Supercomputing:
HPCMP, Performance Measures and Opportunities

Cray J. Henry
August 2004

http://www.hpcmo.hpc.mil
Presentation Outline

- What’s New in the HPCMP
  - New hardware
  - HPC Software Application Institutes
  - Capability Allocations
  - Open Research Systems
  - On-demand Computing

- Performance Measures - HPCMP

- Performance Measures – Challenges & Opportunities
HPCMP Centers

1993

2004

Legend

\[\text{Legend} \]

- MSRCs
- ADCs and DDCs

Total HPCMP End-of-Year Computational Capabilities

Over 400X Growth
## Major Shared Resource Centers

### Army Research Laboratory (ARL)
- **System**: IBM P3, SGI Origin 3800, IBM P4, Linux Networx Cluster, LNX1 Xeon Cluster, IBM Opteron Cluster, SGI Altix Cluster
- **Processors**: 1,280 PEs, 256 PEs, 512 PEs, 768 PEs, 128 PEs, 256 PEs, 2,100 PEs

### Aeronautical Systems Center (ASC)
- **System**: Compaq SC-45, IBM P3, COMPAQ SC-40, SGI Origin 3900, SGI Altix Cluster
- **Processors**: 836 PEs, 528 PEs, 64 PEs, 2,048 PEs, 128 PEs, 32 PEs

### Engineer Research and Development Center (ERDC)
- **System**: Compaq SC-40, Compaq SC-45, SGI Origin 3800, Cray T3E, SGI Origin 3900, Cray X1
- **Processors**: 512 PEs, 512 PEs, 512 PEs, 1,888 PEs, 1,024 PEs, 64 PEs

### Naval Oceanographic Office (NAVO)
- **System**: IBM P4, SV1, IBM P4
- **Processors**: 1,408 PEs, 64 PEs, 3,456 PEs

### FY Dates
- **FY 01 and earlier**
- **FY 02**
- **FY 03**
- **FY 04**
# HPCMP Systems (ADCs)

<table>
<thead>
<tr>
<th>HPC Center</th>
<th>System</th>
<th>Processors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Army High Performance Computing Center (AHPCRC)</td>
<td>Cray T3E</td>
<td>1,088 PEs</td>
</tr>
<tr>
<td></td>
<td>Cray X1, LC</td>
<td>128 PEs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>64 PEs</td>
</tr>
<tr>
<td>Arctic Region Supercomputing Center (ARSC)</td>
<td>Cray T3E</td>
<td>272 PEs</td>
</tr>
<tr>
<td></td>
<td>Cray SV1</td>
<td>32 PEs</td>
</tr>
<tr>
<td></td>
<td>IBM P3</td>
<td>200 PEs</td>
</tr>
<tr>
<td></td>
<td>IBM Regatta P4</td>
<td>800 PEs</td>
</tr>
<tr>
<td></td>
<td>Cray X1</td>
<td>128 PEs</td>
</tr>
<tr>
<td>Maui High Performance Computing Center (MHPCC)</td>
<td>IBM P3 (2)</td>
<td>736/320 PEs</td>
</tr>
<tr>
<td></td>
<td>IBM Netfinity</td>
<td>512 PEs</td>
</tr>
<tr>
<td></td>
<td>Cluster</td>
<td>320 PEs</td>
</tr>
<tr>
<td></td>
<td>IBM P4</td>
<td></td>
</tr>
<tr>
<td>Space &amp; Missile Defense Command (SMDC)</td>
<td>SGI Origins</td>
<td>1,200 PEs</td>
</tr>
<tr>
<td></td>
<td>Cray SV-1</td>
<td>32 PEs</td>
</tr>
<tr>
<td></td>
<td>W.S. Cluster</td>
<td>64 PEs</td>
</tr>
<tr>
<td></td>
<td>IBM e1300 Cluster</td>
<td>256 PEs</td>
</tr>
<tr>
<td></td>
<td>Linux Cluster</td>
<td>256 PEs</td>
</tr>
<tr>
<td></td>
<td>IBM Regatta P4</td>
<td>32 PEs</td>
</tr>
</tbody>
</table>

**FY 01 and earlier**

**FY 02**

**FY 03**

**FY 04 upgrades**

Why is the date important?

