Table of Contents 1
for the Computing Metrics Report Appendices

8. Project Checklist and Metrics of DOD Engineering Projects Presentation 7/18/06 -Doug Post

DOD Project Survey of code, software engineering techniques, etc.

9. Project Presentation Template

Four Quadrant View of a Project including basic information and goals; resources being supplied by centers, code and project management, and scientific outputs

10. Ten Project Showcase Prepared by Centers 7/17/06 from ANL , ORNL, NERSC

Ten projects used the project presentation template for description.

Department of Energy
Office of Science
Washington, DC 20585

March 10, 2006

Dr. Jill P. Dahlburg, Chair
Naval Research Laboratory, Code 1001
4555 Overlook Avenue
Washington, DC 20375

Dear Dr. Dahlburg:

I am requesting that the Advanced Scientific Computing Advisory Committee (ASCAC) convene a sub-panel to examine the issue of science based performance metrics for the present and proposed computational facilities for the Office of Science (SC).

BACKGROUND

The Office of Advanced Scientific Computing Research (ASCR) is undergoing substantial changes over the next two – five years in building out petascale computational facilities for science. These facilities will include two capability systems at Oak Ridge and Argonne national laboratories, with a peak performance of one petaflop and 200 – 500 teraflops, respectively, and a capacity system at the Lawrence Berkeley National laboratory with an aggregate peak capacity of 500 teraflops.

CHARGE

The sub-panel should weigh and review the approach to performance measurement and assessment at these facilities, the appropriateness and comprehensiveness of the measures, and the science accomplishments and their effects on the Office of Science's science programs. Additionally, the sub-panel should consider the evolution of the roles of these facilities and the computational needs over the next three – five years, so that SC programs can maintain their national and international scientific leadership.

In addition to the above, the sub-panel is asked to provide input for the Office of Management and Budget (OMB), evaluation of ASCR progress towards the long-term goals specified in the OMB Program Assessment Rating Tool (PART). See attached enclosure. Note that the OMB guidelines specify ratings of excellent, good, fair, poor, or not acceptable. In addition to these ratings, comments on observed strengths or deficiencies in the management of any component or sub-component of ASCR's portfolio and suggestions for improvement would be very valuable.

I would like a report on the findings and recommendations at the November 2006 ASCAC meeting. I appreciate ASCAC's willingness to undertake this important activity.

Sincerely,

Raymond L. Orbach
Director

Enclosure

ATTACHMENT

ASCR PART Long Term Measures

- By 2015, demonstrate progress toward developing the mathematics, algorithms, and software that enable effective scientifically critical models of complex systems, including highly nonlinear or uncertain phenomena, or processes that interact on vastly different scales or contain both discrete and continuous elements.
    - Definition of "Excellent" – ASCR supported research develops the mathematics needed for effective modeling of complex systems. Algorithms implementing many of these mathematical techniques are developed. The most promising algorithms have been selected, and software deploying these algorithms has been created and disseminated in a number of scientific disciplines.
    - Definition of "Good" – ASCR supported research significantly advances the mathematics needed for effective modeling of complex systems. Algorithms implementing several of the mathematical techniques are developed showing the potential of these techniques to enable new scientific discovery.
    - Definition of "Fair" – ASCR supported research modestly advances the mathematics needed for effective modeling of complex systems.
    - Definition of "Poor" – ASCR supported research leads to limited progress in understanding the mathematics of complex systems.
    - How will progress be measured? – Expert Review every three years will rate progress as "Excellent", "Good", "Fair", or "Poor".

- By 2015, demonstrate progress toward developing, through the Genomes to Life partnership with the Biological and Environmental Research program, the computational science capability to model a complete microbe and a simple microbial community.
    - Definition of "Excellent" – In partnership with BER, develop a computational model that accurately describes the potential of a microbial community to clean up waste, sequester carbon, or produce hydrogen, validated experimentally by the use or reengineering of that community based on model predictions.
    - Definition of "Good" – In partnership with BER, develop a computational model that accurately describes the potential of a microbial community to clean up waste, sequester carbon, or produce hydrogen, validated by its consistency with available data.
    - Definition of "Fair" – In partnership with BER, develop a number of the components of a computational model that could accurately describe the potential of a microbial community to clean up waste, sequester carbon, or produce hydrogen.
    - Definition of "Poor" – In partnership with BER, produce a modest output of computational research that could lead to the development of

components of models to describe the potential of microbial communities to clean up waste, sequester carbon, or produce hydrogen.

- How will progress be measured? – Expert Review every three years will rate progress as "Excellent", "Good", "Fair", or "Poor".

KEY: Processors x K; Processor/P and Computer/C speed (Gflop & Tflop); Mp & Ms = primary & disk memory (Tbytes)
Processor hours per year in Millions;
Processor FLT. PT. operations per year in Peta-flop hours speed adjusted = Processor hours/year In millions x processor speed in GFLOPS
***PF = Petaflop. Petaflop-hr/year = 8736 (a one petaflops computer running for a year). Delivers 31.4 zeta-fl-op = .031 yotta-flop**
One Teraflop-hr year delivers 21.4 exa-flop
Hours/year 8736 128 processors operating for 1 year delivers 1.12 Million hours

Gary/Ward LLNL Ratios per Processor TF
Memory= 0.5 TB; Disk=20 TB; I/O= 1 GBps; Network rae = 0.1 GBps

Ratios…from Gary/Ward LLNL		**nodes/n**	**Proc(K)**	**P.speed(GF)**	**1.0 C.pk(TF)**	**0.5 Mp(TB)**	**20.0 Ms(TB)**	**Proc-hr/yr(M)**	**PF*hr/yr**		**O/S**
DOE Installed Machines											
NERSC	Seaborg	416	6.66	1.4	9.1	6.6	44.0	58.2	79.5	cluster, 16P/n	IBM/AIX
	Bassi	111	0.89	7.5	6.7	3.6	100.0	7.8	58.5	cluster, 7P/n	IBM/AIX
	Jacquard	356	0.71	4.4	3.1	2.1	30.0	6.2	27.1	cluster, 2P/n	Linux
	DaVinci	1	0.03	5.6	0.2	0.2	30.0	0.3	1.6	smP	SGI
			8.29		19.1			72.4	166.7		
ORNL	Phoenix.05	1	1.02	17.6	18.0	2.0	32.0	8.9	157.2	smPv	Cray X-1
	pSeries		0.86	5.2	4.5			7.5	39.3		
	Jaguar.05-06	5212	10.42	2.6	27.1	20.8	120.0	91.1	236.8	cluster	Cray XT3
			12.31		49.6			107.6	433.3		
	Jaguar.06	*11,500*	*23.02*	*4.3*	*100.0*	*45.0*	*900.0*	*201.1*	*873.6*	*cluster, 2Pn*	
	Jaguar.07		*35.61*	*7.0*	*250.0*	*70.0*	*900.0*	*311.1*	*2184.0*		
	Baker late 08		*100.00*	*10.0*	***1000.0***	*200-400*		*873.6*	*8736.0*		
ANL	BG Solution		2.05	2.8	5.7			17.9	49.8		
	2007-08 first				*100.0*						
	ALCF in 1,000	*72,000*	*288.00*	*3.5*	***1000.0***	*288.0*		*2516.0*	*8736.0*		
DOE Total Installed			22.65		74.38			197.89	649.80		
NSF Ceenters Machines											
NCSA	Tungsten		2.50	6.1	15.3			21.8	133.7		
	Teragrid		1.78	5.8	10.3			15.5	90.0		
	Tungsten2		1.02	7.2	7.4			8.9	64.6		
	SGI		1.02	6.0	6.1			8.9	53.3		
			6.324		39.1			55.2	341.6		
SDSC	DataStar	300	2.53	6.2	15.6	7.3	115.0	22.1	136.3	cluster 8+	
	Teragrid Cluster		0.51	8.0	4.1	1.0	40.0	4.5	35.8	cluster Itanium	
	Intimidata		2.05	2.8	5.7	0.5	20.0	17.9	50.1	cluster /I	
			5.09		25.4			44.4	222.2		
PSC	Big Ben		2.09	4.7	9.9			18.3	86.5		
	Alpha		3.02	2.0	6.0			26.3	52.4		
			5.106		15.9			44.6	138.9		
NCAR	bluesky	50	1.6	5.2	8.3	3.3	27.5	14.0	72.9	IBM p690 POWER4	AIX cluster (76x8 LPARs; 25x32 LPARs)
	bluevista	78	0.624	7.6	4.7	1.2	55.0	5.5	41.5	IBM p575 POWER5	Clusster 8P/n
	lightning	128	0.256	4.4	1.1	0.3	6.6	2.2	9.9	IBM e1350 Opteron	SuSE Linux 2P/n
	pegasus	64	0.128	4.4	0.6	0.1	3.7	1.1	4.9	IBM e1350 Opteron	SuSE Linux 2P/n
	frost	1024	2.048	2.8	5.7	0.5	6.5	17.9	50.2	IBM BlueGene/L	CNK & SuSE Linux
	tempest	1	0.128	1	0.1	0.1	4.2	1.1	1.1	SGI Origin3800	Irix
			4.8	25.4		5.5	103.5	41.9	180.6		
								83.8	361.2		
Total NSF			21.3	25.4		5.5	103.5	186.2	1244.4		
NSF Total Installed											

DoD center computers as of July 1, 2006, future computer procurements are determined through competitive acquisition process

Dod ERDC	SGI 3900	1	1.0	1.024	1.0	8.9	9	SGI	smP
Vicksburg,MS	Cray XT3	4.176	5.2	21.7	8.4	36.5	190	Cray	Cluster
	Cray X1	0.256	3.2	0.819	0.3	2.2	7	Cray	smPv
	Compaq SC45	0.512	2.0	1.024	0.5	4.5	9	Compaq	Sierra Cluster
DoD NAVO	IBM P4	1.408	5.2	7.32	1.4	12.3	64	IBM	AIX
BaySt.Louis,MS	IBM P4	2.944	6.8	20.019	6.0	25.7	175	IBM	AIX
	IBM P4	0.512	6.8	3.482	0.7	4.5	30	IBM	AIX
	IBM P5	3.072	7.6	23.347	6.1	26.8	204	IBM	AIX
	IBM P5	1.92	7.6	14.592	3.8	16.8	127	IBM	AIX
DoD ARL	IBM P4	0.128	6.8	0.87	0.1	1.1	8	IBM	AIX
Aberdeen,MD	LNXi Cluster	4.206	12.0	50.3	9.0	36.7	439	LNXi	Woodcrest
	LNXi Cluster	3.368	6.4	21.555	6.7	29.4	188	LNXi	Dempsey
	SGI Cluster	0.256	0.0		0.3	2.2		SGI	Cluster
	IBM Cluster	2.372	4.4	10.437	3.5	20.7	91	IBM	Opteron
	LNXi Cluster	2.356	7.1	17	4.4	20.6	146	LNXi	Xeon
DoD ASC	IBM P4	0.32	0.0		0.3	2.8	0	IBM	AIX
Dayton,OH	SGI 3900	2.176	1.3	2.867	2.2	19.0	25	SGI	smP
	HP Cluster	2.048	5.2	10.65	4.1	17.9	93	HP	Opteron
	SGI Cluster	2.048	6.0	12.288	2.0	17.9	107	SGI	Altix
	Compaq SC45	0.836	2.0	1.672	0.8	7.3	15	Compaq	Sierra Cluster
DoD AHPCRC Minneapolis,MN	Cray X1E	1.02	4.5	4.608	0.3	8.9	40	Cray	smPv
DoD ARSC	Cray X1	0.51	3.2	1.638	0.5	4.5	14	Cray	smPv
Fairbanks,AK	IBM P4	0.80	5.3	4.262	1.7	7.0	37	IBM	AIX
DoD MHPCC Maui,HI	IBM P3/4	1.61	1.7	2.778	0.8	14.1	24	IBM	AIX
DoD SMDC	IBM	0.52	1.6	0.832	0.5	4.5	7	IBM	AIX
Huntsville,AL	Linux Cluster	0.61	3.6	2.176	0.6	5.3	19	Linux	Cluster
	Cray	0.14	4.1	0.592	0.2	1.3	5	Cray	smPv
	SGI	0.87	1.6	1.42	0.7	7.6	12	SGI	smP
total				239		367	2088		

712 or

Joint Recommendations for Facility Metrics and Computational Science Metrics for the ASCAC CFM Subpanel

Ray Bair - Project Director -Argonne Leadership Computing Facility

Al Geist/Doug Kothe – Oakridge National Leadership Facility

Bill Kramer/Francesca Verdier – Berkeley National Energy Research Scientific Computing Facility

July 18, 2006




ALCF

Background

- **Representatives from the three centers met with Gordon Bell in late April to discuss replacing the PART metrics currently used by OMB to judge the success of DOE Computational efforts**
  - **Facilities agreed to submit proposed replacement metrics**
- **Focus was preparing and operating ultrascale facilities with a target of a Petaflop peak in the next 3-4 years.**
- **Facilities worked together and submitted joint recommendations on May 15**




ALCF

Current Metrics

- Acquisitions should be no more than 10% more than planned cost and schedule.
  - *This is no longer being required by OMB for FY 06, but is for other reporting*
- 40% of the computational time is used by jobs with a concurrency of 1/8 or more of the maximum usable compute CPUs.
- Every year several selected applications are expected to increase efficiency by at least 50%.

Joint Recommendations

- Representatives from the three centers met with Gordon Bell in late April to discuss replacing the PART metrics currently used by OMB to judge the success of DOE Computational efforts
  - Facilities agreed to submit proposed replacement metrics
- Focus was preparing and operating ultrascale facilities with a target of a Petaflop peak in the next 3-4 years.
- Facilities worked together and submitted joint recommendations on May 15

Joint Recommendations

- **The primary interest of OMB is whether the computational resources in the Office of Science are facilitating science discovery and the PART metrics should reflect this interest.**
- **Unfortunately, much of the impact of science discovery is impossible to measure quantative, especially over the short term.**
  - **Metrics like publications may be good indicators,**
  - **But many of the most important science discoveries of the past yielded only a small number of seminal papers.**
  - **Backward-looking metrics like citations and awards are also valid but long delayed and hence not as meaning in managing the investment portfolio**
- **Further, we believe three PART metrics are sufficient to demonstrate DOE Office of Science's progress in advancing the state of high performance computing.**

Two Types of Metrics

- **Mission/Science-based metrics – how well do project teams make use of the resources being provided**
  - **E.g – mission/science output, application software creation and improvement, software to improve scalability system, leadership (best in class) science, impact on industry, mission accomplishment**
  - **Facilities can not be held accountable for the misison/science based metrics (many of which the computing facilities do not control)**
- **Facility-based metrics – how well do the facilities provide the resources**
  - **E.g – availability, user satisfaction, assistance, deliver flop/s and bytes…**

Setting Expectations

- At lot of the metric discussion is about setting expectations with all the parties
  - Stakeholders
  - User (and potential users)
  - Overseers
  - Management
  - Vendors
  - Staff
  - Observers

Goals and Metrics

- Should be a few in number
- Should – with a glance – provide the viewer an 80% confidence things are going in the right direction
  - If metrics don't look right, there is typically huge amounts of detail data to peruse to determine
    - *If things are truly not right*
    - *Diagnosis what the cause and correction are*
- Several types of measures
  - Quality (how good)
  - Activity/Quantity (how much)
- Focus should be on quality

Defining "Metric"

- **Distinguish between**
  - **Goal - the behavior being motivated**
  - **Metric - what is measured to judge whether the goal is being achieved**
  - **Value - and the value for the metric that must be achieved.**
- **Confusion is between "activity" based data and "quality" based metrics.**
  - **The most obvious metrics are activity based (number of users, number of jobs, number of calls, etc.)**
  - **The most important metrics are quality based, which are suggested here.**

Joint Recommendation

- **The Facilities believe a small combination of these new metrics should replace #2 and #3 of the existing PART metrics, along with the modified existing metric #1.**
- **We posed several Facility and Mission/Science metrics to the committee with the expectation one of each type would be proposed.**

Acquisition Metric

- Current - Acquisitions should be no more than 10% more than planned cost and schedule.
  - *This is a reasonable metric and is being met but we have the following suggestion:*
- Major computer acquisition is defined as a Development, Modernization and Enhancement (DME) project which is subject to DOE Order 413.3.
- A similar metric is defined in DOE 413.3, and thus it is reasonable to align the 413 and PART metrics
- Best to score this item as follows:
  - *green: 10% or below,*
  - *yellow: between 10% and 25%,*
  - *red: above 25%.*

Facility Metrics

Goal #1: User Satisfaction

- Meeting the metric means that the users are satisfied with how well the facility provides resources and services.
- Metric #1.1: Users find the systems and services of a facility useful and helpful.
  - User feedback is key to maintaining effective resources and services. The survey should assess the quality and timeliness of support functions – including properly resolving user problems and providing effective systems and services. Interpreting survey results is both quantitative and qualitative. For quantitative results, different functions are rated on a numerical scale. If a scale from 1 to 7 is used, then scores above 5.25 are considered successful.
- Value #1.1: The overall satisfaction of an annual user survey is 5.25 or better (out of 7).
- Metric #1.2: Facility responsiveness to user feedback.
  - Possibly a more important aspect is how the facility responds to issues identified in the survey and other user feedback. Does the facility use the information to make improvements and are those improvements reflected in improved scores in subsequent years?
- Value #1.2: There is an improved user rating in areas where previous user ratings had fallen below 5.25 (out of 7).

Goal #1. Rational for User Satisfaction Goal

- DOE Facilities are in the business of enabling science
  - More complex than providing cycles, storage and access
- Computer and storage systems are often considerably larger than dedicated lab and university resources
  - Leadership Centers are 10x (or more) the scale of systems being used by new projects
- Not only do large projects have the usual issues
  - Accounts, files, porting, data transport
  - Compiler/library/tool availability and versioning
  - etc.
- They also have problems that primarily appear at scale
  - Scalability of algorithms, data structures, input/output, etc.
  - Debugging and performance optimization at scale
  - etc.

Goal #1. User Satisfaction

- Many things contribute to a user's satisfaction or dissatisfaction
  - Accessibility and availability of the systems and data on them
  - Computation turnaround time
  - Responsiveness to queries and accuracy/applicability of answers
  - Availability of accurate information about tools and services
  - How successfully computations ran
  - Ease of use of tools and services provided
  - Whether the resources were adequate for their mission/science studies
- Center user surveys provide a direct measure of user satisfaction
  - Can be reduced to simple metrics
  - Can be compared across users (projects) and years
  - Can help identify common areas for improvement, and track user perception of the effectiveness of those improvements

Goal #1. User Satisfaction

- Surveys are a tool that provides part of the picture
- We use other sources of information as well
  - Periodic discussions with project staff
  - Trouble ticket assessments
  - System usage analysis
  - Feedback at Workshops, User Meetings, and Town Halls

Goal #2: Office of Science systems are ready and able to process the user workload.

- **Meeting this metric means the machines are up and available most of the time. Availability has real meaning to users.**
- **<u>Metric #2.1: Availability</u>**
  - **Scheduled availability targets would be determined per-machine, based on the capabilities and mission of that machine. These should apply after an initial period of introductory/early service.**
  - **Scheduled availability is the percentage of time a system is available for users, accounting for any scheduled downtime for maintenance and upgrades.**
    - ***(Σ scheduled hours – Σ outages during scheduled time)/Σ scheduled hours***
  - **Overall availability is the percentage of time a system is available for users, based on the wall clock time of the period.**
    - ***(Σ Wall clock hours – Σ outages)/Σ wall clock hours***
  - **A service interruption is any event or failure (hardware, software, human, environment) that disrupts full service to the client base.**
  - **Degradation of service below the agreed upon level is treated as a service interruption.**
  - **Any shutdown that has less than 24 hours notice is treated as an unscheduled interruption.**
  - **A service outage is the time from when computational processing halts to the restoration of computation (e.g., not when the system was booted, but rather when user jobs are recovered and restarted).**

Goal #2: Office of Science systems are ready and able to process the user workload.

- **<u>Value #2.1</u>: Within 18 months of delivery and thereafter, scheduled availability is > 95%**
- **<u>Value #2.1</u>: Within 18 months of delivery and thereafter, overall availability is > 90% or another value as agreed by the program office.**
  - **Example - ORNL over the next year will be making significant portions of Jaguar (greater than 10%, maybe as much as 20-30%) available for development and testing to Cray, SNL, and ORNL staff to prepare the Multi-core OS for the 250TF and 1000TF systems.**

Goal #2: Office of Science systems are ready and able to process the user workload.

- This is an attractive concept, that can be complicated to assess in practice
- A system is a collection of integrated hardware resources and software services
  - Centers strive to make them all available on a continuous basis
  - Different computations use different services
- Availability and cost are coupled
  - Tradeoffs are made in center capabilities, architectures and support models that impact cost and availability
  - For very large systems, the ability to work around some faulty compute processors is likely to be the optimal approach
- Therefore availability targets should be system-dependent
  - To reflect mission needs and agreed tradeoffs acquisition/operating cost and availability

Goal #3: Facilities provide timely and effective assistance

- Helping users effectively use complex systems is a key role that leading computational facilities supply. Users desire their inquiry is heard and is being worked. Users also need to have most of their problems answered properly in a timely manner.
- Metric #3.1: Problems are recorded and acknowledged
- Value #3.1: 99% of user problems are acknowledged within 4 working hours.
- Metric #3.2: Most problems are solved within a reasonable time
  - Many problems are solved within a short time period in order to help make users effective. Some problems take longer to solve – for example if they are referred to a vendor as a bug report.
- Value #3.2: 80% of user problems are addressed within 3 working days, either by resolving them to the user's satisfaction within 3 working days, or for problems that will take longer, by informing the user how the problem will be handled within 3 working days (and providing periodic updates on the expected resolution).

Goal #3: Facilities provide timely and effective assistance

- **A key component of user productivity (and satisfaction)**
  - Time to solution often directly impacts time to discovery
  - Some projects must stop until queries are addressed
  - Mission/Science project plans may have to be altered, depending on the response
  - Keeping users informed about the resolution status and path is important
- **Many problems are straightforward to address**
  - Technical questions, account management issues, etc.
  - These can be turned around in a few days or less
- **Other problems require longer**
  - Software updates, testing and deployment, by the Center or other parties
  - Difficult bugs, feature requests, new capability requirements
  - These may take a very long time, but the users deserve to know the plans for dealing with them

Goal #4: Facility facilitates running capability problems

- **Major computational facilities have to run capability problems. This is a complex goal that has many aspects which contribute to meeting the metric. While NERSC and NLCF have demonstrated that it is possible to provide the majority of its time to applications of scale with high overall utilization, it is clear there are consequences to other parts of the workload. Several aspects that influence a facility's ability to meet this goal include:**
  - The ability to run at scale is strongly influenced by which projects are provided allocations and the amount of time each project is given.
  - The total number of projects that run on a system.
  - The higher the utilization on systems, the more challenging it is to run large jobs without impacting turnaround of other parts of the workload.
  - The definition of a capability job needs to be defined by agreement between the Program Office and the Facility.
    - *In general, a larger number of computational processors increase the size of capability jobs.*
    - *On the other hand, a larger number of projects decrease the size of capability jobs.*

Goal #4: Facility facilitates running capability problems

- Metric #4.1: The majority of computational time goes to capability jobs.
- Value #4.1: T% of all computational time for jobs that use more than N CPUs (or equivalently, x% of the available resources), as determined by agreement between the Program Office and the Facility.

- Metric #4.2: Capability jobs are provided excellent turnaround
  - Job turnaround is an important metric for the user community and is commonly associated with user productivity. Job turnaround is determined as the ratio of the total amount of time a job requests to run divided by the time the job waited to run. This is called the expansion factor.
    - *It is possible use the actual run time, but consideration has to be made for jobs that run much less than they request. For example, NERSC uses run time, but does not count jobs that run less than several minutes, since they are jobs that fail early in their scripts.*
    - *It may be better to count nodes for capability jobs, rather than processors.*
    - *Facilities would define when a job becomes eligible to run – and the time starts*
- Value #4.2: For jobs defined as capability jobs, the expansion factor is X or more. X ≤ 10 is a potential value that may be appropriate.

Mission and/or Science Metrics

These are metrics for the mission/science projects run at the DOE-SC facilities.

CS Metrics for Application Scientists

- **CS Goal #1: Project Progress**
  - While there are many laudable mission and science goals, it is vital that significant computational progress is made against the Nation's challenges and questions.

- **Metric #CS1.1: Progress is demonstrated toward the scientific milestones in the top 20 projects at each facility based on the computational results planned and promised in their project proposals.**
  - It may be better to specify this by the amount time projects get, for example, rather than using an arbitrary number such as 20, use a limit such as projects receiving more than 5% of a facilities resource.

CS Metrics for Application Scientists

- **Value #CS1.1: For the top 20 projects at each facility, an assessment is made by the related program office regarding how well scientific milestones were met or exceeded relative to plans determined during the review period. For allocations where the research is government funded, the funding office will conduct the review. For allocations where there is no government funding, the review will be conducted by a peer review panel selected by the DOE office of Advanced Scientific Computing Research.**
  - It may be better to specify this by the amount time projects get, for example, rather than using an arbitrary number such as 20, use a limit such as projects receiving more than 5% of a facilities resource.

CS Metrics for Application Scientists

- CS Goal #2: Scalability of Computational Science Applications
- The major challenge facing computational science during the next five to ten years is the increased parallelism needed to use more computational resources.
  - Multi-core chips accelerate the need to respond to this challenge. Moore's Law will continue this trend as the number of CPUs on a chip double every 2 to 3 years.
- This goal could replace the current goal #3 of increasing the efficiency of applications, which is no longer an issue.
  - While this metric applies to science projects rather than facilities that host them, facility staff often provide substantial help to the identified projects for them to be successful. Nonetheless, meeting this goal can not be a facility metric

CS Metrics for Application Scientists

- Metric #CS2.1: Science applications should increase in scalability.
- Value #CS2.1: The scalability of selected applications increase by a factor of 2 every three years. The definition of scalability (strong, weak, etc.) might be domain- and/or code-specific.

Additional Suggestions from the committee

- **All metrics should show be able to show improvement over some time period – reaching an acceptable level**
  - **Hence Green, Yellow, Red or Blue, Green, Yellow, Red value levels should be proposed**
- **Much of the information the committee requested can only be provided by the mission and science projects**
  - **Suggestions to improve the process could be to recommend DOE use a common format for**
    - ***Project Proposals***
    - ***For user requirements (Greenbooks, SCaleS reports, etc)***
  - **Require proposals to include quarterly progress milestones**
    - ***Have quarterly reporting from mission/science projects collected***
    - ***Maybe this only applies to INCITE and very large projects***
    - ***Please recommend the DOE or a third party accumulate these reports – the facilities should not be put into the position of a police officer to the projects – we are there to help, not enforce.***

ASCAC Computer Facilities sub Panel
Centers Facilities Metrics

The following sections are metrics relation to centers management

1. Facility Overview
2. User interface and communication including satisfaction monitoring and metrics
3. Qualitative output
4. Quantitative output
5. Center x User Readiness for 10x processor expansion

1.0 Overview of Resources Provided by the Center

a. Contact information for the project
    i. URL to Staff directory, emails; phones
    ii. URL
b. Organizational structure with staff sizes and functional titles (separate page)
c. FTE's……. Total
    i. overhead and overall management
    ii. operations
    iii. system development tools,
    iv. consulting
    v. user specific support and projects
d. Physical infrastructure
    i. building size,
    ii. power – amount
    iii. cost $Mwhr,
    iv. cooling capability
    v. network access
e. Balance sheet and budget for:
    i. hardware,
    ii. maintenance,
    iii. staff, software,
    iv. utilities,
    v. buildings,
    vi. institutional overhead, etc.
f. Institutional affiliation and degree of institutional support
g. Present and planned hardware
    i. Computers
    ii. Disk memory for cache and on-line datasets or databases
    iii. tertiary storage, e.g. in use peta-bytes versus potentially available
h. Software development and production tools provided top 5 (enumerate on separate pages)
i. Application codes available to the users that are supported by the center (ISVs, open source, etc.) top 5 enumerate with software development tools listing
j. What auxiliary services do you offer your users
    i. Visualization
    ii. Other

2.0 User interface and communication including satisfaction monitoring and metrics

a. How do you measure the success of your facility today in being able to deliver service beyond the user surveys (e.g. the NERSC website)?
b. Do all users-experimenter teams, team members, and any users that the team community provides utilize the survey?
c. Have these surveys been effective at measuring and understanding making changes in operations? (Please cite)
d. Describe your call center – user support function: hours of coverage, online documentation, trouble report tracking, trouble report distribution, informing the users, how do users get information regarding where their job/trouble report is in the queue?
e. What mechanisms are provided for the user with respect to dissatisfaction with how a case is being handled?
f. What mechanisms are provided to support event-driven immediate access to your facility (e.g. Katrina or flu pandemic)

3.0 Qualitative measure of output

a. Do you measure how your facility enables scientific discovery?
b. How are the results of measurement disseminated and how do they further Science and especially DOE Science Programs?
c. What impact have any of your measures had on operation of your facility?
d. What impact have the current PART measures had on your successful operations of your facility?
e. What do you view as the appropriate measures for supercomputing facilities now?
f. During the next 3-5 years?

4.0 Aggregate Projects use profiles by scale

a. How many projects does your center support?
b. How many users that are associated with all the projects?
c. How many additional users who either use project data-sets or other center resources?
d. A What is the project usage profile in terms of processor count? We would like these broken down into jobs that require, or can exploit a concurrency level of (roughly) 50, 200, 400, 1,000, 2,000, and 4,000 processors to obtain the science.
    i. Aggregate required memory per job? (Or memory per node)
    ii. Processor distribution?
    iii. Disk space use?
    iv. Tertiary tape use?
    v. Average wall clock time of jobs?
    vi. Average time of jobs in the queue?
    vii. How do you measure project code performance on your machines?
    viii. Amount of project consulting support utilized?

5.0 Center x User Readiness for 10x processors expansion

The mid-term goals for each facility call for a major expansion from machines with of order 5,000 processors to machines of order 50,000 processors or more.

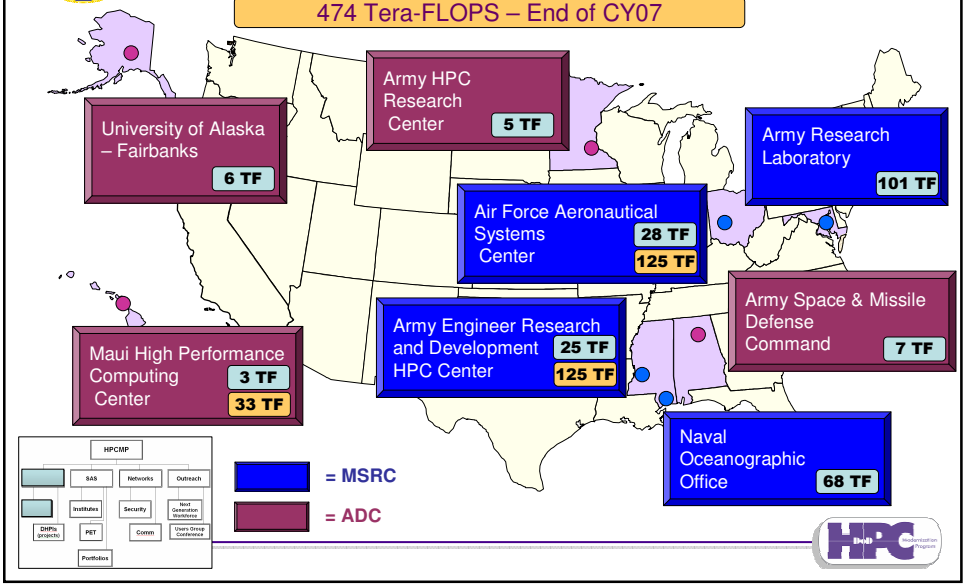
a. Please outline how the center will accommodate this growth over the next 3-5 years.
b. What do you believe is your role in preparing users for this major change?
c. What effort (in terms of personnel) is devoted to code development issues today, and do you view this as adequate coverage as we move to machines with more than 25,000 processors?
d. Are there codes in your user portfolio that will scale today to 10,000, 25,000, or 75,000 processors. What is the nature of these codes (Monte Carlo, CFD, hydro?) Are these codes running today on other systems of comparable size?
e. As machines become more complicated, what do you see as the challenges to your success? For example, are you (or parts of your institution) actively involved in research related to fault-tolerance, memory/bandwidth contention, job scheduling, and etc. on the future machines?
f. How do you determine the path forward for your organization?
g. What do your users want to see in the largest machines now available and those which will be available in the 3 year and 5-7 year time frames? (memory per core/node, number of processors, disk space?)

High Performance Computing

Bradley Comes
DoD HPC Modernization Program

Eight DoD Shared Centers

243 Tera-FLOPS – End of CY06
474 Tera-FLOPS – End of CY07



HPC Systems at the HPCMP Centers

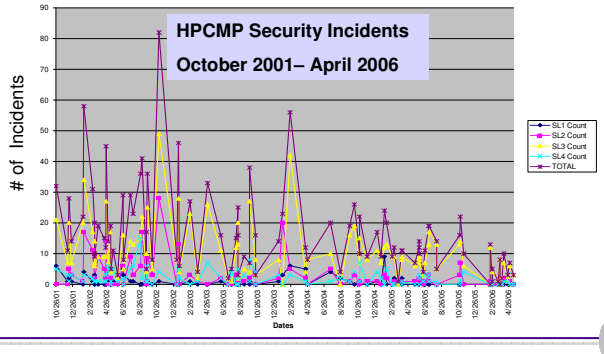
DoD Centers' computers as of December 31, 2006

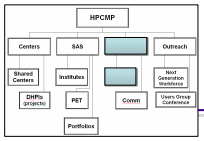
		nodes/n	Proc(K)	P.speed(GF)	C.pk(TF)	Mp(TB)	Ms(TB)	O/S	
DoD ERDC	SGI 3900		1	0.7	1.02	1.0		SGI	smP
Vicksburg, MS	Cray XT3		4.176	2.6	21.70	8.4		Cray	Cluster
	Cray X1		0.256	0.8	0.82	0.3		Cray	smPv
	Compaq SC45		0.512	1.0	1.02	0.5		Compaq	Sierra Cluster
DoD NAVO	IBM P4		1.408	1.3	7.32	1.4		IBM	AIX
Bay St.Louis, MS	IBM P4		2.944	1.7	20.02	6.0		IBM	AIX
	IBM P4		0.512	1.7	3.48	0.7		IBM	AIX
	IBM P5		3.072	1.9	23.35	6.1		IBM	AIX
	IBM P5		1.92	1.9	14.59	3.8		IBM	AIX
DoD ARL	IBM P4		0.128	1.7	0.87	0.1		IBM	AIX
Aberdeen, MD	LNXI Cluster		4.206	3.0	50.30	9.0		LNXI	Woodcrest
	LNXi Cluster		3.368	3.2	21.56	6.7		LNXi	Dempsey
	SGI Cluster		0.256	1.5		0.3		SGI	Cluster
	IBM Cluster		2.372	2.2	10.44	3.5		IBM	Opteron
	LNXI Cluster		2.356	3.6	16.69	4.4		LNXI	Xeon
DoD ASC	IBM P4		0.32	1.3		0.3		IBM	AIX
Dayton, OH	SGI 3900		2.176	0.7	2.87	2.2		SGI	smP
	HP Cluster		2.048	2.8	10.65	4.1		HP	Opteron
	SGI Cluster		2.048	1.5	12.29	2.0		SGI	Altix
	Compaq SC45		0.836	1.0	1.67	0.8		Compaq	Sierra Cluster
DoD AHPCRC	Cray X1E		1.02	1.1	4.61	0.3		Cray	smPv
Minneapolis,MN									
DoD ARSC	Cray X1		0.51	0.8	1.64	0.5		Cray	smPv
Fairbanks,AK	IBM P4		0.80	1.3	4.26	1.7		IBM	AIX
DoD MHPCC	IBM P3/4		1.61	1.3	2.78	0.8		IBM	AIX
Maui,HI									
DoD SMDC	IBM		0.52	2.6	0.83	0.5		IBM	AIX
Huntsville,AL	Linux Cluster		0.61	2.0	2.18	0.6		Linux	Cluster
	Cray		0.14	1.1	0.59	0.2		Cray	smPv
	SGI		0.87	1.6	1.42	0.7		SGI	smP

12 July 2006, R6

HPCMP SECURITY

- **HPCMP Computer Emergency Response Team**
  - 24X7 Monitoring and Response
- **Comprehensive Security Assessments**
  - Physically Visit Approximately 20 sites per year
- **Security Training for Users and HPCMP Staff**
- **Software Protection Initiative**
  - Agents in Software to Protect Where it Can Execute





KEY ELEMENTS
~METRICS~

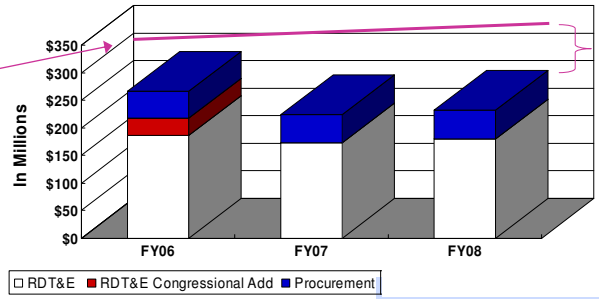
- PET Metrics
- Institute Metrics
- Portfolio Metrics
- Center Metrics
- DREN Metrics

Details on Each of the Above are Available in the Backup Slides in a Section Titled "Metrics"

- Return on Investment (ROI)
  - Success Stories and Annual Report
  - Quantified Cost Avoidance and/or Dollars Saved
    - **2-4 Detailed Project Assessments per year**
    - **"Shared" Contribution to ROI with Labs and Test Centers**

CHALLENGES 1 of 8
Future Budget
FY 2006–FY 2008 Funding Levels

Consumer Price Index Trend Line

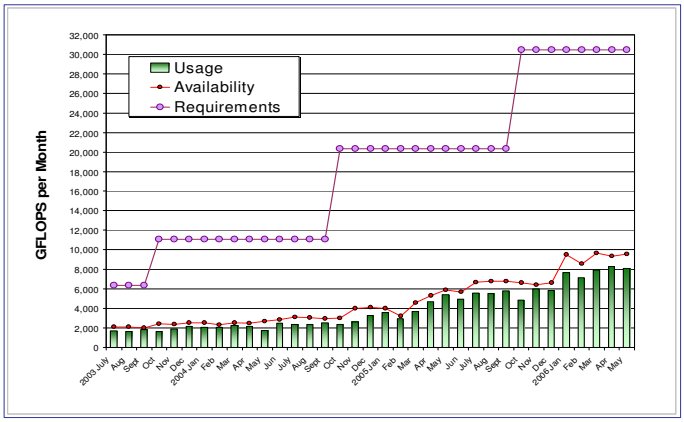


26% Decline since 1997

Procurement Funding Supports the Ability to Deploy Larger and Larger HPC Systems while the Effectiveness of Funding for Human Resources Declines

CHALLENGES (2 of 8)

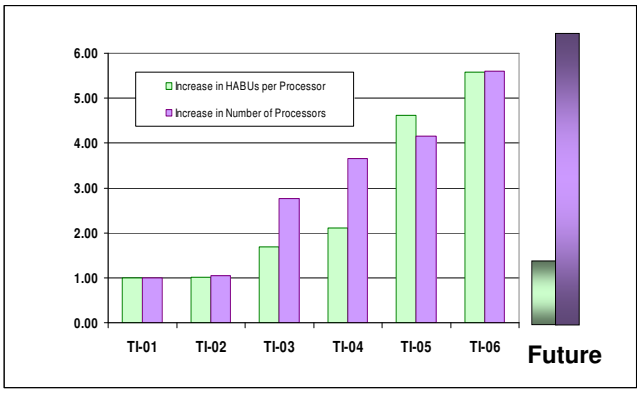
~REQUIREMENTS, ALLOCATIONS, and UTILIZATION~



Requirements Continue to Far Exceed Availability of HPC Resources

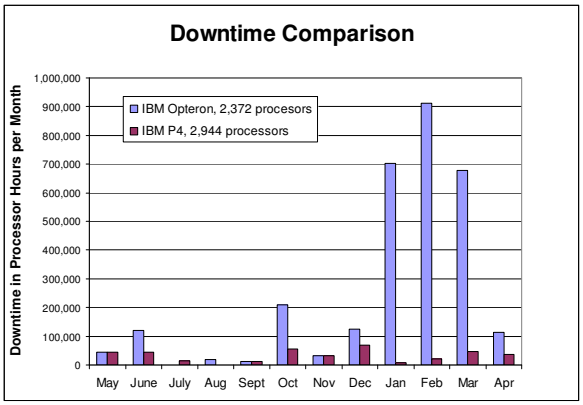
CHALLENGES (3 of 8)

~Application Scalability~



Application codes must be prepared to leverage computational contributions from increased number of processors

CHALLENGES (4 of 8)
~System Reliability~



Large-scale commodity clusters are more vulnerable to overall system failures than well-integrated systems

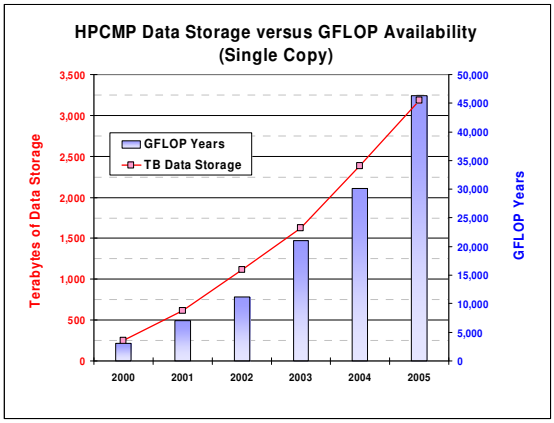
CHALLENGES (5 of 8)
~Job Complexity~

- Existing Challenge Jobs
  - 200 to 1000 processor range
  - Approximately 20 to 30 Challenge Projects per Year
- Existing CAP Jobs
  - 2000 to 4000 processor range
  - Approximately 5 Phase II CAP Projects per Year

