<td>Full-Configuration Test, run a single app over all cores; FFT mimics planewave DFT codes.</td>
</tr>
<tr>
<td>PSNAP</td>
<td>FWQ operating system noise test.</td>
</tr>
<tr>
<td>NAS PB serial &amp; 256-way MPI</td>
<td>Serial application performance on a single packed node; measures memory BW/ computation rate balance and compiler capabilities. Packed means all cores run.</td>
</tr>
<tr>
<td>NAS PB UPC</td>
<td>Measure performance characteristics not visible from MPI for FT benchmark.</td>
</tr>
<tr>
<td>Multipong</td>
<td>NERSC MPI PingPong for “latency” and BW, nearest- and furthest nodes in topology; also intra-node.</td>
</tr>
<tr>
<td>AMR Elliptic</td>
<td>C++/F90 LBNL Chombo code; proxy for AMR Multigrid elliptic solvers; 2 refinement levels; weak scaling with geometry replication; very sensitive to OS noise;</td>
</tr>
</tbody>
</table>
Validation & Benchmark Efforts

- XT4: DC & QC / CNL
- IBM p575 / AIX
- BG/P / LWK
- Sun QC Opteron + InfiniBand / Linux
- IBM Power6 / AIX
- SiCortex MIPS / Linux
- SGI DC Itanium / NumaLink
- Compilers PGI / Intel / PathScale / XLF
- Profiles from CrayPat and NERSC’s IPM
Composite Performance Metrics
Benchmark Hierarchy

Full Workload

- composite tests
  - full application
    - stripped-down app
      - kernels
        - system component tests
          - NPB Serial, NPB Class D, UPC NPB, FCT
          - Stream, PSNAP, Multipong, IOR, MetaBench, NetPerf
          - AMR Elliptic Solve
          - CAM, GTC, MILC, GAMESS, PARATEC, IMPACT-T, MAESTRO
          - SSP, ESP
Sustained System Performance (SSP)

• Aggregate, un-weighted measure of sustained computational capability relevant to NERSC’s workload.

• Geometric Mean of the processing rates of seven applications multiplied by $N$, # of cores in the system.
  – Largest test cases used.

• Uses floating-point operation count predetermined on a reference system by NERSC.

$$SSP \text{ in TFLOPS} = \frac{N \times \sqrt[7]{\prod P_i}}{1000}$$
NERSC-6 Composite SSP Metric

The largest concurrency run of each full application benchmark is used to calculate the composite SSP metric

NERSC-6 SSP

CAM 240p  GAMESS 1024p  GTC 2048p  IMPACT-T 1024p  MAESTRO 2048p  MILC 8192p  PARATEC 1024p

For each benchmark measure
• FLOP counts on a reference system
• Wall clock run time on various systems

36
Key Point - Sustained System Performance (SSP) Over Time

- Integrate the SSP over a particular time period.
- SSP can change due to:
  - System upgrades, Increasing # of cores, Software Improvements
- Allows evaluation of systems delivered in phases.
- Takes into account delivery date.
- Produces metrics such as SSP/Watt and SSP/$

\[ \text{Value} = \frac{\text{Potency}}{\text{Cost}}. \]

SSP Over 3 Year Period for 5 Hypothetical Systems

Area under SSP curve, when combined with cost, indicates system ‘value’
### Example of N6 SSP on Hypothetical System

<table>
<thead>
<tr>
<th>Hypothetical N6 System</th>
<th>Tasks</th>
<th>System Gflopct</th>
<th>Time</th>
<th>Rate per Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAM</td>
<td>240</td>
<td>57,669</td>
<td>408</td>
<td>0.589</td>
</tr>
<tr>
<td>GAMESS</td>
<td>1024</td>
<td>1,655,871</td>
<td>2811</td>
<td>0.575</td>
</tr>
<tr>
<td>GTC</td>
<td>2048</td>
<td>3,639,479</td>
<td>1493</td>
<td>1.190</td>
</tr>
<tr>
<td>IMPACT-T</td>
<td>1024</td>
<td>416,200</td>
<td>652</td>
<td>0.623</td>
</tr>
<tr>
<td>MAESTRO</td>
<td>2048</td>
<td>1,122,394</td>
<td>2570</td>
<td>0.213</td>
</tr>
<tr>
<td>MILC</td>
<td>8192</td>
<td>7,337,756</td>
<td>1269</td>
<td>0.706</td>
</tr>
<tr>
<td>PARATEC</td>
<td>1024</td>
<td>1,206,376</td>
<td>540</td>
<td>2.182</td>
</tr>
</tbody>
</table>

**GEOMETRIC MEAN**

| Rate Per Core = Ref. Gflop count / (Tasks*Time) |

**SSP (TF) = Geo mean of rates per core * # cores in system/1000**

- N6 SSP of 100,000 core system = \(0.7 \times \frac{100,000}{1000} = 70\)
- N6 SSP of 200,000 core system = \(0.7 \times \frac{200,000}{1000} = 140\)

**Allows vendors to size systems based on benchmark performance**
Benchmarking Methodology
• A selected vendor at NERSC is required to meet benchmark performance levels reported in the RFP response as a condition of acceptance…
  – and throughout the life of the subcontract.

• Includes all applications (with all inputs), all lower-level tests, SSP, and other tests, with strict constraints on variability across runs…
  – both in dedicated mode and production mode.
Base Case for Application Runs

- **Primary basis for comparison among proposed systems.**
- **Limits the scope of optimization.**
  - Modifications only to enable porting and correct execution.
- **Limits allowable concurrency to prescribed values.**
- **MPI only for all codes (even if OpenMP directives present).**
- **Fully packed nodes.**
- **Libraries okay (if generally supported).**
- **Hardware multithreading okay, too.**
  - Expand MPI concurrency to occupy hardware threads.
Optimized Case for Application Runs

- Allow the Offeror to highlight features of the proposed system.
- Applies to seven SSP apps only, all test problems.
- Examples:
  - Unpack the nodes;
  - Higher or lower concurrency than corresponding base case;
  - Hybrid OpenMP / MPI;
  - Source code changes for data alignment / layout;
  - Any / all of above.
- Caveat: number of tasks used to calculate SSP must use the total number of processors blocked from other use.
Summary

• Workload-based evaluation.
• Appropriate aggregate metrics.
• Formal methodology for tests, with stringent requirements based on proposed system.
• Wide range of tests from all levels of the benchmark hierarchy.
Scientists Need More Than Flop/s

- **Performance** — How fast will a system process work if everything is working well?
- **Effectiveness** — What is the likelihood that users can get the system to do their work?
- **Reliability** — The system is available to do work and operates correctly all the time.
- **Consistency** — How often will the system process users’ work as fast as it can?
- **Usability** — How easy is it for users to get the system to go as fast as possible?