Generally we see price-performance gains of ~ 1.68
(e.g., 2001 = 1

<table>
<thead>
<tr>
<th>Year</th>
<th>Price-Performance Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>1.68 x</td>
</tr>
<tr>
<td>2003</td>
<td>2.82 x</td>
</tr>
<tr>
<td>2004</td>
<td>4.74 x</td>
</tr>
</tbody>
</table>
## HPCMP Dedicated Distributed Centers

<table>
<thead>
<tr>
<th>Location</th>
<th>System</th>
<th>Description (Processors/Memory)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arnold Engineering Development Center (AEDC)</td>
<td>HP Superdome</td>
<td>32 PEs</td>
</tr>
<tr>
<td></td>
<td>IBM Itanium Cluster</td>
<td>16 PEs</td>
</tr>
<tr>
<td></td>
<td>IBM Regatta P4</td>
<td>64 PEs</td>
</tr>
<tr>
<td></td>
<td>Pentium Cluster</td>
<td>8 PEs</td>
</tr>
<tr>
<td>Air Force Research Laboratory, Information Directorate (AFRL/IF)</td>
<td>Sky HPC-1</td>
<td>384 PEs</td>
</tr>
<tr>
<td>Air Force Weather Agency (AFWA)</td>
<td>IBM Regatta P4</td>
<td>96 PEs</td>
</tr>
<tr>
<td></td>
<td>Heterogeneous HPC</td>
<td>96 PEs</td>
</tr>
<tr>
<td>Aberdeen Test Center (ATC)</td>
<td>Powerwulf</td>
<td>32 PEs</td>
</tr>
<tr>
<td>Fleet Numerical Meterology and Oceanography Center (FNMOC)</td>
<td>SGI Origin3900</td>
<td>256 PEs</td>
</tr>
<tr>
<td>Joint Forces Command (JFCOM)</td>
<td>IBM Regatta P4</td>
<td>96 PEs</td>
</tr>
<tr>
<td></td>
<td>Xeon Cluster</td>
<td>256 PEs</td>
</tr>
</tbody>
</table>

FY 04 new systems and/or upgrades

As of: April 2004
## HPCMP Dedicated Distributed Centers

<table>
<thead>
<tr>
<th>Location</th>
<th>System</th>
<th>Description (Processors/Memory)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Naval Air Warfare Center, Aircraft Division (NAWCAD)</strong></td>
<td>SGI Origin 2000</td>
<td>30 PEs</td>
</tr>
<tr>
<td></td>
<td>SGI Origin 3900</td>
<td>64 PEs</td>
</tr>
<tr>
<td><strong>Naval Research Laboratory-DC (NRL-DC)</strong></td>
<td>SUN Sunfire 6800</td>
<td>32 PEs</td>
</tr>
<tr>
<td></td>
<td>Cray MTA</td>
<td>40 PEs</td>
</tr>
<tr>
<td></td>
<td>SGI Altix</td>
<td>128 PEs</td>
</tr>
<tr>
<td></td>
<td>SGI Origin 3000</td>
<td>128 PEs</td>
</tr>
<tr>
<td><strong>Redstone Technical Test Center (RTTC)</strong></td>
<td>SGI Origin 3900</td>
<td>28 PEs</td>
</tr>
<tr>
<td><strong>Simulations &amp; Analysis Facility (SIMAF)</strong></td>
<td>SGI Origin 3900</td>
<td>24 PEs</td>
</tr>
<tr>
<td></td>
<td>Beowulf Cluster</td>
<td></td>
</tr>
<tr>
<td><strong>Space and Naval Warfare Systems Center-San Diego (SSCSD)</strong></td>
<td>Linux Cluster</td>
<td>128 PEs</td>
</tr>
<tr>
<td></td>
<td>IBM Regatta P4</td>
<td>128 PEs</td>
</tr>
<tr>
<td><strong>Whites Sands Missile Range (WSMR)</strong></td>
<td>Linux Networx</td>
<td>64 PEs</td>
</tr>
</tbody>
</table>

FY 04 new systems and/or upgrades

As of: April 2004
## Center POC’s

<table>
<thead>
<tr>
<th>Name</th>
<th>Org</th>
<th>Web URL</th>
<th>Contact Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brad Comes</td>
<td>HPCMO</td>
<td><a href="http://www.hpcmo.hpc.mil">http://www.hpcmo.hpc.mil</a></td>
<td>703-812-8205, <a href="mailto:bcomes@hpcmo.hpc.mil">bcomes@hpcmo.hpc.mil</a></td>
</tr>
<tr>
<td>Tom Kendall</td>
<td>ARL MSRC</td>
<td><a href="http://www.arl.hpc.mil">http://www.arl.hpc.mil</a></td>
<td>410-278-9195, <a href="mailto:tkendall@arl.army.mil">tkendall@arl.army.mil</a></td>
</tr>
<tr>
<td>Jeff Graham</td>
<td>ASC MSRC</td>
<td><a href="http://www.asc.hpc.mil/">http://www.asc.hpc.mil/</a></td>
<td>937-904-5135, <a href="mailto:Jeff.Graham@wpafb.af.mil">Jeff.Graham@wpafb.af.mil</a></td>
</tr>
<tr>
<td>Chris Flynn</td>
<td>AFRL Rome DC</td>
<td><a href="http://www.if.afrl.af.mil/technology/facilities/HPC/hpcf.html">http://www.if.afrl.af.mil/technology/facilities/HPC/hpcf.html</a></td>
<td>315-330-3249, <a href="mailto:Christopher.Flynn@rl.af.mil">Christopher.Flynn@rl.af.mil</a></td>
</tr>
<tr>
<td>Dr. Lynn Parnell</td>
<td>SSCSD DC</td>
<td><a href="http://www.spawar.navy.mil/sandiego/">http://www.spawar.navy.mil/sandiego/</a></td>
<td>619-553-1592, <a href="mailto:parnell@sscsd.hpc.mil">parnell@sscsd.hpc.mil</a></td>
</tr>
<tr>
<td>Maj Kevin Benedict</td>
<td>MHPCC DC</td>
<td><a href="http://www.mhpcc.edu">http://www.mhpcc.edu</a></td>
<td>808-874-1604, <a href="mailto:Kevin.Benedict@maui.afmc.af.mil">Kevin.Benedict@maui.afmc.af.mil</a></td>
</tr>
</tbody>
</table>
Disaster Recovery