Today's CAP Job is Tomorrow's Challenge Job

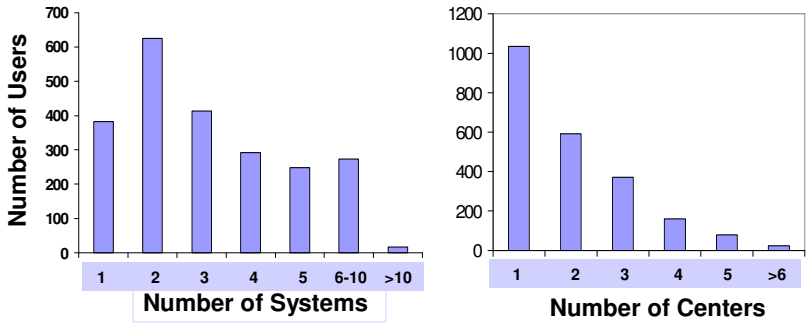
CHALLENGES (6 of 8)
~Data~

- **Data Management**
  - Locality
  - Movement
  - Sharing
  - Duplication
  - Disaster Recovery
  - Storage Technologies
- **Data Analysis**
  - Data -> Information
  - Remote Visualization



Application Code Enhancements Need to Couple Data Generation and Data Analysis into one End Product

CHALLENGES (7 of 8)
~Mobility of User Community~



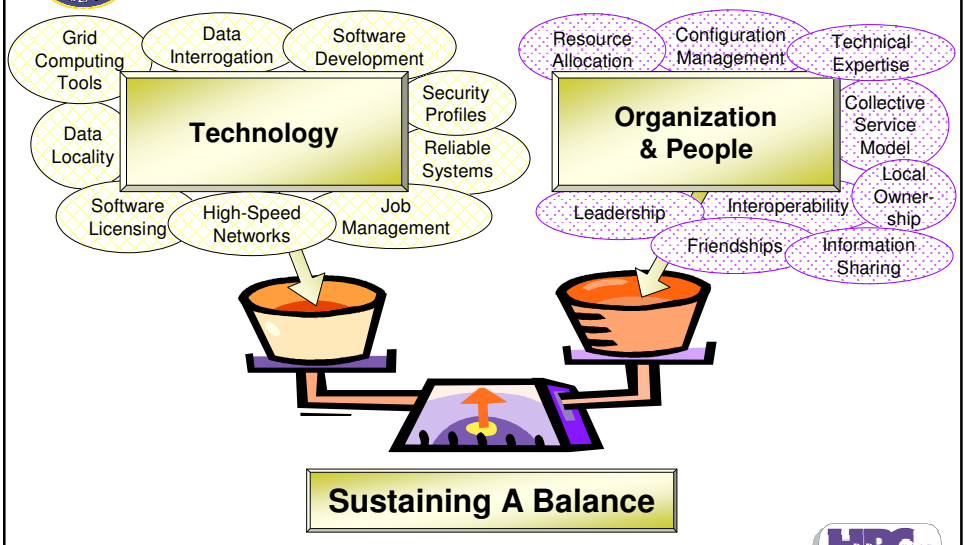
Enhancement of Baseline Configuration and Deployment of Grid Technologies is becoming Increasingly Important

CHALLENGES (8 of 8)
~Facilities~

- **Power**
  - Currently working in the 2MW per Center range
  - Foresee needing to operate at 5+MW per Center
  - May be going back to single large AC/DC Converters
- **Cooling**
  - CFD modeling for facility planning becoming more critical
    - Air movement becoming very sensitive
  - May be moving back to liquid cooled approaches
    - Vendor Integrated Designs
    - Retrofit Designs
  - Availability of Air Handlers and Chillers
- **Space**
  - Under-the-floor and overhead space requirements becoming more critical in support of cooling and cabling requirements
  - Space for Air Handlers and AC/DC Converters
  - Physical Size of large scale systems

Increasing Costs Associated with Facilities are Adversely Impacting Deployment of Capable HPC Systems

CLOSING THOUGHT



Sustaining A Balance

High Performance Computing

DoD HPC Modernization Program

Metric Slides

User Productivity Enhancement and Technology Transfer (PET) Metrics

	1Q	2Q	3Q	4Q
# of s/w applications enhanced	45	46	46	35
# of new technologies transferred	10	22	19	27
# of publications posted to the OKC	39	12	10	26
Results of CTA leader assessments				
Meets cost & schedule objectives	see separate sheet	see separate sheet		
# of training events	19	15	13	11
# of code signatures developed and added to database	Database structure and tools under development			
# of customers assisted with application or science specific support	129	75	78	160
User Satisfaction Survey scores	4.0 range for all			
Assessment of training events by trainees	4.18	4.44	4.26	4.40

17

HPCMP Software Institutes Metrics (as of June 2006)

Institute	# of Codes In Development/ Maintenance/ Assistance	Code Performance Increase (Over Baseline)	Validation & Verification Ongoing/ Complete	Net New Institute Personnel /Users	Stakeholder Assessment/ User Surveys
IMPTS (Vicksburg)	10/4/?	N/A	4/3	5/12	5.0/5.0
BHSAI (Ft Detrick)	6/0/1	>5	3/1	11/3	4.75/4.0
BEI (Stennis)	5/0/0	N/A	3/0	4/15	5.0/4.0
ISSA (AMOS)	5/0/5	>6; >776; >10; >8; N/A	5/5	13/15	4.5/4.0
IHAAA (Eglin)	15/16/33	N/A	14/1	15/4	4.5/4.0
HI-ARMS (Moffett)	-/- FY06 startup	-/-	-/-	-/-	*/*

12 July 2006, R6

Metrics Rollup (4 Most Recent Portfolios)

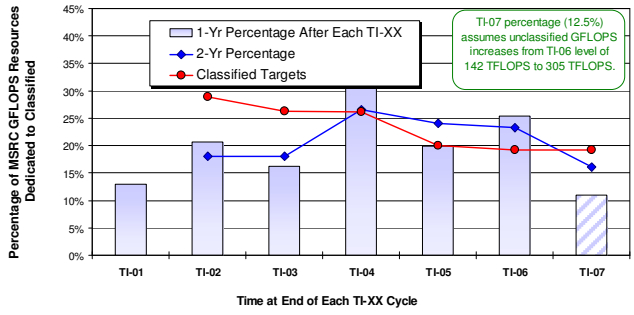
				SOFTWARE METRICS					
Purple are MY GOALS AND METRICS (Andy's) which roll up from your goals/metrics. G3M1a refers to MY goal 3; metric 1a, for example. So I'll be tracking what you're doing.									
			Metric						
	Category			Objective	Threshold	Attained	FY2005	FY2006	FY2007
Schedule Metrics	Planning	Output Measure:	G3M1c: Meet cost and schedule objectives/obligations (your spend plan)	0%	5%	Y			
		Output Measure:	Portfolio Software Development Plan	4	3	3			
		Output Measure:	Portfolio Test and Evaluation Master Plan (Annex)	4	3	3			
		Output Measure:	Interface Control Document	4	3	2			
		Output Measure:	Financial Reports	monthly as specified		Y			
		Output Measure:	Quarterly Reports	quarterly as specified		Y			
Technical Metrics	Products	Output Measure:	G3M1a: Number of successful test events per FY	4	3	3			
		Output Measure:	G3M1b: Number of Critical Technical Parameters (CTP) achieved per test event	44	36	37			
		Output Measure:	G4M1a: Number of projects that successfully integrate 2 science disciplines (FY05-FY06)	4	2	2			
		Output Measure:	G4M1b: Number of projects that successfully integrate 3 science disciplines (FY06-FY07)	4	2	2			
		Output Measure:	G4M1c: Number of projects that successfully integrate 4 science disciplines (FY07)	4	2	na			
	Quality	Output/Outcome Measure:	User satisfaction survey results	5	≥ 4	4.2			
		Output/Outcome Measure:	G3M2a: Number of codes enhanced and in use by DoD or industry	All codes proposed for FY06	80% of codes in proposed for FY06	73?			
		Output/Outcome Measure:	G3M2a: Number of codes enhanced and in use by DoD or industry	All codes proposed for FY07	80% of codes proposed for FY07	na			
		Output/Outcome Measure:	G3M2b: Number of codes enhanced and in use by groups not involved in development	Three documented events by end of FY06	Two documented events by end of FY06	6?			
		Output/Outcome Measure:	G3M2b: Number of codes enhanced and in use by groups not involved in development	Six documented events by end of FY07	Four documented events by end of FY07	na			
	Impact	Outcome Measure: Programs Impacted	Quantitative assessment from stakeholders of mission impact (DoD, Industry)	Two documented events by end of FY06	One documented events by end of FY06	2			
		Outcome Measure: Programs Impacted	Quantitative assessment from stakeholders of mission impact (DoD, Industry)	Four documented events by end of FY07	Three documented events by end of FY07	na			

HPC Centers Metrics

April 2006

Goal 1: Provision Resources to Optimally Address DoD Workload

Metric # 1: Meet Annual ***Classified-versus-Unclassified*** Requirements at the four MSRCs



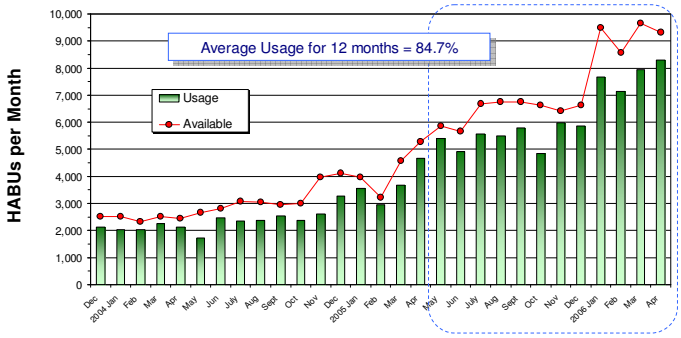
1

HPC Centers Metrics

April 2006

Goal 1: Provision Resources to Optimally Address DoD Workload

Metric # 2: Assess Users' Responsiveness to Utilizing Deployed Resources on allocated systems: Percent of Available Capacity Utilized (Target to be ≥ an Average of 75% over 12 months)

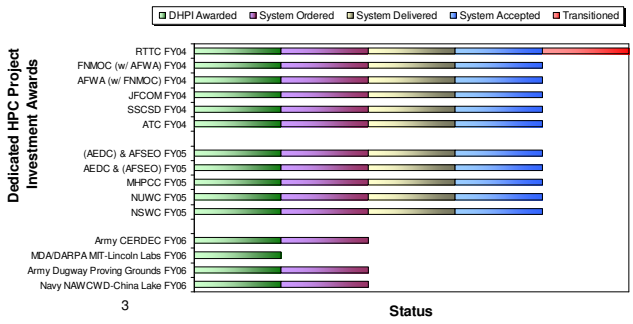


HPC Centers Metrics

April 2006

Goal 1: Provision Resources to Optimally Address DoD Workload

Metric # 3: Dedicated HPC Project Investments (DHPIs) are Deployed Efficiently: Target: System acceptance NLT 1 year from DHPI award



HPC Centers Metrics

April 2006

Goal 1: Provision Resources to Optimally Address DoD Workload

Metric # 4: COTS Software Shared Across Multiple Centers:

Center	Architecture	OS	PEs	Abaqus 6.7	Abaqus CAE	Abaqus Viewer	Ansys	AVS 6.3	AVS 7.0	CFD++	Cobalt	Accelrys/Insight	Fluent	Gasp	Gaussian	Gaussian Linda	LS-Dyna	Matlab	PTC	Totalview	Vampir	IntelTrace
AHPCRC	Cray X1E	Unicos	1024								X				X							
	CRAY T3E-1200E 1/01	Unicos	1088								X					X						
ARL	IBM P4	AIX 5.2	120	X			X	X		X	X	X	X	X		X	X			X		
	Linux Networx Evolocity II	Red Hat 3.0	256	X	X	X	X	X	X	X	X		X	X	X		X	X		X	X	
	SGI Altix 3000 (Itanium2)	SGI Propack 3	256							X	X	X			X		X					
	IBM 1350 (Opteron)	SUSE 9	2304							X	X		X	X	X	X	X	X				
	Intel XEON EM64T	SUSE 9	2048	X	X	X	X			X	X		X	X	X	X	X	X		X	X	
ARSC	Cray X1	Unicos	512								X					X						
	IBM P4	AIX 5.2	800	X			X	X		X	X		X	X		X	X			X		
ASC	SGI Origin 3900	RIX	2048	X				X	X	X	X	X	X	X	X	X	X		X			
	HP Opertron Cluster		2048																			
	SGI Altix Itanium 2	Linux	2048	X			X			X	X		X	X	X	X	X	X		X	X	
	SC 45	TRU64 2.6	836	X			X	X	X	X	X	X		X		X	X	X		X		
	SC 40	TRU64 2.6	64	X			X	X	X	X	X	X		X		X	X	X		X		
ERDC	SC40	TRU64 V5.1	512	X			X	X	X	X	X	X	X	X		X	X	X		X		
	SC45	TRU64 V5.1	512	X			X	X	X	X	X	X	X	X		X	X	X		X		
	Cray XT3 Opteron	Linux/Catamount	4128							?	?									X		
	SGI Origin 3800	IRIX64 6.5	512	X				X	X	X	X	X	X	X		X	X	X	X			
	SGI Origin 3900	IRIX64 6.5	1024	X				X	X	X	X	X	X	X		X	X	X	X			
MHPCC	AIX P3/P4	AIX 5.1	1088	X			X	X		X	X		X	X		X	X			X		
NAVO	IBM P4	AIX 5.1	1408	X			X	X		X	X		X	X		X	X			X		
	IBM P4+	AIX 5.2L	512	X			X	X		X	X		X	X		X	X			X		
	IBM P4+	AIX 5.2L	2944	X			X	X		X	X		X	X		X	X			X		
SMDC	Cray SV-1e	Unicos 10.0.1.2	16							X						X						
	SGI Origin 3900	IN 6.5.25	256	X				X	X	X	X	X	X	X	X	X	X		X			
	SGI Origin 3900	IN 6.5.25	128	X				X	X	X	X	X	X	X	X	X	X		X			
	Altix Linux Cluster	SuSE 9.2	256	X			X			X	X		X	X	X	X	X	X				
	Cray X1 AC - Cray X1E	Unicos/MP 2.5	128								X					X						
Desktop Computing Required					Yes		Yes		Yes	Yes	*Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	

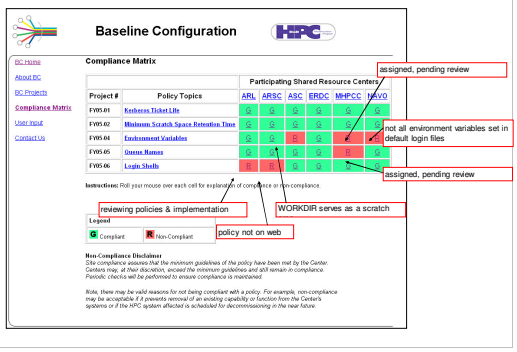
4

HPC Centers Metrics

April 2005

Goal 1: Provision Resources to Optimally Address DoD Workload

Metric # 5: Common Operating Environment:
Number of Sites Running the CBoD Baseline Configuration (Target = Composite Compliance of 70%)



As of May 25, 2006

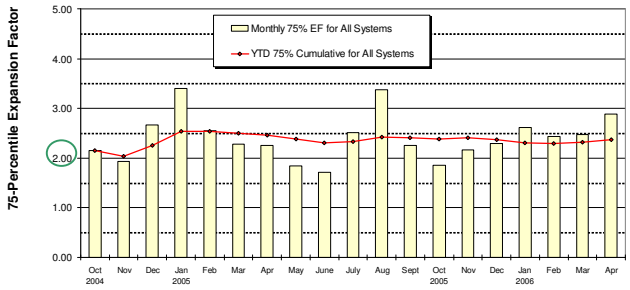
5

HPC Centers Metrics

April 2006

Goal 2: Operate Efficient and Effective Centers

Metric # 1: Provide Environments that Enhance DoD User Productivity:
Sub-metric # 1a: Overall 75-Percentile Expansion Factors for **12 Months** (Target ≤ 2)



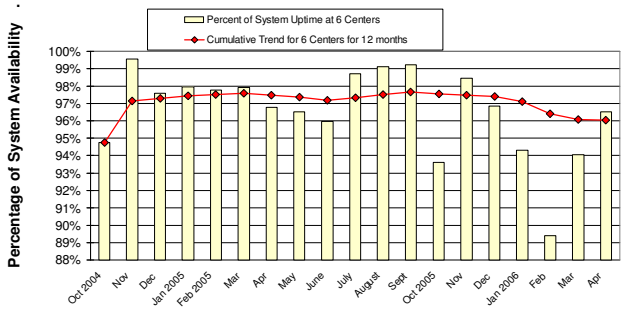
6

HPC Centers Metrics

April 2006

Goal 2: Operate Efficient and Effective Centers

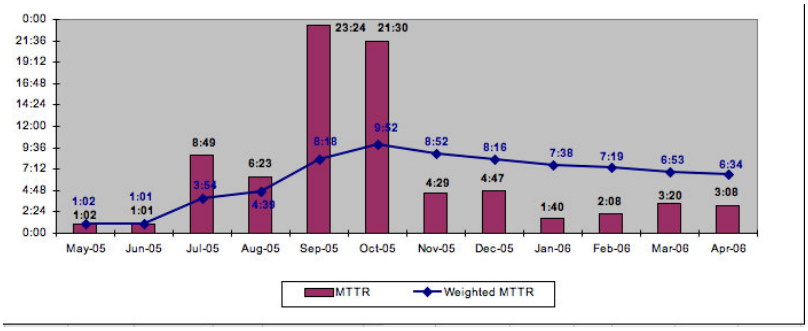
Metric # 1: Provide Environments that Enhance DoD User Productivity:
Sub-metric # 1b: Weighted System Uptime for Fiscal Year-to-Date (Target ≥ 98%)



7

DREN - Mean Time to Repair

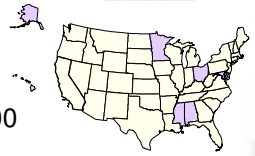
Month	Pri-1 Count	MTTR (Hrs.)	MTTR (Mins.)	Outage Time (# of Pri-1 * MTTR)	Weighted MTTR (Mins)	Weighted MTTR (Hrs)	Availability
May-05	10	1:02	62	620	62	1:02	99.998%
Jun-05	19	1:01	61	1159	61	1:01	99.998%
Jul-05	17	8:49	529	8990	234	3:54	99.954%
Aug-05	20	6:23	383	7660	279	4:39	99.919%
Sep-05	16	23:24	1404	22464	499	8:18	99.131%
Oct-05	11	21:30	1290	14190	592	9:52	99.895%
Nov-05	21	4:29	269	5649	533	8:52	99.936%
Dec-05	20	4:47	287	5740	496	8:16	99.962%
Jan-06	14	1:40	100	1400	459	7:38	99.977%
Feb-06	9	2:08	128	1152	440	7:19	99.998%
Mar-06	19	3:20	200	3800	414	6:53	99.921%
Apr-06	16	3:08	188	3008	395	6:34	99.940%

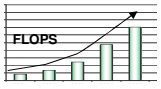


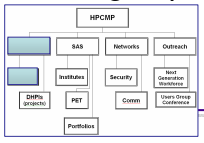
12 July 2006, R6

A Few Statistics

- 8 Large HPC Centers
  - Geographically Distributed
  - Cray, HP, IBM, Linux Clusters, SGI
- 28 HPC Systems
  - 14 existing systems — 1,000 to 4,000 processors
  - ~ 2 large systems at each large center
  - 222 peak Tera-Flops increasing at ~ 50% per year
  - 3 Peta-bytes Data Storage and Growing
- High-speed Wide Area Network
  - Approximately 130 Service Delivery Points







FACILITIES

- Facilities
  - Power Consumption (the HPCMPO is projecting an increase of 25% for each year)
    - Increased levels in the number of cores-per-socket will result in a return to ~135 watts per socket without a proportionate cpu-for-cpu performance increase
    - Increases in memory size to 2 GB per dual-core would add 80 watts of power consumption per CPU
  - Extremely large cost to provide power conditioning, cooling and UPS for 5 to 10 megawatt HPC systems
    - We are investigating a combination of diesel generator and flywheel technologies
  - System power levels in 2009 will exceed our ability to cool with chilled air, which will add 10 – 15% to compute-node costs for additional for cooling technology
  - Increasing pressure to divert HPC procurement funds for power, cooling, and UPS infrastructure – our HPC acquisition in 2014 could require 7 megawatts

ASCAC Computer Facilities sub Panel
Experimental Project Metrics

The following sections are metrics relation to experimental project use and management

1. **Experimental project overview**
2. **Project team resources**
3. **Project code**
4. **Project input from center**
5. **Project software engineering processes**
6. **Project output measurements**
7. **Future**

1.0 Experiment Project Overview

a. Project name
b. Contact information for the project
   i. Principal investigators, emails; phones
   ii. URL
c. DOE Office support: DOE program manager; (SC Office (BES, BER, NP, HEP, ASCR, FES, other)
d. Scientific domain (chemistry, fusion, high energy, nuclear, other.),
e. What are the technical goals of the project?
   i. What problem or "grand challenge" are you trying to solve?
   ii. What is the expect impact of project success? (e.g. better understanding of supernovae explosions, prediction of ITER performance, …)
f. Support for the development of the code
   i. Degree of DOE support to develop the code?
   ii. SciDAC, DOE SC program
   iii. internal institutional funding sources (e.g. LDRD,..),
   iv. industry,
   v. other agencies, ….
g. What is the project profile in total human resources including
   i. trained scientists,
   ii. computational scientists and mathematicians,
   iii. program development and maintenance,
   iv. use(rs) of the team codes?
h. Size of any or all external communities that your code or datasets support.

2. Project Team Resources

a. Team size
b. Team institutional affiliation(s). (e.g. all the institutions involved, including universities, national labs, government agencies,..). I.e. to what extent is the team multi-institutional?
c. To what extent are the code team members affiliated with the computer center institution? (e.g. are the team members also members of the computer center institution?)

d. Team composition and experience total
   i. domain scientists,
   ii. computational scientists, computer scientists, computational mathematicians, database managers
   iii. programmers
   iv. other
e. Team composition by educational level (total)
   i. Ph.D.,
   ii. MS, BS, undergraduate students, graduate students, post-docs, younger faculty, senior faculty, national laboratory scientists, industrial scientists, etc.)
f. Team resources utilization: time spent on code and algorithm development, maintenance, problem setup, production, and results analysis

3. Project Code

a. Problem Type (data analysis, data mining, simulation, experimental design, etc.)
b. Types of algorithms and computational mathematics (e.g. finite element, finite volume, Monte-Carlo, Krylov methods, adaptive mesh refinement, etc.)
c. What platforms does your code run on?
   i. What is your preferred platform?
b. Code size (single lines of code, function points, etc.);
   i. Code age
   ii. Amount of code added per year
c. Computer languages employed,
   i. LOC/ language 1;,
   ii. LOC/ language 2
   iii. LOC/ language 3
   iv. Structure of the codes (e.g. 250,000 SLOC Fortran-main code, 30,000 C++-problem set-up, 30,000 SLOC Python-steering, 10,000 SLOC PERL-run scripts,…)
d. What libraries are used?
   i. What fraction of the effort do they represent?
e. Code Mix:
   i. To what extent does your team develop and use your own codes?
   ii. Codes developed by others in the DOE and general scientific community?
   iii. Application codes provided by the center?
f. What is the present parallel scalability
   i. Projected or maximum scalability
   ii. How is measured?
   iii. Is the code massively parallel?
g. What memory/processor ratio do your project require? (e.g. Gbytes/processor)
h. Parallelization model (e.g. MPI, OpenMP, Threads, UPC, Co-Array Fortran, etc.) E.g. Does your team use domain decomposition and if so what tools do you use?
i. What is the "efficiency" of the code
   i. how is it measured?
j. What are the major bottlenecks for scaling your code?

k. What is the split between interactive and batch use?
    i. Why, and is interactive more productive
l. What is the split between code development on the computer center computers and on computers at other institutions?

4.0 Project resources input from the centers

a. Steady state user of resources on a production basis per month
    i. Processor number
    ii. Processor time
    iii. Disk
    iv. Tertiary rate of change
b. Annual use of resources
    i. Processor time
    ii. Disk
    iii. Tertiary storage rate of change
c. Software provided by center
d. Consulting
e. Direct project support as a team member
f. What is the size of their jobs in terms of memory, concurrency (processors), disk, and tertiary store?
g. What is the scalability of these codes
h. What is the wall-clock time for typical runs?

5.0 Software Engineering, Development, Verification and Validation Processes

a. Software development tools used (
    i. parallel development,
    ii. debuggers,
    iii. visualization,
    iv. production management and steering
b. Software engineering practices. Please list the specific tools or processes used for
    i. configuration management,
    ii. quality control,
    iii. bug reporting an tracking,
    iv. code reviews,
    v. project planning,
    vi. project scheduling an tracking
c. What is your verification strategy?
d. What use do you make of regression tests?
e. What is your validation strategy?
f. What experimental facilities do you use for validation?
g. Does your project have adequate resources for validation?

6.0 Project output (t) and user metrics

Enumerate project output.
In addition provide:

a. # Publications?

b. Citations?
c. Dissertations?
d. Prizes and other honors?
e. Residual and supported, living datasets and/or databases that are accessed by a community?
    i. Describe size of the external user community for the datasets
f. Change in code capabilities and quality (t)
g. Code contributed to the centers
h. Code contributed to the scientific community at large
i. Company spin-offs based on code or trained people and/or CRADAs
j. Corporation, extra-agency, etc. use
k. Scientist output: Increase in trained scientists during 2001-2005,
l. Program Developers: Increase in trained code developers capable of writing project-level codes during 2001-2005

7.0 Project Future (qualitative)

a. What is today's greatest impediment in terms of your use of the center's computational facilities?
b. With the projected increases resources over next 3 yrs?
c. What do you believe the proposed increases in capacity at the facilities will provide (e.g. based on observations of historical increases)?
    i. Better turn-around time for the project
    ii. More users and incremental improvement in use with little or no change in scale or quality
    iii. Reduced granularity, resulting in constant solution time, though more accurate results
    iv. New applications permitting in new approaches and new science
b. How, specifically, has your use changed with specific facilities increases?
c. How is the project x effort projected to change in the next 5 years?
d. What is your plan for utilizing increased resources?

Questionnaire for code project history

Please fill out the short questionnaire below for your code. We need the information to address questions about what we need to do to prepare for the use of the next generation of computer platforms. The purpose is to gather some information on the size and types of codes that run on our systems. Where there are choices, please circle the appropriate choice or choices. Don't agonize over the answers. Usually one or two significant digits of accuracy are more than adequate. If you don't have data for all the questions, do the best you can. If you want to attach additional information, we would welcome it as well. Please return the questionnaire to Doug Post when you have completed it. We need it back by July 4, 2006.

Doug Post, Chief Scientist, DoD High Performance Computing Modernization Program
post@hpcmo.hpc.mil

Date: Month, Day, Year

1. Name of code:

2. Development Group
   - Institution(s)
   - Size of development team (FTEs)
   - Point of Contact Name
   - Address
   - Email
   - Telephone

3. Maintenance/user Group (if different from developers)
   - Institution(s)
   - Size of maintenance team (FTEs)
   - Point of Contact Name
   - Address
   - Email
   - Telephone

4. Domain Science Area(s)

5. Purpose of code

6. Number of users

7. Funding Sponsor(s)

DoD: Army Navy Air Force DTRA DARPA MDA
DOE: NNSA ASCR BES BER FES HE NP CSGEB FE NEST SMSE
NSF NIST NOAA NASA Other_____

8. Approximate size of code in single lines of code (sloc)
   - Total
   - Fortran 77
   - Fortran 90 or 95
   - C
   - C++
   - Python
   - Java
   - PERL
   - Other (List)

9. History and dates:
   - Development started (month/year)
   - First usable version(month/year)
   - First significant applications(month/year)
   - Reasonably mature(month/year)
   - Expected retirement(month/year)

10. Platforms that the code runs on

11. Degree of parallelism
   - Typical number of processors for a run
   - Largest number of processors that the code has run on

12. Estimate of the computer time used last year by your code (GFLOP/s—years)

13. Memory Requirements
   - Are you seriously limited by memory?
   - How much memory would you like?
   - Total memory (GBytes)
   - Memory per processor(GBytes)

14. List the Algorithms (CFD, FEM, MC, CCG, etc.,)	
15. A few key references for the code (published papers or reports, web site url, etc.)	

Appendix 7. Computational Science and Engineering Software Development Issues-D. Post, DoD HPCMP

Computational science and engineering utilizing peta-flop computers offers tremendous promise for playing a transformational role in the success of the Department of Energy Office of Science programs. The key to realizing this potential will be the successful development of the many different types of computational applications (Table 7.1) that can run effectively and efficiently on the DOE SC planned peta-flop computers as well as the development of those computing systems.

Table 7.1 Taxonomy of Computational Science and Engineering Application Projects

- **Scientific discovery—**study of new scientific phenomena such as calculating the trade-off many different effects to determine the most important mechanisms; or calculation of the non-linear behavior of a complex system such as the generation of a high-resolution first principles turbulence simulation dataset
- **Experimental analysis and design—**the analysis of experimental data from DOE research facilities; or the design of a new high energy particle detectors
- **Prediction of operational conditions—**path of a hurricane, evolution of space weather, path of a satellite, exploration of potential operating modes for a tokamak reactor experiment, …
- **Scientific design and analysis—**analysis of large datasets (e.g. screening of all known microbial drug targets against the known chemical compound libraries, design of materials with specific properties), analysis of large datasets of turbulence simulations,..
- **Engineering design and analysis—** Design of a passively safe reactor core for the Advanced Burner Reactor, tokamak reactors, high energy accelerators,…

The panel and the DOE SC computer centers surveyed the DOE SC and other computational science and engineering communities to characterize the state of development of these applications. These surveys and case studies identified many of the challenges that the application development teams will need to address (Table 2.2).

Table 7.2 Peta-flop application scaling challenges

- Scale from 100 to 10,000 GigaFlops to 1,000,000 GigaFlops
- Scale from 10s to 1,000s of processors to 10,000 to 100,000s of processors
- Evolution from small code development teams to large code development teams
- Increased emphasis of multi-disciplinary and multi-institutional code development teams
- Greater utilization of software engineering practices and metrics
- Greater employment of software project management practices

- Calculating the trade-off of many different strongly interacting effects across many more orders of magnitude of multiple time and distance scales
- Verification and validation of applications of growing complexity
- Development of problem generation and setup methods for larger and more complex problems
- Analysis and visualization of larger and more complex datasets
- Achievement of adequate levels of code performance and efficiency
- Relatively immature tools for developing and running massively parallel applications
- Developing applications to run on computers that don't yet exist.

The general characteristics and metrics (Appendix 8) for existing Tera-flop applications (Table 7.3) help define the scale of the challenge. Development of codes with either a lot of users (e.g. commercial scientific codes) or that calculate many multiple effects (e.g. weather, climate, nuclear explosions, chemistry,..) requires relatively large teams. Smaller development teams are required for codes with fewer effects or few users. The successful large teams had team members from many disciplines, i.e. team members who were domain scientists, scientific programmers, software engineers, project managers, etc. Almost all of the teams were led by domain scientists with strong computational science and leadership skills as well as domain science expertise. The larger code teams generally found it useful to adopt greater degrees of software project management and software engineering. Most computational science and engineering codes are fairly large (100s of thousands of lines of code) and took 10 years or more to develop. Fortran is the dominant language, but the number of the newest codes had significant portions of C and C++. Almost all codes utilize a number of languages, including several scripting languages (Python, PERL, etc.). C and Fortran are fairly interchangeable and pose similar challenges. Object oriented languages (e.g. C++) are also slowly gaining acceptance. The successful C++ codes generally use only a few levels of inheritance or templating. Otherwise memory latency and intercommunication kills performance. In addition, the challenge of writing clear, understandable C++ code is much greater and the learning curve is much steeper for C++ compared to C or Fortran. MPI is the dominant parallelization model by far. The average age of these projects is between 15 and 20 years. Almost all of the codes have been under continual development for their whole life. They started being used to deliver results within a few years of the start of the project, and have been productive from that point forward. Codes that didn't deliver some useful capability within a few years of project start were usually unsuccessful. Codes that cease being developed usually cease being used and die within a very few years.

For the DoD survey, the "average" code runs on 7 platforms so that the ability to port to different platforms is a high priority. This usually results in sub-optimal utilization of any particular system. The age of the codes is much longer than the life time of computer platforms (3 to 6 years), so that performance optimization above what is necessary to achieve adequate performance is a lower priority than portability. Many of the codes have been able to scale to ~ 1000 to 3000 processors, although some exhibit poor scaling above 10 to 100 processors. Interconnect latency is one of the main reasons for poor

parallel scaling. Most DoD applications typically use between 128 and 292 processors for a typical job as measured by job count, even if they have demonstrated that they can run with higher levels of processor counts. The bulk of computer times is used for larger jobs. Codes that scale well will typically use almost all the processors the scheduling system will let them use, especially those that are close to "pleasingly" parallel. The typical memory per processor varies form 0.75 to 4 GBytes. The Blue Gene L memory (c2006) of 512 Mbytes/2-processor node is a limitation for many applications; future Blue Gene plan improved memory/processor ratios.

Table 7.3 Characteristics of c2006 Tera-flop computational science and engineering applications taken from a DOD application survey of top 40 codes

Metric	Mean	median
Team size (FTEs)	38	6
Number of users	5000	27
Code size (Single lines of code)	820k	257k
Dominant Language		
Fortran	58%	
C	17%	
C++	13%	
Other	12%	
Parallelization model	Almost 100% MPI	
Project age (years)	20	17.5
Production version age (years)	15	15
Number of platforms	7	7
Largest degree of parallelization	1000 to 3000	1000 to 3000
Typical minimum of processors	225	128
Typical maximum of processors	292	128
Typical memory per processor	0.75 to 4 GBytes	

Analyses of these and other projects indicates that project success is enhanced by attention to verification and validation, software project management and software engineering, and risk minimization[1]. Almost all of the projects use some level of automated version control like CVS. Regression testing is not as common. Almost none of the projects have formal validation programs, or dedicated experimental support for

[1] *Lessons Learned From ASCI*, D. E. Post and R. P. Kendall, The International Journal of High Performance Computing Applications, **18**(2004), pp. 399-416.

validation. Validation is mostly done by comparing with the results of published data that were often taken long before the code was written and is not connected to the code project. Success for the code projects was measured by published results, customer satisfaction when there were external customers, invited papers, citations, and professional society and sponsoring organization prizes and awards, as well as grant or contract renewal.

This context and "lessons learned" identify the major ingredients of successful large-scale computational science and engineering projects, particularly the ingredients that will be important for success of the DOE SC peta-flop program applications. The DOE SC computational science and engineering program will not succeed unless the applications are successful in producing significant scientific results. Each project represents a significant investment by DOE SC. Even small code development teams will consume significant resources. Including the cost of the computer and computer center support, a six-member project using $1/20^{th}$ of a petaflop computer will cost up to $30M over a 5 year period on the leadership class facility at ORNL. It is thus essential that the application projects be well supported and well managed.

We identified six key measures and checklist items that can be tracked through peer review and DOE oversight:

1. Continual scientific and engineering output
2. Verification and Validation
3. Software project risk and management
4. Parallel scaling and parallel performance
5. Portability
6. Software engineering

It is essential that a balanced and graded approach be employed when applying these measures and checklists. Small projects work well with relatively few formal processes and would be crippled if forced to follow all the procedures necessary for much larger projects. However, even smaller projects need to organize their work and follow basic software engineering principles such as configuration management, testing, etc.

1. Continual scientific and engineering output

Successful code projects need to be continually applied to the solution of important and challenging problems. This provides a continuous set of reality checks for the application and the application development team. It ensures that the project tracks changing and evolving requirements (i.e. tracks the evolution of the emerging scientific progress in the scientific domain), and that the team members continue to be motivated and productive scientists. Measures for this include the normal ones for scientific output, e.g. publications, invited papers, patents, significant discoveries, design accomplishments, citations, etc.

2. Verification and Validation

Without verification and validation, there is little or no assurance that the code is free of important errors and defects and includes accurate treatments of all the important effects. Indeed, without validation and verification there is no assurance that computational results have any validity at all. Measures for verification include the frequency of

regression tests, the fraction of the code tested, the number and types of verification tests (symmetry, predictable behaviors, truncation error convergence with grid size, comparison with analytic test problems, benchmarks with similar codes, etc.). Measures for validation include detailed numerical and statistical comparison of code results with experimental data for conditions as close to the problems of interest as possible. Every code project should have a validation plan. When possible, it should include collaboration with relevant experimental groups. Surveys and the case studies and there is a paucity of experimental data for the relevant regimes. The best validation data is ideally obtained from experiments designed specifically for validation, especially experiments conducted after the computational result has been obtained.

3. Software project risk and management

The history of the development of large-scale scientific codes, just as for the development of industrial software, indicates that code development has many risks. As many as one-half (or more) of large-scale scientific code projects fail to achieve their initial goals. A significant portion of those never produced significant results and were abandoned without ever achieving significant results. Like all complicated endeavors involving teams, it is important to organize the collective efforts of individuals to achieve a successful outcome. The code development tasks must be planned and organized. Progress needs to be tracked and periodically reported to management. The level of organization and planning depends on the size of the code team and scale of the project. A graded approach for the level and formality of software project management is essential. Teams with only a few individuals at a single institution require relatively little planning and organization. Success for teams with many individuals from several institutions developing a complex, multi-effect code will require significant levels of planning and organization. The team leaders will need to monitor progress and adjust the project schedule and task plans accordingly. The most successful projects placed a strong emphasis on identifying, minimizing and mitigating project risks. Appropriate measures include successful reviews by internal and external monitors, completion of milestones, successful delivery of code capability, continual scientific progress reflected by the usual measures of scientific results (published papers, invited papers, citations, …). While software project management is important, it is essential to realize that scientific code development is a research activity, and requires agile processes that provide a proper balance between an organized development process and a flexible development process that can change based on the technical progress made during the project.

4. Parallel scaling and parallel performance

Most of the increased performance from the teraflop range of present computers to the petaflop range will be obtained through parallelization. Codes that can take advantage of petaflop computers will need to be able to incorporate algorithms that scale well from hundreds or thousands of processors to tens of thousand or hundreds of thousands of processors. Since this will be accomplished in stages for most codes, it will be necessary for the codes to exhibit continual progress in scaling. In addition, the code development teams will have to emphasize identifying and exploiting algorithms that have improved parallel scaling. The DOE SC should aggressively promote and support the development of such algorithms as well. Given the investment in computing the DOE is making, efficient use of the DOE petaflop computing facilities should be strongly emphasized in

allocating computer time. Appropriate measures include continual demonstration of good parallel scaling and achievement of a reasonable fraction of peak performance.

5. Portability

DOE SC will be fielding at least 3 major computer facilities in the 500 to 1000 Teraflop range. Many, if not most, of the computer applications to be able to run on at least two, and possibly even all three of these platforms. Most of the applications will also run on other platforms as well. Much of the code development and problem setup and testing will be carried out on smaller scale platforms. Code portability is thus absolutely essential. Appropriate measures include demonstration of reasonable levels of performance on key platforms, including both large scale and smaller scale platforms.

6. Software engineering

The DOE SC petaflop applications represent substantial investments by the Department of Energy (as much as $30M/project or more over a 5 year period). Attention to efficient and effective code development procedures can improve the likelihood and level of the scientific success of the code applications. Fewer defects and early detection of those defects will improve the accuracy of the scientific results. Since most of the code development will be accomplished by multi-institutional teams, procedures to facilitate coordinated code development will also need to be emphasized.

Effective software engineering practices include utilization of the best software development tools (including tools for configuration management, defect tracking, parallel profiling and optimization, static analysis, etc.), use of effective development processes such as software architecture design, code review, definition of common interface specifications and uniform code styles to facilitate module development and integration, use of collaboration tools to facilitate development by multi-institutional teams, use of problem setup tools, remote and local visualization and data analysis tools, efficient and effective archiving of datasets of results, and sharing of datasets with other groups when appropriate,

Measures include review of the appropriate level for the use of these procedures by each team, the degree to which the teams use them and demonstrations of their effectiveness.

Computational Science and Engineering Applications with Emphasis on DoD Applications

Douglass Post, Chief Scientist

DoD High Performance Computing Modernization Program
(IPA from CMU Software Engineering Institute)
With Richard Kendall (SEI), Andy Mark (HPCMP), Jeff Carver (MSU), Susan Squires (SUN), Bob Lucas (ISI), Jeremy Kepner (LL-MIT) & Tobi McFarland (HPCMP).

DOE SC Review Panel Workshop
San Francisco, July 2006

What are the characteristics of CSE applications and what are the requirements for success?

What's my Background for such an assessment?

- Development and application of CSE for astrophysics (1967), for nuclear weapons and ICF at LLNL (1968-1973) and for controlled fusion, plasma physics, atomic & molecular physics and engineering design at PPPL and ITER (1975- 1998).
- Leadership of ICF and secondary nuclear weapon code development at LLNL 1998-2000, Leadership of LANL nuclear weapon code development 2001-2003.
- Leadership of code analysis group for DARPA HPCS 2003-present.
- Leadership role in DoD High Performance Computing Modernization Program as an IPA from the CMU Software Engineering Institute

- Studied conditions for success for CSE in nuclear weapons, fusion and plasma physics, atomic and molecular physics, materials, nuclear engineering, ASCI and other fields and programs-DARPA HPCS
- Documented case studies of approximately 10 large-scale CSE projects, informal case studies of many more
- Conclusion of case studies and surveys: Domain science competence, good algorithms, V&V, software project management, and sound software engineering are the key elements for success.

Surveyed DoD codes to verify characterizations of CSE codes.