Retain third-copy of critical data at a hardened backup site so users can access their files from an alternate site in the event of disruption of their primary support site.

- **Status:**
  - All MSRCs, MHPCC, and ARSC will have “off-site” third-copy backup storage for critical data.
  - On-going initiative.

- Working with centers to document the kinds of data that would need to be recovered.

- Implementation to begin Q1 FY05.
User Interface Toolkit

Provide an API-based toolkit to the user community and developers that facilitates the implementation of web-based interfaces to HPC

Facilitates Information Integration
Baseline Configuration

Implement and Sustain a Common Set of Capabilities and Functions Across the HPCMP Centers

Enables Users to Easily Move Between Centers Without the Requirement to Learn and Adapt to Unique Configurations
Software Applications Support

- Lasting impact on services
- High value service programs
- Tightly integrated software
- Address top DoD S&T and T&E problems
- Transfer of new technologies from universities
- On-site support
- Training
- Assure software intended use/user
- Protect software through source insertion

Fellows
NDSEG Interns
Growing Our Future

HPC Software Applications Institutes
HPC Software Portfolios
PET Partners
Software Protection
DoD Modernization Program
5–8 HPC Software (Applications) Institutes

- HPCMP chartered
- Service managed
- 3–6 year duration
  - Ends with Transition to Local Support
- $0.5–3M annual funding for:
  - 3–12 computational and computer scientists
  - Support development of new and existing codes
  - Adjust local business practice to use science-based models & simulation
- Integrated with PET

$8-12M
HPC Computational Fellowships

- Patterned after successful DOE fellowship program
- National Defense Science and Engineering Graduate Fellowship Program (NDSEG) chosen as vehicle for execution of fellowships
  - HPCMP added as fellowship sponsor along with Army, Navy, and Air Force
  - Computer and computational sciences added as possible discipline
- HPCMP is sponsoring 11 fellows for 2004 and similar numbers each following year
- HPCMP fellows are strongly encouraged to develop close ties with DoD laboratories or test centers, including summer research projects
- User organizations have responded to DUSD (S&T) memo with fellowship POCs to select and interact with fellows
HPCMP Resource Allocation Policy

Capability Allocations

Goal: Support the top capability work

How:

- New TI-XX resources generally are implemented for a few months before the end of the current fiscal year without formal allocation

- Dedicate major fractions of large new systems to short-term, massive computations that generally cannot be addressed under normal shared resource operations for the first 2–3 months of life

- HPCMP issued call for short-term Capability Application Project (CAP) proposals

- Capability Application Projects will be implemented between October and December on large new systems each year
  
  Proposals are required to show that the application efficiently used on the order of 1,000 processors or more and would solve a very difficult, important short-term computational problem
Call released to HPCMP community on 22 April 2004 with responses sent to HPCMPO by 1 June 2004

- 21 proposals received across all large CTAs (CSM, CFD, CCM, CEA, and CWO)

CAPs will be run on new 3,000 processor Power4+ at NAVO, 2,100 processor Xeon and 2,300 processor Opteron clusters at ARL

CAPs will be run in two phases:

- Exploratory phase designed to test scalability and efficiency of application codes to significant fractions of systems (5-15 projects on each system)

- Production phase designed to accomplish significant capability work with efficient, scalable codes (1-3 projects on each system)

Production phase of CAPs will be run after normal acceptance testing and pioneer work on these systems
“Open Research” Systems

- In response to customer demand: ~50% of Challenge Project leaders prefer to use an “open research” system

- “Open Research” systems concentrate on basic research allowing better separation of sensitive and non-sensitive information
  - minimal background check facilitating graduate student and foreign national access

- For FY05 the systems at ARSC will transition into an “open research” mode of operation
  - Eliminate the requirement for users of that system to have NACs
  - Customers would have to “certify” that their work is unclassified non-sensitive (e.g., open literature, basic research)
  - All other operational and security policies apply, such as all users of HPCMP resources must be valid DoD users assigned to a DoD computational project
  - Consistent with Uniform Use-Access Policy

- The account application process for “open research” centers or systems require certification by government program manager that computational work is cleared for open literature publication
  - Component of FY 2005 account request