- Identify general characteristics
- Preamble (anonymity guaranteed)

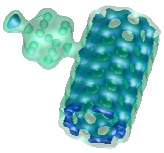
Questionnaire asked for:

- Contact information
- Code purpose
- Team size, number of users
- Domain Science area and sponsor
- Code size (slocs)
  - Total and for each language
- Code history
  - How long did the code take to develop and how old is it now?)
- Platforms
- Degree of parallelism
- Computer time usage
- Memory requirements
- Algorithms

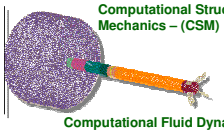
A Large, Diverse DoD User Community

- **587 projects and 2,262 users at approximately 144 sites**
- **Requirements categorized in 10 Computational Technology Areas (CTA)**

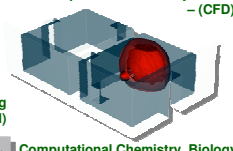
Electronics, Networking, and Systems/C4I – (ENS)



Computational Structural Mechanics – (CSM)

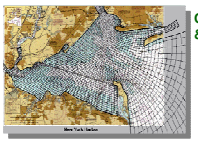


Computational Fluid Dynamics – (CFD)

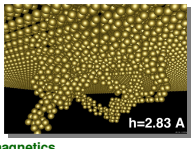


Environmental Quality Modeling & Simulation – (EQM)

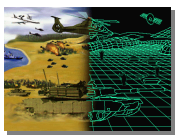
Climate/Weather/Ocean Modeling & Simulation – (CWO)

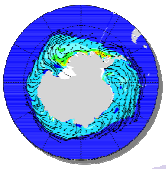


Computational Chemistry, Biology & Materials Science – (CCM)

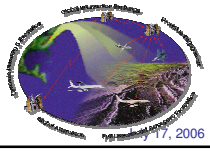


Forces Modeling & Simulation – (FMS)

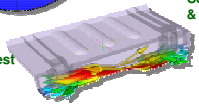




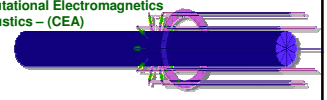
Signal/Image Processing – (SIP)



Integrated Modeling & Test Environments – (IMT)



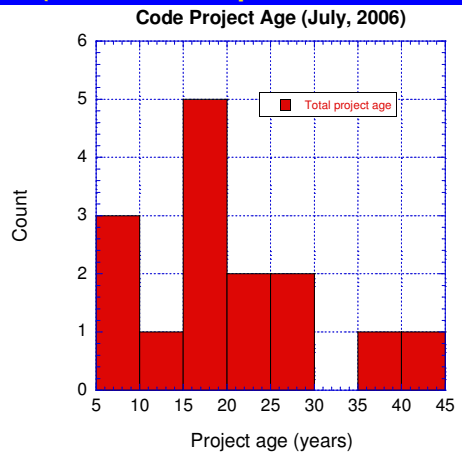
Computational Electromagnetics & Acoustics – (CEA)



We sent surveys to our top 40 codes (ordered by time requested), with 15 responses so far.

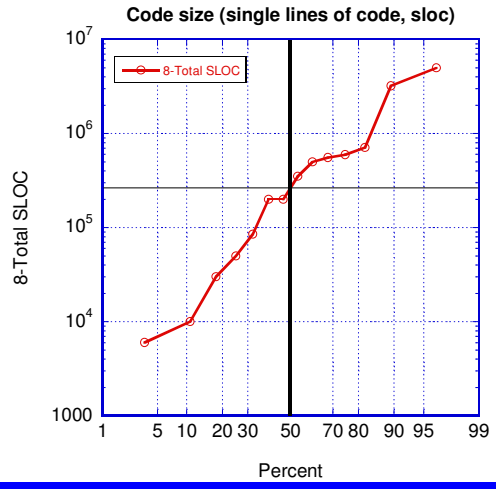
Application Code	Hours	Application Code	Hours
CTH (SNL)	93,435,421	DMOL	5,200,100
HYCOM (30% DoD)	89,005,100	ICEM (commercial)	4,950,000
GAUSSIAN (Commercial)	49,256,850	CFD++ (commercial)	5,719,000
ALLEGRA (SNL)	32,815,000	ADCIRC (DoD + academia)	4,100,750
ICEPIC (100% DoD)	26,500,000	MATLAB (commercial)	4,578,430
CAML (100% DoD)	21,000,000	NCOM	5,080,000
ANSYS (Commercial)	17,898,520	Loci-Chem	5,500,000
VASP (U.ofVienna)	18,437,500	GAMESS (Iowa State)	5,142,250
Xflow (Commercial)	15,165,000	STRIPE	4,700,000
ZAPOTEC (SNL)	12,125,857	USM3D	4,210,000
XPATCH (DoD commercial)	23,462,500	FLUENT (commercial)	3,955,610
MUVES	10,974,120	GASP	4,691,000
MOM	18,540,000	Our DNS code (DNSBLB)	2,420,000
OVERFLOW (NASA)	8,835,500	ParaDis	4,000,000
COBALT (commercial)	14,165,750	FLAPW	4,050,000
ETA	11,700,000	AMBER	4,466,000
CPMD (MPI & IBM)	5,975,000	POP (LANL)	3,800,000
ALE3D (LLNL)	5,864,500	MS-GC	3,500,000
PRONTO (SNL)	5,169,100	TURBO	3,600,600
		Freericks Solver	2,600,000

Most projects are at least 15 years old (and had predecessors).



- Almost all the codes that will run on platforms delivered within the next 5 years **exist now**.

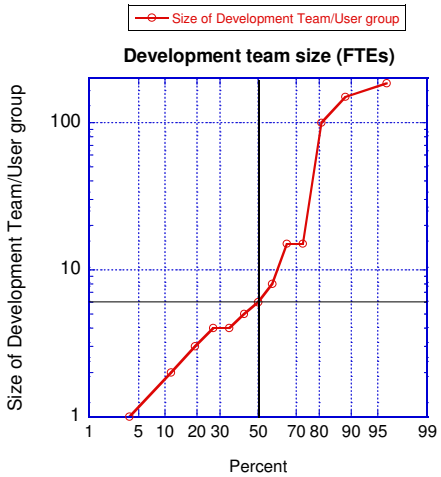
Median code size is ~ 300,000 slocs.



- Most codes will take 5 years or more to develop[1].

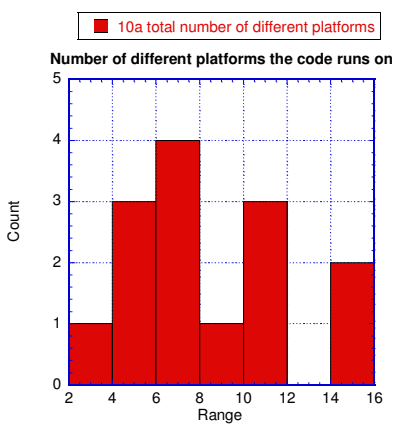
[1]D. E. Post and R. P. Kendall, *International Journal of High Performance Computing Applications*, 18(2004), pp. 399-416

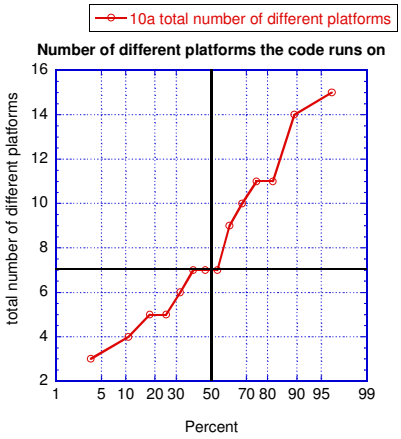
Median team size is 6 FTEs.



- Teamwork will be essential for new codes, especially for petaflop computing.

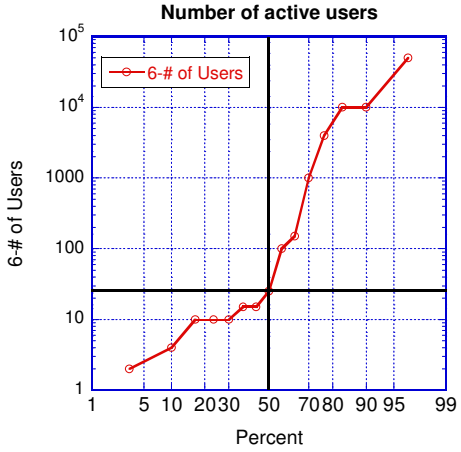
Median code runs on 7 different platforms.





- Code portability is a key, if not dominant, priority for code developers.

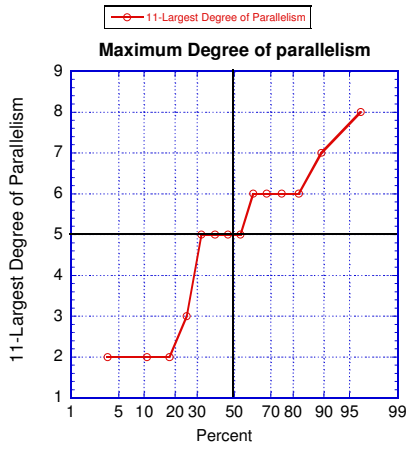
Median code has ~ 25 users.



- User support and acceptance will be essential for success
- Support for code maintenance will be essential!

Median code is fairly parallel.

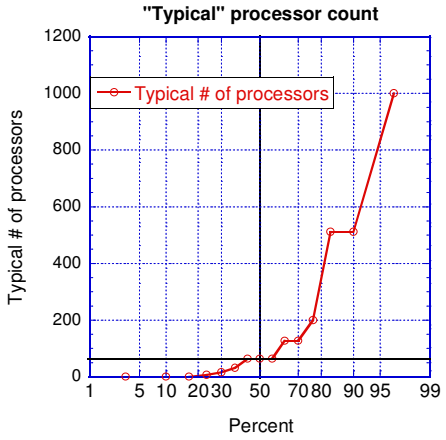
8. > 30,000 processors
7. 10,001 to 30,000 processors
6. 3,001 to 10,000 processors
5. 1,001 to 3000 processors
4. 300 to 1,000 processors
3. 101 to 300 processors
2. 11 to 100 processors
1. Less than 10 processors



- We have to scale from 100-3,000 processors to 50,000-200,000 processors in two years to achieve petaflop performance.

11

"Routine" processor count is much less than peak.



- We have to scale from 30-200 processors to 20,000-200,000 processors in two years to achieve petaflop performance.

12

58% of the codes are predominantly written in Fortran.

	Team size FTEs	# users	Total sloc(k)	SLOC Fortran 77 (k)	SLOC Fortran 90, 95 (k)	SLOC C (k)	SLOC C++ (k)	other
Mean	38	5,038	820	24%	34%	17%	13%	13%
Median	6	27	275					

- New languages with higher levels of abstraction are attractive, but they will have to be compatible and inter-operable with Fortran with MPI.

Most runs don't use a lot of processors.

	Total project age	age production version	total number of different platforms	Largest Degree of Parallelism	Typical minimum # of processors	Typical Maximum # of processors	Is memory a limitation?	Memory processor GBytes /proc
Mean	19.8	15.1	6.9	1000 to 3000	225	292	Sometimes	0.75-4
Median	17.5	15.5	7.0	1000 to 3000	128	128		

- Most users want at least 1 GByte / processor of memory.

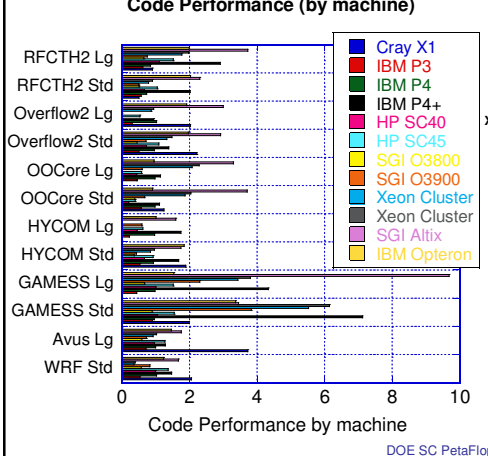
Code performance varies among platforms. HPCMP TI-05 Application Benchmark Codes perform differently on different platforms.

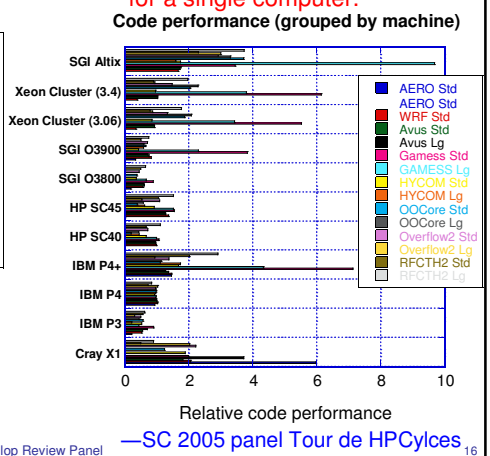
- **Studied performance of 9 DoD HPCMP benchmark codes on 12 different HPCMP platforms**
- **Aero** – Aeroelasticity CFD code (Fortran, serial vector, 15,000 lines of code)
- **AVUS** (Cobalt-60) – Turbulent flow CFD code (Fortran, MPI, 19,000 lines of code)
- **GAMESS** – Quantum chemistry code (Fortran, 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)
- **CTH** – Shock physics code (SNL) (~43% Fortran/~57% C, MPI, 436,000 lines of code)
- **WRF** – Multi-Agency mesoscale atmospheric modeling code (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)

Performance depends on the computer and on the code.

- Normalized Performance = 1 on the NAVO IBM SP3 (HABU) platform with 1024 processors (375 MHz Power3 CPUs) assuming that each system has 1024 processors.
- GAMESS had the most variation among platforms.

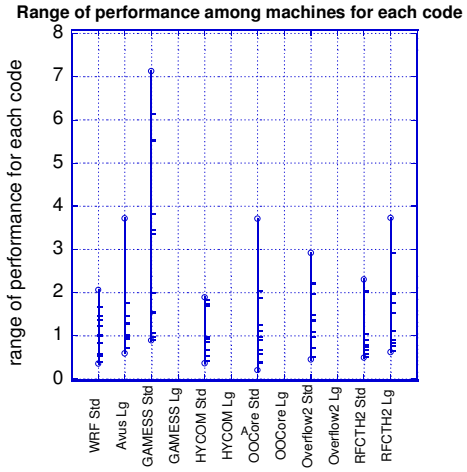
Substantial variation of codes for a single computer.





—SC 2005 panel Tour de HPCylces

Performance range of codes is large.



General conclusions

- Performance depends on application and on the computer
  - No computer works best for all applications
  - A suite of applications requires a suite of computer types
- Tuning for a platform can pay off in a big way
- Shared memory is really good for some codes

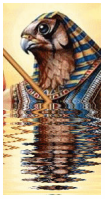
5 detailed DARPA HPCS case studies of CSE codes begin to span CSE space.

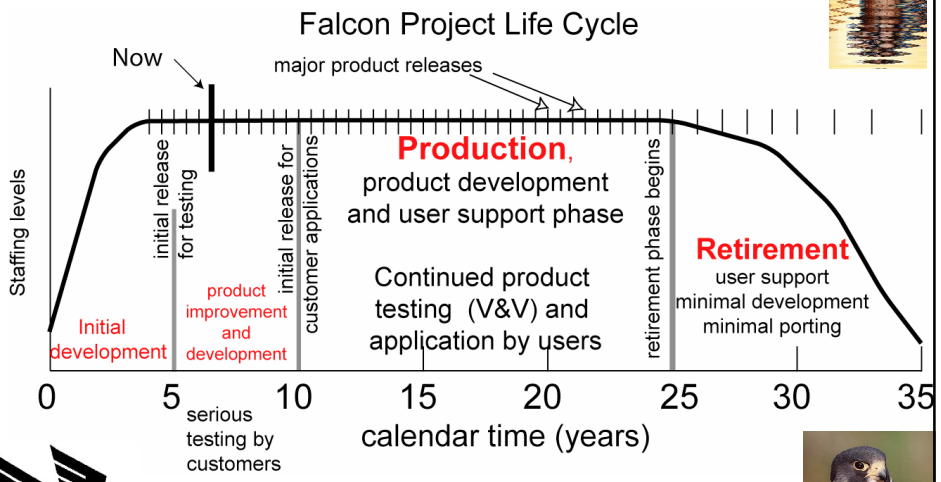
	Falcon	Hawk	Condor	Eagle	Nene
Application Domain	Product Performance	Manufacturing	Product Performance	Signal Processing	Process Modeling
Project Duration	~10 years (since 1995)	~6 years (since 1999)	~20 years (since 1985)	~3 years	~25 years (since 1982)
Number of Releases	9 Production	1	7	1	> 20
Earliest Predecessor	1970s	early 1990s	1969	?	1977-78
Staffing	15 FTEs	3 FTEs	3-5 FTEs	3FTEs	~10FTEs+100s of contributors
Customers	<50	10s	100s	Demonstration code	~100,000
Nonimal Code Size	~405,000	~134,000	~200,000	<100,000	760,000
Primary Languages	F77 (24%), C (12%)	C++ (67%), C (18%)	Fortran 77 (85%)	C++, Matlab	Fortran 77 (95%)
Other Languages	F90,Python,Perl,ksh/ csh/sh	Python, Fortran 90	Fortran 90, C, Slang	Java Libraries(~70%)	C (1%)
Target Hardware	Parallel Supecomputers	Parallel Supercomputers	PCs to Parallel Supercomputers	Embedded App	PCs to Parallel Supercomputers
Status	Production	Production ready	Production	Demonstration code	Production
Sponsors	DOE	DoD	DoD	DoD	DoD, DOE, NSF

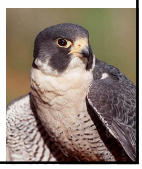
DARPA HPCS Team Identified Key Characteristics from Detailed CSE Case Studies

- General project properties
- Life cycle
- Workflows
- Observations and comparisons
- Tools
- Lessons learned

Requirements for Computers and Application Codes Strongly Influenced by Code Project Life Cycle and Workflows*







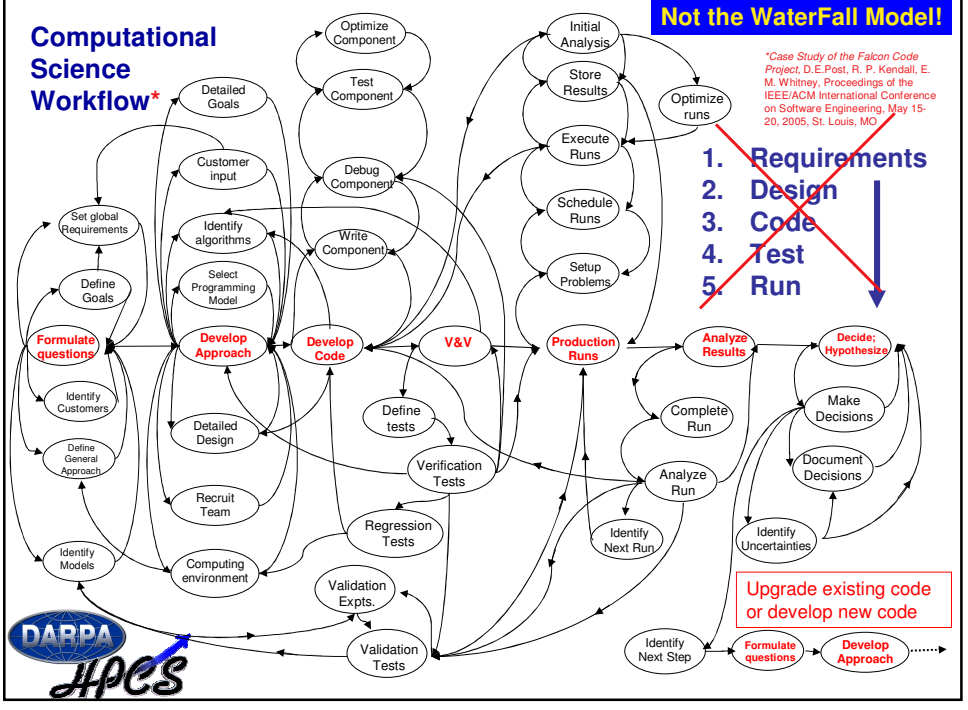
*Case Study of the Falcon Code Project, D.E.Post, R. P. Kendall, E. M. Whitney, Proceedings of the IEEE/ACM International Conference on Software Engineering, May 15-20, 2005, St. Louis, MO

Computational Science Workflow*

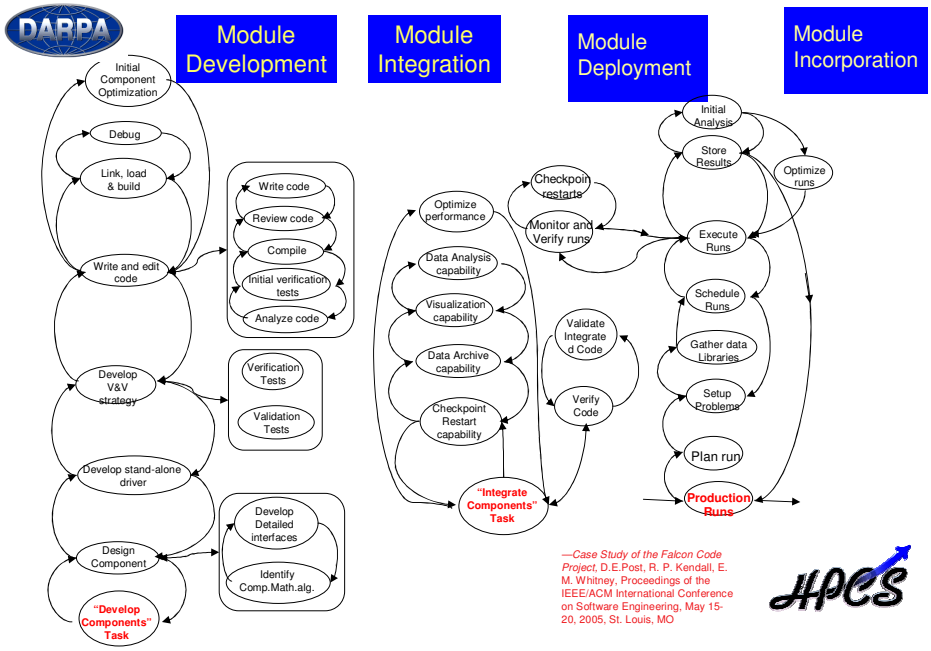
Not the WaterFall Model!

*Case Study of the Falcon Code Project, D.E.Post, R. P. Kendall, E. M. Whitney, Proceedings of the IEEE/ACM International Conference on Software Engineering, May 15-20, 2005, St. Louis, MO

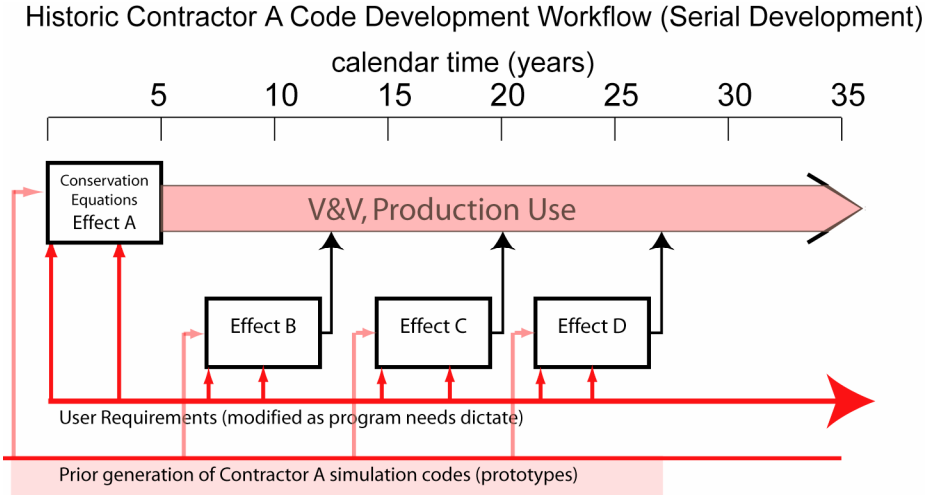
1. ~~Requirements~~
2. ~~Design~~
3. ~~Code~~
4. ~~Test~~
5. Run



The Nene Workflow Model



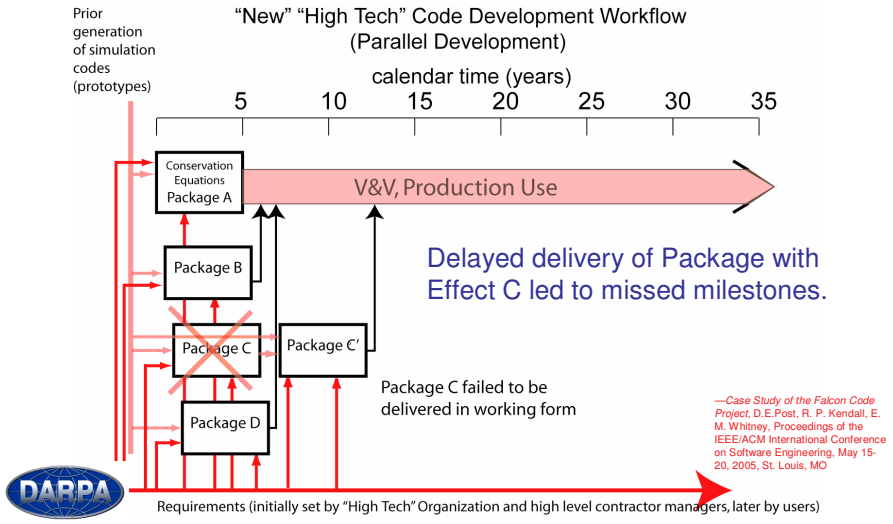
Proto-FALCON Workflows were initially serial



DARPA

—Case Study of the Falcon Code Project, D.E.Post, R. P. Kendall, E. M. Whitney, Proceedings of the IEEE/ACM International Conference on Software Engineering, May 15-20, 2005, St. Louis, MO

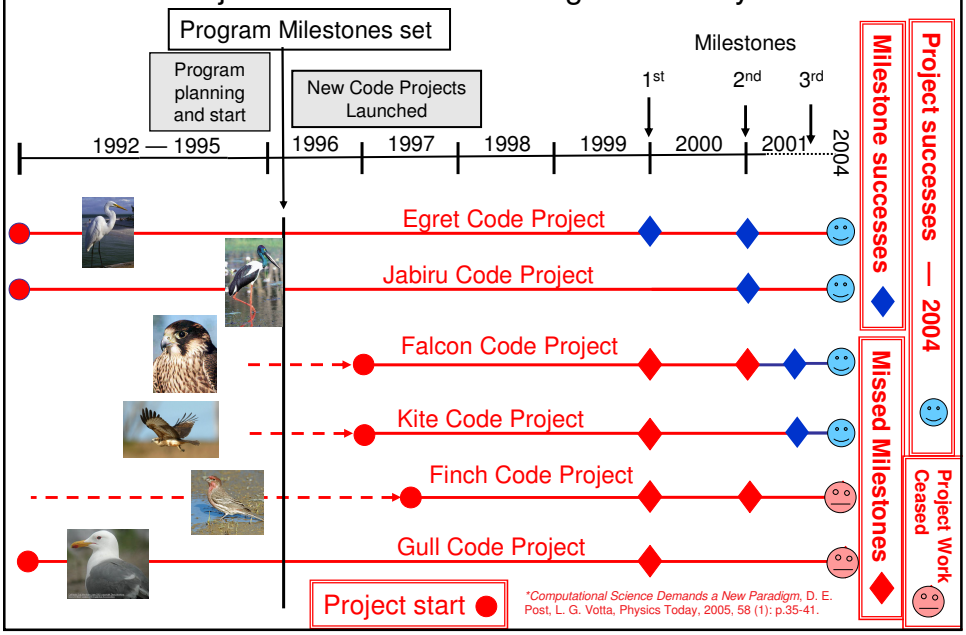
Ambitious schedule required parallel development with no contingency.



—*Case Study of the Falcon Code Project*, D.E.Post, R. P. Kendall, E. M. Whitney, Proceedings of the IEEE/ACM International Conference on Software Engineering, May 15-20, 2005, St. Louis, MO

Large scale code development is risky*.

Code Project Schedule for Six Large-scale Physics Codes



Computational Science Demands a New Paradigm, D. E. Post, L. G. Votta, Physics Today, 2005, 58 (1): p.35-41.

We studied these projects to identify the *"Lessons Learned*"*

The Successful projects emphasized:

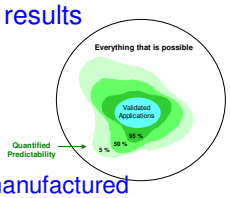
- Conservative approach - Minimize Risks!
  - Building on successful code development history and prototypes
  - Better physics and computational mathematics over better "computer science"
  - The use of proven Software Engineering rather than new Computer Science
    - Don't let the code project become a Computer Science research project!
- Sound Software Project Management - Plan and Organize the Work!
  - Highly competent and motivated people in a good team
  - Development of the team
  - Software Project Management: Run the code project like a project
  - Determining the Schedule and resources from the requirements
  - Identifying, managing and mitigating risks
  - Focusing on the customer
    - For code teams and for stakeholder support
  - Software Quality Engineering: Best Practices rather than Processes
- Verification and Validation – Correct Results are Essential!
  - Need for improved V&V methods became very apparent

The unsuccessful projects didn't emphasize these!

Lessons Learned From ASCI, D. E. Post and R. P. Kendall, The International Journal of High Performance Computing Applications, 18(2004), pp. 399-416.

Verification and Validation

- Customers want to know why they should believe code results
- Codes are only a model of reality
- Verification and Validation are essential
- Verification
  - Verify equations are solved correctly
  - Regression suites of test problems, convergence tests, manufactured solutions, analytic test problems, code comparisons and benchmarks
- Validation
  - Ensure models reflect nature, check code results with experimental data
  - Specific validation experiments are required
    - Federal sponsor is funding multi-billion dollar validation experiments for V&V,…
- V&V experience with thes and other codes indicates that a stronger intellectual basis is needed for V&V
- More intense efforts are needed in both types of V&V if computational science is to be credible



—Computational Science Demands a New Paradigm, D. E. Post, L. G. Votta, Physics Today, 2005, 58 (1): p.35-41.

Roach, 1998; Roache, 2002; Salari and Knupp, 2000; Lindl, 1998; Lewis, 1992; Laughliin, 2002)

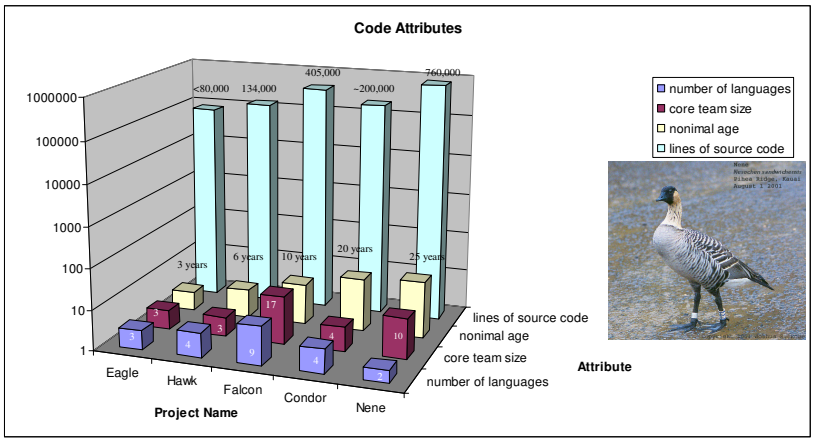
DARPA HPCS team made 9 observations based on detailed case studies.

- We made 9 observations from the five detailed case studies (Falcon, Hawk, Condor, Eagle, Nene).
  - These observations and conclusions were consistent with our prior, less detailed case studies.
- These 9 observations help identify the issues to focus on for petaflop applications.

Nine Cross-Study Observations

1. Once selected, the primary languages (typically Fortran) adopted by existing code teams do not change.
2. The use of higher level languages (e.g. Matlab) has not been widely adopted by existing code teams except for "bread-boarding" or algorithm development.
3. Code developers in existing code teams like the flexibility of UNIX command line environments.
4. Third party (externally developed) software and software development tools are viewed as a major risk factor by existing code teams.
5. The project goal is scientific discovery or engineering design. "Speed to solution" and "execution time" are not highly ranked goals for our existing code teams unless they directly impact the science.
6. All but one of the existing code teams we have studied have adopted an "agile" development approach.
7. For the most part, the developers of existing codes are scientists and engineers, not computer scientists or professional programmers.
8. Most of the effort has been expended in the "implementation" workflow step.
9. The success of all of the existing codes we have studied has depended most on keeping their customers (not always their sponsors) happy.

Summary of Code Attributes



Codes primarily use one or two programming languages, but utilize many others for special purposes.

	Falcon	**Hawk**	**Condor**	**Eagle**	Nene
Application Domain	Product Performance	Manufacturing	Product Performance	Signal Processing	Process Modeling
Project Duration	~10 years (since 1995)	~6 years (since 1999)	~20 years (since 1985)	~3 years	~25 years (since 1982)
Number of Releases	9 Production	1	7	1	?
Earliest Predecessor	1970s	early 1990s	1969	?	1977-78
Staffing	15 FTEs	3 FTEs	3-5 FTEs	3FTEs	~10FTEs+100s of contributors
Customers	<50	10s	100s	Demonstration code	~100,000
Nonimal Code Size	~405,000	~134,000	~200,000	<100,000	750,000
Primary Languages	**F77 (24%), C (12%)**	**C++ (67%), C (18%)**	**Fortran 77 (85%)**	**C++, Matlab**	**Fortran 77 (95%)**
Other Languages	F90,Python,Perl,ksh/csh/sh	Python, Fortran 90	Fortran 90, C, Slang	Java Libraries(~70%)	C (1%)
Target Hardware	Parallel Supecomputers	Parallel Supercomputers	PCs to Parallel Supercomputers	Embedded App	PCs to Parallel Supercomputers
Status	Production	Production ready	Production	Demonstration code	Production
Sponsors	DOE	DoD	DoD	DoD	DoD, DOE, NSF

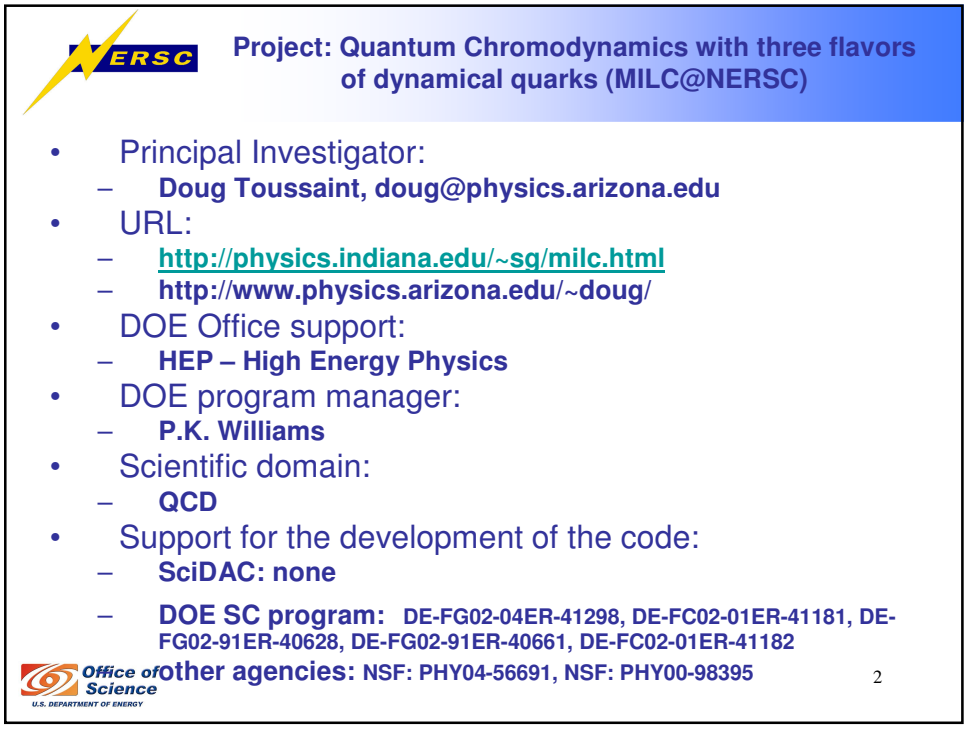
What do teraflop applications tell us?

- Need measures for applications:
  - V&V
  - Software engineering, project planning and management, software quality
  - Incremental delivery, risk minimization and avoidance
  - Time to solution: code project and centers
  - Success and effectiveness of application
  - Life cycle sustainment
    - Invest 100s of $M, how will DOE preserve capability that has been developed?
  - Is the SciDAC funding adequate, or is more support needed to ensure successful code development?
  - Does peer review process need to be enlarged to assess software engineering issues?

What do teraflop applications tell us?

- Need measures for centers:
  - Productivity
    - Time to Solution
  - Programming and production efficiency (not Linpack performance)
    - Better Benchmarks
  - Software development and production tools
  - User support
  - User requirements
  - Utilization effectiveness


BERKELEY LAB
NERSC
NATIONAL ENERGY RESEARCH
SCIENTIFIC COMPUTING CENTER
NERSC 4 Showcase Projects
Francesca Verdier
Associate Manager, NERSC Services
fverdier@lbl.gov
ASCAC Metrics Sub Panel Meeting
July 17, 2006
Office of Science
U.S. DEPARTMENT OF ENERGY


NERSC
Project: Quantum Chromodynamics with three flavors of dynamical quarks (MILC@NERSC)
• Principal Investigator:
– Doug Toussaint, doug@physics.arizona.edu
• URL:
– http://physics.indiana.edu/~sg/milc.html
– http://www.physics.arizona.edu/~doug/
• DOE Office support:
– HEP – High Energy Physics
• DOE program manager:
– P.K. Williams
• Scientific domain:
– QCD
• Support for the development of the code:
– SciDAC: none
– DOE SC program: DE-FG02-04ER-41298, DE-FC02-01ER-41181, DE-FG02-91ER-40628, DE-FG02-91ER-40661, DE-FC02-01ER-41182
other agencies: NSF: PHY04-56691, NSF: PHY00-98395
Office of Science
U.S. DEPARTMENT OF ENERGY
2

Project: Quantum Chromodynamics with three flavors of dynamical quarks

- What problem are you trying to solve?
  - **This research addresses fundamental questions in high energy and nuclear physics, and is directly related to major experimental programs in these fields. In particular we are simulating systems which test the portion of the standard model of high energy physics that describes the strong interactions.**
- What is the expect impact of project success?
  - **Non-perturbative QCD can determine the correctness of the Standard Model as well as establish agreement between theory and a variety of experimental results. The U.S. spends 750 million dollars per year on HEP experiments computational validation and cross checking of that work is crucial.**
- External communities & sizes that code and/or datasets support:
  - **The MILC Collaboration is engaged in a broad research program in Quantum Chromodynamics (QCD). This research addresses fundamental questions in high energy and nuclear physics, and is directly related to major experimental programs in these fields. It includes studies of the mass spectrum of strongly interacting particles, the weak interactions of these particles, and the behavior of strongly interacting matter under extreme conditions.**
  - **Data is contributed to "The Gauge Connection" at http://qcd.nersc.gov.**

Office of Science U.S. DEPARTMENT OF ENERGY

3

2. MILC@NERSC Project Team Resources

- Team institutional affiliations:
  - **University of Arizona**
  - **Indiana University**
  - **University of California, Santa Barbara**
  - **Washington University**
  - **Boston University**
- To what extent are the code team members affiliated with the computer center institution? (e.g. are the team members also members of the computer center institution?)
  - ***Team members are largely based at Universities. None currently at NERSC.***
- Team composition and experience:
  - **domain scientists: 6**
  - **graduate students and postdocs: 6-10**
  - **computer scientists: n/a**
  - **computational mathematicians: n/a**
  - **database managers: n/a**
  - **programmers: n/a**

Office of Science U.S. DEPARTMENT OF ENERGY

4

2. MILC@NERSC Project Team Resources

- Team composition by educational level:
  - **senior faculty: 6**
  - **Grad students and Postdocs: 6-10**

- Team resources utilization:
  - **time spent on code and algorithm development:**
    - **Significant ongoing development, but largely not done at NERSC for efficient use of allocation. Can test/debug most changes at small scale.**
  - **code maintenance:**
    - **Ongoing, but done by a limited set of the team**
  - **problem setup:**
    - **Relatively straightforward**
  - **production runs:**
    - **Predominant use of NERSC allocation**
  - **results analysis:**
    - **A variety of codes are used by different researchers to analyze the quark configurations we produce. Those are analyzed by members of the collaboration and potentially flow to the larger QCD community. Significant use of data output from large scale runs is done both at NERSC and provided to the QCD community for a variety of physics analysis.**
    - **See "The Gauge Connection" at http://qcd.nersc.gov**
  - **publications:**
    - **All team members participate in publishing**

Office of Science U.S. DEPARTMENT OF ENERGY

3. Project Code: MILC (su3_rmd)

- Problem Type:
  - **Simulation and physics analysis of simulation results**
- Types of algorithms and computational mathematics:
  - **Lattice Monte Carlo, Large sparse matrix inversion (CG)**
- What platforms does your code routinely run on?
  - **IBM SPs, Linux Clusters, and specialized QCD hardware**
- Code size (single lines of code, function points, etc.);
  - **Close to 18 years, very stable in terms of the size of code.**
- Computer languages employed:
  - **C , Assembly language, and MPI**
- What libraries are used? And What fraction of the codes does it represent?
  - **None. Code is self contained.**
- Code Mix:
  - **To what extent does your team develop and use your own codes? 100%**
  - **Codes developed by others in the DOE and general scientific community? no**
  - **Commercial application codes provided by the center? no**

Office of Science U.S. DEPARTMENT OF ENERGY

3. Project Code: MILC

- What is the present parallel scalability on each of the computers the code operates on
  - **Projected or maximum scalability: The scale at which runs are done is determined:**
    - **In principle by the performance of global reductions**
    - **in practice by queue structure/policy and its impact on turn around time (turn around time is what is most important)**
  - **How is measured? wall clock time.**
  - **Is the code massively parallel? MILC runs well on thousands of processors and is expected to keep pace with the scale of future MPP resources.**
- What memory/processor ratio do your project require? (e.g. Gbytes/processor)
  - ***1-2 GB per processor is an upper bound on the current calculations***
- Parallelization model: **MPI**
- Does your team use domain decomposition and if so what tools do you use?
  - **Spatial decomposition. Regular lattice and one temporal dimension.**

Office of Science U.S. DEPARTMENT OF ENERGY

3. Project Code: MILC

- What is the "efficiency" of the code and how is it measured:
  - **The code emits wall clock timings for each section. Code profiling for more detailed performance analysis.**
- What are the major bottlenecks for scaling your code?
  - **On some architectures at very large concurrency the performance of global reductions (MPI_Allreduce) suffers due to scaling bottlenecks. These have been studied in detail and the bottlenecks are inherent in the MPI library not the MILC code.**
- What is the split between interactive and batch use? Why this split? Is interactive use more productive?
  - ***Most runs at NERSC are batch (insignificant interactive use)***
- What is the split between code development on the computer center computers and on computers at other institutions?
  - ***Nearly all development and testing is done on local (researcher owned) computers. QCD can be tested and debugged at small scale and there is no point burning allocated time for that work.***

Office of Science U.S. DEPARTMENT OF ENERGY

4. MILC@NERSC project resources input from the centers

- Plan with benchmarks & milestones:
  - ***In the next year, we expect to generate several hundred archived gauge configurations in each of these ensembles. We plan to divide the work of analyzing these configurations between NERSC and other centers where we have also applied for time. For next year we ask to analyze 100 configurations from each of these ensembles, out of 300 total that we hope to generate.***
- Steady state user of resources on a production basis per month (desired):
  - **Processor number: 1024 and 2048 way**
  - **Processor time 300-400 K IBM SP POWER3-equivalent hours**
  - **Disk : 250 GB**
  - **Tertiary amount and rate of change: 10GB**
- Annual use of resources (actual):
  - **Processor time (IBM SP POWER3-equivalent hours):**
    - **2002: 1.3M allocated; 1.8M used**
    - **2003: 1.5M allocated; 2.3M used**
    - **2004: 2.5M allocated; 3.1M used (14 months)**
    - **2005: 2.2M allocated; 2.1M used**
    - **2006: 2.4M allocated; 4.4M used (7 months)**
  - **Disk: 250 GB**
  - **Tertiary storage rate of change: 4.6 TB AY05; 3.2 TB so far this year**

Office of Science U.S. DEPARTMENT OF ENERGY

9

4. MILC@NERSC project resources input from the centers

- Software provided by center:
  - **C, MPI, performance profiling tools (IPM)**
- Consulting provided by center:
  - **Root caused scaling bottlenecks in MILC. Primary scaling barrier is inherent performance bottlenecks in MPI_Allreduce.**
- Direct project support from center acting as a team member:
  - ***none***
- What is the size of their jobs in terms of:
  - **memory: 500 MB aggregate**
  - **concurrency (processors): 2048**
  - **disk: 4-5 GB per run**
  - **tertiary store: insignificant**
- What is the scalability of these codes:
  - ***4096 cpus would be a good target concurrency, but for some NERSC architectures running at 2048 is preferred due to MPI_Allreduce scaling issues when running in a non-dedicated mode.***
- What is the wall-clock time for typical runs?
  - ***8 hours or as determined by best turn around given queue policies***

Office of Science U.S. DEPARTMENT OF ENERGY

10

5. MILC@NERSC Software Engineering, Development, Verification and Validation Processes

- Software / Project development tools used:
  - **Most software development is done off-site. E.g. through the Lattice Gauge Theory SciDAC project or by Doug T. NERSC does aid our project management by hosting a web based data repository for the QCD configurations that are the product of large scale computation. That allows other researchers to use the data that is generated from our work. http://qcd.nersc.gov**
- Software engineering practices. Please list the specific tools or processes used for:
  - **configuration management:**
  - **quality control:**
  - **bug reporting and tracking:**
  - **code reviews:**
  - **project planning:**
  - **project scheduling and tracking:**

These are all handled through the MILC collaboration

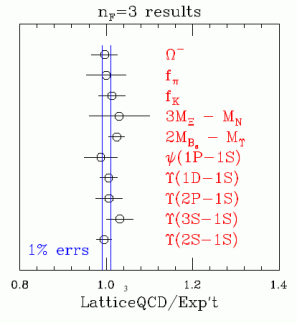
Office of Science U.S. DEPARTMENT OF ENERGY

11

5. MILC@NERSC Software Engineering, Development, Verification and Validation Processes

- What is your verification strategy (correctness)?
- What use do you make of regression tests?
- What is your validation strategy (functionality and operability)?
- What experimental facilities do you use for validation?
- Does your project have adequate resources for validation?

There is a large body of experimental data with which we can compare our results. Our calculations can be checked against very accurately known experimental values, such as the mass of the proton. The comparison at right shows error estimates that make up a more detailed comparison of simulation to experiment.



Office of Science U.S. DEPARTMENT OF ENERGY

5a. MILC@NERSC project code productivity & scalability

- Measures of experiment productivity and performance including scalability of runs:
  - ***Scalable performance up to 2048 seaborg CPUs has been demonstrated in previous ERCAP requests.***
  - ***Most of the productivity targets have to do with scientific understanding being conveyed through peer-reviewed journals.***
- Scaling limits including i/o, node memory size, interconnect b/w or latency, algorithm:
  - ***Small message latency and the scaling of MPI collectives. The latter has been shown to be sensitive to architectural issues and the quality of the MPI_Allreduce implementation. I/O and memory demands tend to be modest.***
- Projected scalability:
  - ***MILC and other QCD codes should be able to fully scale on tomorrow's large scale multi-core machines. As long as system architects keep small message latency low and provide scalable global reductions, MILC should be able to make efficient use of even the largest systems.***

Office of Science U.S. DEPARTMENT OF ENERGY

5a. MILC@NERSC Project code scalability (history)

	16-256 CPUs	512 CPUs	1,024 CPUs	1,280 CPUs	1,536CPUs	2,048CPUs	3,072 CPUs
AY2006	0.3%	13.7%	0.1%	-	3.7%	81.4% (max)	0.8% (max)
AY2005	0.6%	6.0%	13.1%	1.0%	79.4% (max)	-	-
AY2004	6.4%	3.4%	18.0%	67.7%	~0	3.1% (max)	-
AY2003	60.9%	0.9%	26.4%	5.5%	0.8%	0.1% (max)	-
AY2002	98.9%	1.1% (max)	-		-	-	-

Office of Science U.S. DEPARTMENT OF ENERGY

6. MILC@NERSC Scientific | Engineering Output

- The scientific accomplishments 2000 to present:

  We have just completed a study of the equation of state of high temperature QCD at zero baryon density on lattices with four and six time slices. Final results will be presented at the Lattice 2006 conference. This is the first study with an improved action and a realistic set of quarks with such small lattice spacings.
- The effect on the Office of Science programs:

  The results of this work combines with outputs from experimental HEP experiments and programs to provide an increasingly detailed understanding of the Standard Model.
- Publications:
  - ***See appendix***
- Citations (last 5 years):
  - ***(haven't received yet)***
- Dissertations:
  - ***1-2 students per year***
- Prizes and other honors:
  - ***(haven't received yet)***
- Residual and supported, living datasets and/or databases that are accessed by a community? Size of the community?
  - **Yes. The web based QCD data repository hosted by NERSC. http://qcd.nersc.gov.**
  - Change in code capabilities and quality:
  - ***The MILC code is improved through the Lattice Guage Theory SciDAC project and through direct implementation of new algorithms by Doug Toussaint.***

Office of Science U.S. DEPARTMENT OF ENERGY

6. MILC@NERSC Scientific | Engineering Output

- Code and/or data contributed to the centers:
  - ***None, we give them performance feedback***
- Code and/or data, results, contributed to the scientific and engineering community at large:
  - ***Yes through the Gauge Connection, a widely used web based repository of QCD quark configurations. http://qcd.nersc.gov***
- Company spin-offs based on code or trained people and/or CRADAs:
  - **N/A**
- Corporation, extra-agency, etc. use:
  - ***none***
- Production of scientists & computational scientists during 2001-2005:
  - ***Roughly one to two students a year.***
- Production of trained software engineers 2001-2005: ***N/A***

Office of Science U.S. DEPARTMENT OF ENERGY

MILC Publications 2000-2006

Publications of the MILC Collaboration : Refereed Journals (2000-2006)

Critical Behavior in Nt=4 Staggered Fermion Thermodynamics, C. Bernard, C. DeTar, S. Gottlieb, U.M. Heller, J. Hetrick, K. Rummukainen, R. Sugar and D. Toussaint, Phys. Rev. D61, 054503, (2000) [arXiv:hep-lat/9908008].

Scaling tests of the improved Kogut-Susskind quark action, C. Bernard, T. Burch, T.A. DeGrand, C. DeTar, S. Gottlieb, U.M. Heller, J. Hetrick, K. Orginos, R. Sugar and D. Toussaint, Phys. Rev. D 61, 111502, (2000) [arXiv:hep-lat/9912018].

The static quark potential in three flavor QCD, C. Bernard, T. Burch, T.A. DeGrand, C. DeTar, S. Gottlieb, U.M. Heller, J. Hetrick, K. Orginos, R. Sugar and D. Toussaint, Phys. Rev. D62, 034503, (2000).

The QCD spectrum with three quark flavors, C. Bernard, T. Burch, T. DeGrand, S. Datta, C. DeTar, S. Gottlieb, U.M. Heller, K. Orginos, R. Sugar and D. Toussaint), Phys. Rev. D64, 054506, (2001) [arXiv:hep-lat/0104002].

Zero temperature string breaking in lattice quantum chromodynamics, C. Bernard, T. DeGrand, C. DeTar, S. Gottlieb, U.M. Heller, J. Hetrick, P. Lacock, K. Orginos, R. Sugar and D. Toussaint), Phys. Rev. D64, 074509, (2001) [arXiv:hep-lat/0103012].

Measurement of hybrid content of heavy quarkonia using lattice NRQCD, T. Burch, K. Orginos and D. Toussaint, Phys. Rev. D64, 074505, (2001) [arXiv:hep-lat/0103025].

Lattice results for the decay constant of heavy-light vector mesons, C. Bernard, P. Williams, S. Datta, S. Gottlieb, C. DeTar, U. M. Heller, C. McNeile, K. Orginos, R. Sugar and D. Toussaint, Phys. Rev. D65, 014510, (2002) [arXiv:hep-lat/0109015].

Chiral Logs in the Presence of Staggered Flavor Symmetry Breaking, C. Bernard, Phys. Rev. D65, 054031, (2002) [arXiv:hep-lat/0111051].

Lattice Calculation of Heavy-Light Decay Constants with Two Flavors of Dynamical Quarks, C. Bernard, S. Datta, T. DeGrand, C. DeTar, Steven Gottlieb, Urs M. Heller, C. McNeile, K. Orginos, R. Sugar and D. Toussaint, Phys. Rev. D66, 094501, (2002) [arXiv:hep-lat/0206016].

Witten-Veneziano Relation, Quenched QCD, and Overlap Fermions, Thomas DeGrand and Urs M. Heller (The MILC Collaboration), Phys. Rev. D65, 114501, (2002) [arXiv:hep-lat/0202001].

Lattice calculation of 1−+ hybrid mesons with improved Kogut-Susskind fermions, C. Bernard, T. Burch, C. DeTar, Steven Gottlieb, E.B. Gregory, U.M. Heller, J. Osborn, R. Sugar, and D. Toussaint, Phys. Rev. D68, 074505 (2003) [arXiv:hep-lat/0301024].

Pion and Kaon masses in Staggered Chiral Perturbation Theory, C. Aubin and C. Bernard, Phys. Rev. D68, 034014 (2003) [arXiv:hep-lat/0304014].

36. High-Precision Lattice QCD Confronts Experiment, The Fermilab Lattice, HPQCD, MILC and UKQCD Collaborations: C. T. H. Davies, E. Follana, A. Gray, G. P. Lepage, Q. Mason, M. Nobes, J. Shigemitsu, H. D. Trottier, M. Wingate, C. Aubin, C. Bernard, T. Burch, C. DeTar, Steven Gottlieb, E. B. Gregory, U. M. Heller, J. E. Hetrick, J. Osborn, R. Sugar, D. Toussaint, M. Di Pierro, A. El-Khadra, A. S. Kronfeld, P. B. Mackenzie, D. Menscher, J. Simone, Phys. Rev. Lett. 92, 022001 (2004) [arXiv:hep-lat/0304004].

37. Hybrid configuration content of heavy S-wave mesons, Tommy Burch and Doug Toussaint (The MILC Collaboration), Phys. Rev. D68, 094504 (2003) [arXiv:hep-lat/0305008].

38. Topological Susceptibility with the Improved Asqtad Action, C. Bernard, T. Burch, T. DeGrand, C. DeTar, Steven Gottlieb, E. Gregory, A. Hart, A. Hasenfratz, U.M. Heller, J. Hetrick, J. Osborn, R.L. Sugar, D. Toussaint, Phys.Rev. D68, 114501 (2003) [arXiv:hep-lat/0308019].

39. First determination of the strange and light quark masses from full lattice QCD, The HPQCD, MILC and UKQCD Collaborations: C. Aubin, C. Bernard, C. Davies, C. DeTar, S. Gottlieb, A. Gray, E. Gregory, J. Hein, U.M. Heller, J. Hetrick, G. Lepage, Q. Mason, J. Osborn, R. Sugar, D. Toussaint), Phys. Rev. D 70 031504(R) (2004) [arXiv:hep-lat/0405022].

40. QCD Thermodynamics with Three Flavors of Improved Staggered Quarks, C. Bernard, T. Burch, C. DeTar, Steven Gottlieb, E.B. Gregory, U.M. Heller, J. Osborn, R. Sugar, D. Toussaint, Phys. Rev. D 71, 034504 (2005) [arXiv:hep-lat/0405029].

41. Light hadrons with improved staggered quarks: approaching the continuum limit, C. Aubin, C. Bernard, T. Burch, C. DeTar, Steven Gottlieb, E.B. Gregory, U.M. Heller, J. Osborn, R. Sugar, D. Toussaint, Phys. Rev. D 70, 094505 (2004) [arXiv:hep-lat/0402030].

42. Light pseudoscalar decay constants, quark masses, and low energy constants from three-flavor lattice QCD, C. Aubin, C. Bernard, C. DeTar, Steven Gottlieb, E.B. Gregory, U.M. Heller, J.E. Hetrick, J. Osborn, R. Sugar, D. Toussaint, Phys. Rev. D 70, 114501 (2004) [arXiv:hep-lat/0407028].

43. Topological susceptibility in staggered fermion chiral perturbation theory, B. Billeter, C. DeTar and J. Osborn, Phys. Rev. D 70, 077502 (2004 [arXiv:hep-lat/0406032])

44. Semileptonic decays of D mesons in three-flavor lattice QCD, The Fermilab Lattice, MILC, and HPQCD Collaborations: C. Aubin, C. Bernard, C. DeTar, M. Di Pierro, A. El-Khadra, Steven Gottlieb, E. B. Gregory, U. M. Heller, J. Hetrick, A. S. Kronfeld, P. B. Mackenzie, D. Menscher, M. Nobes, M. Okamoto, M. B. Oktay, J. Osborn, J. Simone, R. Sugar, D. Toussaint, H. D. Trottier, Phys. Rev. Lett. 94, 011601 (2005) [arXiv:hep-ph/0408306].

45. Charmed meson decay constants in three flavor lattice QCD, The Fermilab Lattice, MILC, and HPQCD Collaborations: C. Aubin, C. Bernard, C. DeTar, M. Di Pierro, E.D. Freeland, Steven Gottlieb, E.B. Gregory, U.M. Heller, J.E. Hetrick, A.X. El-Khadra, A.S. Kronfeld, L. Levkova, P.B. Mackenzie, F. Maresca, D. Menscher, M. Nobes, M. Okamoto, D. Renner, J.N. Simone, R.L. Sugar, D. Toussaint, H.D. Trottier, Phys. Rev. Lett. 95 122002 (2005) [arXiv:hep-lat/0506030].

46. Staggered lattice artifacts in 3-flavor heavy baryon chiral perturbation theory: J. Bailey and C. Bernard, Proceedings of Science (Lattice 2005) 047 (2005) [arXiv:hep-lat/0510006].
47. Staggered chiral perturbation theory for heavy-light mesons: C. Aubin and C. Bernard, Phys. Rev. D 73, 014515 (2006) [arXiv:hep-lat/0510088].
48. Staggered chiral perturbation theory and the fourth-root trick: C. Bernard, arXiv:hep-lat/0603011, to be published in Phys. Rev. D.
49. Comment on 'Flavor extrapolations and staggered fermions': C. Bernard, M. Golterman, Y. Shamir and S. Sharpe, arXiv:hep-lat/0603027.
50. Observations on staggered fermions at non-zero lattice spacing: C. Bernard, M. Golterman, and Y. Shamir, arXiv:hep-lat/0604017, submitted to Phys. Rev. D.

Publications in Conference Proceedings

52. Semileptonic Decays of Heavy Mesons with the Fat Clover Action, C. Bernard, T.A. DeGrand, C. DeTar, S. Gottlieb, U.M. Heller, J. Hetrick, C. McNeile, K. Orginos, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 83-84, 274, (2000) [arXiv:hep-lat/9909076].
53. Improved flavor symmetry in Kogut-Susskind fermion actions, K. Orginos, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 83-84, 878, (2000) [arXiv:hep-lat/9909087].
54. Heavy-light decay constants with Dynamical Gauge Configurations and Wilson or Improved Valence Quark Actions, C. Bernard, T.A. DeGrand, C. DeTar, S. Gottlieb, U.M. Heller, J. Hetrick, C. McNeile, K. Orginos, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 83-84, 289, (2000) [arXiv:hep-lat/9909121].
55. Perturbation Theory for Fat-link Fer mion Actions, C. Bernard and T. DeGrand, Nucl. Phys. B (Proc. Suppl.) 83-84, 845, (2000) [arXiv:hep-lat/9909083].
56. fB for Various Actions : Approaching the Continuum Limit with Dynamical Fermions, C. Bernard, S. Datta, C. DeTar, S. Gottlieb, U.M. Heller, J. Hetrick, C. McNeile, K. Orginos, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc.Suppl.) 94, 346, (2001) [arXiv:hep-lat/0011029].
57. Zero Temperature String Breaking with Staggered Quarks, C. Bernard, T. Burch, T. DeGrand, C. DeTar, S. Gottlieb, U.M. Heller, P. Lacock, K. Orginos, R. Sugar, and D. Toussaint), Nucl. Phys. B (Proc.Suppl.) 94, 546, (2001) [arXiv:hep-lat/0010066].
58. Quark Loop Effects with an Improved Staggered Action, C. Bernard, T. Burch, T. DeGrand, C. DeTar, S. Gottlieb, U.M. Heller, K. Orginos, R. Sugar and D. Toussaint), Nucl. Phys. B (Proc.Suppl.) 94, 227, (2001) [arXiv:hep-lat/0010065].

59. Thermodynamics with 2 + 1 and 3 Flavors of Improved Staggered Quarks, C. Bernard, T. Burch, S. Datta, T.A. DeGrand, C.E. DeTar, S Gottlieb, U.M. Heller, K. Orginos, R. Sugar and D. Toussaint, Nucl. Phys. A702, 140, (2002) [arXiv:hep-lat/0110030]. 7
60. Thermodynamics with 3 and 2+1 Flavors of Improved Staggered Quarks, C. Bernard, T. Burch, S. Datta, T.A. DeGrand, C.E. DeTar, Steven Gottlieb, U.M. Heller, K. Orginos, R. Sugar and D. Toussaint), Nucl. Phys. B (Proc. Suppl.), 106, 429, (2002) [arXiv:hep-lat/0110067].
61. Heavy-light decay constants with three dynamical flavors, C. Bernard, T. Burch, S. Datta, T. DeGrand, C. DeTar, Steven Gottlieb, Urs M. Heller, K. Orginos, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 106, 412, (2002) [arXiv:hep-lat/0110072].
62. Determining hybrid content of heavy quarkonia using lattice nonrelativistic QCD Tommy Burch, Kostas Orginos and Doug Toussaint, Nucl. Phys. B (Proc. Suppl.) 106, 382, (2002) [arXiv:hep-lat/0110001].
63. Light hadron properties with improved staggered quarks, C. Bernard, T. Burch, T. DeGrand, C. DeTar, Steven Gottlieb, E.B. Gregory, Urs M. Heller, J. Osborn, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 119, 257, (2003) [arXiv:hep-lat/0208041].
64. Topological susceptibility with the improved Asqtad action, C. Bernard, T. Burch, T. DeGrand, C. DeTar, Steven Gottlieb, E.B. Gregory, A. Hasenfratz, Urs M. Heller, J. Hetrick, J. Osborn, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 119, 991, (2003) [arXiv:hep-lat/0209050].
65. Static hybrid quarkonium potential with improved staggered quarks, C. Bernard, T. Burch, T. DeGrand, C. DeTar, Ziwen Fu, Steven Gottlieb, E.B. Gregory, Urs M. Heller, J. Hetrick, J. Osborn, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 119, 598, (2003) [arXiv:hep-lat/0209051].
66. Chiral logs with staggered fermions, C. Aubin, C. Bernard, C. DeTar, Steven Gottlieb, Urs M. Heller, K. Orginos, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 119, 233, (2003) [arXiv:hep-lat/0209066].
67. High temperature QCD with three flavors of improved staggered quarks, C. Bernard, T. Burch, C. DeTar, Steven Gottlieb, E.B. Gregory, Urs M. Heller, J. Osborn, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 119, 523, (2003) [arXiv:hep-lat/0209079].
68. Exotic hybrid mesons from improved Kogut-Susskind fermions, C. Bernard, T. Burch, C. DeTar, Steven Gottlieb, E.B. Gregory, Urs M. Heller, J. Osborn, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 119, 260, (2003) [arXiv:hep-lat/0209097].
69. Heavy–light meson decay constants with N f = 3, C. Bernard, T. Burch, S. Datta, C. DeTar, Steven Gottlieb, E.B. Gregory, U.M. Heller, R. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 119, 613, (2003) [arXiv:hep-lat/0209163].
70. Quark Loop Effects in Semileptonic Form Factors for Heavy-Light Mesons, C. Bernard, C. DeTar, Steven Gottlieb, E. Gregory, U.M. Heller, C. McNeile, J. Osborn, R. Sugar, and D. Toussaint, Nucl. Physics B (Proc. Suppl.), 129& 130, 364 (2004) [arXiv:hep-lat/0309055].

71. Pion and kaon physics with improved staggered quarks, C. Aubin, C. Bernard, C. DeTar, Steven Gottlieb, E. Gregory, U.M. Heller, J.E. Hetrick, J. Osborn, R. Sugar, and D. Toussaint, Nucl. Physics. B (Proc. Suppl.), 129& 130, 227 (2004) [arXiv:hep-lat/0309088].
72. The Phase Diagram of High Temperature QCD with Three Flavors of Improved Staggered Quarks, C. Bernard, T. Burch, C. DeTar, Steven Gottlieb, E. Gregory, U.M. Heller, J.E. Hetrick, J. Osborn, R. Sugar, and D. Toussaint, Nucl. Phys. B (Proc. Suppl.), 129& 130, 626 (2004) [arXiv:hep-lat/0309118].
73. Excited States in Staggered Meson Propagators, C. Bernard, T. Burch, C. DeTar, Steven Gottlieb, E. Gregory, U.M. Heller, J.E. Hetrick, J. Osborn, R. Sugar, and D. Toussaint, Nucl. Physics. B (Proc. Suppl.), 129& 130, 230 (2004) [arXiv:hep-lat/0309117].
74. Three Flavor QCD at High Temperatures, C. Bernard, T. Burch, C. DeTar, Steven Gottlieb, E.B. Gregory, U.M. Heller, J. Osborn, R.L. Sugar, D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 140, 538 (2005) [arXiv:hep-lat/0409097].
75. Topological susceptibility with three flavors of staggered quarks, C. Aubin, C. Bernard, Brian Billeter, C. DeTar, Steven Gottlieb, E. Gregory, U.M. Heller, J.E. Hetrick, J. Osborn (2), R.L. Sugar, D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 140, 600 (2005) [arXiv:hep-lat/0409051].
76. Results for light pseudoscalars from three-flavor simulations, C. Aubin, C. Bernard, C. DeTar, Steven Gottlieb, E.B. Gregory, Urs M. Heller, J.E. Hetrick, J. Osborn, R.L. Sugar, D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 140, 231 (2005) [arXiv:hep-lat/0409041].
77. Heavy-light decay constants using clover valence quarks and three flavors of dynamical improved staggered quarks, C. Bernard, S. Datta, C. DeTar, Steven Gottlieb, E.B. Gregory, U.M. Heller, J. Osborn, R.L. Sugar and D. Toussaint, Nucl. Phys. B (Proc. Suppl.) 140, 449 (2005) [arXiv:hep-lat/0410014].
78. Leptonic decay constants fDs and fD in three flavor lattice QCD, The Fermilab Lattice, HPQCD and MILC Collaborations: J.N. Simone, C. Aubin, C. Bernard, C. DeTar, M. di Pierro, A.X. El-Khadra, Steven Gottlieb, E.B. Gregory, U.M. Heller, J.E. Hetrick, A.S. Kronfeld, P.B. Mackenzie, D.P. Menscher, M. Nobes, M. Okamoto, M.B. Oktay, J. Osborn, R. Sugar, D. Toussaint, H.D. Trottier, Nucl. Phys. B (Proc. Suppl.) 140, 443 (2005) [arXiv:hep-lat/0410030].
79. Semileptonic D → π/K and B → π/D decays in 2+1 flavor lattice QCD, The Fermilab Lattice, HPQCD and MILC Collaborations: M. Okamoto, C. Aubin, C. Bernard, C. DeTar, M. Di Pierro, A.X. El-Khadra, Steven Gottlieb, E.B. Gregory, U.M. Heller, J. Hetrick, A.S. Kronfeld, P.B. Mackenzie, D.P. Menscher, M. Nobes, M.B. Oktay, J. Osborn, J.N. Simone, R. Sugar, D. Toussaint, H.D. Trottier Nucl. Phys. B (Proc. Suppl.) 140, 461 (2005) [arXiv:hep-lat/0409116].
80. The scaling dimension of low lying Dirac eigenmodes and of the topological charge density, C. Aubin, C. Bernard, Steven Gottlieb, E.B. Gregory, Urs M. Heller, J.E. Hetrick, J. Osborn, R. Sugar, D. Toussaint, Ph. de Forcrand, and O. Jahn, Nucl. Phys. B (Proc. Suppl.) 140, 626 (2005) [arXiv:hep-lat/0410024].

81. The Ω– and the strange quark mass, D. Toussaint and C. Davies (MILC and UKQCD Collaborations), Nucl. Phys. B (Proc. Suppl.) 140, 234 (2005) [arXiv:hep-lat/0409129].
82. The Quenched Continuum Limit, C.T.H. Davies, G.P. Lepage, F. Niedermayer and D. Toussaint (HPQCD, MILC and UKQCD Collaborations), Nucl. Phys. B (Proc. Suppl.) 140, 261 (2005) [arXiv:hep-lat/0409039].
83. Properties of light quarks from lattice QCD simulations, C. Aubin, C. Bernard, C. DeTar, Steven Gottlieb, E.B. Gregory, Urs M. Heller, J.E. Hetrick, L. Levkova, F. Maresca, J. Osborn, D. Renner, R.L. Sugar and D. Toussaint, Journal of Physics: Conference Proceedings, 16 160 (2005).
84. The Equation of State for QCD with 2+1 Flavors of Quarks, The MILC Collaboration: C. Bernard, T. Burch, C. DeTar, S. Gottlieb, U.M. Heller, J. Hetrick, L. Levkova, F. Maresca, D. Renner, R. Sugar and Doug Toussaint, Proceedings of Science (Lattice 2005) 156 (2005) [arXiv:hep-lat/0509053].
85. Update on pi and K Physics, The MILC Collaboration: C. Bernard, C. DeTar, S. Gottlieb, U.M. Heller, J. Hetrick, L. Levkova, F. Maresca, J. Osborn, D. Renner, R. Sugar and D. Toussaint, Proceedings of Science (Lattice 2005) 025 (2005) [arXiv:hep-lat/0509137].
86. Predictions from Lattice QCD, The FNAL and MILC Collaborations: Andreas S. Kronfeld, I.F. Allison, C. Aubin, C. Bernard, C.T.H. Davies, C. DeTar, M. Di Pierro, E.D. Freeland, Steven Gottlieb, A. Gray, E. Gregory, U.M. Heller, J.E. Hetrick, A.X. El-Khadra, L. Levkova, P.B. Mackenzie, F. Maresca, D. Menscher, M. Nobes, M. Okamoto, M.B. Oktay, J. Osborn, D. Renner, J.N. Simone, R. Sugar, D. Toussaint, and H.D. Trottier, Proceedings of Science (Lattice 2005) 206 (2005) [arXiv:hep-lat/0509169].
87. The locality of the fourth root of staggered fermion determinant in the interacting case, The MILC Collaboration: C. Bernard, Ph. de Forcrand, Steven Gottlieb, U.M. Heller, J.E. Hetrick, O. Jahn, L. Levkova, F. Maresca, D.B. Renner, R. Sugar, D. Toussaint, Proceedings of Science (Lattice 2005) 299 (2005) [arXiv:hep-lat/0510025].
88. More evidence of localization in the low-lying Dirac spectrum, The MILC Collaboration: C. Bernard, Ph. de Forcrand, Steven Gottlieb, U.M. Heller, J.E. Hetrick, O. Jahn, L. Levkova, F. Maresca, D.B. Renner, R. Sugar, D. Toussaint, Proceedings of Science (Lattice 2005) 299 (2005) [arXiv:hep-lat/0510025].

Project: Cosmic Microwave Background Data Analysis (CMB)

- Principal Investigator:
  - **Julian Borrill, jdborrill@lbl.gov**
- URL:
  - **http://crd.lbl.gov/~borrill/cmb/nersc/**
- DOE Office support:
  - **HEP – High Energy Physics**
- DOE program manager:
  - **Jeffrey Mandula**
- Scientific domain:
  - **Astrophysics**
- Support for the development of the codes:
  - **SciDAC: none**
  - **DOE SC program: KAA401 411210 Project Number 4192-0**
  - **other institutional funding: Brazil (INPE), Canada (NRC/CNRC), Finland (SA/AF), France (CNRS), Germany (MPI), Italy (ASI), Norway (NF), UK (PPARC)**
  - **industry: none**
  - **other agencies: NASA, NSF**

Office of Science U.S. DEPARTMENT OF ENERGY

23

Project: Cosmic Microwave Background Data Analysis

- What problem are you trying to solve?
  - **To obtain precise measurements (including statistical and systematic uncertainties) of the fundamental parameters of cosmology from the analysis of ground-, balloon- and satellite-based observations of the tiny fluctuations in the temperature and polarization of the Cosmic Microwave Background radiation.**
  - **To develop the massively parallel algorithms and their implementations, together with the infrastructure for the management of irreducible O(10 - 100) TB datasets, needed for the next generation of CMB polarization observations such as the joint ESA/NASA Planck satellite mission.**
- What is the expected impact of project success?
  - **To enable the most exact analysis of CMB polarization datasets possible given the inevitable computational constraints, in particular minimizing the uncertainties on the resulting cosmological parameters.**
  - **To provide an integrated data analysis resource to the CMB community as a whole and thereby to avoid the re-invention of the wheel by each experiment.**

Office of Science U.S. DEPARTMENT OF ENERGY

24

Project: Cosmic Microwave Background Data Analysis

- External communities & sizes that code and/or datasets support:
  - **At any time over the last 10 years we have been supporting O(100) analysts from O(10) experiments, with new teams joining the project as others are completed.**
  - **The results of these analyses support the entire world-wide theoretical cosmology and ultra high energy physics communities.**

Office of Science U.S. DEPARTMENT OF ENERGY

25

2. CMB Project Team Resources

- Team institutional affiliation(s):
  - **US: Berkeley Lab, UC Berkeley, UC Davis, UC Irvine, UC Santa Barbara, UI Urbana-Champaign, U Hawaii, Brown, CalTech, Chicago, Columbia, Harvard, Princeton, NASA JPL**
  - **Brazil: INPE Sao Jose dos Campos**
  - **Canada: U British Columbia, U Toronto**
  - **Finland: U Helsinki**
  - **France: U Paris IAP, U Paris APC, U Paris CdF, U Paris LAL, CEA Saclay**
  - **Germany: MPI Garching**
  - **Italy: U Roma La Sapienza, U Roma Tor Vergata, SISSA Trieste, U Milano INFN, U Milano IASF-CNR**
  - **Norway: U Oslo**
  - **UK: U Cambridge, U Cardiff, Imperial College London, U Oxford, U Sussex**
  - **In all, about 40 institutions**
- To what extent are the code team members affiliated with the computer center institution?
  - **The project PI works closely with the NERSC Center, although nobody from NERSC is on their team.**

Office of Science U.S. DEPARTMENT OF ENERGY

26

2. CMB Project Team Resources

- Team composition and experience:
  - **domain scientists: 80**
  - **computational scientists: 5**
  - **computer scientists: 0 team members, 10+ consultants**
  - **computational mathematicians: 0 team members, 5 consultants**
  - **database managers: 0 team members, 1 consultant**
  - **programmers: 10**
  - **program development and maintenance: 10**
  - **users of the team codes: 200+**
- Team composition by educational level:
  - **senior faculty: 2**
  - **national laboratory scientists: 5**
  - **industrial scientists: 0**
  - **younger faculty: 5**
  - **Ph.D: 70**
  - **MS: 0**
  - **BS: 1**
  - **post-docs: 10**
  - **graduate students: 0 (currently)**
  - **undergraduate students: 0 (currently)**

Office of Science U.S. DEPARTMENT OF ENERGY

27

2. CMB Project Team Resources

- Team resources utilization:
  - **code and algorithm development: 30%**
  - **code maintenance: 10%**
  - **problem setup: 10%**
  - **production runs: 10%**
  - **data management: 10%**
  - **inter- & intra-systems management: 10%**
  - **results analysis: 5%**
  - **publications: 5%**
  - **grant management: 5%**
  - **project management: 5%**

Office of Science U.S. DEPARTMENT OF ENERGY

28

3. CMB Project Codes

- Code classes:
  - **Dataset simulation: LevelS**
  - **Data abstraction: M3**
  - **Noise Estimation: MADnes, MADping**
  - **GLS map-making: MADmap, Mapcumba, ROMA**
  - **Destriping map-making: POLAR, Springtide**
  - **ML power spectrum estimation: MADspec**
  - **MC power spectrum estimation: MASTER, FASTER, POLspice**
  - **Gibbs sampling methods: MAGIC, Commander**
- Problem Type:
  - **Data simulation & data analysis**
- Types of algorithms and computational mathematics:
  - **FFTs, SHTs (Spherical Harmonic Transforms), Monte Carlo Methods, Iterative (PCG) Solvers, Dense Linear Algebra**

3. CMB Project Codes

- What platforms does your code routinely run on ?
  - **At NERSC AY 2006 (Dec 2005 through July 9, 2006):**
    - **65.9% on the IBM Power3, Seaborg**
    - **26.1% on the IBM Power5, Bassi**
    - **7.9% on the Opteron linux cluster, Jacquard**
    - **Some use of the SGI visualization server, DaVinci**
  - **Some code development & small runs at:**
    - **NASA Ames (Project Columbia - recently abandoned as unusable due to slow read I/O rate)**
    - **NASA JPL Clusters**
    - **NCSA**
    - **CITA, Toronto**
    - **CSC, Helsinki**
    - **CEA, Saclay & Planck HFI-DPC, Paris**
    - **MPI, Garching**
    - **CINECA, Bologna & Planck LFI-DPC, Trieste**
    - **BSC, Barcelona (Mare Nostrum - early benchmarking phase)**
    - **COSMOS, Cambridge**

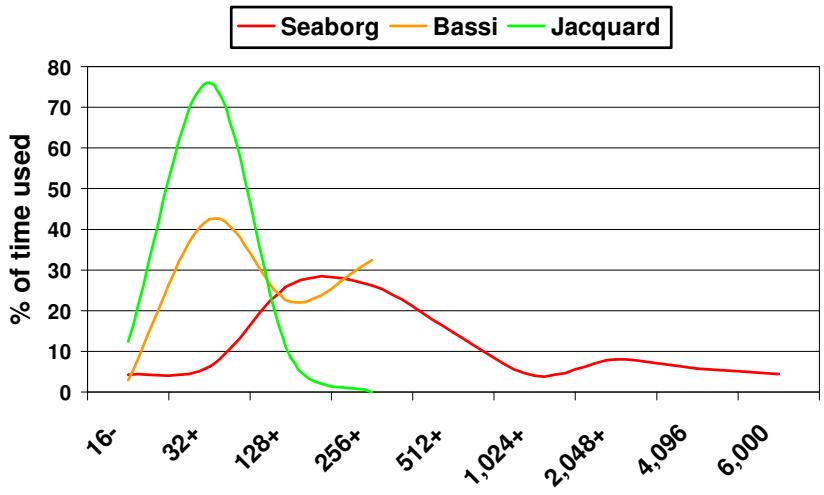
3. CMB Project Codes

- Code size (single lines of code, function points, etc.):
  - **MADCAP code suite (including M3): 50,000 lines**
  - **Other codes: 100,000 lines**
  - **Code ages and yearly growth: 1 - 10 years old, continual development (major revisions may reduce length).**
- Computer languages employed:
  - **C, C++, Fortran77, Fortran90**
  - **Structure of the codes: all main code**
- What libraries are used? And what fraction of the codes does it represent?
  - **FFTW: 10%**
  - **ccSHT: 5%**
  - **LAPACK: 5%**
  - **ScaLAPACK: 5%**
  - **CFITSIO: 1%**
  - **libXML: 1%**
  - **HEALPix: 1%**
  - **CMBfast/CAMB: 1%**

3. CMB Project Codes

- Code Mix:
  - To what extent does your team develop and use your own codes ?
    - **The great majority of our code is self-developed**
  - Codes developed by others in the DOE and general scientific community ?
    - **Some application-specific libraries (HEALPix, CMBfast/CAMB)**
  - Commercial application codes provided by the center ?
    - **Some general libraries provided by center (FFTW, ccSHT, LAPACK/ScaLAPACK)**
    - **Some general libraries self-installed (CFITSIO, libXML)**

2006 CMB Project Scalability



Office of Science U.S. DEPARTMENT OF ENERGY

33

3. CMB Project Codes – current scalability data (Dec 1 2005 – early July 2006)

- Seaborg: 6080 processor Power3
  - 4.4% of the time on 6,000 processors
  - 5.8% of the time on 4,096 processors
  - 7.5% of the time on 2,048 processors
  - 3.9 % of the time on 1,024-1,584 processors
  - 18.4% of the time on 384-976 processors
  - 22.6% of the time on 256 processors
  - 25.8% of the time on 128 processors
  - 10.5% of the time on 1 - 112 processors
- Bassi: 888 processor Power5
  - 2.2% of the time on 384 processors
  - 30.2% of the time on 256 processors
  - 22.5% of the time on 120-248 processors
  - 20.4% of the time on 64-96 processors
  - 24.7% of the time on 8-32 processors
- Jacquard: 712 processor Opteron
  - 0.1% of the time on 256 processors
  - 8.4% of the time on 128 processors
  - 10.2% of the time on 48-64 processors
  - 65.7% of the time on 32 processors
  - 14.2% of the time on 16-30 processors
  - 1.4% of the time on 2-12 processors

Office of Science U.S. DEPARTMENT OF ENERGY

34

3. CMB Project Codes

- Projected or maximum scalability & how is measured ?
  - **All codes projected to scale to any concurrency consistent with:**
    - **Minimum memory requirement for a particular code & datasets**
    - **Efficiency degradation (particularly poor I/O scaling) for very large concurrencies**
  - **Codes have successfully run at up to 6000-way concurrency.**
- Is the code massively parallel ?
  - **Analysis codes are massively parallel.**
  - **Simulation code is serial, but large enough datasets can be split into independent pieces and run with embarrassing parallelism.**
- Parallelization model:
  - **MPI**
- Does your team use domain decomposition ?
  - **No**
- What is the "efficiency" of the code and how is it measured:
  - **Depending on the code, 3 - 80 % of theoretical peak performance, measured by external/Center (IPM) and internal/code-specific profiling.**

Office of Science U.S. DEPARTMENT OF ENERGY

3. CMB Project Codes

- What are the major bottlenecks for scaling your code?
  - **I/O performance for very large concurrencies (although Seaborg does remarkably well in this regard).**
  - **Cache misses for necessarily linear & log-linear algorithms.**
- What memory/processor ratio do your project require ?
  - **1GB/CPU is a minimum, and systems with 4GB/CPU have proven very useful; however since we are constrained by the need to deliver tens of TB data from disk overall system balance is much more important than any single feature though.**
- What is the split between interactive and batch use ? Why this split ? Is interactive use more productive ?
  - **Total interactive hours Dec 2005 – June 2006: 15,223**
  - **Total hours used: 839,941**
  - **Percent interactive use: 1.8%**
  - **Limits on interactive job sizes (rightly) preclude its use for production computing which accounts for most of our usage.**
- What is the split between code development on the computer center computers and on computers at other institutions?
  - **75/25, largely by locality (i.e., most development done in home nation)**

Office of Science U.S. DEPARTMENT OF ENERGY

4. CMB Project resources input from the centers

- Plan with benchmarks & milestones:
  - **As the Planck satellite launch approaches, our required NERSC resources (cycles & storage) will increase significantly. Fromthe most recent NERSC Greenbook:**
    - **O(10–100) exaflops of total processing capacity,**
    - **O(100) TB of archival file storage for primary data and derived data products.**
    - **O(10) TB of scratch file storage at any one time to support a particular analysis,**
    - **O(1–10) GB of local tmp file storage on each processor or node to stage intermediate data products and enable out-of-core computations**
    - **Scalable, massively parallel I/O supporting the simultaneous transfer of very large volumes of data across the entire processor set being used; currently much of the Planck-scale CMB data analysis is I/O bound.**
    - **An inter-processor communication system supporting the fast global reductions of gigabytes of distributed data.**

Office of Science U.S. DEPARTMENT OF ENERGY

37

4. CMB Project resources input from the centers

- Steady state use of resources on a production basis per month (desired):
  - **Number of processors: 32 - 256 uniformly distributed; occasional 1024+**
  - **Processor time: 100,000+ SP POWER3 hours**
  - **Disk: 20 TB**
  - **Tertiary amount and rate of change: 50 TB + 50 TB/yr**
- Annual use of resources (actual):
  - **Processor time (IBM SP POWER3-equivalent hours):**
    - **2002: 508K**
    - **2003: 809K**
    - **2004: 1,071K (14 months)**
    - **2005: 971K**
    - **2006: 840K (7 months)**
  - **Disk: 5 TB (current)**
  - **Tertiary storage rate of change: 10 TB (current)**

Office of Science U.S. DEPARTMENT OF ENERGY

38

4. CMB Project resources input from the centers

- Software provided by center:
  - **F90, C, C++, MPI, ESSL, LAPACK, cfitsio, FFTW, ccSHT, ScaLAPACK**
- Consulting provided by center:
  - **Phone & email support for system, library, compiler & filesystem issues.**
- Direct project support from center acting as a team member:
  - **Evolving, with possible Planck buy-in to key resources (e.g. NGF).**
- What is the size of their jobs in terms of:
  - **memory: 1 GB - 1 TB**
  - **concurrency (processors): described on previous slide**
  - **disk: up to 2 TB**
  - **tertiary store: 10 TB**
- What is the wall-clock time for typical runs?
  - **Avg for 16-112 CPU jobs: 0h27m (2h55m for jobs > 35m)**
  - **Avg for 128-240 CPU jobs: 1h9m (8h15)**
  - **Avg for 256-496 CPU jobs: 0h40m (2h22)**
  - **Avg for 512-1,008 CPU jobs: 1h51m (3h52)**
  - **Avg for 1,024-2,032 CPU jobs: 0h29m (1h18)**
  - **Avg for 2,048+ CPU jobs: 2h25m (3h41)**