- Operations on all other systems remain under current policies
On-demand (Interactive) Systems

- "Real-time" community has asked for "guaranteed" or on-demand service from shared resource centers
  - Request is aimed at ensuring quick response time from shared resource when system is being used interactively
  - Results needed now — can't wait

- Current policy requires that all Service/Agency work, be covered by an allocation
  - Note: "On-demand" system will have lower utilization but fast turn around
  - Service "valuation" of this service demonstrated by FY05 allocations — need sufficient allocation to dedicate a system to this mode of support

- Anticipating the Services/Agencies will allocate sufficient time to dedicate one 256 processor cluster at ARL
On-Demand Application
--Distributed Interactive HPC Testbed

- **Goal:** Assess the potential value and cost of providing greater interactive access to HPC resources to the DoD RDT&E community and its contractors.

- **Means:** Provide both unclassified and classified distributed HPC resources to the DoD HPC community in FY05 for interactive experimentation exploring new applications and system configurations.
Distributed Interactive HPC Testbed

Legend
- Remote Users
- Networked HPC’s
  Unclassified
  System in Black
- Classified
  Systems in Red

- MHPCC Koa Cluster Koa Cluster
- SSCSD Seahawk Seafarer
- AFRL Coyote Wile
- ASC Mach 2 Glenn
- ARL Powell

- Distributed HPC’s
- Accessed by authorized users anywhere on the DREN and Internet
- Interactive and time critical problems
- Low latency support for interactive and real-time applications—proper HPC configuration?
- Cohabitation of interactive and batch jobs?
- Web-based access to network of HPC’s with enhanced usability
- Consistency with HPCMP approved secure environment using DREN and SDREN
- Information management system supporting distributed HPC applications
- Demonstrating new C4ISR applications of HPC
- Expanding FMS use beyond Joint experimentation to include training and mission rehearsal
Objectives: to provide SIP users with a High Productivity Interactive Parallel MATLAB environment (it will provide the user-friendly MATLAB high-level language syntax plus the computational power of the interactive HPCs)

To allow interactive experiments for demanding SIP problems: problems that take too long to finish on a single Workstation, or that require more memory than what is available on a single computer, or systems with both constrains in which users’ research may benefit by an interactive modus-operandi.

Approach: to use MatlabMPI or other Parallel MATLAB viable approaches to deliver parallel execution but keeping the familiar MATLAB interactive environment