Office of Science U.S. DEPARTMENT OF ENERGY

39

5. CMB Software Engineering, Development, Verification and Validation Processes

- Software development tools used:
  - **parallel development:**
  - **debuggers: pdbx, totalview**
  - **visualization: idl**
  - **production management and steering:**
- Software engineering practices - please list the specific tools or processes used for:
  - **configuration management: CVS, autoconf**
  - **quality control: cross-code comparison & re-analysis of standard datasets**
  - **bug reporting and tracking: individual email (mostly single author codes)**
  - **code reviews, project planning & project scheduling and tracking: These are very experiment/team specific. In the case of the Planck team (by far the largest) these include a number of weekly telecons and annual face-to-face meetings.**

Office of Science U.S. DEPARTMENT OF ENERGY

40

5. CMB Software Engineering, Development, Verification and Validation Processes

- What is your verification strategy?
  - **Cross-code comparison & re-analysis of standard datasets**
- What use do you make of regression tests?
  - **None**
- What is your validation strategy?
  - **Cross-code comparison & re-analysis of standard datasets**
- What experimental facilities do you use for validation?
  - **None - as a data analysis project we use simulated data with know inputs.**
- Does your project have adequate resources for validation?
  - **We could always use more resources - cycles, storage & people.**

Office of Science U.S. DEPARTMENT OF ENERGY

41

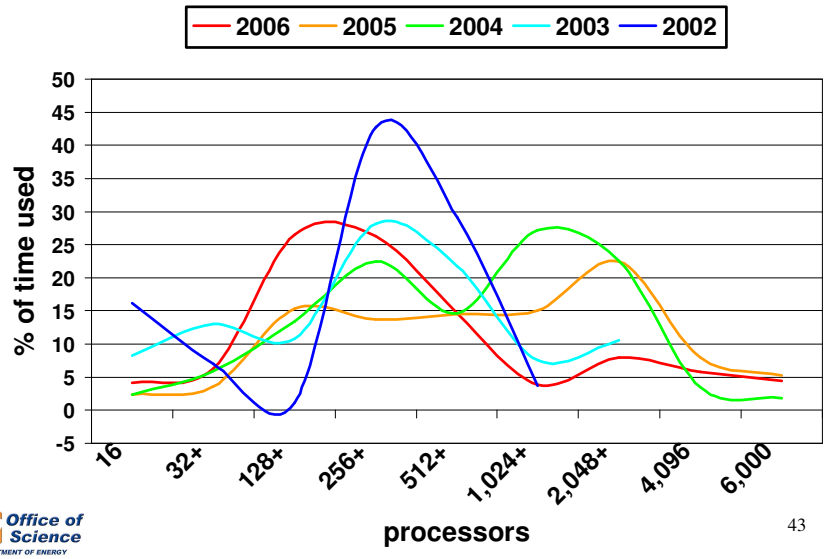
5a. CMB Project code productivity & scalability

- Measures of experiment productivity and performance including scalability of runs**:**
  - **The fundamental measure of productivity is the successful analysis of a dataset.**
  - **Scaling is driven by the size of the datasets being analyzed as the algorithms used for the analysis are determined by constraints on the available computational resources.**
- Scaling limits
  - **IO scaling, algorithm scaling**
- Projected scalability:
  - **No immediate changes from current scalability (see next slides)**
  - **This year we have been focusing a lot of energy on scaling back the concurrency requirements of the codes to make them fit on the Planck Data Processing center clusters, which are in the 128 CPU range.**

Office of Science U.S. DEPARTMENT OF ENERGY

42

CMB Project Scalability - History



Office of Science U.S. DEPARTMENT OF ENERGY

43

5a. CMB Project code scalability (history)

	16 CPUs	32+ CPUs	128+ CPUs	256+ CPUs	512+ CPUs	1,024+ CPUs	2,048+ CPUs	4,096 CPUs	6,000 CPUs
AY 2006	4.2%	6.3%	26.5%	26.3%	14.7%	3.9%	7.9%	5.8%	4.4%
AY 2005	2.4%	3.6%	15.3%	13.7%	14.5%	15.1%	22.4%	7.9%	5.2%
AY 2004	2.4%	5.9%	13.4%	22.4%	14.6%	27.2%	22.6%	3.4%	1.8%
AY 2003	8.3%	13.0%	10.8%	28.2%	21.7%	7.6%	10.5%	-	-
AY 2002	16.2%	7.0%	1.2%	42.8%	29.1%	3.8%	-	-	-

Office of Science U.S. DEPARTMENT OF ENERGY

44

6. CMB Scientific | Engineering Output

- Scientific accomplishments 2000 to present:
  - **Supported the analysis of around 20 past, present and future CMB experiments, highlights including:**
    - **The first detailed measurements of the CMB temperature anisotropy power spectrum (BOOMERanG & Maxima) demonstrating the flatness of the Universe.**
    - **The most detailed measurements of the CMB temperature anisotropy on the very smallest scales (ACBAR), showing Silk damping and a possible Sunyaev-Zeldovich excess. These results were also a crucial small-scale complement to WMAP.**
    - **The first analysis of a simulated Planck dataset, demonstrating (a) its computational tractability on the largest massively parallel systems, and (b) that Planck's 1/f noise will not significantly impact its large-scale polarization anisotropy measurements.**
    - **Re-analysis of the WMAP dataset and correction of its results.**
- The effect on the Office of Science programs:
  - **CMB observations are one of the cornerstones for developing our understanding of the cosmos, including the nature of dark energy. In addition, since the early Universe is the ultimate particle accelerator, they provide a unique window onto the ultra-high energy physics needed to move beyond the current Standard Model.**

Office of Science U.S. DEPARTMENT OF ENERGY

6. CMB Scientific | Engineering Output

ACBAR Publications:

"High Resolution Observations of the CMB Power Spectrum with ACBAR", Kuo et al, astro-ph/0212289

" Estimates of Cosmological Parameters Using the CMB Angular Power Spectrum of ACBAR", Goldstein et al, astro-ph/0212517

BEAST Publications:

"A Map of the Cosmic Microwave Background from the BEAST Experiment", Meinhold et al, astro-ph/0302034

"The CMB Power Spectrum from the Background Emission Anisotropy Scanning Telescope (BEAST) Experiment", O'Dwyer et al, astro-ph/0312610

BOOMERanG Publications:

"Measurement of a Peak in the Cosmic Microwave Background Power Spectrum from the Test Flight of Boomerang", Mauskopf et al, astro-ph/9911444

"A Measurement of Omega From the Boomerang 1997 Test Flight", Melchiorri et al, astro-ph/9911445

"A flat universe from high-resolution maps of the cosmic microwave background radiation", de Bernardis et al, astro-ph/0004404

"First Estimations of Cosmological Parameters From BOOMERANG", Lange et al, astro-ph/0005004

"A measurement by BOOMERANG of multiple peaks in the angular power spectrum of the cosmic microwave background", Netterfield et al, astro-ph/0104460

"Multiple Peaks in the Angular Power Spectrum of the Cosmic Microwave Background: Significance and Consequences for Cosmology", de Bernardis et. al, astro-ph/0105296

"Improved Measurement of the Angular Power Spectrum of Temperature Anisotropy in the CMB from Two New Analyses of BOOMERANG Observations", Ruhl et al, astro-ph/0212229

Office of Science U.S. DEPARTMENT OF ENERGY

6. CMB Scientific | Engineering Output

BOOMERanG Publications (cont):

"Instrument, method, brightness and polarization maps from the 2003 flight of Boomerang", Masi et al, astro-ph/0507509

"A measurement of the angular power spectrum of the CMB temperature anisotropy from the 2003 flight of Boomerang", Jones et al, astro-ph/0507494

"A measurement of the polarization-temperature angular cross power spectrum of the cosmic microwave background from the 2003 flight of Boomerang", Piacentini et al, astro-ph/0507507

"A measurement of the CMB <EE> Spectrum from the 2003 flight of Boomerang", Montroy et al, astro-ph/0507514

"Cosmological Parameters from the 2003 flight of Boomerang", MacTavish et al, astro-ph/0507503

MAXIMA Publications:

"MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on angular scales of 10 arcminutes to 5 degrees", Hanany et al, astro-ph/0005123

"Constraints on Cosmological Parameters from MAXIMA-1", Balbi et al, astro-ph/0005124

"A High Resolution Analysis of the MAXIMA-1 Cosmic Microwave Background Anisotropy Data", Lee et al, astro-ph/0104459

"Cosmological Implications of the MAXIMA-1 High Resolution Cosmic Microwave Background Anisotropy Measurement", Stompor et al, astro-ph/0105062

Planck Publications:

"Comparison of map-making algorithms for CMB experiments", Poutanen et al, astro-ph/0501504

"Making sky maps from Planck data", Ashdown et al, astro-ph/0606348

TOPHAT Publications:

"The Spectrum of Integrated Millimeter Flux of the Magellanic Clouds and 30-Doradus from TopHat and DIRBE Data", Aguirre et al, astro-ph/0306425

6. CMB Scientific | Engineering Output

WMAP Publications:

"Testing for Non-Gaussianity in the Wilkinson Microwave Anisotropy Probe Data: Minkowski Functionals and the Length of the Skeleton", Eriksen et al, astro-ph/0401276

"On Foreground Removal from the Wilkinson Microwave Anisotropy Probe Data by an Internal Linear Combination Method: Limitations and Implications", Eriksen et al, astro-ph/0403098

"Bayesian Power Spectrum Analysis of the First-Year WMAP data", O'Dwyer et al, astro-ph/0407027

"The N-point correlation functions of the first-year Wilkinson Microwave Anisotropy Probe sky maps", Eriksen et al, astro-ph/0407271

Multi-Experiment Publications:

"Cosmology from Maxima-1, Boomerang and COBE/DMR CMB Observations", Jaffe et al, astro-ph/0007333

"Correlations Between the WMAP and MAXIMA Cosmic Microwave Background Anisotropy Maps", Abroe et al, astro-ph/0308355

Methodological and other Publications:

"MADCAP - The Microwave Anisotropy Dataset Computational Analysis Package", Borrill, astro-ph/9911389

"Making Maps Of The Cosmic Microwave Background: The MAXIMA Example", Stompor et al, astro-ph/0106451

"Asymmetric Beams in Cosmic Microwave Background Anisotropy Experiments", Wu et al, astro-ph/0007212

"Power spectrum estimation from high-resolution maps by Gibbs sampling", Eriksen et al, astro-ph/0407028

"Separating cosmological B modes from foregrounds in cosmic microwave background polarization observations", Stivoli et al, astro-ph/0505381

"The angular power spectrum of NVSS radio galaxies", Blake et al, astro-ph/0404085

6. CMB Scientific | Engineering Output

- Citations (last 5 years):
  - O(5000) : ACBAR (300) + BOOMERanG (2400) + MAXIMA (1400)
- Dissertations:
  - **O(20) PhD theses**
- Residual and supported, living datasets and/or databases that are accessed by a community ? Size of the community ?
  - **NASA's Legacy Archive for Microwave Background Data Analysis (LAMBDA) provides data and data products from a large number of CMB missions, including those analyzed at NERSC. LAMBDA supports a community of 500+ experimental and theoretical cosmologists and theoretical physicists.**
- Change in code capabilities and quality:
  - **Codes evolving to handle datasets whose size is growing much faster than the computational resources, and which now include CMB polarization as well as temperature information.**

6. CMB Scientific | Engineering Output

- Code and/or data contributed to the centers:
  - **ccSHT parallel spherical harmonic transform code**
  - **MADbench scientific application benchmark code**
- Code and/or data, results, contributed to the scientific and engineering community at large:
  - **See above.**
- Company spin-offs based on code or trained people and/or CRADAs**:**
  - **None**
- Corporation, extra-agency, etc. use:
  - **None**
- Production of scientists & computational scientists during 2001-2005:
  - **O(10) PhDs completed**
- Production of trained software engineers during 2001-2005**:**
  - **none**

Project: First-Principles Catalyst Design for Environmentally Benign Energy Production

- Principal Investigator:
  - **Manos Mavrikakis, manos@engr.wisc.edu**
  - **Team member Lars Grabow answered this survey**
- URL:
  - **http://www.engr.wisc.edu/che/faculty/mavrikakis_manos.html**
- DOE Office support:
  - **BES - Chemical Sciences**
- DOE program manager:
  - **Raul Miranda**
- Scientific domain:
  - **Chemistry**
- Support for the development of the code:
  - **SciDAC: none**
  - **DOE SC program: DE-FG02-03ER15469, DE-FG02-05ER15731**
  - **other institutional funding: University of Wisconsin-Madison**
  - **industry: S.C. Johnson**
  - **other agencies: NSF-CAREER Award(CTS-0134561), DOE-NETL(DE-FC26-03NT41966), NSF-EPA(CTS-0327959)**

Office of Science U.S. DEPARTMENT OF ENERGY

51

Project: First-Principles Catalyst Design for Environmentally Benign Energy Production

- What problem are you trying to solve?
  - **Design improved catalysts for low temperature fuel cells: anode catalysts with increased CO tolerance, and more efficient cathode catalysts for oxygen reduction**
  - **Investigate detailed reaction mechanism for CO_2 hydrogenation in order to design catalysts for CO_2 fixation (use CO_2 to produce useful chemicals, such as methanol)**
  - **Fundamental studies of Fischer-Tropsch catalysis ($CO+H_2$→ alkanes) for synthesis of liquid fuels from synthesis gas**
- What is the expected impact of project success?
  - **Develop new environmentally benign technologies for energy production**
  - **Train PhD students in the field of ab-initio design of new materials with tailored properties, as needed by several sections of the chemical industry**
- External communities & sizes that code and/or datasets support:
  - ***N/A***

Office of Science U.S. DEPARTMENT OF ENERGY

52

2. Catalyst Design Project Team Resources

- Team institutional affiliation:
  - **University of Wisconsin-Madison**
- To what extent are the code team members affiliated with the computer center institution? (e.g. are the team members also members of the computer center institution?)
  - ***Does not really apply. Code is developed at CAMP, DTU, Denmark. http://www.camp.dtu.dk/English/Software.aspx. Nobody from Madison or Denmark is affiliated with the computer center to my knowledge.***
- Team composition and experience (11 team members):
  - **domain scientists: 11**
  - **computational scientists: 1**
  - **computer scientists: n/a**
  - **computational mathematicians: n/a**
  - **database managers: n/a**
  - **programmers: n/a**
  - **program development and maintenance: 1 student with 5yrs experience**
  - **users of the team codes? ~ 10 users, 0-5 yrs of experience**

Office of Science U.S. DEPARTMENT OF ENERGY

2. Catalyst Design Project Team Resources

- Team composition by educational level (11-12 team members):
  - **senior faculty: 1**
  - **national laboratory scientists: 0**
  - **industrial scientists: 0**
  - **younger faculty: 0**
  - **Ph.D: 1**
  - **MS: 1**
  - **BS: 7**
  - **post-docs: 1**
  - **graduate students: 7**
  - **undergraduate students: 2-3**
- Team resources utilization:
  - **time spent on code and algorithm development: 0%**
  - **code maintenance: 2%**
  - **problem setup: 30%**
  - **production runs: 20%**
  - **results analysis: 20%**
  - **publications: 15%**
  - **grant management: 10%**
  - **other (describe): 3% (administration of local computing resources)**

Office of Science U.S. DEPARTMENT OF ENERGY

3. Catalyst Design Project Code: DACAPO

- Problem Type: **simulation**
- Types of algorithms and computational mathematics:
  - **Iterative Solver, sparse linear algebra, FFTs, Energy Minimization**
- What platforms does your code routinely run on?
  - **At NERSC: almost 100% of the time is on the IBM Power5, Bassi**
- Code size (single lines of code, function points, etc.);
  - **~ 51,000 single lines of code**
  - **Code age: 15 yrs; yearly growth: variable**
- Computer languages employed:
  - **Fortran90 & MPI, Python**
  - **50,000 SLOC Fortran - main code; 80,000 SLOC Python - steering**
- What libraries are used? **And what fraction of the codes does it represent?**
  - **ESSL: 30-40% (estimated)**
  - **MASS: 20-50%**
  - **NetCDF: 2-5%**
- Code Mix:
  - **To what extent does your team develop and use your own codes? N/A**
  - **Codes developed by others in the DOE and general scientific community? N/A**
  - **Commercial application codes provided by the center? N/A**

Office of Science U.S. DEPARTMENT OF ENERGY

3. Catalyst Design Project Code: DACAPO

- What is the present parallel scalability on each of the computers the code operates on
  - **Degree of efficient parallelization depends on number of k-points in the system (function of unit cell size, crystal structure of catalyst, etc.)**
  - **NEB (Nudged Elastic Band algorithm) calculations offer an extra degree of parallelization depending on the number of intermediate images in the path (requires several independent total energy calculations) = higher number of CPUs can be used**
  - **Current scaling (on 888 processor Power5 system):**
    - **2.3% of the time on 224 processors (activation barrier, NEB)**
    - **21.4% of the time on 112 processors (activation barrier, NEB)**
    - **7.4% of the time on 64 processors (activation barrier, NEB)**
    - **2.3 % of the time on 56 processors (activation barrier, NEB)**
    - **3.8% of the time on 40 processors (activation barrier, NEB)**
    - **0.6% of the time on 32 processors (total energy calculations)**
    - **24.0% of the time on 24 processors (total energy calculations)**
    - **31.7% of the time on 16 processors (total energy calculations)**
    - **6.6% of the time on 8 processors (total energy calculations)**
  - **Projected or maximum scalability: For current systems, max. 560 CPUs**
  - **How is measured? Parallelization over k-points is assumed to be ideal. Plane-wave parallelization is acceptable up to ~ 66% efficiency.**
  - **Is the code massively parallel? - No**

Office of Science U.S. DEPARTMENT OF ENERGY

3. Catalyst Design Project Code: DACAPO

- Parallelization model**: MPI**
- Does your team use domain decomposition**? No**
- What is the "efficiency" of the code and how is it measured:
  - **Efficiency depends on degree of parallelization. We usually run with 70-80% efficiency.**
  - **Efficiency is measured by comparison of plane-wave parallelized to k-point parallelized runs.**
- What are the major bottlenecks for scaling your code? (From 2003 Scaling Report)
  - **Scalability to higher number of CPUs per job is limited by the physics of the problems typically encountered. For 20-30 atoms, for example, DACAPO becomes communication bound for more than 64 tasks.**
  - **For this research scaling to 1,024 tasks only achieves 30% efficiency. The same research can be better accomplished in the range of 100 tasks where near linear speedup is possible.**
  - **The code incorporates 2 dimensions of parallelism: k-points and planewaves. k-point parallelization yields better performances than planewave parallelization. It is natural that the efficiency decreases with planewave parallelism, for two reasons: relatively small matrices means that communications time will become important; and the subspace eigenvalue problem is an unavoidable algorithmic bottleneck.**

3. Catalyst Design Project Code: DACAPO

- What memory/processor ratio do your project require? **(e.g. Gbytes/processor)**
  - ***In most cases 1GB/CPU is sufficient. For 5-10% of the jobs 2-8GB/CPU are necessary.***
- What is the split between interactive and batch use? **Why this split? Is interactive use more productive?**
  - **Total interactive hours Dec 2005 – June 2006: 3.6**
  - **Total hours used: 819,143**
  - **Ratio interactive/batch use: insignificant**
- What is the split between code development on the computer center computers and on computers at other institutions?
  - ***N/A***

4. Catalyst Design Project resources input from the centers

- Plan with benchmarks & milestones:
  - **N/A**
- Steady state use of resources on a production basis per month (desired):
  - **Number of processors: > 32**
  - **Processor time: 120K (IBM SP POWER3-equivalent hours)**
  - **Disk: 500 GB**
  - **Tertiary amount and rate of change: insignificant**
- Annual use of resources (actual):
  - **Processor time (IBM SP POWER3-equivalent hours):**
    - **2002: 278K allocated; 382K used**
    - **2003: 181K allocated; 240K used**
    - **2004: 351K allocated; 529K used (14 months)**
    - **2005: 370K allocated; 360K used**
    - **2006: 485K allocated; 819K used (7 months)**
  - **Disk: 500 GB**
  - **Tertiary storage: insignificant**

Office of Science U.S. DEPARTMENT OF ENERGY

59

4. Catalyst Design Project resources input from the centers

- Software provided by center:
  - **Fortran 90, MPI, ESSL, MASS, NetCDF, BLAS, python, VTK**
- Consulting provided by center:
  - **User support via email mainly for compilation issues.**
- Direct project support from center acting as a team member:
  - **None**
- What is the size of their jobs in terms of:
  - **memory: typically 0.5-1 GB / CPU, sometimes up to 6GB / CPU**
  - **concurrency (processors): described on previous slide**
  - **disk: total energy calculations: ~120MB, NEBs: 1.5 – 2.5 GB**
  - **tertiary store: insignificant**
- What is the wall-clock time for typical runs?
  - **Average wall-clock for 1-56 CPU jobs: 4h36m**
  - **Average wall-clock for 64-120 CPU jobs: 3h54m**
  - **Average wall-clock for 128-248 CPU jobs: 4h40m**

Office of Science U.S. DEPARTMENT OF ENERGY

60

5. Catalyst Design Software Engineering, Development, Verification and Validation Processes

- Software development tools used:
  - **parallel development: N/A**
  - **debuggers: N/A**
  - **visualization: N/A**
  - **production management and steering: N/A**
- Software engineering practices. Please list the specific tools or processes used for:
  - **configuration management: N/A**
  - **quality control: N/A**
  - **bug reporting and tracking: mailing list** **https://listserv.fysik.dtu.dk/mailman/listinfo/campos**
  - **code reviews: N/A**
  - **project planning: N/A**
  - **project scheduling and tracking: N/A**

Office of Science U.S. DEPARTMENT OF ENERGY

61

5. Catalyst Design Software Engineering, Development, Verification and Validation Processes

- What is your verification strategy?
  - **N/A**
- What use do you make of regression tests?
  - **N/A**
- What is your validation strategy?
  - **N/A**
- What experimental facilities do you use for validation?
  - **Theoretical results are validated in collaborations with experimental groups at UW and other places (e.g.: BNL, U of Aarhus in Denmark, LBNL)**
- Does your project have adequate resources for validation?
  - **Could definitely use more CPU/year, if available.**

Office of Science U.S. DEPARTMENT OF ENERGY

62

5a. Catalyst Design Project code productivity & scalability

- Measures of experiment productivity and performance including scalability of runs:
  - **Don't know**
- Scaling limits including i/o, node memory size, interconnect b/w or latency, algorithm:
  - **Scalability is mostly limited by physical nature of research. Do not know the scaling limits of the algorithms used in Dacapo, but for high levels of plane-wave parallelization the code becomes communication limited.**
- Projected scalability:
  - **Scalability is mostly limited by physical nature of research. Scalability may not increase significantly.**

Office of Science U.S. DEPARTMENT OF ENERGY

63

5a. Catalyst Design Code Scalability (history)

	8-16 CPUs (1 node)	24 CPUs	32 CPUs	48 CPUs	64 CPUs	80-96 CPUs	112 - 128 CPUs	224 - 240 CPUs	1,024 CPUs
AY2006 (Bassi)	38.0%	24.5%	0.6%		7.3%		21.4%	2.3% (max)	-
AY2005	27.8%	-	35.3%	18.3%	6.6%	9.3%	2.4%	0.2% (max)	-
AY2004	41.1%	-	13.3%	34.3%	10.7%	0.7% (max)	-	-	-
AY2003	29.1%	-	8.6%	33.6%	4.5%	16.9%	-	-	5.4% (max)
AY2002	72.9%	-	12.5%	14.6% (max)	-	-	-	-	-

Office of Science U.S. DEPARTMENT OF ENERGY

64

6. Catalyst Design Scientific | Engineering Output

- The scientific accomplishments 2001 to present:
  - ***30 scientific papers produced in high impact journals, including: Nature Materials, JACS, Angewandte Chemie, Journal of Catalysis, JPC-B, PCCP, etc.***
- The effect on the Office of Science programs:
  - ***Among recent results: (1) Alloy catalysts designed from first-principles: selected as one of the DOE-BES milestones for 2005, (2) Increased cathode catalyst performance for Fuel Cells by a factor of 4.***

Office of Science U.S. DEPARTMENT OF ENERGY

6. Catalyst Design Scientific | Engineering Output

- Publications (2004-2006):

1. "Lattice strain effects on the CO oxidation on Pt(111)", L.C. Grabow, Y. Xu and M. Mavrikakis, **Phys. Chem. Chem. Phys.**, **DOI**: 10.1039/b606131a (2006) – including cover page image

2. "Prediction of Experimental Methanol Decomposition Rates on Platinum from First-Principles", S. Kandoi, J. Greeley, M. Sanchez-Castillo, St. T. Evans, A. A. Gokhale, J. A. Dumesic, M. Mavrikakis, Topics in Catalysis, 37(1), 17-28 (2006).

3. "Near Surface Alloys for Hydrogen Fuel Cell Applications", J. Greeley, M. Mavrikakis, Catalysis Today, 111, 52-58 (2006).

4. "Effect of Subsurface Oxygen on the Reactivity of the Ag(111) Surface", Y. Xu, J. Greeley, M. Mavrikakis, Journal of the American Chemical Society, 127, 12823 (2005).

5. "Mixed-Metal Pt Monolayer Electrocatalysts for Enhanced Oxygen Reduction Kinetics", J. Zhang, M.B. Vukmirovic, K. Sasaki. A.U. Nilekar, M. Mavrikakis, R.R. Adzic, **Journal of the American Chemical Society (Communication), 127, 12480 (2005).**

6. "Controlling the Catalytic Activity of Platinum Monolayer Electrocatalysts for Oxygen Reduction with Different Substrates", J. Zhang, M.B. Vukmirovic, Y. Xu, M. Mavrikakis, R. R. Adzic, Angewandte Chemie International Edition 44, 2132 (2005).

7. "Surface and Subsurface Hydrogen: Adsorption Properties on Transition Metals and Near-Surface Alloys", J. Greeley, M. Mavrikakis, Journal of Physical Chemistry B 109, 3460 (2005).

8. "Trends of Low Temperature Water Gas Shift Reactivity on Transition Metals", N. Schumacher, A. Boisen, S. Dahl, A. A. Gokhale, S. Kandoi, L. C. Grabow, J. A. Dumesic, M. Mavrikakis, I. Chorkendorff, Journal of Catalysis 229, 265 (2005).

9. "A New Class of Alloy Catalysts Designed from First-Principles", J. Greeley, M. Mavrikakis, Nature Materials 3, 810 (2004).

10. "Molecular-level Descriptions of Surface Chemistry in Kinetic Models using Density Functional Theory", with A. Gokhale, S. Kandoi, J. Greeley, M. Mavrikakis, J. A. Dumesic, Chemical Engineering Science 59, 4679 (2004).

11. "Effect of Sn on the reactivity of Cu surfaces", A. A. Gokhale, G. Huber, J. A. Dumesic, M. Mavrikakis, Journal of Physical Chemistry B 108, 14062 (2004).

12. "Strain-Induced Formation of Subsurface Species in Transition Metals", J. Greeley, W. P. Krekelberg, M. Mavrikakis, Angewandte Chemie International Edition 43, 4296 (2004).

13. "Adsorption and dissociation of O2 on Pt-Co and Pt-Fe alloys", Y. Xu, A. Ruban, M. Mavrikakis, Journal of the American Chemical Society 126, 4717 (2004).

14. "Competitive Paths for Methanol Decomposition on Pt(111)", J. Greeley, M. Mavrikakis, Journal of the American Chemical Society 126, 3910 (2004).

15. "Why Au and Cu are more selective than Pt for Preferential Oxidation of CO at low temperature", S. Kandoi, A. A. Gokhale, L. C. Grabow, J. A. Dumesic, M. Mavrikakis, Catalysis Letters 93, 93 (2004).

16. "On the origin of the catalytic activity of nanometer gold particles for low temperature CO oxidation", N. Lopez, T. V. W. Janssens, B. S. Clausen, Y. Xu, M. Mavrikakis, T. Bligaard, J. K. Nxrskov, Journal of Catalysis - Priority Communication, 223, 232 (2004).

Office of Science U.S. DEPARTMENT OF ENERGY

6. Catalyst Design Scientific | Engineering Output

- Publications (2001-2004):

17. "Atomic and molecular adsorption on Ir(111)", W. Krekelberg, J. Greeley, M. Mavrikakis, Journal of Physical Chemistry B 108, 987 (2004).

18. "Adsorption and Dissociation of O_2 on Gold surfaces: Effect of Steps and Strain", Y. Xu, M. Mavrikakis Journal of Physical Chemistry B, 107, 9298 (2003).

19. "A first-principles study of surface and subsurface hydrogen on and in Ni(111): Diffusional Properties and Coverage-Dependent behavior", J. Greeley, M. Mavrikakis, Surface Science 540, 215 (2003).

20. "The adsorption and dissociation of O_2 molecular precursors on Cu: The effect of Steps", Y. Xu, M. Mavrikakis Surface Science 538, 219 (2003).

21. "Atomic-Scale Evidence for an Enhanced Catalytic Reactivity of Stretched Surfaces", J. Wintterlin, T. Zambelli, J. Trost, J. Greeley, M. Mavrikakis, Angewandte Chemie International Edition (frontispiece) 42, 2849-2853 (2003).

22. "DFT studies for cleavage of C-C and C-O bonds in surface species derived from ethanol on Pt(111)", R. Alcalá , M. Mavrikakis, J.A. Dumesic, Journal of Catalysis 218, 178 (2003).

23. "CO Vibrational Frequencies on Methanol Synthesis Catalysts: a DFT study", with: J. Greeley, A. Gokhale, J. Kreuser, J.A. Dumesic, H. Topsoe, N-Y. Topsoe, M. Mavrikakis, Journal of Catalysis 213, 63 (2003).

24. "Atomic and Molecular Adsorption on Rh(111)", M. Mavrikakis, J. Rempel, J. Greeley, L. B. fHansen, J.K. Norskov, J. Chem. Phys. , 117, 6737 (2002).

25. "A First-Principles Study of Methanol Decomposition on Pt(111)", J. Greeley, M. Mavrikakis, Journal of the American Chemical Society 124, 7193 (2002).

26. "Adsorption and dissociation of O2 on Ir(111)", Y. Xu, M. Mavrikakis, Journal of Chemical Physics 116, 10846 (2002).

27. "Methanol Decomposition on Cu(111): A DFT Study", J. Greeley, M. Mavrikakis, Journal of Catalysis, 208, 291 (2002).

28. "DFT studies of Acetone and Propanal Hydrogenation on Pt(111)", R. Alcala, J. Greeley, M. Mavrikakis, J. A. Dumesic, Journal of Chemical Physics 116, 8973 (2002).

29. "Electronic Structure and Catalysis on Metal Surfaces", J. Greeley, J. K. Norskov, M. Mavrikakis, Annual Reviews of Physical Chemistry, 53, 319 (2002).

6. Catalyst Design Scientific | Engineering Output

- Distinctions and Honors:

1. CAREER Award, National Science Foundation (2002-2006).
2. Samuel C. Johnson Distinguished Fellowship (2005-2008).
3. 3M Non-tenured Faculty Award, 3M (2002-2003).
4. Shell Oil Company Foundation Faculty Career Initiation Award (2000).
5. SCIENCE Magazine: quoted in the March 14, 2003, issue (SCIENCE 299, 1684, 2003).
6. Angewandte Chemie International Edition (frontispiece) 42, 2849 (2003).
7. Featured in: Nanotechnology Now, 12/29/03.
8. Featured in: Chemical & Engineering News, Nov. 29, 2004, Vol. 82, Issue 48, pp. 25-28.
9. Cited in: Chemical & Engineering News, Aug. 22, 2005, Vol. 83, Issue 84, pp. 42-47.
10. Most viewed article the March 2004 issue of *Catalysis Letters:* "Why Au and Cu Are More Selective than Pt for Preferential Oxidation of CO at Low Temperature".

6. Scientific | Engineering Output

- Distinctions and Honors (cont.):

11. Press Release by NATURE MATERIALS: October 17, 2004 – Designer Catalysts for Hydrogen Chemistry.

12. Featured in Italian Newspaper: IL-SOLE 24 ORE (p. 11, 10/20/2004)

13. "Hot Paper of the Week" by ChemWeb.com, Member News Bulletin, Feb. 19, 2000.

14. Featured in Reactive Reports, March 2005 issue: http://www.reactivereports.com/44/44_1.html

15. Featured in EMSL – Pacific Northwest Laboratory Research Highlights (January/February 2005): http://www.emsl.pnl.gov/new/highlights/200502/

16. Highlighted in DOE-BES Weekly Report (March 28, 2005).

17. Featured in Council on Competitiveness: High Performance Computing and Competitiveness-Grand Challenge Case Study: *Customized Catalysts to Improve Crude Oil Yields: Getting More bang from Each Barrel* (April 2005): http://www.compete.org/pdf/HPC_Customized_Catalysts.pdf

18. Featured in the *Nanotechnology* section of MIT Technology Review (June 2005): http://www.technologyreview.com/articles/05/06/issue/ftl_nano.asp?p=2

19. Featured in DOE-BES Computational Research 2005 Greenbook: http://www.nersc.gov/news/greenbook/N5greenbook-print.pdf

20. Invited Participation in the NAE 2006 German-American Frontiers of Engineering Symposium (GAFOE), Murray Hill, NJ, 5/06.

Office of Science U.S. DEPARTMENT OF ENERGY

6. Catalyst Design Scientific | Engineering Output

- Citations (last 5 years):
  - **413**
- Dissertations:
  - **Jeff Greeley (2004) - PhD**
  - **Ye Xu (2004) - PhD**
  - **Amit Gokhale (2005) - PhD**
  - **Jacob Schieke (2002) - MS**
- Prizes and other honors:
  - **See list on previous slide**
- Residual and supported, living datasets and/or databases **that are accessed by a community? Size of the community?**
  - **N/A**
- Change in code capabilities and quality:
  - **N/A**

Office of Science U.S. DEPARTMENT OF ENERGY

6. Catalyst Design Scientific | Engineering Output

- Code and/or data contributed to the centers:
  - **N/A**
- Code and/or data, results, contributed to the scientific and engineering community at large:
  - **See previous list of publications**
- Company spin-offs based on code or trained people and/or CRADAs:
  - **None**
- Corporation, extra-agency, etc. use:
  - **None**
- Production of scientists & computational scientists during 2001-2005:
  - **5**
- Production of trained software engineers during 2001-2005:
  - **N/A**

Office of Science U.S. DEPARTMENT OF ENERGY

Project: Particle in Cell Simulation of Laser Wakefield Particle Acceleration

- Principal Investigator:
  - **Cameron Geddes, cgrgeddes@lbl.gov**
- URL:
  - **http://geddes.lbl.gov**
- DOE Office support:
  - **HEP – Accelerator Physics**
- DOE program manager:
  - **Philip Debenham, Bruce Strauss**
- Scientific domain:
  - **Accelerator Physics**
- Support for the development of the code:
  - **SciDAC: SciDAC Advanced Computing for 21st Century Accelerator Science and Technology**
  - **DE-AC03-76SF0098, DE-FG03-95ER40926, DE-FC02-01ER41178, DE-FG02-03ER83857, DE-AC03-76SF00098, DE-FG02-04ER84097**
  - **industry: Tech-X Corporation**
  - **other agencies: NSF: 0113907, AFOSR: FA9550-04-C-0041**

Office of Science U.S. DEPARTMENT OF ENERGY

Project: Particle in Cell Simulation of Laser Wakefield Particle Acceleration

- What problem are you trying to solve?
  - **Detailed and three-dimensional modeling of laser-driven wakefield particle accelerators.**
- What is the expected impact of project success?
  - **Plasma-based compact accelerators may allow access to new energy frontiers for high energy physics, and revolutionize applications of accelerators to radiation sources, chemistry and biology by providing small sources (Nature cover story on 30 Sep 2004)**
  - **High-resolution particle-in-cell simulations provide guidance to design the next accelerators, Account for three - dimensional physics.**
  - **High-resolution runs are vital to the development and validation of reduced computational models.**
- External communities & sizes that code and/or datasets support:
  - ***Plasma accelerator community, future high energy physics experiments***

Office of Science U.S. DEPARTMENT OF ENERGY

2. Wakefield Accelerators Project Team Resources

- Team institutional affiliations:
  - **Lawrence Berkeley National Laboratory, Tech-X Corporation, University of Colorado**
- To what extent are the code team members affiliated with the computer center institution?
  - **The code team members are affiliated with Berkeley Lab's Accelerator and Physics Division and/or Tech-X and Univ. Colorado, but not with NERSC.**
- Team composition and experience:
  - **domain scientists (non-computational): 3**
  - **computational scientists: 5 (these are also domain scientists)**
  - **computer scientists: 0**
  - **computational mathematicians: 0**
  - **database managers: 0**
  - **programmers: 0**
  - **program development and maintenance: 2**
  - **users of the team codes? Above, and wide use in community**

Office of Science U.S. DEPARTMENT OF ENERGY

2. Wakefield Accelerators Project Team Resources

- Team composition by educational level:
  - **senior faculty: 1**
  - **national laboratory scientists: 4**
  - **industrial scientists: 2**
  - **post-docs: 1**
  - **graduate students: 0**
  - **undergraduate students: 0**
  - **Ph.D: 8 domain Ph.D, 1 non-domain**
  - **MS: 0**
  - **BS: 1**
- Team resources utilization: **Incite 7 only, integrated over all scientists**
  - **time spent on code and algorithm development: 0 for this project**
  - **code maintenance: 0.05 FTE**
  - **problem setup: 0.6 FTE**
  - **production runs: 0.1**
  - **results analysis: 0.6**
  - **publications: none so far**

Office of Science U.S. DEPARTMENT OF ENERGY

75

3. Wakefield Accelerators Project Code: VORPAL

- Problem Type: **Simulation, Experimental Design**
- Types of algorithms and computational mathematics:
  - **Particle-in-cell.**
- What platforms does your code routinely run on?
  - **At NERSC: Seaborg (IBM SP), Bassi (IBM p575), Jacquard (Opteron/Infiniband)**
- Code size (single lines of code, function points, etc.);
  - **~200,000**
- Computer languages employed:
  - **C++ & MPI**
  - **200,000 SLOC C++ main code, -problem set-up, 5,000 SLOC Python-steering, 6000 SLOC IDL, 4000 lines BASH, 41,000 SLOC OpenDX)**
- What libraries are used? And What fraction of the codes does it represent?
  - **Serial, Parallel HDF5**
  - **MPI**
  - **Aztec**
- Code Mix:
  - **To what extent does your team develop and use your own codes? 100%**
  - **Codes developed by others in the DOE and general scientific community? 0%**
  - **Commercial application codes provided by the center? IDL visualization**

Office of Science U.S. DEPARTMENT OF ENERGY

76

3. Wakefield Accelerators Project Code: VORPAL

- What is the present parallel scalability on each of the computers the code operates on
  - **Projected or maximum scalability: Up to 4096 Seaborg processors.**
  - **How is this measured? Scaling up to 5000 processors on Seaborg has been demonstrated**
  - Is the code massively parallel? **Yes.**
- What memory/processor ratio do your project require? (e.g. Gbytes/processor)
  - ***15Mbyte/processor (Seaborg), 100Mbyte/proc (jacquard). Scales to large number of processors - listed values are minimum per processor (e.g. max # processors) used so far on large parallel runs.***
- Parallelization model: **MPI**
- Does your team use domain decomposition and if so what tools do you use?
  - ***Sub-domains, message passing.***
- What is the "efficiency" of the code and how is it measured:
  - **VORPAL has demonstrated almost linear speedup at 5,000 processors relative to a 256-processor run. Speedup is measured by the relative run times for a fixed size problem on Seaborg.**
- What are the major bottlenecks for scaling your code?
  - ***Under study.***
- What is the split between interactive and batch use? Why this split? Is interactive use more productive?
  - **All batch.**
- What is the split between code development on the computer center computers and on computers at other institutions?
  - ***All done at other institutions.***

Office of Science U.S. DEPARTMENT OF ENERGY

77

4. Wakefield Accelerators project resources input from the centers

- Plan with benchmarks & milestones: Incite 7
  - ***Simulation in 3d of laser wakefield accelerators at 100 MeV - 1 GeV***
  - ***Detailed 2d simulations to understand parameter dependence, convergence***
  - ***Benchmarking of new models, understanding of what new models may be warranted.***
  - ***Runs:***
    - ***3d short pulse run for essential 3d physics (& accompanying 2d runs) - done***
    - ***2d convergence to understand numerical and physical parameter sensitivity - nearly complete***
    - ***10 TW self modulated simulation in 3d - 100 MeV gain, based on above.***
    - ***1 GeV simulation in 3d***
- Steady state user of resources on a production basis per month (desired): **n/a, not steady state use: intend to do two very large runs**
- Annual use of resources (actual):
  - **Processor time (IBM SP POWER3-equivalent hours):**
    - **2002: 13 K**
    - **2003: 19K**
    - **2004: 266K (14 months)**
    - **2005: 206K**
    - **2006: 366K (7 months)**
  - **Disk: 3 TB**
  - **Tertiary storage: not used yet**

Office of Science U.S. DEPARTMENT OF ENERGY

78

4. Wakefield Accelerators project resources input from the centers

- Software provided by center:
  - **C++, MPI, HDF5, Subversion, IDL, Python (?)**
- Consulting provided by center:
- Direct project support from center acting as a team member:
  - ***Consulting, support, visualization***
- What is the size of their jobs in terms of:
  - **memory: 40 GB**
  - **concurrency (processors): 2000 (seaborg)**
  - **disk: 1 TB**
  - **tertiary store: not used**
- What is the wall-clock time for typical runs?
  - ***40 hours spent over 15 days of run time***

Office of Science U.S. DEPARTMENT OF ENERGY

79

5. Wakefield Accelerators Software Engineering, Development, Verification and Validation Processes

- Software development tools used:
  - **parallel development:**
  - **debuggers:**
  - **visualization: IDL, OpenDX, GnuPlot; remote difficult**
  - **production management and steering: svn, cvs**
- Software engineering practices. Please list the specific tools or processes used for:
  - **configuration management: autoconf, autotools**
  - **quality control: regression tests**
  - **bug reporting and tracking: informal**
  - **code reviews: periodic**
  - **project planning: monthly developer meeting**
  - **project scheduling and tracking: cvs, yearly releases**

Office of Science U.S. DEPARTMENT OF ENERGY

80

5. Wakefield Accelerators Software Engineering, Development, Verification and Validation Processes

- What is your verification strategy?
  - ***Solve problems with known analytic solutions, comparison with other models and community codes.***
- What use do you make of regression tests?
  - ***Nightly regression suite, email to developers. Code integrity checks.***
- What is your validation strategy?
  - ***LBNL experiments***
- What experimental facilities do you use for validation?
  - ***LOASIS laser facility, LBNL***
- Does your project have adequate resources for validation?
  - ***Yes - tight integration with LOASIS experiments.***