It may serve as a vehicle to collect experimental data about productivity issues: are SIP users really more productive on such an Interactive HPC MATLAB platform? (versus the traditional batch oriented HPCs)
<table>
<thead>
<tr>
<th>Site</th>
<th>Computer</th>
<th>Memory and I/O</th>
<th>Online</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARL MSRC</td>
<td><strong>Unclass- Powell:</strong> 128 node Dual 3.06MHz Xeon Cluster</td>
<td>2 GB DRAM and 64 GB disk/node, Myrinet &amp; GigEnet/100MB Backplane</td>
<td>Est. 10/04 w/batch; 4/05 share with batch,</td>
</tr>
<tr>
<td>Aberdeen, MD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC MSRC</td>
<td><strong>Unclass- Mach2:</strong> 24 node Dual 2.66 GHz Xeon, Linux Class- Glenn: 128 node dual Xeon, Linux</td>
<td>4 GB DRAM and 80 GB disk/node, dual GigEnet 4 GB DRAM and local disks</td>
<td>Est. 10/04</td>
</tr>
<tr>
<td>Dayton, OH</td>
<td></td>
<td></td>
<td>Est. Spring/05</td>
</tr>
<tr>
<td>AFRL</td>
<td><strong>Unclass- Coyote:</strong> 26 node Dual 3.06GHz Xeon, Linux Class- Wile: 14 node Dual 2.66/3.06 GHz Xeon, Linux</td>
<td>6 GB DRAM and 400 GB disk/node, dual GigEnet 6 GB DRAM and 200 GB disk/node, dual GigEnet</td>
<td>Yes</td>
</tr>
<tr>
<td>Rome, NY</td>
<td></td>
<td></td>
<td>Est. 12/04</td>
</tr>
<tr>
<td>SSCSD</td>
<td><strong>Unclass- Seahawk:</strong> 16 node 1.3GHz Itanium2, Linux Class- Seafarer: 24 node Dual 3.06 GHz</td>
<td>2 GB DRAM and 36 GB disk/node, dual GigEnet 4 GB DRAM and 80 GB disk/node, dual GigEnet</td>
<td>Est. 12/04</td>
</tr>
<tr>
<td>San Diego, CA</td>
<td></td>
<td></td>
<td>Yes (U) til 3/05</td>
</tr>
<tr>
<td>MHPCC</td>
<td><strong>Unclass/Class- Koa:</strong> 128 node dual Xeon, Linux (system moves between environments)</td>
<td>4 GB DRAM and 80 GB disk/node, shared file system, dual GigEnet</td>
<td>Yes</td>
</tr>
<tr>
<td>Maui, HI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Program</td>
<td>Contact Information</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------------------------------------</td>
<td>--------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Dr. Richard Linderman</td>
<td>HPC for Information Management</td>
<td>315-330-2208, <a href="mailto:Richard.Linderman@rl.af.mil">Richard.Linderman@rl.af.mil</a></td>
<td></td>
</tr>
<tr>
<td>Dr. Bob Lucas</td>
<td>USJFCOM J9</td>
<td>310-448-9449, <a href="mailto:rflucas@isi.edu">rflucas@isi.edu</a></td>
<td></td>
</tr>
<tr>
<td>Dr. Stan Ahalt</td>
<td>PET- SIP CTP</td>
<td>614-292-9524, <a href="mailto:ahalt@osc.edu">ahalt@osc.edu</a></td>
<td></td>
</tr>
<tr>
<td>Dr. Juan Carlos Chaves</td>
<td>Interactive Parallel MATLAB</td>
<td>410-278-7519, <a href="mailto:jchaves@arl.army.mil">jchaves@arl.army.mil</a></td>
<td></td>
</tr>
<tr>
<td>Dr. Dave Pratt</td>
<td>SBA Force transformations</td>
<td>407-243-3308, <a href="mailto:David.R.Pratt@saic.com">David.R.Pratt@saic.com</a></td>
<td></td>
</tr>
<tr>
<td>Rob Ehret</td>
<td>Grid-based Collaboration</td>
<td>937-904-9017, <a href="mailto:Robert.Ehret@sensors.wpafb.af.mil">Robert.Ehret@sensors.wpafb.af.mil</a></td>
<td></td>
</tr>
<tr>
<td>Bill McQuay</td>
<td></td>
<td>937-904-9214, <a href="mailto:William.Quay@sensors.wpafb.af.mil">William.Quay@sensors.wpafb.af.mil</a></td>
<td></td>
</tr>
<tr>
<td>Dr. John Nehrbass</td>
<td>Web enabled HPC</td>
<td>937-904-5139, <a href="mailto:John.Nehrbass@wpafb.af.mil">John.Nehrbass@wpafb.af.mil</a></td>
<td></td>
</tr>
<tr>
<td>Dr. Keith Bromley</td>
<td>Signal Image Processing</td>
<td>619-553-2535, <a href="mailto:bromley@spawar.navy.mil">bromley@spawar.navy.mil</a></td>
<td></td>
</tr>
<tr>
<td>Dr. George Ramseyer</td>
<td>Hyperspectral Image Exploitation</td>
<td>315-330-3492, <a href="mailto:George.Ramseyer@rl.af.mil">George.Ramseyer@rl.af.mil</a></td>
<td></td>
</tr>
<tr>
<td>Richard Pei</td>
<td>Interactive Electromagnetics Sim</td>
<td>732-532-0365, <a href="mailto:Richard.Pei@us.army.mil">Richard.Pei@us.army.mil</a></td>
<td></td>
</tr>
<tr>
<td>Dr. Ed Zelnio</td>
<td>3-D SAR Radar Imagery</td>
<td>937-255-4949 ext.4214, <a href="mailto:Ed_Zelnio@mbvlab.wpafb.af.mil">Ed_Zelnio@mbvlab.wpafb.af.mil</a></td>
<td></td>
</tr>
<tr>
<td>John Rooks</td>
<td>Swathbuckler SAR Radar Imagery</td>
<td>315-330-2618, <a href="mailto:John.Rooks@rl.af.mil">John.Rooks@rl.af.mil</a></td>
<td></td>
</tr>
</tbody>
</table>
Performance Measurement Goals

- Provide Quantitative measures to support selection of computers in annual procurement process (TI-XX)

- Develop an understanding of our key application codes for the purpose of guiding code developers and users toward more efficient applications and machine assignments

- Replace the current application benchmark suite with a judicious choice of synthetic benchmarks that could be used to predict performance of any HPC architecture on the program’s key applications
Resource Management
— Integrated Requirements/Allocation/Utilization Process

Requirements Process
- Bottoms-up survey
- Includes only approved funded S&T/T&E projects
- Reviewed and validated by S&T/T&E executives

Capacity Allocation Process
- 75% Service/Agency, 25% DoD Challenge Projects
- Services/Agencies decide allocation resources for each project
- Reconcile capacity with requirements (first-order prioritization)

Capability Allocation Process
- Services/Agencies decide allocation resources for each project
- Reconcile capacity with requirements (first-order prioritization)

Operations Decisions
Acquisition Decisions

Utilization Tracking
- Track utilization by project
- Monitor turnaround time for timely execution

User Feedback
- Direct feedback from PI and individual users
- Summary report sent to each HPC Center
- Issue addressed and resolved
- User satisfaction impacts requirements, allocation, and utilization statistics
Technology Insertion (TI) Flow Chart