Office of Science U.S. DEPARTMENT OF ENERGY

81

6. Wakefield Accelerators Scientific | Engineering Output (mp278 and incite7)

- The scientific accomplishments 200x to present: Wakefield Acceleration
  - ***Physics behind newly observed monoenergetic electron bunches from laser wakefield accelerators (Nature 2004).***
  - ***Particle simulations of colliding pulse injection.***
  - ***Guiding at relativistic intensities, compensation for self guiding (PRL2004).***
  - ***Modeling of laser ionization blue shifting of laser pulses.***
  - ***Mode coupling and ionization effects in plasma channels (in preparation).***
  - ***Other applications outside plasma accelerators (electromagnetics, etc. Not included here).***
- The effect on the Office of Science programs:
  - **Plasma-based compact accelerators may allow access to new energy frontiers for high energy physics, and revolutionize applications of accelerators to radiation sources, chemistry and biology by providing small sources (Nature cover story on 30 Sep 2004)**
  - **High-resolution particle-in-cell simulations provide guidance to design the next accelerators. Simulations also account for three - dimensional physics.**
  - **High-resolution runs are vital to the development and validation of reduced computational models.**
  - **Broad applications of code to other programs: electromagnetics, etc.**

Office of Science U.S. DEPARTMENT OF ENERGY

82

6. Wakefield Accelerators Scientific | Engineering Output (mp278 and incite7)

- Publications:
  - ***Over 10 on this project***
- Selected Citations (last 5 years):

C.G.R. Geddes, Cs. Toth, J. van Tilborg, E. Esarey, C.B. Schroeder, J. Cary , W.P. Leemans, "Guiding of Relativistic Laser Pulses by Preformed Plasma Channels," Phys. Rev. Lett., volume 95, issue 14, 2005, pp. 145002-1 to 4. LBNL-57058.[Geddes guiding PRL2005.pdf]

C.G.R. Geddes, Cs. Toth, J. van Tilborg, E. Esarey, C.B. Schroeder, D. Bruhwiler, C. Nieter, J. Cary & W.P. Leemans, "Production of high quality electron bunches by dephasing and beam loading in channeled and unchanneled laser plasma accelerators," Physics of Plasmas, vol. 12, 2005, pp. 056709-1 to 10. LBNL-57062 [Geddes dephasing PoP2005.pdf]

C.G.R. Geddes, Cs. Toth, J. van Tilborg, E. Esarey, C.B. Schroeder, D. Bruhwiler, C. Nieter, J. Cary & W.P. Leemans, "High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding," Nature, Sept 30 2004, pp. 538-41. LBNL-55732. [Geddes Guided Accel Nature 2004.pdf]

Office of Science U.S. DEPARTMENT OF ENERGY

6. Wakefield Accelerators Scientific | Engineering Output (mp278 and incite7)

- Dissertations:
  - ***C.G.R. Geddes, 'Plasma Channel Guided Laser Wakefield Accelerator,' UC Berkeley, 2005. Dissertation presented experiments as well as simulations done under this project.***
- Prizes and other honors:
  - ***Hertz foundation dissertation award, 2005; Rosenbluth dissertation award 2006 for the above thesis.***
- Residual and supported, living datasets and/or databases that are accessed by a community? Size of the community?
  - ***N/A***
- Change in code capabilities and quality:
  - ***New code 2000, geared towards problem. Many capabilities added since, including absorbing boundaries, ionization, variable weighting.***

Office of Science U.S. DEPARTMENT OF ENERGY

6. Wakefield Accelerators Scientific | Engineering Output

- Code and/or data contributed to the centers:
  - ***N/A***
- Code and/or data, results, contributed to the scientific and engineering community at large:
  - ***Results as above.***
  - ***code installed at LBL, Argonne, JLab, Fermi and others (free)***
- Company spin-offs based on code or trained people and/or CRADAs:
  - *N/A*
- Corporation, extra-agency, etc. use:
  - ***Code sold commercially to non - DOE, revenue $40,000 2006.***
- Production of scientists & computational scientists **during 2001-2005:**
  - ***Cameron Geddes, Estelle Michel, Amar Hakim***
- Production of trained software engineers **during 2001-2005:**
  - ***Victor Przebinda; Greg Warner (in training)***

A Closer Look at Four Selected Projects on the Leadership Systems at the ORNL National Center for Computational Sciences

Douglas B. Kothe
NCCS Director of Science

Al Geist
NCCS Chief Technology Officer

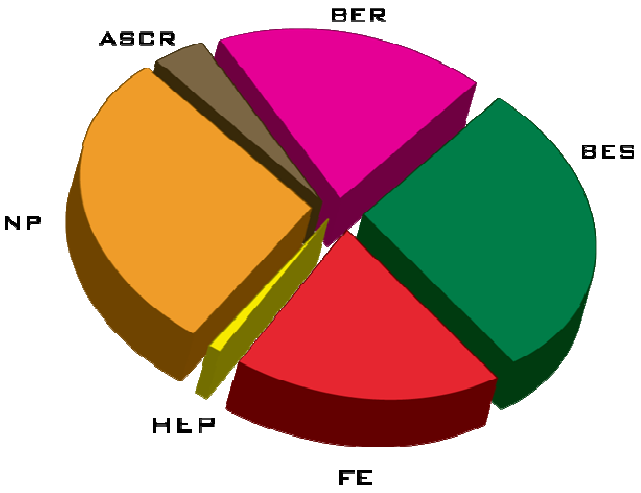
THE CENTER FOR COMPUTATIONAL SCIENCES

OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY

Outline

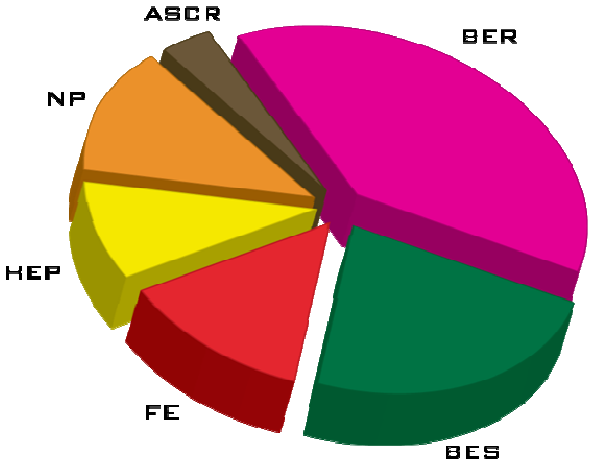
- **NCCS gets allocations for two capability systems**
- **Project showcases (deep dive)**
  - Choosing representative projects across science domains
  - Fusion
  - Combustion
  - Climate
  - Nanoscience
- **Petascale readiness**

NCCS Allocations for the Cray XT3



ASCR	1,000,000	4%
BER	4,996,856	19%
BES	7,500,000	29%
FE	5,000,000	19%
HEP	30,000	<1%
NP	7,550,000	29%
Total	26,076,856	100%

NCCS Allocations for the Cray X1E

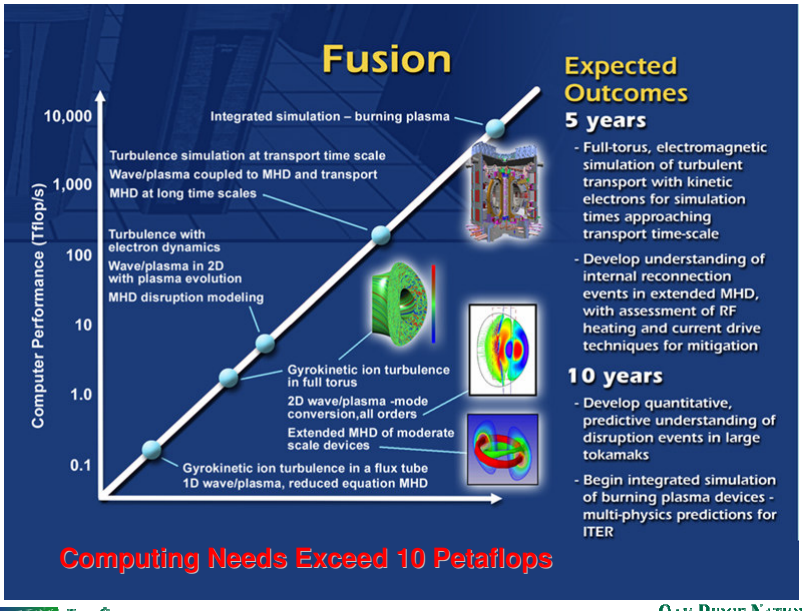


ASCR	200,000	4%
BER	2,029,000	38%
BES	1,200,000	23%
FE	665,240	13%
HEP	500,000	9%
NP	700,000	13%
Total	5,294,240	100%

4 Representatives from FY06 Allocated Projects

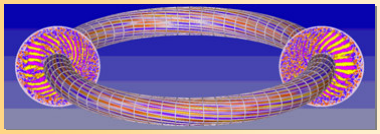
Project	Jaguar Allocation	Percent of Jaguar	Phoenix Allocation	Percent of Phoenix	Type	Description	Domain Science	PI
AST003	1,250,000	4.1%	0	0.0%	LCF	Multi-dimensional Simulations of Core-Collapse Supernovae	Astrophysics	Burrows
AST004	3,000,000	9.9%	0	0.0%	LCF	Ignition and Flame Propagation in Type Ia Supernovae	Astrophysics	Woosley
AST005	3,550,000	11.7%	700,000	11.9%	LCF	Multi-dimensional Simulations of Core-Collapse Supernovae	Astrophysics	Mezzacappa
BIO014	500,000	1.7%	0	0.0%	LCF	Next Generation Simulations in Biology	Biology	Agarwal
BIO015	1,484,800	4.9%	0	0.0%	INCITE	Molecular Dynamics Simulations of Molecular Motors	Biology	Karplus
CHM022	1,000,000	3.3%	300,000	5.1%	LCF	Rational Design of Chemical Catalysts	Chemistry	Harrison
CLI016	0	0.0%	29,000	0.5%	LCF	Role of Eddies in Thermohaline Circulation	Climate	Cessi
CLI017	3,000,000	9.9%	2,000,000	33.9%	LCF	**Climate-Science Computational End Station**	**Climate**	Washington
CLI018	1,496,856	4.9%	0	0.0%	LCF	Studies of Turbulent Transport in the Global Ocean	Climate	Peacock
CSC023	1,000,000	3.3%	200,000	3.4%	LCF	PEAC End Station	Computer Science	Worley
CSC026	950,000	3.1%	0	0.0%	INCITE	Real-Time Ray-Tracing	Computer Science	Smyth
EEF049	3,500,000	11.6%	300,000	5.1%	LCF	**Simulations in Strongly Correlated Electron Systems**	**Materials Science**	Schulthess
EEF050	0	0.0%	200,000	3.4%	INCITE	Large Scale Computational Tools for Flight Vehicles	Engineering	Hong
EEF051	500,000	1.7%	0	0.0%	INCITE	Numerical Simulation of Brittle nd Ductile Materials	Materials Science	Ortiz
FUS011	2,000,000	6.6%	225,000	3.8%	LCF	**Gyrokinetic Plasma Simulation**	**Fusion**	Lee
FUS012	0	0.0%	440,240	7.5%	LCF	Tokamak Operating Regimes Using Gyrokinetic Simulations	Fusion	Candy
FUS013	3,000,000	9.9%	0	0.0%	LCF	Wave-Plasma Interaction and Extended MHD in Fusion Systems	Fusion	Batchelor
FUS014	0	0.0%	400,000	6.8%	INCITE	Interaction of ETG and ITG/TEM Gyrokinetic Turbulence	Fusion	Waltz
HEP004	30,000	0.1%	0	0.0%	LCF	Reconstruction of CompHEP-produced Hadronic Backgrounds	High Energy Physics	Newman
HEP005	0	0.0%	500,000	8.5%	LCF	Design of Low-loss Accelerating Cavity for the ILC	Accelerator Physics	Ko
NPH004	1,000,000	3.3%	0	0.0%	LCF	Ab-inito Nuclear Structure Computations	Nuclear Physics	Dean
SDF022	3,000,000	9.9%	600,000	10.2%	LCF	**High-Fidelity Numerical Simulations of Turbulent Combustion**	**Combustion**	Chen
Total	30,000,656		5,894,240					

Showcase Fusion Energy



Future Energy Security – Fusion Simulation

- **The Problem**
  Understanding the physics of plasma behavior is essential to designing reactors to harness clean, secure, sustainable fusion energy.
- **The Research**
  Controlling turbulence is essential because it causes plasma to lose the heat that drives fusion. Realistic simulations determine which reactor scenarios promote stable plasma flow.
- **Impact of Achievement**
  High-resolution computer simulations are needed to set up experiments and engineers will use the simulations to design equipment for efficient reactor operation.



A twisted mesh structure is used in the GTC simulation.

Principal Investigator
Wei-li Lee
Princeton Plasma Physics Laboratory

Gyrokinetic Plasma Simulation
Fusion Project FUS011

- **Stakeholders: PI and Clients (pays for product development)**
  - PI: Wei-li Lee, Princeton Plasma Physics Laboratory (wwlee@pppl.gov)
  - Clients: DOE SC/FES (Rostom Dagazian), DOE SC/SciDAC (Michael Strayer), DOE/SC/ASCR/MICS (Anil Deane)
- **Code development support (DOE support: 100%)**
  - SciDAC-1 Project: Center for Gyokinetic Particle Simulation of Turbulent Transport
    - $0.8M (MFES), $0.2M (ASCR)
  - SciDAC-1 Fusion Simulation Project: Center for Plasma Edge Simulation
    - $1M (MFES), $1M (ASCR)
  - MICS Multi-Scale Math & Education Project: Multi-Scale Gyrokinetics Project
    - $0.55M (ASCR)
  - Value of computer time: ~$1.94M
- **Technical goals**
  - Understand turbulent transport in magnetic fusion core & edge plasmas & its interactions with low frequency MHD modes & high frequency cyclotron waves.

Gyrokinetic Plasma Simulation
Fusion Project FUS011

- **Grand challenge problems**
  - ITG turbulence on ITER-size plasmas
  - Neoclassical neutral edge transport
  - Wave heating and effects on MHD profiles
- **Expected impact of project success**
  - A greater understanding of the energy and transport issues in core and edge ITER plasmas

Gyrokinetic Plasma Simulation
Project Team Resources

- **Team size**
  - 19 (core), other extended team members
- **Team institutional affiliations**
  - PPPL, NYU, U. Irvine, U. Colorado, General Atomics, Columbia University, Rutgers University, Cal Tech, MIT, Lehigh University
- **Team computer center institution affiliation**
  - ORNL NCCS liaison on team (SciDAC project member; former PPPL staff)
- **Team composition and experience**
  - Domain scientists: 19
  - Computational/computer scientists & applied mathematicians: 12
- **Team composition by educational level**
  - Ph.D.: 19 (current)
  - Mix of Jr/Sr faculty, national lab scientists, & industrial scientists
  - Graduate students: 5
- **Team composition by WBS activity**
  - Production: 48%; Results analysis: 30%; Code/algorithm development: 15%; Maintenance: 5%; Problem setup: 2%

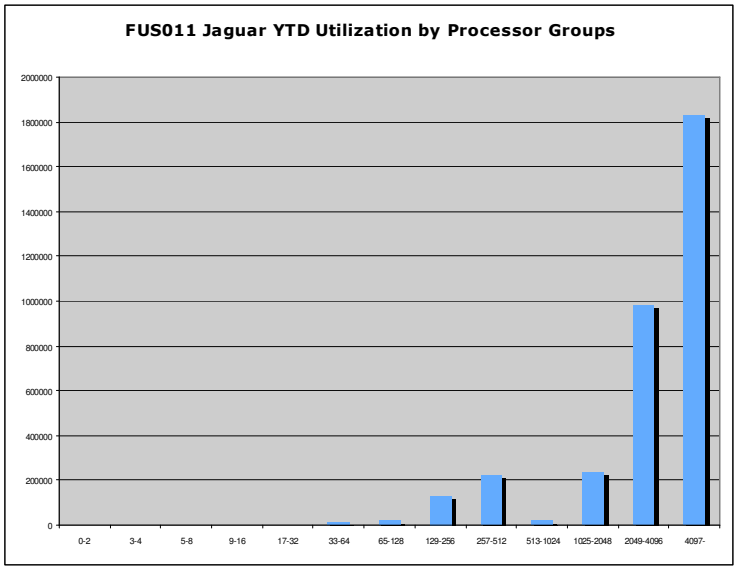
Gyrokinetic Plasma Simulation
Project Resources: NCCS Input

- **CY06 milestones at NCCS**
  - GTC
    - Convergence studies on ETG and ITG simulations.
    - Studying the trends of size scaling and isotope scaling of ITG turbulence with adiabatic electrons and with realistic electron dynamics.
    - Studying realistic electron effects and finite-beta effects on ITG turbulence
  - XGC-ET
    - Study neoclassical component of the pedestal scaling law for various existing tokamak devices, compare with experimental results, and make predictions for ITER
    - Study neoclassical flow dynamics in the edge plasma under phenomenological L and H mode conditions for various devices.
    - Study Divertor heat load under quiescent H-mode condition and under simulated ELM conditions for various devices and plasma conditions.
    - The primary code development activities in the immediate future will include turbulence physics implementation, self-consistently with the neoclassical and neutralwall physics, and the corresponding code optimization and parallelization
  - MSPC
    - The initial code development activities include 1) the optimization and parallelization of the existing 3D gyrokinetic particle in slab geometry to increase its efficiency, 2) the implementation of finite-beta effects using the split-weight scheme [Lee01], 3) the integration with the MHD modes [Lee03] along with the mesh refinement methodology, and 3) the schemes for retaining ion cyclotron waves in gyrokinetic particle simulation.

Gyrokinetic Plasma Simulation
Project Resources: NCCS Input

- **Software provided by NCCS**
  - Compilers, editors, debuggers, communication/math libraries, viz tools, performance tools
- **Size of typical NCCS jobs (concurrency (processors), memory, local, and archival storage)**
  - Typically (for GTC) a 4800 PE job requiring 9.6 TB memory and 5 TB of local storage
- **What is the scalability of these codes**
  - Excellent (GTC): Good 65K PE scaling observed on BG/L
- **What is the wall-clock time for typical runs?**
  - Typically 100 hours
- **Steady state (production) monthly resource use**
  - Processor number: 4800 on Cray XT3
  - Processor time: 100 wall clock hours per simulation
  - Local storage: 1 TB
  - Archival storage: 10 TB annually, or ~1 TB per month

Gyrokinetic Plasma Simulation
NCCS Resource Usage



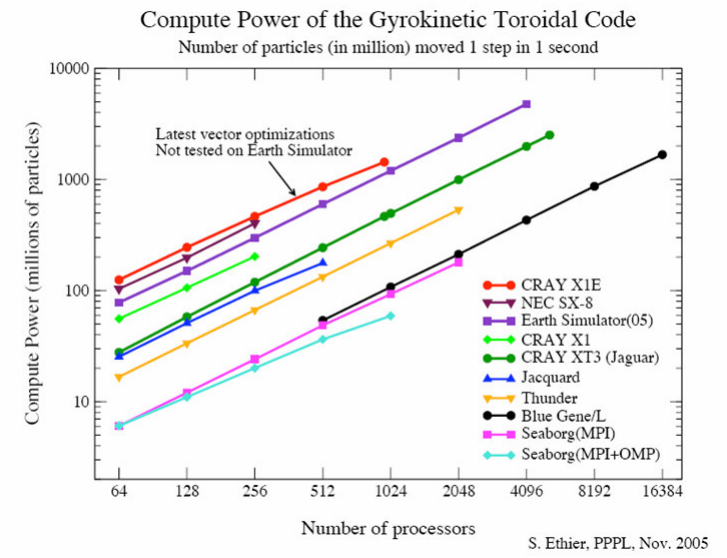
Gyrokinetic Plasma Simulation
Project Codes

- **Problem Type**
  - Scientific simulation: magnetic fusion plasma physics embodied in GTC, GEM, XGC, Degas-2, M3D, NIMROD
- **Algorithm types and computational mathematics**
  - Particle-in-cell + finite element, AMR, finite difference, Krylov iterative solvers
- **Platforms used for routine execution**
  - Cray XT-3, Cray X1E, Earth Simulator, IBM SP, SGI Altix, Linux clusters, NEC, Blue Gene (Cray XT-3 preferred)
- **Code statistics (GTC)**
  - Size (function points): 2,000
  - Age: 7 years
  - Annual growth: 5%
- **Computer languages employed**
  - Computer languages employed: Object-based F90, (some C and some C++)
  - Structure of the codes: mostly Fortran
- **Libraries**
  - Libraries used: MPI, PETSC, HDF5, NetCDF
  - Library extent: <10% (for most codes)
- **Code Mix**
  - Team internally develops and uses all codes that are needed
  - Using PETSC, HYPRE, Prometheus, SuperLU developed by TOPS
  - No commercial application codes provided by NCCS are used

Gyrokinetic Plasma Simulation
Project Codes

- **Present parallel code (GTC) scalability on all relevant platforms**
  - Projected or maximum scalability
    - Executed on BG/L up to 65K PEs; realized a speedup of 1.9/2 for dual core PEs
  - Scalability is measured with simple execution timings
  - Code is massively parallel
- **What memory/processor ratio do your project require? (e.g. Gbytes/processor)**
  - 2 GB/PE generally sufficient; some codes require up to 4 GB/PE
- **Parallelization model**
  - Toroidal domain decomposition of grid with MPI, use OpenMP when available
- **Code (GTC) "efficiency"**
  - Hardware-centric (classical) measure: % of peak (e.g., 16% on Cray XT3)
  - Physics-centric measure: # of particles per second pushed in one time step
- **Major scaling bottlenecks**
  - Particle-mesh operations (gather-scatter), linear solvers (XGC), spline operations (XGC)
- **Split between interactive and batch use**
  - Production: 100% batch (a small amount of debugging work)
  - Interactive: Exclusively for development/debugging
- **What is the split between code development on the computer center computers and on computers at other institutions?**
  - 99% of NCCS resource usage is for production; 99% of development work is on Linux clusters

Gyrokinetic Plasma Simulation
GTC Scalability



Gyrokinetic Plasma Simulation
Software Engineering, Development, V&V

- **Software development tools**
  - Parallel development: Cray PAT performance tools
  - Debuggers: TotalView
  - Visualization: Internally-developed tool using AVS/Express and IDL
  - Production management and steering: batch submission scripts moving to Kepler-based workflow
- **Software engineering practices**
  - Configuration management: CVS, SVN
  - Quality control: regression tests
  - Bug reporting and tracking: CVSlogs & emails
  - Code reviews: informal
  - Project planning: proposals, reviews
  - Project scheduling and tracking:
- **Verification strategy**
  - Exhaustive benchmarking with other codes linearly and nonlinearly (in particular, with FULL, GEM, GS2, GYRO)
  - GTC is the de facto standard gyrokinetic code
- **Regression test use**
  - Test against benchmark code (FULL for linear) when GTC is ported to ensure match of several derived quantities
  - GTC (and all the fusion codes) always use the CYCLONE parameters for test runs
- **Validation strategy?**
  - Not ready to validate code - still in the process of fully verifying
- **What experimental facilities do you use for validation?**
  - Compared some results to NSTX and DIIID, but need to better understand the physics before proceeding
- **Does your project have adequate resources for validation?**
  - No

Gyrokinetic Plasma Simulation
Science Output

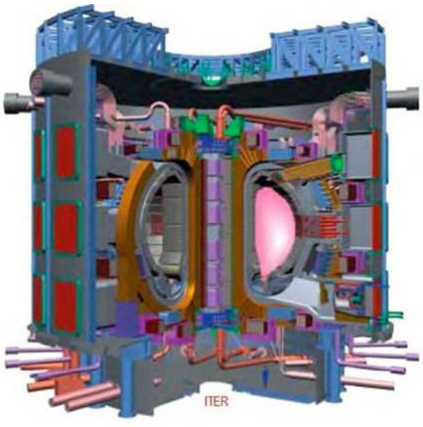
- **Recent Scientific accomplishments**
  - Evolving simulation tool suite for ITER design and analysis
  - ITG drift instabilities are a principal cause of turbulent transport in tokamaks
  - Nonlinearly generated zonal flows associated with ITG turbulence break up eddies and reduce turbulent transport
  - Large ITER-like values of a/ρ (e.g., 1000) indicate a transition from Bohm to GyroBohm ion diffusivity scaling (good for ITER!)
  - Velocity-space nonlinearities in ITG turbulence further enhance zonal flow, further reducing turbulent transport
  - Particle convergence (measured by numerical particle noise) demonstrated for ITG simulations
  - ETG drift instabilities may not be relevant for tokamak confinement
  - Turbulence spreading the cause for Bohm scaling in small devices
- **Impact on Office of Science programs**
  - Understand, quantify, and control how turbulence causes heat, particles, and momentum to escape from plasmas
  - ITER design guidance, scaling laws
- **Publications**
  - Average of 12/year in journals

Gyrokinetic Plasma Simulation
Science Output

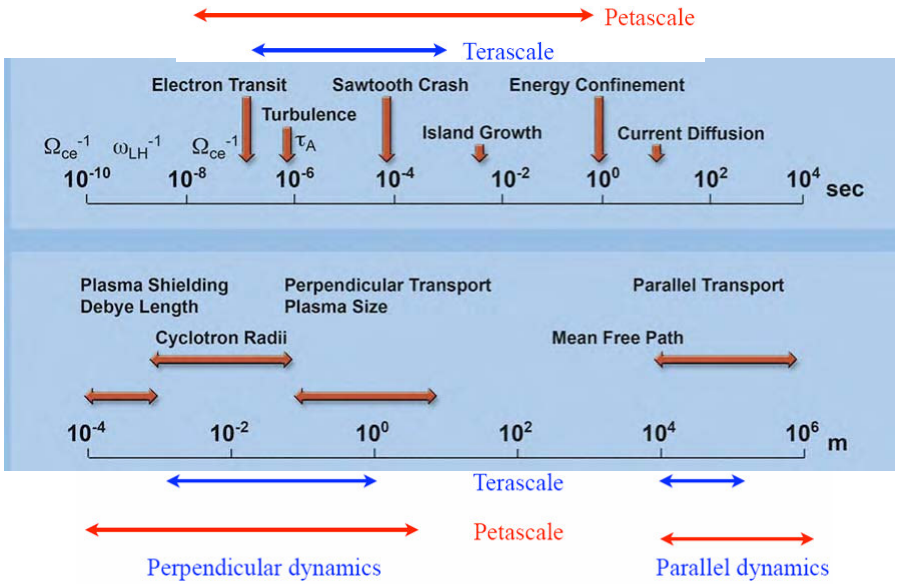
- **Dissertations**
  - 6 PhD dissertations based in part upon project simulation tools
- **Prizes and other honors**
  - 6 APS fellows, 10 invited talks at major meetings, one Gordon Bell Prize winner.
- **Change in code capabilities and quality over time**
  - GTC has gone from adiabatic electrons capability in 1999 to kinetic electrons capability in 2006. From large aspect ratio circular geometry to shaped plasmas in full general geometry. GTC scalability increased from 64 to 64,000 processors
- **Code and/or data contributed to the centers**
  - ETG/ITG simulation datasets, end-to-end analysis tools, evolved versions of all fusion codes
- **Code and/or data, results, contributed to the scientific and engineering community at large**
- **Company spin-offs based on code or trained people and/or CRADAs**
  - TechX is now pursuing gyrokinetic simulation
- **Corporation, extra-agency, etc. use**
  - General Atomics and TechX
- **Training, education, outreach**
  - Production of scientists & computational scientists during 2001-2005: 10
  - Production of trained software engineers during 2001-2005: 2

Gyrokinetic Plasma Simulation
Science Possible at the Petascale

- **Explore burning ITER-size plasmas with electromagnetic (Alfven) physics**
  - electron transport associated with electron skin depth
  - size scaling and isotope scaling with electromagnetic perturbations
- **Integrated modeling of**
  - Core-edge simulation
  - Transport time scale simulation
  - Heating and turbulence simulation
  - Turbulence and MHD simulation
- **Example petascale simulation**
  - ITER-type plasma with a grid size of the order of the electron skin depth
  - One trillion particles on a 10,000x10,000x100 grid (100 particles/cell)
  - Assume half the memory for particle data and the other half for grid data
  - 10^8 elements per plane; toroidal and radial domain decomposition

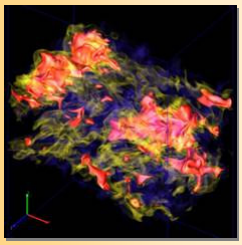


Gyrokinetic Plasma Simulation
Science Probed at the Petascale



Showcase - Combustion Simulation

- **The Problem**
  Detailed computer models are needed for design of cleaner, more efficient and environmentally friendly combustors.
- **The Research**
  The first 3-dimensional direct numerical simulation of a non-premixed flame with detailed chemistry.
- **Impact of Achievement**
  Advancing basic understanding of turbulent combustion and developing predictive combustion models are essential to deliver reliable data for manufacturer design of combustors and to limit hardware testing costs.



Principal Investigator
Jackie Chen
Sandia National Laboratories

High-Fidelity Numerical Simulations of Turbulent Combustion
Combustion Project SDF022

Stakeholders: PI and Clients (pays for product development)

PI: Jackie Chen, SNL/Livermore (jhchen@sandia.gov)

Clients: DOE SC/BES/Chemical Sciences (John Miller), DOE SC/SciDAC (Michael Strayer)

Code development support

DOE support: 85% SC/SciDAC, 15% SC/BES

Value of computer time: ~$3.74M

Technical goals

- Understanding the coupling between turbulence and chemistry in combustion
- Validate experimental techniques and chemical mechanisms in the presence of transport
- Advance predictive model development for design of combustion devices
- Fully characterize operating parameter space of devices (not viable w/ experiment)

Grand challenge: Understand how turbulent mixing affects

- Extinction and reignition
- Autoignition with compression heating
- Flame structure
- Flame propagation
- Coupling of aerodynamic stretch and intrinsic flame instabilities
- Pressure effects on amplification of flame instabilities
- Pressure affects on autoignition, NOx/CO emissions, & soot production/destruction & transport.

High-Fidelity Numerical Simulations of Turbulent Combustion
Combustion Project SDF022

Impact

- Improved realizable fuel efficiencies of devices
  - A 50 % increase in fuel efficiency in automobiles translates into a 21% reduction in oil used for transportation, where transportation accounts for 2/3 of U.S. oil consumption

External communities: sizes that code and/or datasets support

- Simulation data will be shared with the combustion community via a web portal and biannual workshops targeted at specific modeling issues in the community.
- Data already shared with two modeling groups at U. Iowa/Ames Lab (Fox/Smith) and Stanford (Pitsch)
- In discussion with several other modeling groups about use of data
- Currently in the process of enhancing capabilities for analysis, visualization and data sharing at Sandia Combustion Research Facility

High-Fidelity Numerical Simulations of Turbulent Combustion
Project Team Resources

- **Team size**
  - 5 (core), 10 (extended)
- **Team institutional affiliations**
  - SNL/Livermore, Univ. of Utah
- **Team computer center institution affiliation**
  - 3 NCCS liaisons on team (1 SciDAC project member; former SNL staff)
- **Team composition and experience**
  - Domain scientists: 8
  - Computational/computer scientists & applied mathematicians: 2
- **Team composition by educational level**
  - Ph.D.: all (currently) but one, which is pending (graduate student)
  - Mix of faculty, postdocs/students, and national lab scientists
- **Team composition by WBS activity**
  - Production: 70%; Results analysis: 10%; Code/algorithm development: 5%; Maintenance: 5%; Problem setup: 10%

High-Fidelity Numerical Simulations of Turbulent Combustion
Project Resources: NCCS Input

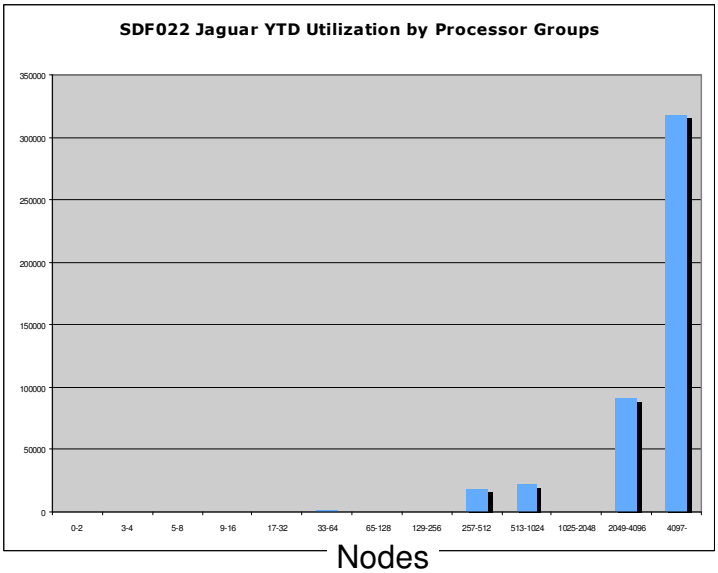
- **CY06 milestones at NCCS**
  - *Lean limit turbulent premixed combustion for stationary power generation*
    - Parametric 3D DNS of a canonical slot-burner Bunsen flame configuration with detailed CH4-air chemistry to understand turbulence/flame interactions in the thin-reaction zones regime
    - Perform simulations long enough to achieve statistical stationarity
    - The dependency of turbulent flame characteristics on the ratios of turbulence intensity-to-flame speed and integral length scale-to-flame thickness (Reynolds & Karlovitz numbers) will be determined
    - Investigate the role of curvature dissipation and provide statistics to improve mean flame stretch model predictions
    - Investigate attenuation of flame response to high turbulence intensities
    - ~2M node-hours for a series of 4 parametric runs
  - *Extinction/reignition of turbulent methane-air jet flames*
    - Build on success of FY05 simulations of turbulent nonpremixed CO/H2/air temporal jet flames by extending the study to more complicated methane/air kinetics, which may exhibit a qualitatively different behavior as extinction is approached
    - Perform 3D DNS of a methane/air temporal jet flame to study the effect of Reynolds number and fuel kinetics on local extinction and re-ignition
    - Provide new understanding of the dynamics of extinction and reignition, and to provide a numerical benchmark for model development
    - ~1.6 million hours for a parametric study of 3 runs

High-Fidelity Numerical Simulations of Turbulent Combustion
Project Resources: NCCS Input

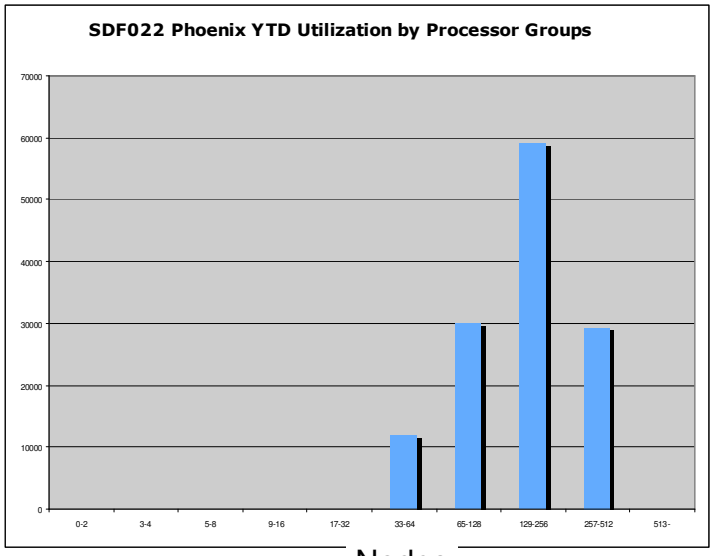
- **Steady state (production) monthly resource use**
  - Processor number: 4800 on Cray XT3
  - Processor time: 168 wall clock hours (in 24 hour chunks)
  - Local storage: 10-20 TB
  - Archival storage: 25 TB
- **Software provided by NCCS**
  - Compilers, editors, debuggers, communication/math libraries, viz tools, performance tools
- **Size of typical NCCS jobs (concurrency (processors), memory, local, and archival storage)**
  - Typically a 4800 PE job requiring 9.6 TB memory, ~3 TB of local storage, ? of archival storage
- **What is the scalability of these codes**
  - Good scaling observed out to max available PEs (5K)
- **What is the wall-clock time for typical runs?**
  - Typically max allowed (24 hours)

High-Fidelity Numerical Simulations of Turbulent Combustion
NCCS Resource Usage

CPU hours



High-Fidelity Numerical Simulations of Turbulent Combustion
NCCS Resource Usage



High-Fidelity Numerical Simulations of Turbulent Combustion
Project Code

- **Problem Type**
  - Scientific simulation: S3D solves a fully coupled system of time-varying PDEs governing the full compressible reacting Navier-Stokes, total energy, species continuity and continuity equations coupled with detailed chemistry. The PDEs are supplemented with constitutive relationships for the ideal gas EOS and models for reaction rate, molecular transport, & thermodynamic properties.
- **Algorithm types and computational mathematics**
  - High-order accurate, non-dissipative numerical scheme: 4th-order explicit Runge-Kutta time integration, eighth-order (with tenth-order filters) finite spatial differences on a Cartesian, structured grid, and Navier-Stokes Characteristic Boundary Condition
  - The coupling of high-order finite difference methods with explicit R-K time integration make very effective use of the available resources, obtaining spectral-like spatial resolution without excessive communication overheads and allowing scalable parallelism
- **Platforms used for routine execution**
  - Ports easily to all platforms
  - Preferred machine: Cray XT3 (90% parallel efficiency on 5K PEs) or Cray X1E
- **Code statistics**
  - Size: Lines - 101,320; Functions - ~350 (10% growth annually)
  - Age: 16 years
- **Computer languages employed**
  - Mix of Fortran 77 and mostly Fortran 90
- **Libraries**
  - Libraries used: MPI
  - Library extent: <1%

High-Fidelity Numerical Simulations of Turbulent Combustion
Project Code

- **Code Mix**
  - Team internally develops and uses all codes that are needed
  - No codes developed by others in the DOE and general scientific community are used
  - No commercial application codes provided by NCCS are used
- **Present parallel code scalability on all relevant platforms**
  - Projected/maximum scalability: >5K processors on XT3, but can go higher (projected maximum is 100K PEs)
  - How measured
    - Shock physics DNS code at LLNL (Miranda) has similar algorithms and has scaled to 100K BG/L PEs
    - S3D should scale better than Miranda given the larger ratio of work load per processor to communication.
  - Massively parallel: yes
- **Memory/processor ratio (minimum) required**
  - ~1 GB/PE generally sufficient
- **Parallelization model**
  - Domain decomposition with MPI
- **Code "efficiency"**
  - Hardware-centric measure: 90% parallel efficiency on 5K XT3 PEs for weak scaling test
  - Physics-centric measure: minimize the CPU time per grid point per time step
- **Major scaling bottlenecks**
  - None easily identifiable up to 100K PEs
- **Split between interactive and batch use**
  - Production: 100% batch (a small amount of debugging work)
  - Interactive: Exclusively for development/debugging
- **What is the split between code development on the computer center computers and on computers at other institutions?**
  - 100% of NCCS resource usage is for production; 100% of development work is on Linux clusters

High-Fidelity Numerical Simulations of Turbulent Combustion
S3D Scaling

S3D's parallel scaling has been tested both on X1E and XT3 and is found to scale extremely well on thousands of processors as shown in Figure 2. On X1E, 75% parallel efficiency is observed on 900 processors, and on XT3, 98% efficiency is observed at 2048 processors, and 90% efficiency is observed on 5120 processors.

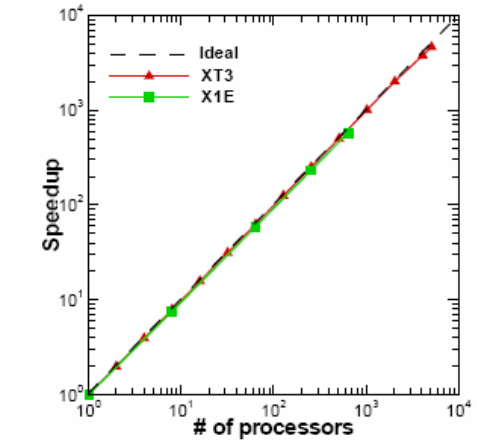


Figure 2: Parallel speedup on X1E and XT3

High-Fidelity Numerical Simulations of Turbulent Combustion
Code Productivity & Scalability

Productivity needs and measures
- Getting data in & out of center
- Better turn-around time for the project

Performance needs and measures
- Generate more accurate results in a constant solution time (with larger problems)
- Perform larger runs with more detailed chemistry
  - 25 TF: Chemical mechanism for CO/Hw and reduced mechanism for CH_4 and molecular transport model. There would be 2.5 decades of time and length scales resolved for reactive turbulent flow.
  - 100 TF: Same as above, but would increase Reynolds number or Damkohler number.
  - 250 TF: Same as above, but would increase Reynolds number or Damkohler number and would also increase chemical mechanism size to describe sooting flames like ethylene, transport simplified 2-equation soot model and optically thick thermal radiation.
  - 1 PF: Same as above, but would increase Reynolds number or Damkohler number and would also increase pressure from ambient to 10-20 atmospheres where greater resolution is required. Would consider chemical mechanisms that include multi-stage ignition characteristics, like n-heptane.