1. Requirements Update
   - Update Acquisition Plan

2. Update Selection Criteria
   - Benchmark Performance and Price/Performance
   - Usability

3. Update Benchmarks
   - Applications
   - Synthetics

4. Issue call to HPC vendors

5. Vendors prepare bids

6. Evaluate results and build possible solution sets

7. Invite solution set bids and guaranteed benchmark results

8. System(s) Delivered

9. Benchmark Tests

10. Vendors prepare bids including benchmark performance

11. Evaluate results and negotiate final deal

12. System(s) Accepted
Types of Benchmark Codes

- Synthetic codes
  - Basic hardware and system performance tests
  - Meant to determine expected future performance
  - Scalable, quantitative synthetic tests will be used for scoring and others will be used as system performance checks by Usability Team

- Application codes
  - Actual application codes as determined by requirements and usage
  - Meant to indicate current performance
### Percentage of Unclassified Non-Real-Time Requirements, Usage, and Allocations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CFD</td>
<td>[35.5%] (36.9%) {38.6%}</td>
<td>48.3% (37.2%)</td>
<td>40.7% (44.4%)</td>
<td>[43.3%] (41.6%) {41.2%}</td>
</tr>
<tr>
<td>CCM</td>
<td>[15.5%] (18.6%) {16.2%}</td>
<td>16.4% (21.2%)</td>
<td>14.2% (12.6%)</td>
<td>[14.2%] (15.9%) {15.7%}</td>
</tr>
<tr>
<td>CWO</td>
<td>[21.9%] (19.2%) {20.8%}</td>
<td>21.3% (23.1%)</td>
<td>21.9% (17.6%)</td>
<td>[23.3%] (21.1%) {19.8%}</td>
</tr>
<tr>
<td>CEA</td>
<td>[4.1%] (4.0%) {4.8%}</td>
<td>5.1% (4.8%)</td>
<td>8.2% (6.6%)</td>
<td>[4.9%] (6.4%) {5.7%}</td>
</tr>
<tr>
<td>CSM</td>
<td>[11.4%] (11.8%) {11.7%}</td>
<td>3.5% (7.5%)</td>
<td>9.6% (11.0%)</td>
<td>[8.3%] (8.6%) {10.3%}</td>
</tr>
<tr>
<td>EQM</td>
<td>[3.0%] (3.2%) {2.1%}</td>
<td>0.6% (1.6%)</td>
<td>4.0% (3.1%)</td>
<td>[2.3%] (3.0%) {2.4%}</td>
</tr>
<tr>
<td>SIP</td>
<td>[1.0%] (1.4%) {1.4%}</td>
<td>1.2% (1.1%)</td>
<td>0.2% (0.4%)</td>
<td>[0.4%] (0.7%) {0.8%}</td>
</tr>
<tr>
<td>CEN</td>
<td>[0.5%] (0.4%) {0.6%}</td>
<td>1.3% (1.2%)</td>
<td>0.1% (1.2%)</td>
<td>[1.4%] (0.5%) {1.1%}</td>
</tr>
<tr>
<td>IMT</td>
<td>[2.9%] (0.8%) {0.8%}</td>
<td>2.1% (0.7%)</td>
<td>0.7% (1.9%)</td>
<td>[0.9%] (1.1%) {1.3%}</td>
</tr>
<tr>
<td>Other</td>
<td>[1.3%] (1.2%) {0.2%}</td>
<td>0.1% (0.8%)</td>
<td>0.2% (0.7%)</td>
<td>[0.4%] (0.4%) {0.6%}</td>
</tr>
<tr>
<td>FMS</td>
<td>[2.9%] (2.6%) {2.9%}</td>
<td>0.2% (0.8%)</td>
<td>0.2% (0.4%)</td>
<td>[0.7%] (0.8%) {1.1%}</td>
</tr>
</tbody>
</table>
TI-05 Application Benchmark Codes

- Aero – Aeroelasticity CFD code (single test case)
  (Fortran, serial vector, 15,000 lines of code)
- AVUS (Cobalt-60) – Turbulent flow CFD code
  (Fortran, MPI, 19,000 lines of code)
- GAMESS – Quantum chemistry code
  (Fortran, MPI, 330,000 lines of code)
- HYCOM – Ocean circulation modeling code
  (Fortran, MPI, 31,000 lines of code)
- OOCore – Out-of-core solver
  (Fortran, MPI, 39,000 lines of code)
- RFCTH2 – Shock physics code
  (~43% Fortran/~57% C, MPI, 436,000 lines of code)
- WRF – Multi-Agency mesoscale atmospheric modeling code (single test case)
  (Fortran and C, MPI, 100,000 lines of code)
- Overflow-2 – CFD code originally developed by NASA
  (Fortran 90, MPI, 83,000 lines of code)
<table>
<thead>
<tr>
<th>CTA</th>
<th>Benchmark</th>
<th>Size</th>
<th>Unclassified %</th>
<th>Classified %</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSM</td>
<td>RF-CTH</td>
<td>Standard</td>
<td>a%</td>
<td>A%</td>
</tr>
<tr>
<td>CSM+CFD</td>
<td>RF-CTH</td>
<td>Large</td>
<td>b%</td>
<td>B%</td>
</tr>
<tr>
<td>CFD</td>
<td>Cobalt60</td>
<td>Standard</td>
<td>c%</td>
<td>C%</td>
</tr>
<tr>
<td>CFD</td>
<td>Cobalt60</td>
<td>Large</td>
<td>d%</td>
<td>D%</td>
</tr>
<tr>
<td>CFD</td>
<td>Aero</td>
<td>Standard</td>
<td>e%</td>
<td>E%</td>
</tr>
<tr>
<td>CEA+SIP</td>
<td>OOCore</td>
<td>Standard</td>
<td>f%</td>
<td>F%</td>
</tr>
<tr>
<td>CEA+SIP</td>
<td>OOCore</td>
<td>Large</td>
<td>g%</td>
<td>G%</td>
</tr>
<tr>
<td>CCM+CEN</td>
<td>GAMESS</td>
<td>Standard</td>
<td>h%</td>
<td>H%</td>
</tr>
<tr>
<td>CCM+CEN</td>
<td>GAMESS</td>
<td>Large</td>
<td>i%</td>
<td>I%</td>
</tr>
<tr>
<td>CCM</td>
<td>NAMD</td>
<td>Standard</td>
<td>j%</td>
<td>J%</td>
</tr>
<tr>
<td>CCM</td>
<td>NAMD</td>
<td>Large</td>
<td>k%</td>
<td>K%</td>
</tr>
<tr>
<td>CWO</td>
<td>HYCOM</td>
<td>Standard</td>
<td>l%</td>
<td>L%</td>
</tr>
<tr>
<td>CWO</td>
<td>HYCOM</td>
<td>Large</td>
<td>m%</td>
<td>M%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
Emphasis on Performance

- Establish a DoD standard benchmark time for each application benchmark case
  - NAVO IBM Regatta P4 (Marcellus) chosen as standard DoD system for TI-04 (Initially IBM SP3 – HABU)

- Benchmark timings (at least three on each test case) are requested for systems that meet or beat the DoD standard benchmark times by at least a factor of two (preferably up to four)

- Benchmark timings may be extrapolated provided they are guaranteed, but at least one actual timing on the offered or closely related system must be provided
2004 HPEC Conference

CTH Standard

NAVO IBM SP P3 — 1288 Processors

y = 4.57590E-05x^{7.15387E-01}

R^2 = 9.94381E-01

x = Number of Processors
y = 1/Time

"Slope"
"Curvature"
"Goodness of Fit"
HPCMP System Performance
(Unclassified)

FY 2003
FY 2004

Normalized Habu Equivalents

~40

= number of application test cases not included
(out of 13 total)

Cray T3E  IBM P3  SGI O3800  IBM P4  HP SC40  HP SC45  Cray X1  SGI O3900

System
### How the Optimizer Works:

**Problem Description**

#### KNOWN

<table>
<thead>
<tr>
<th>Prices</th>
<th>Application Score Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>S S S S S S S S S</td>
</tr>
<tr>
<td>$</td>
<td>S S S S S S S S S</td>
</tr>
<tr>
<td>$</td>
<td>S S S S S S S S S</td>
</tr>
<tr>
<td>$</td>
<td>S S S S S S S S S</td>
</tr>
<tr>
<td>$</td>
<td>S S S S S S S S S</td>
</tr>
<tr>
<td>$</td>
<td>S S S S S S S S S</td>
</tr>
<tr>
<td>$</td>
<td>S S S S S S S S S</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Application Test Case Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>% % % % % % % % %</td>
</tr>
</tbody>
</table>

#### UNKNOWN

<table>
<thead>
<tr>
<th>Optimal Quantity Set</th>
<th>Workload Distribution Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td># # # # # # # #</td>
<td>% % % % % % % % % % % % % % % % % % % % % % % %</td>
</tr>
<tr>
<td># # # # # # # #</td>
<td>% % % % % % % % % % % % % % % % % % % % % % %</td>
</tr>
<tr>
<td># # # # # # # #</td>
<td>% % % % % % % % % % % % % % % % % % % % % %</td>
</tr>
<tr>
<td># # # # # # # #</td>
<td>% % % % % % % % % % % % % % % % % % % % %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Application Test Case Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>% % % % % % % % %</td>
</tr>
</tbody>
</table>

**Budget Limits**

| % % % % % % % % % |

**Optimize Total Price/Performance**
## Price Performance Based Solutions

The optimizer produces a list of system solutions in rank order based upon Performance / Life Cycle Cost.