High-Fidelity Numerical Simulations of Turbulent Combustion
Science Output

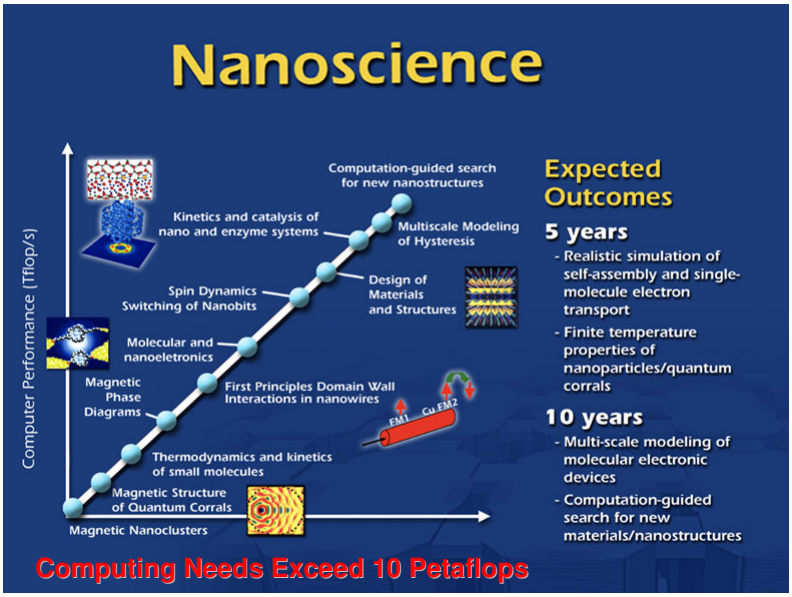
- **Recent scientific accomplishments**
  - 3D 500M-grid point DNS of turbulent dynamic plane CO/H2 jet flames performed with detailed chemistry at Re up to 9000
    - Re-dependence on turbulent mixing properties and flame structure quantified
  - Determined Turbulence-to-Mixing time scale ratio for reactive flows
    - Quantity widely used in combustion models, e.g. transported PDF model
  - Understand flame structure in stationary lean premixed flames under intense turbulence
    - First DNS of a stationary turbulent Bunsen flame with detailed chemistry
    - Flame structure is penetrated by small scale eddies leading to thickening of preheat zone, but conditional mean reaction rates still resemble a strained laminar flame
- **Impact on Office of Science programs**
  - Achieving lean premixed combustion in land-based stationary gas turbines
    - High thermal efficiency
    - Low NOx emissions due to lower flame temperatures
  - DNS-enlightened understanding of premixed flame propagation and structure increases simulation predictability and likelihood of meeting engineering goals
- **Publications**
  - Average of 5/year in journals

High-Fidelity Numerical Simulations of Turbulent Combustion

Science Possible at the Petascale

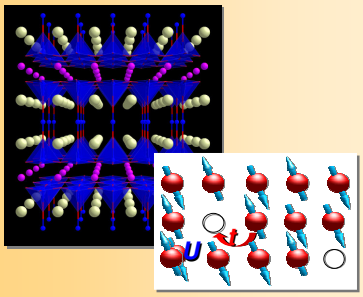
- **Study turbulent combustion at higher Reynolds and Damkohler number (ratio of smallest turbulent scale to flame scale, thinner flames)**
  - Increase pressure from ambient to 10-20 atmospheres where greater resolution is required
  - Consider chemical mechanisms that include multi-stage ignition characteristics, e.g. n-heptane
  - Continue to study turbulent transport with improved physics models
- **Petascale requirements driver**
  - To treat the multiscale problem of turbulence, the number of grid points required is huge and scales as Reynold number ~ N^9/4.
  - Need to simulate for long times to achieve statistical stationarity for model development (several 100,000 time steps per realization). The number of transported variables is also large ~20-30 to describe the simplest hydrocarbon fuels like methane).

Showcase Nanoscience



Materials and Nanoscience

- **The Problem**
  Functional nanostructures and strongly correlated materials created extraordinarily promising materials that can revolutionize our way of life.
- **The Research**
  New insights from large-scale computer simulations can greatly accelerate scientific progress.
- **Impact of Achievement**
  - Understanding the nature of HiTc
  - Design of high density storage
  - Self assembling molecular devices
  - Mechanisms to control DNA damage



First solution of 2D Hubbard model for predicting superconductivity transition temperature

Principal Investigator
Thomas Schultess
Oak Ridge National Laboratory

EEF049 – Predictive simulations in strongly correlated electron systems and functional nanostructures

- **Stakeholders: PI and Clients**
  - PI: Thomas Schultess, Oak Ridge National Laboratory
  - Clients: DOE SC/BES, DOE SC/OSCAR, NSF
- **Code development support**
  - DOE 50%, NSF 40%, Internal 10%
  - Value of computer time: ~$3.11M
- **Technical goals**
  - Develop computational instrumentation that will allow us to push the envelope in electronic structure calculations for functional nanostructures as well as perform quantum many-body simulations for material-specific models of strongly correlated electron systems.
- **Grand challenge:**
  - Predicting superconductivity transition temperatures
  - Properties of nanoparticles for ultra high density storage medium
  - Simulation of molecular devices in natural conditions
  - Understanding mechanisms that control damage to DNA

EEF049 – Predictive simulations in strongly correlated electron systems and functional nanostructures

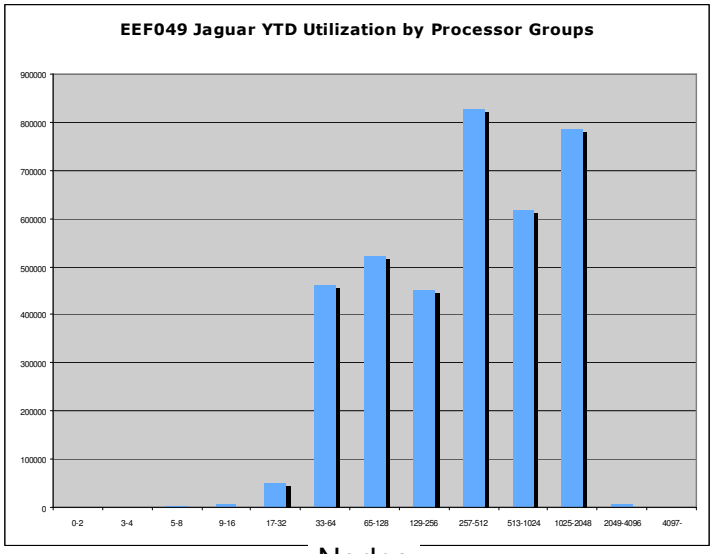
- **Team size**
  - 24
- **Team institutional affiliations**
  Oak Ridge National Laboratory, NCSU, Vanderbilt, University of Tennessee, University of Cincinnati, Georgia Tech., Pittsburg Supercomputer Center
- **Team computer center institution affiliation**
  - 1 NCCS liaison on team
- **Team composition and experience**
  - Domain scientists: 23
  - Computational/computer scientists & applied mathematicians: 1
- **Team composition by educational level**
  - Ph.D.: 22 (currently) Plus two graduate students
  - Mix of faculty, postdocs/students, and national lab scientists
- **Team composition by WBS activity**
  - Production: 70%; Results analysis: 10%; Code/algorithm development: 5%; Maintenance: 5%; Problem setup: 10%

Codes: Predictive simulations in strongly correlated electron systems and functional nanostructures

- **QMC/DCA** (Dynamical cluster approximation) 5 yr old
  30,000 lines of Fortran 90
  Libraries: MPI, BLAS, LAPACK, SPRNG
  Scaling: Demonstrated O(1000) expected O(10,000)
  Efficiency: almost perfect parallel speedup
- **LSMS** (Large Scale Multiple scattering) 15 yr old
  82,000 lines of Fortran 90 plus C/C++ for I/O
  Libraries: MPI, BLAS, LAPACK, HDF5
  Scaling: Demonstrated O(10,000)
  Efficiency: >90%
- **SPF** (Spin Phonon Fermion) 2 yr old
  13,500 lines of C/C++
  Libraries: MPI, BLAS, LAPACK
  Scaling: Demonstrated O(1000) expected O(10,000)
  Efficiency: N/A (code still under development)

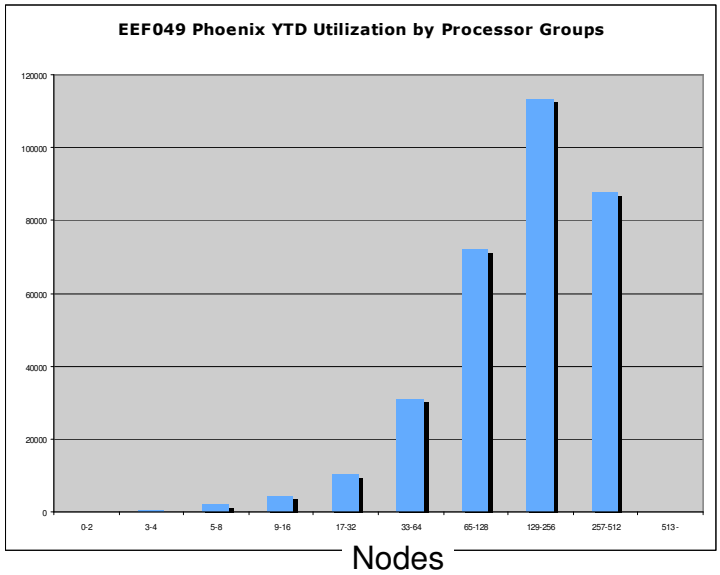
Predictive simulations in strongly correlated electron systems and functional nanostructures

NCCS Resource Usage (Cray XT3)



Predictive simulations in strongly correlated electron systems and functional nanostructures

NCCS Resource Usage (Cray X1E)



Project Highlights
Prediction of giant tunneling magnetoresistance

Magnetoresistance applications today:

- Recording head in computer hard discs
- Magnetic random access memory

Typical TMR for amorphous aluminum oxide barrier:

- 1995: ~10% (discovery @ MIT)
- 2005: ~70% (after a big experimental effort)
- **LSMS is the only code presently capable of performing the fully relativistic all electron LSDA calculations for non-collinear magnetic systems with several thousand atoms**

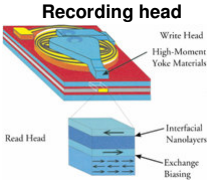
Computational prediction: TMR of 1000% is possible for crystalline MgO barrier, if interfaces are good enough

- **Butler, Zhang, Schulthess, and MacLaren (ORNL), *Phys. Rev. B* (2001)**

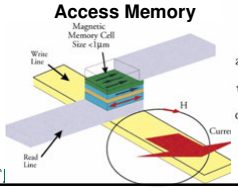
By 2004, MgO-based heterostructures with >300% TMR **discovered experimentally**

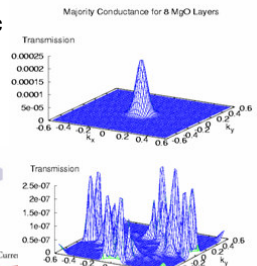
- **Parkin *et al.*, *Nature Materials* (2004)**
- **Yasa et al., *Nature Materials* (2004)**

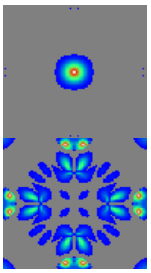
Recording head



Magnetic Random Access Memory







Project Highlights
Modeling high-T_c Superconductors

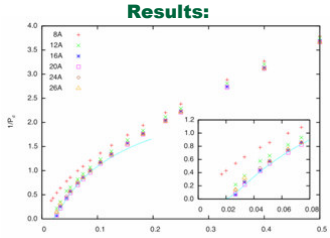
2D Hubbard model for cuprates:

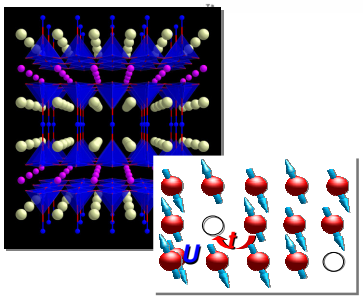
- **Most studied model in this field**
- **No known solution, these simulations are first known results**

QMC/DCA algorithm/code

- **Treats strong non-local correlations in a cluster using quantum Monte Carlo (QMC)**
- **Embedded in an effective medium – dynamical cluster approximation (DCA)**

Results:



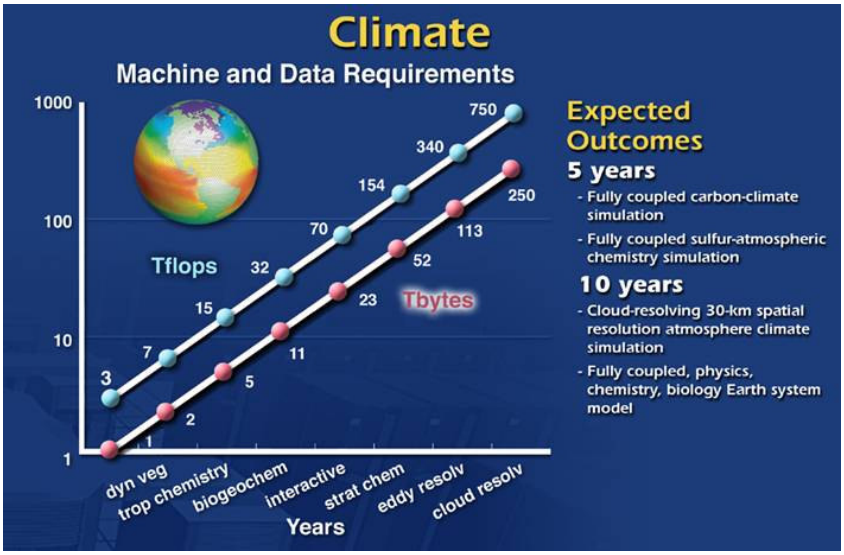


References

Maier TA, Jarrell MS, Scalapino DJ PHYSICAL REVIEW LETTERS 96 (4): Art. No. 047005 FEB 3 2006

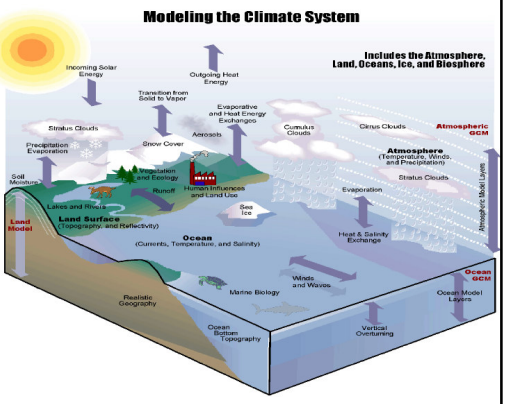
Maier TA, Jarrell M, Schulthess TC, et al. PHYSICAL REVIEW LETTERS 95 (23): Art. No. 237001 DEC 2 2005

Showcase Climate



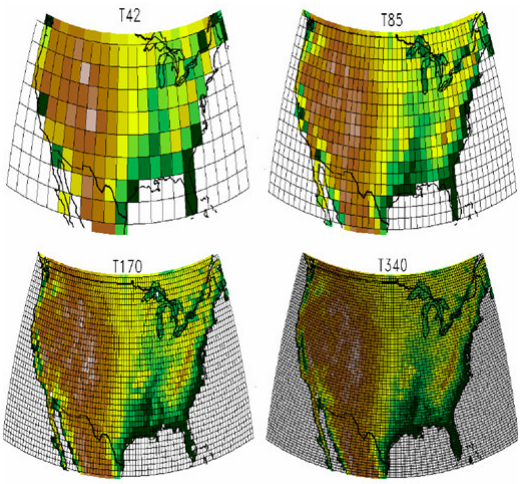
Climate End Station Goal: Understand and Predict the Earth's Climate System

- **Simulate the dynamic ecological and chemical evolution the climate system.**
  - Biogeochemical feedbacks in the global climate system,
  - Document, understand and correct the "biases" or systematic errors
  - Understand internal variability and abrupt transitions of the climate system
  - Focus on processes having an impact on the global carbon cycle.
- **Deliver a next-generation climate model in three years.**
  - Integrate Biogeochemistry, Dynamic Vegetation, Atm Chemistry, New Dynamical Core
  - Input emissions of carbon dioxide and other greenhouse gases
- **Develop, and support the CCSM for use in climate simulation experiments.**
  - Capability tools & simulation frameworks to advance climate-change science
  - High-priority simulations that require NLCF high-end modeling capability
  - Outreach through simulations, analysis of model results and workshops.

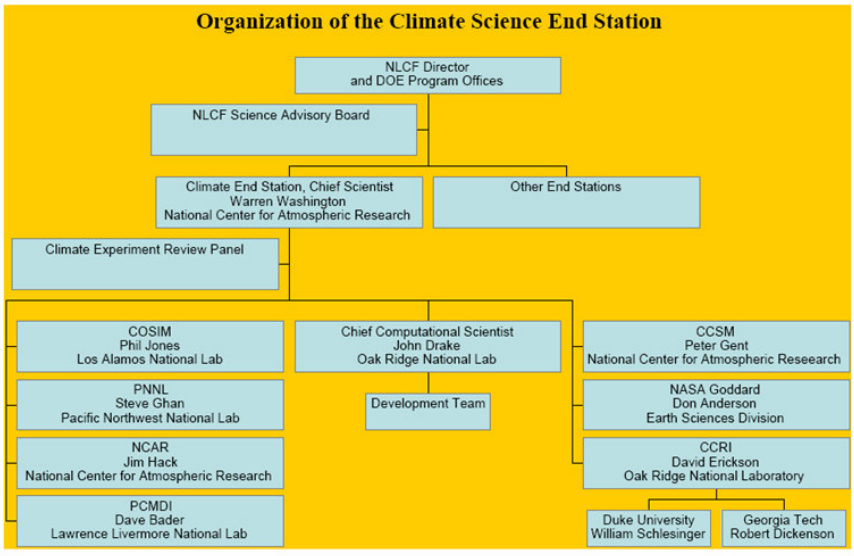


Climate-Science CES Development & Grand Challenge Team
Leadership Computing Enables Aggressive Milestones

- **FY06 Milestones**
  - **High resolution ocean and sea ice , POP2 and CICE**
  - **High resolution atmosphere model bias studies,**
  - **Biogeochemical intercomparison simulations from C4MIP**
  - **Climate Change scenarios stabilization with CCSM3.0 at T85**



Climate-Science CES Development & Grand Challenge Team



Climate-Science CES Development & Grand Challenge Team
Climate Project CLI017

- **Stakeholders: PI and Clients (pays for product development)**
  - PI: Warren Washington, NCAR (wmw@ucar.edu)
  - Clients: DOE SC/BER (Anjuli Bamzai), DOE SC/SciDAC (Michael Strayer)
- **Code development support**
  - DOE support: 95% (BER)
  - Other support: 5% (NSF)
  - Value of computer time: ~$8.12M
- **Technical goal**
  - Predict future climates based on scenarios of anthropogenic emissions (derived from human activities) and other changes resulting from options in energy policies
- **Grand challenge problems**
  - Predictive simulation of biogeochemical (carbon and chemical) cycles in the Earth's system
  - Predictive simulation of global as well as regional aspects of the physical climate system
  - Predictive simulation of the atmosphere-land and ocean-ice system

Climate-Science CES Development & Grand Challenge Team
Climate Project CLI017

- **Expected impact of project success**
  - Understand if and how human activities might alter the climate in major and irreversible ways
  - Influence energy policy and associated R&D directions due to simulated attribution of climate change to different emission scenarios
  - Influence geopolitical relations & regulation because of simulated ecological & air quality impacts on the century timescale
  - Improve ability to accurately predict climate on regional scales
  - Improve ability to simulate biogeochemical cycles in the Earth's system

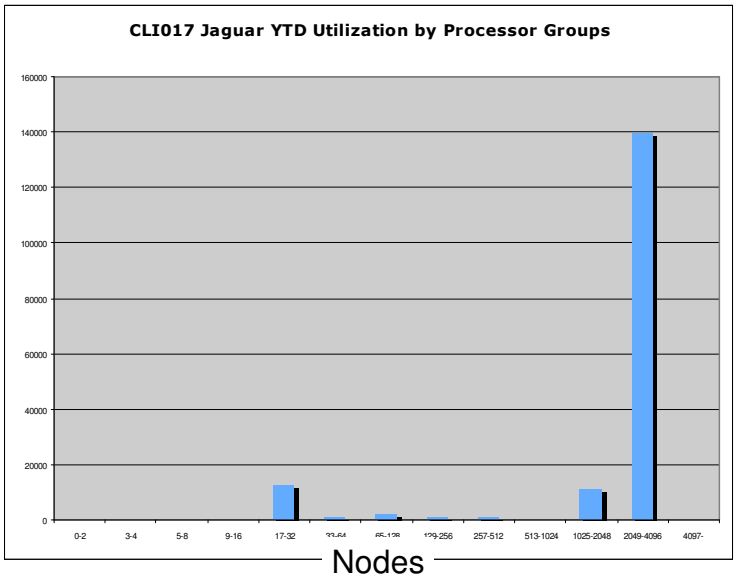
Climate-Science CES Development & Grand Challenge Team
Project Team Resources

- **Team size: >40**
- **Team institutional affiliations**
  - NCAR, ORNL, LANL, LLNL, LBNL, PNNL, ANL, Georgia Tech, Duke, NASA, NOAA
- **Team computer center institution affiliation**
  - ORNL NCCS liaison on team (SciDAC project member)
  - Roughly ¼ of the team members are affiliated with ORNL
- **Team composition and experience**
  - Domain scientists: 12
  - Computational/computer scientists & applied mathematicians: 11
  - Programmers: 12
  - Other: 9
- **Team composition by educational level**
  - Ph.D.: 25 (current)
  - Mix of Jr/Sr faculty, national lab scientists
- **Team composition by WBS activity**
  - Production: 10%; Results analysis: 15%; Code/algorithm development: 70%; Maintenance: 5%

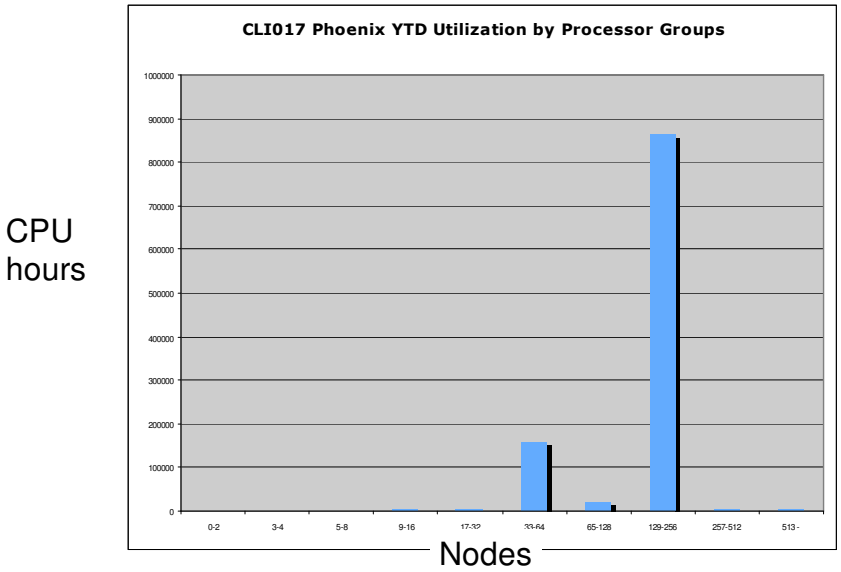
Climate-Science CES Development & Grand Challenge Team
Project Resources: NCCS Input

- **Software provided by NCCS**
  - MPI and NetCDF libraries; NCL, NCO, Ferret, CDAT, and IDL for data analysis; Subversion for version control; Totalview debugger.
- **Size of typical NCCS jobs (concurrency, memory, local & archival storage)**
  - CCSM job: 220 X1E PEs, 0.4 TB memory, 1 TB disk, 5 TB tertiary storage
  - POP job: 1200 XT3 PEs, 2.4 TB memory, 5 TB disk, 10 TB tertiary storage
- **What is the scalability of these codes**
  - CCSM currently scales to 500 PEs for production runs
  - POP (ocean component) scales to 10K PEs for high-resolution stand-alone runs
- **What is the wall-clock time for typical runs?**
  - 10-30 days, in job increments of 12-24 hours
- **Steady state (production) monthly resource use**
  - Processor number: 1200 on Cray XT3, 500 on Cray X1E
  - Processor time: 500K processor-hours
  - Local storage: 1-5 TB of work space
  - Archival storage: 5 TB

Climate-Science CES Development & Grand Challenge Team
NCCS Resource Usage (Cray XT3)



Climate-Science CES Development & Grand Challenge Team
NCCS Resource Usage (Cray X1E)



Climate-Science CES Development & Grand Challenge Team
Project Codes

- **Problem type**
  - CCSM is a fully-coupled, global climate model that provides state-of-the-art computer simulations of the Earth's past, present, and future climate states
- **Algorithm types and computational mathematics**
  - Semi-implicit finite difference, semi-Lagrangian finite volume, and Eulerian spectral
- **Platforms used for routine execution**
  - Cray X1E, Cray XT3, IBM Power clusters, SGI Altix, Earth Simulator, Opteron Linux clusters
  - Preferred: Currently Cray X1E, moving to Cray XT3
- **Code statistics**
  - Size (LOC): >700,000
  - Age: Initial release of the coupled model in 1996 – some components date back to 1982
  - Annual growth: ~50,000 LOC/year.
- **Computer languages employed**
  - LOC: 690,000 Fortran main; 16,700 C utilities; 25,000 C-shell build & run scripts; 32,600 TeX docs; 30,000 text read-me files; 13,700 HTML docs; 7400 "make" scripts; 1300 Perl build scripts
- **Libraries**
  - Libraries used: MPI, NetCDF, MCT, ESMF timers, MPH, PILGRIM.
  - Library extent: MCT, MPH, PILGRIM, and ESMF timers are maintained - represent 90,000 LOC
- **Code Mix**
  - The team develops the CCSM code
  - The Model Coupling Toolkit (MCT) and the Multi-Program-Components Handshaking (MPH) utilities are general-purpose libraries developed as part of the CCSM project.
  - NCO, CDAT, Ferret, and IDL are supplied by Center for data analysis

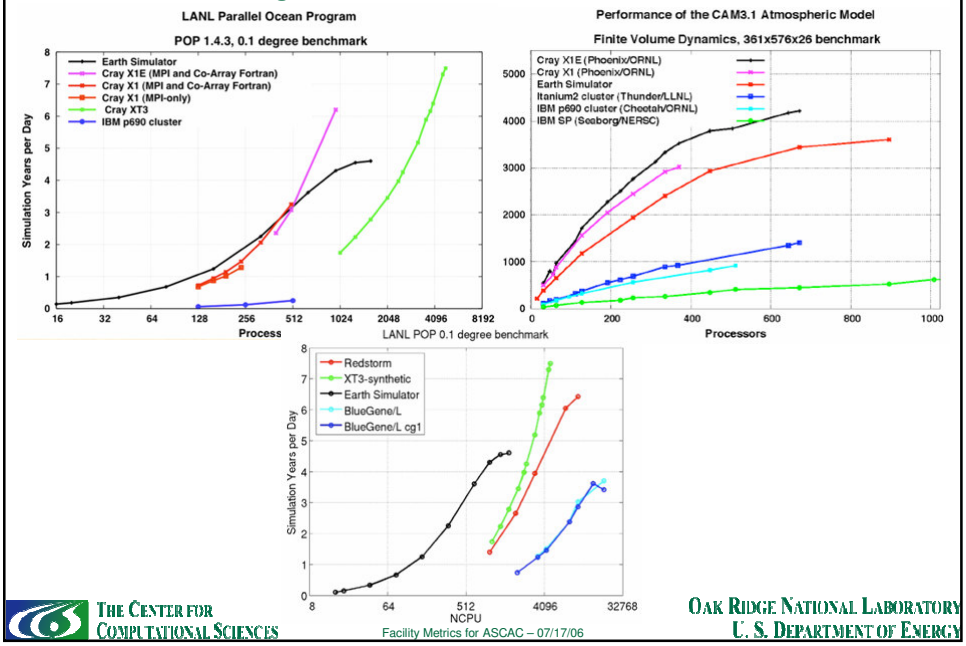
Climate-Science CES Development & Grand Challenge Team
Project Codes

- **Present parallel code scalability on all relevant platforms**
  - At current resolutions, CCSM scales to hundreds of MPI tasks. Scalability has hard limits from the data-distribution algorithm, not from parallel inefficiency
  - Current development will enable scaling to thousands of processors by increasing resolution, adding computational complexity, and implementing more-scalable data distributions
- **Memory/processor ratio (e.g. Gbytes/PE)**
  - 2 GB/processor is adequate
- **Parallelization model**
  - MPI with 2D domain decomposition is the primary mechanism for parallelism
  - OpenMP parallelism is also implemented and used on systems for which it is appropriate.
- **Code "efficiency"**
  - The metric of interest is simulated years of the Earth's climate per real-time day (years/day)
  - Scientists require >5 years/day of throughput for adequate scientific progress
  - **As computers grow larger and more capable, science runs grow in complexity and fidelity up to 5 years/day limit**

Climate-Science CES Development & Grand Challenge Team
Project Codes

- **Major scaling bottlenecks**
  - Science is pushing CCSM to grow more in computational complexity than in resolution, so parallelism in terms of grid points is limited, and computational cost per grid point is growing
  - Since processor speeds are likely to increase less than in the recent past, more parallelism must be identified. Opportunities exist for more-distributed algorithms and task parallelism.
- **Split between interactive and batch use**
  - A single climate simulation runs for weeks, generating history output that is made publicly available and analyzed over years by scientists across the world. The computation phase of this workflow is strictly batch. Software development often requires quick turnaround for debugging, similar to interactive use
- **Split between code development on the computer center computers and on computers at other institutions**
  - Center computers are primary targets for code development
  - Initial development may be performed on workstations or workgroup clusters before integration and testing on the center computers.

Climate-Science CES Development & Grand Challenge Team
POP and CAM Scaling



Climate-Science CES Development & Grand Challenge Team
Code Productivity & Scalability

- **Productivity needs and measures**
  - Developer resources, given the ambitious development goals of CCSM
  - Growing need to commit limited developer resources to code scalability
- **Performance needs and measures**
  - Adequate turn-around (5 years/day) for simulations with dramatically increased computational expense
  - First-ever simulations of the full Earth system at scientifically relevant resolutions, allowing input of real-world emissions instead of prescribed atmospheric concentrations.
  - Facility increases have allowed dramatic increases in model complexity and fidelity while maintaining traditional rates of throughput
- **History of scaling and projected scalability**
  - Beginning an ambitious development phase to enhance the capability of the model, making it a true Earth-system model, in preparation for the next IPCC report.
  - Dramatically increasing the # of PEs used for a single run, to allow the much-more-expensive Earth-system model at adequate resolution and adequate throughput.

Climate-Science CES Development & Grand Challenge Team
Software Engineering, Development, V&V

- **Software development tools**
  - Parallel development: Cray PAT performance tools
  - Debuggers: TotalView
  - Visualization: NCL, NCO, Ferret, IDL, CDAT
  - Production management and steering: shell scripts, Wiki web service
- **Software engineering practices**
  - Configuration management: SVN
  - Quality control: Requirements documents, coding standards, standardized test suites
  - Bug reporting and tracking: Changelogs, Wiki web tool
  - Code reviews: Change review board, Software Engineering Working Group
  - Project planning: Climate Change Working Group, Climate End Station Board of Directors
  - Project scheduling and tracking: MS Project, web pages
- **Verification strategy**
  - Unit testing (?), error-growth tests (chaotic system), standardized regression tests
- **Regression test use**
  - Used before any library or compiler change
- **Validation strategy?**
  - No version of the model is used for science before a 200+ year validation run to confirm that it produces realistic climate
  - Analyze features of the simulated climate that represent yearly and decadal patterns in the observed climate
- **What experimental facilities do you use for validation?**
  - Historic climate data from the ARM program, ground- and sea-based weather measurements, and satellite data.
- **Does your project have adequate resources for validation?**
  - Limited availability of qualified experts
  - Increased priority of the project and introduction of a more-hierarchical validation process has mitigated the issue.

Climate-Science CES Development & Grand Challenge Team
Science Output – The Model and the Data

- **Basic science output: model and data**
  - The CCSM model provides state-of-the-science simulation of the Earth's climate and is freely available to scientists
  - Output from control runs and century-scale future-climate runs under a variety of emission scenarios are made available to scientists through the Earth Systems Grid.
- **Representative recent accomplishments (Cray X1E)**
  - Production CCSM runs in IPCC configuration
    - Four ensemble runs with natural CO2 forcing completed in May
    - Four new ensemble runs started for anthropogenic forcing, each using 248 processors
  - First results of C-LAMP
    - Carbon LAnd Model intercomparison Project
    - Results from equilibrium runs of CASA' and CN carbon-cycle models
  - First-ever control runs of CCSM with finite-volume dynamical core
    - FV dycore critical for chemistry and full carbon cycle
    - Completed first 300-year run
    - Started new run with science refinements to ocean viscosity
- **Representative recent accomplishments (Cray XT3)**
  - Scaled up POP production runs
    - High-resolution ocean simulation, scales to full system
    - "Sweet spot" production runs now using 2400 processors each, up from 1152
    - Two such runs now active
  - Porting of full CCSM
    - Successfully passed initial test suite; Now testing multi-year runs
  - New inter-agency work starting
    - NASA carbon assimilation; NOAA performance assessment

Climate-Science CES Development & Grand Challenge Team
Science Output

- **Publications**
  - ~10/year
- **Living datasets and/or databases accessed by community; size of community**
  - Thousands of years of simulated climate made available through the Earth Systems Grid and used for hundreds of publications in support of the reports of the Intergovernmental Panel on Climate Change (IPCC)
  - Hundreds of scientists worldwide.
- **Change in code capabilities and quality over time**
  - CCSM maintains state-of-the-science capability
  - Current development could make it the first true global Earth system model through the addition of the carbon cycle and fully coupled chemistry.
- **Code and/or data contributed to the centers**
  - CCSM executables are maintained by the project at the Center
- **Code and/or data, results, contributed to the scientific and engineering community**
  - CCSM is freely available in regular public releases
- **Training, education, outreach**
  - Production of scientists & computational scientists during 2001-2005: 4
  - Production of trained software engineers during 2001-2005: 4

What Would You Do With 1 PF for One Month

- **Astrophysics (VH1/RADHYD)**
  - High resolution MHD simulations in general relativistic gravity to explore neutron star spin-up and natal kicks in detail for a variety of progenitor masses. Ray-by-ray MGFLD RHD simulations using RadHyd, allowing detailed exploration of the effects of progenitor asymmetries
- **Nanoscience (LSMS/VASP)**
  - Current electronic structure simulations focus on individual configurations (magnetic, structural, molecular), good enough for bulk systems (long length scales in the thermodynamic limit) but not in nanoscience where temperature fluctuations are important and entropic effects have to be considered explicitly by calculating the free energy at finite temperature.
- **Combustion (S3D)**
  - Cleaner and more efficient combustion. Turbulence and chemical mechanisms for multi-stage ignition of n-heptane at 10-20 atmosphere pressure

What Would You Do With 1 PF for One Month

- **Materials Science (DCA)**
  - Hubbard is simplified model – make model more realistic: include more DOFs, more electronic orbitals. Hubbard cannot distinguish between different HT materials. Want to drive up transition temperatures (between 40K to 150K), but Hubbard model gives one transition temperature. In real materials, (e.g., Hg compounds) Hubbard model needs to be more realistic. Taken into account. Large chain good for more measurements, reduces your statistical error.
- **Fusion (GTC)**
  - Size and isotope scaling studies of core turbulence transport for ITER. Goal of ultimate integrated simulation combining wave heating, turbulence, MHD, and neoclassical physics
- **Climate (POP)**
  - Details of the north Atlantic circulation, critical to the stability of the polar ice cap, are currently not modeled accurately, with questions about how the formulation of the model (isopicnal or height) might change the frequency and strength of warm water incursions under the arctic sea ice. A series of POP/CICE high resolution (1/20 degree) simulations would determine how soon the cap is likely to disappear

31 May 2006

6

To: Dr. Gordon Bell

Subject: Response to information request by the ASCAC sub-panel on Computing Facilities Measurement (CFM)

From: Bill Kramer (NERSC), Francesca Verdier (NERSC)

Dear Gordon

Please consider this and the attached spreadsheet the NERSC specific response to your request for input regarding appropriate metrics for the OMB and Office of Science to use for "performance measurement and assessment at [Office of Science Computational] facilities, the appropriateness and comprehensiveness of the measures, and the science accomplishments and their effects on SC's science programs….[T]he sub-panel is asked to provide input for the Office of Management and Budget (OMB), evaluation of ASCR progress towards the long-term goals specified in the OMB Program Assessment Rating Tool (PART)" .

The first part of the response is covered in a separate memo prepared jointly by the NERSC, ORNL and Argonne facilities.

Part 2: Detailed data response

1. Facility overview "balance sheet"

a. Organizational structure with staff sizes and functional titles (single page)

http://www.nersc.gov/about/org.php

b. Contacts or URL to key staff contacts

For the purpose of this document, Bill Kramer, the NERSC General Manager is the contact. His information is 510-486-7577 and kramer@nersc.gov. General contact information can be found at

http://www.nersc.gov/about/contact.php

c. Physical infrastructure (building size, power – amount & cost $Mwhr, cooling capability, network access, etc.)

NERSC is housed at the LBNL Oakland Scientific Facility in downtown Oakland CA (approximately 4.5 miles from LBNL proper). OSF supports multiple activities including the HEP/NP funded PDSF, ESnet and some laboratory systems. The OSF has approximate 18,000 sf of3' raised computer floor and a 1,000 sf operations area. The facility has a maximum feed of 12 MW of power. It has two 10 Gbps links to the Bay Area Metropolitan Area Network (MAN). One link is the production connection to ESnet and the other link is used for other purposes.

The maintenance and operation of OSF are integrated with LBNL, so NERSC pays the same charges (space, electrical, etc.) as it would if OSF were on the lab proper . NERSC does not pay direct costs.

d. Balance sheet and budget: hardware, maintenance, staff, software, utilities, buildings, institutional overhead, etc.

http://www.nersc.gov/news/reports/LBNL-57582.pdf

The budget planfro FY 06 is for $38M per year, which is compatible with DOE Office of Science plans. Before 2005 NERSC's budget average was between $28–29M per year, with some additional investment above the original plan such as $3.5M for ~2 TB additional memory on NERSC-3. The investment strategy for 2006–2010 is very consistent with the past five years, as shown in Table 4.

Table 4
Investment Strategy

Computational systems	*~35%*
Staff costs	*~24%*
Infrastructure	*~16%*
NERSC balance investments	*~7%*
Overhead/Lab costs	*~18%*

e. Institutional affiliation and degree of institutional support

NERSC is operated under the general contract for Lawrence Berkeley National Laboratory. LBNL is part of the University of California. The NERSC program is fully funded by DOE..

f. Present and planned computers, storage, etc. their properties & utilization e.g. in use petabytes versus potentially available

Please see our 2006-2010 5 year plan at http://www.nersc.gov/news/reports/LBNL-57582.pdf

This 2006-2010 5 year plan was peer reviewed. NERSC does all major computational acquisitions using the openly competitive Best Value Source Selection process so we plan for specific properties or architectures.

g. Software development and production tools provided

Please see http://www.nersc.gov/nusers/resources/software/

h. Application codes available to the users that are supported by the center (ISVs, open source, etc.)

Please see http://www.nersc.gov/nusers/resources/software/apps/

i. What auxiliary services do you offer your users (visualization, data storage and retrieval, consulting)?

NERSC provides services in the areas of consulting, applications software, web documentation, training, account and allocations management, collaborative scientific team support, system and network monitoring and support, security, outreach, analytics and visualization. For further detail, please see pages 36 to 53 of our 2006-2010 5 year plan at http://www.nersc.gov/news/reports/LBNL-57582.pdf

j. How many FTE's are involved in consulting and support of all support? What is the ratio or number of consultants to projects and/or experimenters?

*In FY 06, NERSC is authorized approximately 61 technical FTEs. Of these approximately 55 FTE are involved in direct support of systems and users. NERSC has approximately **2,500 users and between 350 and 400 projects every year.** Consulting has about 8 FTEs (including one PDSF consultant) and Analytics is an additional 4.5 FTE.*

2. User interface and communication including satisfaction monitoring and metrics

a. How do you measure the success of your facility today in being able to deliver service beyond the user surveys (e.g. the NERSC website)?

This consists of achieving the metrics in our 5 year plan, user satisfaction (survey, NERSC User Group, etc.) and other things. See the Part 1 response regarding Facility Metrics for this information.

b. Do all users-experimenter teams, team members, and any users that the team community provides utilize the survey?

We are not sure we understand what this mean. The number of respondents to each survey is provided in that survey description.

c. Have these surveys been effective at measuring and understanding making changes in operations? (Please cite)

Yes, in the Response Summary to every annual survey is a section called "Survey Results Lead to Changes at NERSC" which lists improvements made based on last year's survey. Such improvements include reorganizing the NERSC website, improving the relationship with IBM's compiler support group to improve compiler bug resolution time, developing the remote license server, and implementing queue priority scheduling changes. NERSC also talks about the survey response activities with the NERSC User Group, which meets months via teleconference and semi-annually face to face.

Please see the individual survey results http://www.nersc.gov/news/survey/ for further detail.

d. Describe your call center – user support function: hours of coverage, online documentation, trouble report tracking, trouble report distribution, informing the users, how do users get information regarding where their job/trouble report is in the queue?

NERSC User Services provides with live consulting and scientific support from 8 a.m. until 5 p.m. (Pacific time) Monday through Friday. Users can contact NERSC via telephone, email or web submissions. From 8 to 5, there are several consultants on phone coverage. NERSC staff respond to the users within 4 work hours. The NERSC Computer Operations and Network Support provide basic help desk user support around the clock, including password change requests and management of system problems

Online documentation is at: http://www.nersc.gov/nusers/

User trouble report tracking is done using the RightNow Web product, which includes incidence escalation procedures so that we can track our user assistance metrics. Users can track their trouble tickets on the web.

All users are kept informed of important facility changes via email. In addition, they can subscribe to a status email list to get informed be email of all "down" and "back up" announcements. Archives are kept for past information, for example the Systems Availability Log (http://www.nersc.gov/nusers/status/nstat.php) and the Announcements email archive (http://www.nersc.gov/nusers/announcements/index.php?list=all-announcements). NERSC also maintains a regular meeting schedule with the NERSC User Group – with a monthly teleconference and semi-annual meetings. Both long term and short term issues are discussed with the NUG.

e. What mechanisms are provided for the user with respect to dissatisfaction with how a case is being handled?

Unresolved trouble tickets are escalated first to the consultant assigned to manage the ticket to resolution, then to management. If a user is unsatisfied with how their case was handled, they will typically send email to management, who review the case and respond. In addition, NERSC staff review trouble tickets for patterns and closure.

f. What mechanisms are provided to support event-driven immediate access to your facility (e.g. Katrina or flu pandemic)

NERSC can provide event-driven immediate access through rapid processing of a new allocation request (within several hours of completing the request), with queue priority mechanisms that allow "boosted" jobs to start with very short delays, with proactive contacts from the consultants to rapidly get "special" users up to speed in using the facility. Other special services, depending on need, include customized access to the mass storage system, special tape handling, network tuning, and whatever else it takes to meet the special needs.

2a. Qualitative output measures and metrics

Do you measure how your facility enables scientific discovery?

For the past several years, NERSC has documented a list of peer reviewed publications that result at least in part from work done at NERSC. For details, please see http://www.nersc.gov/news/reports/

How are these disseminated and how do they further Science and especially DOE Science Programs?