<table>
<thead>
<tr>
<th>System</th>
<th>Total # Proc</th>
<th>Opt # 1</th>
<th>Opt # 2</th>
<th>Opt # 3</th>
<th>Opt # 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>64</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>188</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>128</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>256</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>256</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>D</td>
<td>512</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>256</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td><strong>Performance / Life Cycle</strong></td>
<td><strong>3.03</strong></td>
<td><strong>3.02</strong></td>
<td><strong>2.97</strong></td>
<td><strong>2.95</strong></td>
<td></td>
</tr>
</tbody>
</table>
Capturing True Performance

Benchmarks

Top 500 or Peak GFlops is not a Measure of Real Performance

At the end of TI-03
The slope of this semi-log plot for the entire set of data equates to a constant factor of $1.76 \pm 0.26$, although the slopes for the last two years have been 1.42 and 1.48, respectively.
HPCMP Benchmarking and Performance Modeling

Challenges & Opportunities

http://www.hpcmo.hpc.mil
Benchmarks

**Today**

- Dedicated Applications
  - 80% weight
  - Real codes
  - Representative data sets

- Synthetic Benchmarks
  - 20% weight
  - Future look
  - Focus on key machine features

**Tomorrow**

- Synthetic Benchmarks
  - 100% weight
  - Coordinated to application “signature”
  - Performance on real codes accurately predicted from synthetic benchmark results
  - Supported by genuine “signature” databases

*Next 1–2 years key — must prove that synthetics benchmarks and application “signatures” can be coordinated*
How -- Application Code Profiling Plan

- Began at behest of HPC User Forum in partnership with NSA
- Has evolved to multi-year plan -- how key application codes perform on HPC systems
  - Maximizing use of current HPC resources
  - Predicting performance of future HPC resources
- Performers include
  - Programming Environment and Training (PET) partners
  - Performance Modeling and Characterization Laboratory (PMaC) at SDSC
  - Computational Science and Engineering Group at ERDC
  - Instrumental, Inc.
- Research and production activities include
  - Profiling key DoD application codes at several different levels
  - Characterizing HPC systems with a set of system probes (synthetic benchmarks)
  - Predicting HPC system performance based on application profiles
  - Determining a minimal set of HPC system attributes necessary to model performance
  - Constructing the appropriate set of synthetic benchmarks to accurately model the HPCMP computational workload to use in system acquisitions
Support for TI-05 (Scope and Schedule)

- **Level 3 application code profiling**
  - Eight application codes – 14 unique test cases
  - Each test case to be run at 3 different processor counts

- **Predictions for existing systems**
  - 21 systems at 7 centers (some overlap possible in predictions)
  - Benchmarking POCs identified for each center
  - Goal: benchmarking results and predictions complete by Dec 2004

- **Predictions for offered systems**
  - Goal: benchmarking results finalized by 19 November 2004; all predictions completed by 31 December 2004

- **Sensitivity Analysis**
  - Goal: Determine how accurate a prediction do we need.
Should We Do Uncertainty Analysis?
Overall goal: Understand and accurately estimate uncertainties in performance predictions

Determine functional form of performance prediction equations and develop uncertainty equation

Determine uncertainties in underlying measured values from system probes and application profiling and use uncertainty equation to estimate uncertainties

Compare results of performance prediction to measured timings and uncertainties of these results to predicted uncertainties

Assess uncertainties in measured timings and determine whether acceptable agreement is obtained

Eventual goal: Propagate uncertainties in performance prediction to determine uncertainties in acquisition scoring
Assumption: Uncertainties in measured performance values can be treated as uncertainties in measurements of physical quantities.

For small, random uncertainties in measured values $x, y, z, \ldots$, the uncertainty in a calculated function $q (x, y, z \ldots)$ can be expressed as:

$$
\delta q = \sqrt{\left(\frac{\partial q}{\partial x} \delta x\right)^2 + \cdots + \left(\frac{\partial q}{\partial z} \delta z\right)^2}
$$

Systematic errors need careful consideration since they cannot be calculated analytically.
Propagation of Uncertainties in Benchmarking and Performance Modeling

- Benchmark Times
  - \( \delta T \)
  - \( \frac{1}{T} \)
  - Optimizer
  - Total Performance for Solution Set
  - \( \sigma_{TS} \)

- Benchmark Performance
  - \( \delta P, \sigma_P \)
  - Least Squares Fit
  - Price/Performance for Solution Set
  - Averaging over spans of Solution Sets
  - \( \sigma_{\$/TS} \)

- Benchmark Scores
  - \( \sigma_S \)
  - Power Law
  - Rank Ordering of Solution Sets
  - \( \sigma_{\%} \)
U (EXIST+LC) Architecture % Selection by Processor Quantity for Varying Spans (TI-04)
Performance Measurement – Closing Thoughts

- Clearly identify your goals
  - Maximize the amount of work given fixed $ and time.
  - Alternative goals: power consumption, weight, volume

- Define Work Flow
  - Production (run) time
  - Alternative goals: development time, problem set-up time, result analysis time

- Validate Measures
  - Understand the error bounds

- Don’t rely on “Marketing” specifications!