These publications get disseminated via the usual scientific process. DOE will have to answer the question about how they further their science programs.

a. What impact have any of your measures had on operation of your facility?

For an example, see the answer to 2.c, above.

b. What impact have the current PART measures had on your successful operations of your facility?

- *Acquisitions should be no more than 10% more than planned cost and schedule. This is defined as a Development, Modernization and Enhancement (DME) project. It is likely use to OMB and with the current definitions it is reasonable from the center's viewpoint.*
- *40% of the computational time is used by jobs with a concurrency of 1/8 of more of the maximum usable compute CPUs. For NERSC this is 760 CPUs on the IBM SP-3.*

  This metric has positive and negative effects. NERSC has consistency meet this metric, with 70% of the time being used by 512 way jobs and more than 45% by jobs 760 and larger.

  On the positive side, it motivated a major increase in scalability by many science projects and demonstrated that significant increases in scalability are possible. Now, many of the projects that ran at scale at NERSC are qualified to run at the NLCF. Thus, the metric motivated a change in user behavior in a direction the Office of Science wants and needs.

  On the negative side, this metric has nothing to do with the quality of science of a project, and some very important projects have very valid reasons that large scale jobs are not appropriate. Also, in order to encourage this, the small long running jobs of the NERSC workload have experiences significantly longer queue times.
- *Every year several science applications are expected to increase efficiency by at least 50%. While not directly a NERSC metric, NERSC staff have provided significant help to the identified projects.*

  This metric was motivated by the desire to increase the percent of peak performance applications have. It probably is no longer as important.

c. What do you view as the appropriate measures for supercomputing facilities now and during the next 3-5 years?

See the Facility Metrics discussion in Part 1 of the response in a separate memo.

3. Aggregate Projects use profiles by scale

What is your usage profile in terms of processor count? We would like these broken down into jobs that require, or can exploit a concurrency level of (roughly) 50, 200, 400, 1,000, 2,000, and 4,000 processors to obtain the science.

Percent of total job time by concurrency level, 11/04 through 4/06

(Note: Bassi data start 11/10/05 and Jacquard data starts 06/08/05.)

Concurrency	***Seaborg***	***Bassi***	***Jacquard***
1-48	8.36%	19.20%	65.71%
49-208	18.99%	34.97%	27.60%
209-400	6.96%	34.42%	4.60%
401- 1,008	30.03%	11.39%	1.96%
1,009-2,000	23.43%	-	-
2,001-4,000	9.97%	-	-
> 4,000	2.23%	-	-

a. Aggregate required memory per job? (Or memory per node)

Not available.

b. Processor distribution?

At NERSC, currency is always mapped to CPUs, so see the table above.

c. Disk space use?

Not available.

d. Tertiary tape use?

NERSC uses Storage Resource Units (SRUs) to measure tertiary storage use. SRUs are computed as a weighted sum of space used (highest weight for most projects), I/O (can be the highest weight for some projects) and number of files (low weight):

yearly user SRUs = 0.01436 x Avg files + 4.787 x Avg space (GB) + 4.0 x I/O (GB)

User SRUs are by default charged to projects in proportion to the user's allocation in each project. The user can change this formula if the defaults don't match with real use.

http://www.nersc.gov/nusers/resources/hpss/hpss-charging.php

Allocation Year 2005 Project SRU use distribution

AY2005 SRUs	*Num Projects*	*Percent of total use*
500K – 2M	5	64.3%
100K – 500K	10	19.9%
25K – 100K	21	9.8%
5K – 25K	34	4.4%
1K – 5K	48	1.3%
< 1K	168	0.2%

e. Average wall clock time of jobs?

Average wall clock for regular priority jobs that ran for more than 35 minutes, 11/04 through 4/06

(Note: Bassi data start 11/10/05 and Jacquard data starts 06/08/05; their wall clocks are for all jobs)

Concurrency	*Seaborg*	*Concurrency*	*Bassi*	*Concurrency*	*Jacquard*
1-112	05:32:05	1-56	01:16:48	1-14	01:33:32
113-240	07:56:59	57-120	01:36:46	15-30	02:25:49
241-496	06:42:10	121-248	02:23:46	31-62	03:23:21
497-1,008	12:20:21	249-504	02:32:56	63-126	00:45:17
1,009-2,032	13:30:27	505+	05:11:16	127-254	01:11:26
2,033+	10:28:55			255+	00:49:54

f. Average time of jobs in the queue?

Average wait time for regular priority jobs, 11/04 through 4/06

(Note: Bassi data start 11/10/05 and Jacquard data starts 06/08/05.)

Concurrency	*Seaborg*	*Concurrency*	*Bassi*	*Concurrency*	*Jacquard*
1-112	09:20:39	1-56	01:27:10	1-14	02:33:06
113-240	35:01:13	57-120	04:55:29	15-30	04:12:27
241-496	32:49:23	121-248	08:04:02	31-62	04:55:46
497-1,008	46:49:56	249-504	20:18:29	63-126	04:39:07
1,009-2,032	65:57:23	505+	09:41:11	127-254	06:08:50
2,033+	79:16:59			255+	29:20:53

g. How do you measure project code performance on your machines?

NERSC provides IPM and other tools for users to do the measures, and DOE requires IPM or other performance data with the project proposals, but we do not regularly monitor user codes. Users can use IPM whenever they wish with very little overhead.

h. Amount of project consulting support utilized?

NERSC will provide a summary total calls with a break down of type – but not by project.

4. Top 20 Project profiles usage

NERSC will provide what is possible based on the project proposals submitted. See the appended excel spreadsheet. We will also offer the committee access to our proposal data base for them to review the proposals directly

a. Project name

Spreadsheet column: *"Project name"*

b. Contact information?

Spreadsheet column: *"PI email"*

c. Brief description of size and shape of project team and the projects user community

Spreadsheet column: *"2006 NERSC team members"*
We do not have any other project team information.

d. Briefly describe characterize the size, shape, and age of the codes

NERSC does not have this information.

e. Computing resources utilized by the teams: machines, disk, tertiary

- ***Spreadsheet column:*** *"2006 Project Machine use"*
- ***Spreadsheet column:*** *"TB stored in HPSS May 2006"*
- *Disk usage is not available at the project level (only at the user level).*

f. Software provided by center

Spreadsheet column: *"2006 NERSC application software / tools used"*
Spreadsheet columns: *"Code[1,2,3] libraries"*

g. Consulting and direct team support by your center

This information is not in the proposal; NERSC cannot provide.

h. What is the size of their jobs in terms of memory, concurrency (processors), disk, and tertiary store?

a. Memory – not available
b. Concurrency - ***Spreadsheet column:*** *"2006 Project scaling on most heavily used machine"- shows the most frequently used processor counts*
c. Disk usage is not available at the project level (only at the user level).
d. Tertiary Storage - ***Spreadsheet column:*** *"TB stored in HPSS May 2006"*

i. What is the scalability of these codes

Spreadsheet column: *"2006 Project scaling on most heavily used machine"- shows the most frequently used processor counts*

j. What is the wall-clock time for typical runs?

Spreadsheet column: *"Typical wall time (hours)"*

5. (Center x User) Readiness for 10x processors expansion

The mid-term goals for each facility call for a major expansion from machines with of order 5,000 processors to machines of order 50,000 processors or more.

a. Please outline how the center will accommodate this growth over the next 3-5 years.

There are several sources for this information. DOE has studies such as the Scales I and II reports that document future computational needs. Speciifc to NERSC, the NERSC User Group's Greenbook (http://www.nersc.gov/news/greenbook/2005greenbook.pdf) lays out computational and scientific goals for the next several years. Some of this information is summarized on pages 21-32 and 56-57 of the NERSC 2006-2010 5 year plan at http://www.nersc.gov/news/reports/LBNL-57582.pdf

b. What do you believe is your role in preparing users for this major change?

Make wise choices in technology acquisition, assure that systems and software are stable, provide consulting help to users, provide scaling incentive programs, provide training for new systems.

c. What effort (in terms of personnel) is devoted to code development issues today, and do you view this as adequate coverage as we move to machines with more than 25,000 processors?

NERSC FTEs for scientific code development has been essentially eliminated, the SciDAC program having partially picked up this role. NERSC believes it is extremely beneficial for HPC centers to be actively engaged in code development and will re-engage in this effort if sufficient funding can be provided. The consultants assist with algorithmic improvements for a small number of projects, but this amounts to less than 1 FTE.

The response to the second part of the question depends on how SciDAC-II is implemented and how well targets machines of the scale. It is likely, without increased motivation (see the computational science goals of the Part 1 response) the computational science community will have difficulty fully utilization 25,000 processors at high levels of concurrency.

d. Are there codes in your user portfolio that will scale today to 10,000, 25,000, or 75,000 processors. What is the nature of these codes (Monte Carlo, CFD, hydro?) Are these codes running today on other systems of comparable size?

Application performance in the absence of an architectural context is hard to predict. Monte Carlo will in general scale very well given that, except in the case of dynamic load balance, there is little room for cross CPU contention. Problems with communication

topologies that map naturally to the switch topology can come close to Monte Carlo type scalability, e.g. Molecular Dynamics codes, which in 3D map perfectly to the BlueGene torus

We do know that several NERSC codes currently in use will scale to 10,000 and possibly more processor. The first four of these codes are benchmarks used in the evaluation and selection of NERSC-5, and are somewhat representative of they respective science areas:

a. *GTC will scale to 10,000 processors, partly due to its weak scaling needs*
b. *PMEND will scale to 10,000 processors given the right system.*
c. *MILC will scale to 10,000 processors if the allreduce is fast*
d. *MADCAP might scale to 10,000 processors if the system I/O can support its requirements*
e. *LBMHD will scale to 10,000 and maybe 25,000 processors, except that the memory and wall time requirements for grid sizes that would use that many processors may be prohibitive.*

e. As machines become more complicated, what do you see as the challenges to your success? For example, are you (or parts of your institution) actively involved in research related to fault-tolerance, memory/bandwidth contention, job scheduling, and etc. on the future machines?

Yes , refer to our 5 year plan. And yes, LBNL and UCB researchers are engaged in all the areas.

f. How do you determine the path forward for your organization?

Please see our 2006-2010 5 year plan at
http://www.nersc.gov/news/reports/LBNL-57582.pdf

g. What do your users want to see in the largest machines now available and those which will be available in the 3 year and 5-7 year time frames? (memory per core/node, number of processors, disk space?)

Please see the NERSC User Group Greenbook at
http://www.nersc.gov/news/greenbook/2005greenbook.pdf. Past versions of the greenbook can be found at http://www.nersc.gov/about/NUG/ We use a Best Value source selection process and do not specify the architectural details such as memory per core.

Top 20 Team metrics evaluation

We selected the top 20 projects – those which had used the most MPP computational time in allocation years 2005 and 2006 to date (12/1/04-05/10/06). In addition we provide information on the top 5 HPSS projects (one of which, mp111, overlaps with the top 20 computational projects).

1. Project (background)

a. Code name and contact information for the project principal investigator (name, institution, mailing address, phone/fax, email, URL for code)

a. ***Spreadsheet columns:*** *"Code[1,2,3] name"*
b. ***Spreadsheet column:*** *"PI name"*
c. ***Spreadsheet column:*** *"PI email"*
d. ***Spreadsheet column:*** *""Major team institutions (lead first)"*

b. DOE Office that supports the team and the name and contact information of the DOE program manager: (breakdown by SC Office funding (BES, BER, NP, HEP, ASCR, FES, other)

Spreadsheet column: *"DOE Office"*

c. Scientific domain (chemistry, fusion, high energy, nuclear, etc.)

Spreadsheet column: *"Science domain"*

d. What are the technical goals of the project? What problem are you trying to solve? What is the impact of your project success? (e.g. better understanding of supernovae explosions, prediction of ITER performance, …)

Spreadsheet column: *"Project goals"*

e. How did you get the resources to develop the code? SciDAC, DOE SC program, internal institutional funding sources (e.g. LDRD,..), industry, other agencies, …

NERSC can not provide this, but we do indicate the project's overall funding source.

Spreadsheet column: *"Project funding"*

f. What is the project profile in human resources including trained scientists, computational scientists, program maintenance, and use(rs) of you codes? (see also output)

NERSC can not provide this.

g. Size of any external communities that your code or datasets support

NERSC can not provide this.

2. Project Team Resources (balance sheet)

a. Team size (small teams of 1-3, medium 4-10, or large 11-20).

Spreadsheet column: *"2006 NERSC team members"*

b. Team institutional affiliation(s). (e.g. all the institutions involved, including universities, national labs, government agencies,..). I.e. to what extent is the team multi-institutional?

Spreadsheet column: *"Major team institutions (lead first)"*
The list was cut off after 10 for the largest projects.

c. To what extent are the code team members affiliated with the computer center institution? (e.g. are the team members also members of the computer center institution?)

   NERSC can not provide this.

d. Team composition and experience by discipline (domain scientists, computer scientists, computational mathematicians, computational scientists, database managers, programmers, etc.)

   NERSC can not provide this.

e. Team composition by educational level (Ph.D., MS, BS, undergraduate students, graduate students, post-docs, younger faculty, senior faculty, national laboratory scientists, industrial scientists, etc.)

   NERSC can not provide this.

f. Team resources utilization: time spent on code and algorithm development, maintenance, problem setup, production, and results analysis

   NERSC can not provide this.

g. Code Mix: To what extent does your team develop and use your own codes? Codes developed by others in the DOE and general scientific community? Application codes provided by the center?

   NERSC can not provide this.

3. Project Code (balance sheet)

Information from the ERCAP allocation requests is provided for the project's top 3 codes.

a. Problem Type (data analysis, data mining, simulation, experimental design, etc.)

   Most of the codes run on the MPP machines at NERSC are simulations. The codes run on the PDSF are a mixture of simulations and data mining/analysis codes. The information provided is NERSC's best guess since it is not collected in ERCAP.

   Spreadsheet columns: *"Code[1,2,3] problem type"*

b. Type of algorithms and computational mathematics (e.g. finite element, finite volume, Monte-Carlo, Krylov methods, adaptive mesh refinement, etc.)

   Spreadsheet columns: *"Code[1,2,3] algorithms"*

c. What platforms does your code run on? What is your preferred platform?

   We provide the NERSC platforms used by the projects but not machines used elsewhere. We assume that percent use indicates which platform is preferred. We provide this information only at the project level, not at the code level.

Spreadsheet column: *"2006 Project Machine use"*

d. Code size (single lines of code, function points, etc.); Code age and level of maturity

NERSC can not provide this.

e. Computer languages employed, LOC/ language, reason for the language choices (e.g. 250,000 SLOC Fortran-main code, 30,000 C++-problem set-up, 30,000 SLOC Python-steering, 10,000 SLOC PERL-run scripts,…)

Languages provided, but not the rest of the information.
Spreadsheet columns: *"Code[1,2,3] languages"*

f. What libraries are used? What fraction of the effort do they represent?

Libraries provided, but not the level of effort.
Spreadsheet columns: *"Code[1,2,3] libraries"*

g. What memory/processor ratio do your problems require? (e.g. Gbytes/processor)

NERSC can not provide this.

h. What is the use of resources on a per use and aggregated basis? Range of aggregate processor time, memory footprint, disk, tape, etc., for typical code runs and aggregate use

NERSC can not provide this.

i. Parallelization model (e.g. MPI, OpenMP, Threads, UPC, Co-Array Fortran, etc.) E.g. Does your team use domain decomposition and if so what tools do you use?

Parallel model provided; no information on domain decomposition.
Spreadsheet columns: *"Code[1,2,3] parallel model"*

j. What is the "efficiency" of the code and how is it measured?

NERSC can not provide this.

k. <u>What is the codes present and projected parallel scalability and how is measured?</u>

We provide the most frequently used processor counts at the project level (not the code level) and only on the machine most heavily used by the project. We also provide the highest processor count used on Seaborg in the last 2 allocation periods. We can not provide projected scalability.

   a. ***Spreadsheet column****: "2006 Project scaling on most heavily used machine"*
   b. ***Spreadsheet column****: "2005/2006 Seaborg largest processor count and its use"*

l. What are the major bottlenecks for scaling your code?

Spreadsheet columns: *"Code[1,2,3] performance limits"*

m. What is the split between interactive and batch use? Why

The most common reasons for interactive use at NERSC are: visualization, code development and testing, parameter space testing. We cannot provide the reasons for individual projects.

Spreadsheet column: *"Ratio of interactive use"*

n. What is the split between code development on the computer center computers and on computers at other institutions.

NERSC can not provide this.

4. Software Engineering, Verification and Validation Code Processes

NERSC can not provide any of this data.

a. Software development tools used (debuggers, visualization, parallel development, production management and steering)
b. Software engineering practices (configuration management, quality control, code review, project planning, project organization, project tracking, schedule estimation, etc.)
c. What is your verification strategy?
d. What use do you make of regression tests?
e. What is your validation strategy?
f. What experimental facilities do you use for validation?
g. Does your project have adequate resources for validation?

5. Project input: facilities recourses utilization (cross-check on facilities)

This is the same as above since NERSC is providing the information.

This cross-checks with the centers output and includes machine time, data and tertiary stores, consulting and support people, software libraries, and all support from a user's perspective

Enumerate all the resources that the project receives from the center.

6. Project output (t) and user metrics

NERSC can not provide this information except for number of publications reported on the ERCAP request form (where the PIs were requested to list no more than 15).

Spreadsheet column: *"Number pubs reported"*

Enumerate project output.
In addition provide:

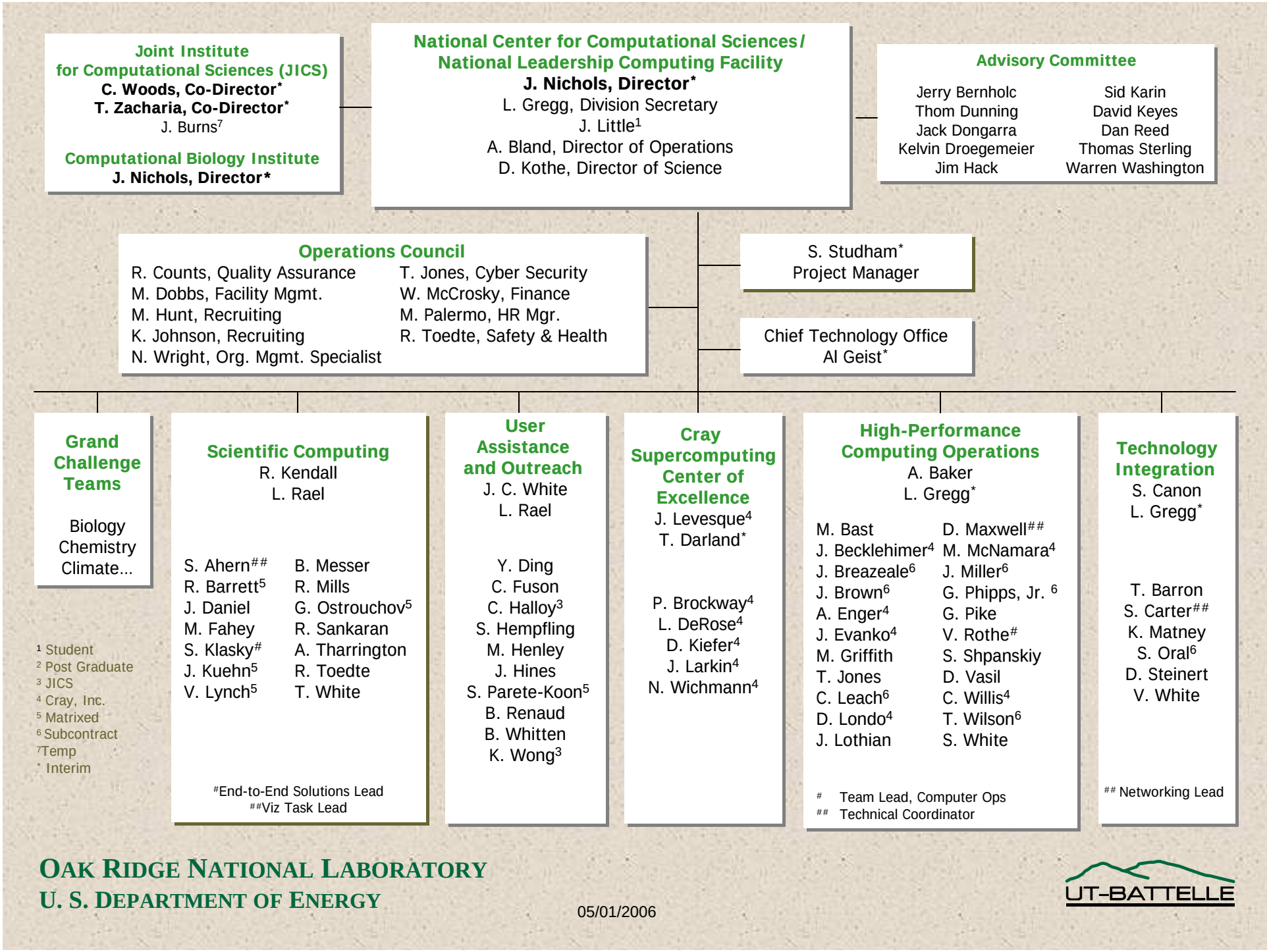
a. Publications? Citations? Dissertations? Prizes?

b. Residual and supported, living datasets and/or databases that are accessed by a community?
   Size, shape, and user community for the datasets
c. Change in code capabilities and quality (t)
d. Code contributed to the centers or to the scientific community at large
e. Company spinoffs based on code or trained people and/or CRADAs
f. Corporation, extra-agency, etc. use
g. Changes in trained scientists, developers, users..

7. The Future

NERSC can not provide this information.

a. What is today's greatest impediment in terms of your use of the center's computational facilities?
b. With the projected increases resources over next 3 yrs?
c. What do you believe the proposed increases in capacity at the facilities will provide (e.g. based on observations of historical increases)?
    a. Better turn-around time for
    b. More users and incremental improvement in use with little or no change in scale or quality
    c. Reduced granularity, resulting in constant solution time, and more accurate answers
    d. New applications permitting in new approaches and new science
d. How, specifically, has your use changed with specific facilities increases?
e. How is the project x effort projected to change in the next 5 years?
f. What is your plan for utilizing increased resources?



OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY

05/01/2006

UT-BATTELLE

ASCAC Computer Facilities sub Panel
Facilities and Experimental Project Metrics

1.0 Overview of Resources Provided by the Center

a. Contact information for the project

Thomas Zacharia, Associate Laboratory Director, Computing and Computational Sciences, 865-574-4897, ZachariaT@ornl.gov

Jeffrey Nichols, Interim Director, Center for Computational Sciences, 865-574-6224, NicholsJA@ornl.gov

b. Organizational structure with staff sizes and functional titles (separate page)

The organization chart of the Leadership Computing Facility at ORNL can be viewed at: http://nccs.gov/aboutus/organization/pdf/NCCS_Org_Chart.pdf

c. FTE's

i. Overhead and Overall Management
  a. Management: 4.3
  b. Administrative: 2.4
ii. Operations
  a. 8 technical staff
  b. 12 vendor and contract operators
iii. System development tools
  a. 7.7 technical staff
  b. 1 contractor
iv. Consulting
  a. 10 technical staff
  b. 4 part-time contract staff
v. User Specific Support and Projects
  a. 13.7 technical staff

d. Physical infrastructure

The Leadership Computing Facility at ORNL is housed within the Center for Computational Sciences Building on the ORNL campus. This state-of-the-art computing facility has 16 MW of electrical and cooling capacity for the systems, with planned upgrade to 40MW, and with easy "designed in" expansion of the capacity to accommodate future CCS systems. With 40,000 ft^2 of floor space, the CCS can simultaneously deploy multiple petascale systems; space is available so that next-generation systems can be installed and brought into service before shutting down current generation systems, thereby allowing an orderly transition from one system to the next.

Construction of the CCS began in March 2002. The entire 300,000 ft^2 computer center and office complex was financed by and built to ORNL's specifications by a private developer who leases the building to UT-Battelle, the managing contractor for ORNL, who then leases the space to the U.S. Department of Energy. The building was completed in April 2003. After final checkout and commissioning, the Leadership computer center was moved into the CCS

building over a six day period in June 2003. The facility was designed from the ground up to be a leadership-class computing center

Power and Cooling

The first rule of center design is that modern computers are power hungry and getting more so. Today, the CCS has 8 megawatts (MW) of power installed for the computer systems, and another 8 MW for the rest of the building, including the cooling plant. ORNL is currently installing a new 70 MW substation on the campus and will increase the computer center power to 25 MW in 2008 and has plans to take the power up to 40 MW by 2010. The 161,000 volt power feeds from the Tennessee Valley Authority, who supplies power throughout the region, into the ORNL substation have a mean time to interrupt of over 10 years each, resulting in extremely highly reliable power for the computer center. Nevertheless, the center has a 500 KW uninterruptible power system and a 750 KW generator to supply non-stop power for the networks, disks, and storage system.

Whatever goes in as power must be removed as heat. The CCS has three chillers, each with 1,200 tons of chilling capacity. This gives us enough cooling capacity for up to 12 MW of computers in the center. The chiller plant was designed with expansion in mind. There are additional flanges and pad space to allow another chiller to be installed without disrupting the operation of the computer center, if the demand requires. The piping was designed to allow larger chillers to be installed should we need to expand to even more capacity. The chillers operate in an N+1 configuration, with one spare always available should we need to perform maintenance, or have a failure. For additional capacity and redundancy, ORNL is installing a new connection from the computer center to the laboratory chilled water plan, where cooling capacity for up to 30 MW is available, and additional expansion capability is available.

The CCS pays 5.4 cents per kilowatt hour for power from TVA.

Links to the World

The CCS is well connected to major national and international research and production networks, providing high-speed connectivity to partners, collaborators, and users of the facility around the world. The CCS is connected to DOE's primary production and research network ESnet at 10 gigabits per second (Gb/s). The CCS is also connected to the Internet2 network at 10 Gb/s. The CCS is part of the TeraGrid network, linked at 10 Gb/s. In addition, the CCS is leading the development of the DOE Ultra science network with connections at 20 Gb/s, and is linked to the NSFs experimental Cheetah network at 10 Gb/s. All these connections are possible because ORNL purchased its own fiber optic communications links connecting ORNL to Atlanta and Chicago where major network hubs terminate. These connections give ORNL network capability as high as 4 terabits per second, if needed.

e. Balance sheet and budget for:

See Appendix A – LCF 2006 Budget. The average FTE rate for FY06 is $307,920

f. Institutional affiliation and degree of institutional support

The CCS is part of the Oak Ridge National Laboratory which is managed by UT-Battelle, LLC. The $70M computational science facility was funded privately by UT-Battelle. In addition, the State of Tennessee built a $10M Joint Institute for Computational Sciences

building for collaboration between the CCS and Academia and has further funded 40 joint faculty/ORNL staff members in computational sciences.

g. Present and planned hardware
   i. Computers

   ***Phoenix** – Cray X1E, 1,024 multi-streaming vector processors, 2 TB memory, 32 TB scratch disk (increasing to 44 TB this year), 18 TF peak performance.*
   ***Jaguar** – Cray XT3, 5,212 compute processors, 82 service and I/O processors, 10.5 TB of memory, 120 TB of scratch disk, 25 TF peak performance.*

   In 2006, Jaguar will be expanded to 100 TF peak performance by replacing the single-core processors with dual-core processors and then adding 68 additional cabinets. The system will then have 23,016 compute processors, 45 TB of memory and 900 TB of scratch disk.

   In 2007, Jaguar will be further upgraded to 250 TF by replacing the dual-core processors with multi-core processors resulting in a system with 35,608 compute processors, 70 TB of memory, and 900 TB of scratch disk.

   ii. Disk memory for cache and on-line datasets or databases

   The CCS today has a shared home-directory file system available to our users located on NFS servers. This file system provides 5 TB of space for persistent storage of small files.

   The CCS is building a replacement for the NFS storage that will provide a high performance file system linking all of the computers. The system has 10 TB of disk space today, but will be increased to 100 TB later this year as the system is put into production. Our plans are to further increase this to approximately one petabyte over the next 2-3 years.

   iii. tertiary storage, e.g. in use peta-bytes versus potentially available

   The CCS uses the High Performance Storage System (HPSS) for long-term storage of files. Today the system has approximately 920 TB of data stored in the system and is growing at about 1-2 TB per day. The capacity of our HPSS system is 5 PB. We plan to add additional tape libraries and tape drives to increase the bandwidth and capacity each year as driven by the demand from users.

h. Software development and production tools provided top 5 (enumerate on separate pages)

1. *Totalview; debugger from Etnus [all platforms]*
2. *CrayPAT; performance monitoring and profiling [jaguar, phoenix]*
3. *Subversion; version control system*
4. *ID; from RSI, scripting/analysis/visualization [all platforms]*
5. *VisIT; LLNL visualization application*

i. Application codes available to the users that are supported by the center (ISVs, open source, etc.) top 5 enumerate with software development tools listing

1. *CCSM The Community Climate System Model is a fully-coupled, global climate model that provides state-of-the-art computer simulations of the Earth's past, present, and future climate states. [phoenix]*
2. *NWChem is a computational chemistry package designed to run on high-performance parallel supercomputers. Code capabilities include the calculation of molecular electronic energies and analytic gradients using Hartree-Fock self-consistent field (SCF) theory, Gaussian density function theory (DFT), and second-order perturbation theory. For all methods, geometry optimization is available to determine energy minima and transition states. Classical molecular dynamics capabilities provide for the simulation of macromolecules and solutions, including the computation of free energies using a variety of force fields. [phoenix,ram]*
3. *VASP is a package for performing ab-initio quantum-mechanical molecular dynamics (MD) using pseudopotentials and a plane wave basis set. [jaguar, ram]*
4. *GAMESS, the General Atomic and Molecular Electronic Structure System is a general ab initio quantum chemistry package. GAMESS can compute SCF wavefunctions ranging from RHF, ROHF, UHF, GVB, and MCSCF. Correlation corrections to these SCF wavefunctions include Configuration Interaction, second order perturbation theory, and Coupled-Cluster approaches, as well as the Density Functional Theory approximation. Analytic gradients are available, for automatic geometry optimization, transition state searches, or reaction path following. Computation of the energy Hessian permits prediction of vibrational frequencies. [ram]*
5. *NAMD is a molecular dynamics program designed for parallel computation. Full and efficient treatment of electrostatic and van der Waals interactions are provided via the Particle Mesh Ewald algorithm. (O(N Log N) NAMD interoperates with CHARMM and X-PLOR as it uses the same force field and includes a rich set of MD features (multiple time stepping, constraints, and dissipative dynamics). [jaguar]*

j. What auxiliary services do you offer your users
   i. Visualization

The CCS visualization facility provides a variety of visualization libraries, tools, and display devices ranging from the desktop to a 216 ft^2, 35 megapixel display wall. The visualization engine for the CCS is a 128 processor Opteron cluster linked by an Elan3 Quadrics interconnect and by gigabit Ethernet to the computer systems and storage environment of the center. The CCS provides high-end visualization at ORNL, and to the desktops of our user community, wherever they may be.

   ii. Data Analysis

The CCS provides two separate systems for data analysis. "Ram" is a 256 processor SGI Altix system with 2 TB of shared memory. "Ewok" is a 160 processor EM64T Xeon cluster.

2.0 User Interface and Communication Including Satisfaction Monitoring and Metrics

a. How do you measure the success of your facility today in being able to deliver service beyond the user surveys (e.g. the NERSC website)?

The CCS assigns a member of the Scientific Computing staff to act as a liaison for each project. These staff members work closely with the project and bring their needs/concerns forth to the rest of CCS. Additionally, the CCS User Meeting provides a forum in which the users can express both positive feedback and concerns about the center. Information from

these sources, when combined with user survey responses and general feedback (in tickets), gives us robust insight into the user view of our facility.

b. Do all users-experimenter teams, team members, and any users that the team community provides utilize the survey?

We did not receive survey replies from all users. However, the notice of availability of the survey was sent to all users. The survey was available on our website, and was open and available to any users that wanted to complete it.

c. Have these surveys been effective at measuring and understanding making changes in operations? (Please cite)

The survey period has only recently closed, and we are in the process of evaluating its results. The responses that we did receive were generally positive. Additionally, our users were offered an additional opportunity to ask questions/make comments/offer suggestions during our user meeting earlier this year.

d. Describe your call center – user support function: hours of coverage, online documentation, trouble report tracking, trouble report distribution, informing the users, how do users get information regarding where their job/trouble report is in the queue?

The CCS User Assistance Center (UAC) is staffed from 9AM-5PM (Eastern Time) Monday through Friday, exclusive of ORNL holidays. User trouble reports can come in via email, telephone, or walk-in. Email reports, both during and after hours, are automatically logged in our ticketing system. Phone calls during the 'shift' are answered by one of the on-duty consultants. After hours, phones are forwarded to the HPC operators. If a situation is critical, they can notify the appropriate people that action needs to be taken. Otherwise, they can take a problem report and forward it to the UAC.

Problems are tracked via RT. The 'owner' of a ticket contacts appropriate CCS/Vendor staff in troubleshooting the problem. When issues are forwarded to vendors for support, the owner notifies the user and places the ticket in a 'vendorWait' state.

Initial trouble reports go to all members of the help@nccs.gov email list. Further emails about a specific problem go only to the owner. However, staff members have access to all tickets in the consult queue and can therefore check the progress of other tickets, provide information for those issues, etc. In general, most tickets spend their lifetime owned by the consultant, so there is very little tracking involved. Any emails sent by the user will go to the appropriate person so they can provide any necessary updates. In cases where issues are handed off to vendors, the users are notified of what has occurred. If a user requires further information on the status of their trouble report, they need only contact help@nccs.gov.

e. What mechanisms are provided for the user with respect to dissatisfaction with how a case is being handled?

The CCS website contains contact information for all groups at the center.

f. What mechanisms are provided to support event-driven immediate access to your facility (e.g. Katrina or flu pandemic)

Currently, the Resource Utilization Council (RUC) considers requests for reservations and dedicated time on the systems. The RUC only meets once a week, however in an extreme case a special meeting is called to approve such a reservation. If the users currently exist on the system they could then begin running. Users that do not exist must be reviewed for export control purposes and we must ensure that they have the appropriate paperwork (user agreement, appendix B, etc) on file before allowing them on the system. At present, a procedure to obtain exemption from this policy is not in place.

3.0 Qualitative Measure of Output

a. Do you measure how your facility enables scientific discovery?

The CCS is currently exploring the best ways to measure how it enables scientific discovery, and several methods are currently in place. Examples of current activities are:

- *Regularly (monthly) track the progress of each project's simulation milestones as articulated in the original project allocation proposals*
- *Require that each project submit quarterly update reports in the form of short (8-10 slide) presentations covering recent accomplishments, impact of the accomplishments, next steps, challenges, issues, uncertainties, requirements, and output (publications, presentations, etc.)*
- *Solicit user feedback on the ability of CCS to facilitate individual and project science endeavors, e.g., from regular Application Requirements Council (ARC) teleconference calls, an annual User Meeting, a User Survey, and regular phone conversations and email exchanges*
- *Publish an annual Application Requirements Document and an Annual Report on the activities and computational science accomplishments in CCS.*
- *Understand and articulate the project requirements imposed on CCS necessary for higher fidelity and more productive science output; measure the evolution of CCS facilities against these requirements*
- *Be aware and stay abreast of other computational science (code) capabilities and results generated for similar science endeavors in other facilities throughout the world; take advantage of any improvements or advancements where possible*
- *Breakthrough scientific discovery is most probable if CCS facilities follow a strict leadership usage model; establish and use a Resource Utilization Council (RUC) to manage and enforce leadership usage;*
- *Track science output: number and quality (citations) of publications, number of invited and keynote presentations, the volume of scientific software released outside of the project user community, and extent of sharing of simulation datasets*
- *Maintain a vibrant, active scientific computing group within CCS consisting of expert PhD computational science "liaisons" assigned to one or more projects. These liaisons, accomplished researchers in their own right, not only help the projects but also scrutinize them for their science quality and quantity (including scalability, etc.).*

b. How are the results of measurement disseminated and how do they further Science and especially DOE Science Programs?

The measurement results are disseminated in the form of an Annual Report, an annual Application Requirements Document, quarterly update slides and updates, regular highlights, countless presentations to scientists, stakeholders, etc., and regularly updated externally-visible web pages. The results can help to further science and DOE science through (1) estimations of impact and return on investment for each science result, (2) attraction and

retention of new and established talented scientists by using the science results, and (3) bringing together (e.g., at focused workshops) and fostering the collaboration of groups of scientists who would not otherwise work together on a common problem.

c. What impact have any of your measures had on operation of your facility?

By attempting to do a better job in measuring scientific discovery and making sure discoveries have a high probability of occurrence, CCS is now:

- *Actively gathering, analyzing, and validating application requirements*
- *Holding regular (weekly) RUC meetings to ensure facilities are used in a leadership mode and in a manner that favors quality and productivity of science output*
- *Actively tracking project usage and project job distribution usage*
- *Extracting quarterly updates from projects*
- *Engaging communities (e.g., biology) needing computational science assistance to better position them for scientific discovery*
- *Aware and concerned for applications in their ability to efficiently use next-generation architectures and making plans for how to tackle the need for hybrid parallelism*
- *More regularly contacting projects with requests for information*
- *Attempting to take on a more active role in the project allocation proposal process*

d. What impact have the current PART measures had on your successful operations of your facility?

- *Acquisitions should be no more than 10% more than planned cost and schedule. This is defined as a Development, Modernization and Enhancement (DME) project. It is likely use to OMB and with the current definitions it is reasonable from the center's viewpoint.*
- *40% of the computational time is used by jobs with a concurrency of 1/8 of more of the maximum usable compute CPUs.*

Since Jan, the two LCF platforms at the NLCF have the following utilization: 43% of utilization invoked 2048 processors and 61% of utilization invoked 1024 processors on Jaguar (a 5212 processor XT3), and 36% of utilization invoked 256 processors and 66% of utilization invoked 128 processors on Cheetah (a 1024 processor X1E). The job size distribution for the NLCF platforms essentially defines a capability machine, being skewed and peaked around jobs utilizing approximately half of the available processors.

Of the 17 projects with allocations on the NLCF Jaguar system, data for 3 projects is currently not available (usage remains low), but of the remaining 14 projects, 9 have utilized the system in a capability (half machine) manner and the other 5 have not yet run at scale. Of the 14 projects with allocations on the NLCF Phoenix system, data for 1 project is currently not available, but of the remaining 13 projects, 11 have utilized the system in a capability manner and 3 have not yet run at scale. Some of the codes given allocations on these systems were not yet ready to scale to the NLCF magnitude, so NLCF staff members are currently working to identify algorithmic scaling problems with these codes and the plans required to address these problems.

On the negative side, this metric has nothing to do with the quality of science of a project, and some very important projects have very valid reasons that large scale jobs are not appropriate. For example, anticipated break-through science workloads in nanoscience (magnetic nanoparticles, molecular bio-physics) will require ab initio calculations using runs

*of at least 100 parallel electronic structure runs of at least 1000 tasks each, with the runs having to communicate at each step. This requires a capability resource, because it necessitates order 10*5 processors, but through the execution of 100 simultaneous jobs. The notion of "capability" or "leadership" computing, therefore, must be carefully defined. It is not a particular size of computer, but it is a definition of the computer's usage model. Capability usage does entail using a substantial amount of a given resource, but not necessarily for a single calculation. Also, in order to encourage this, the small long running jobs of the NERSC workload have experienced significantly longer queue times. Every year several science applications are expected to increase efficiency by at least 50%. This metric was motivated by the desire to increase the percent of peak performance applications have. It probably is no longer as important. A more apt measure is strong and weak scaling, as both types of scaling impact the science quality and productivity realized by current projects.*

e. What do you view as the appropriate measures for supercomputing facilities now?

There are several:

- *User satisfaction*
- *Stakeholder satisfaction*
- *Facilities are available and adequately utilized*
- *Facility staff provides timely and effective support*
- *Facility staff assist projects in need of computer & computational science improvement*
- *Facility usage model consistent with leadership (capability mode) mission*
- *Quality of science output*
- *Quantity of science output*
- *Productivity of science conducted (end-to-end workflow)*
- *World leadership and visibility*

f. During the next 3-5 years?

Same as above, with these additions:

- *Application scaling to 100K tasks, each task potentially possessing multiple threads; CCS will help to obtain proactive solutions for hybrid parallelism and insert them into key applications*
- *CCS systems maintain MTTIs well within requirements imposed by applications*

4.0 Aggregate Projects Use Profiles by Scale

a. How many projects does your center support?

The CCS supports about 41 allocated and non-allocated projects. The CCS supports 22 allocated INCITE and LCF projects. Of those 22 projects, 17 have allocations on the Cray XT3 and 12 have allocations on the X1E. There are a total of 36,155,896 allocated cpu-hours on the CCS systems. 30,261,656 have been allocated on the Cray XT3, with the remaining 5,894,240 allocated on the Cray X1E.

b. How many users are associated with all the projects?

There are 367 users who have current accounts on the SGI Altix, Cray XT3 or Cray X1E.

c. How many additional users who either use project data-sets or other center resources?

There is data archived in HPSS that is accessed by web portals (e.g., CDIAC, ARM, etc). Those projects individually track their web access but the computer center does not.

d. What is the project usage profile in terms of processor count? We would like these broken down into jobs that require, or can exploit a concurrency level of (roughly) 50, 200, 400, 1,000, 2,000, and 4,000 processors to obtain the science.
    i. Aggregate required memory per job? (Or memory per node)

    We do not reliably keep this data per job. We may be able to obtain the data from the users, if we do not already have it.

    ii. Processor distribution?

    Number of jobs run and number of hours charged per processor grouping for jobs run between October 01, 2005 and April 30, 2006.

	Allocated		*Non-Allocated*		*Total*	
Number of processors	*Jobs*	*Hours*	*Jobs*	*Hours*	*Jobs*	*Hours*
0-50	*75,511*	*1,813,891*	*21,471*	*105,730*	*96,982*	*1,919,622*
51-200	*7,082*	*2,546,411*	*15,086*	*407,812*	*22,168*	*2,954,223*
201-400	*2,431*	*2,124,886*	*7,916*	*389,028*	*10,347*	*2,513,914*
401-1000	*773*	*1,514,654*	*3,698*	*949,521*	*4,471*	*2,464,175*
1001-2000	*480*	*1,434,909*	*3,354*	*1,391,967*	*3,834*	*2,826,877*
2001-4000	*292*	*1,556,807*	*2,168*	*1,122,315*	*2,460*	*2,679,122*
4001-	*139*	*2,597,406*	*2,061*	*1,487,243*	*2,200*	*4,084,649*

    iii. Disk space use?

    We do not reliably keep this data per job. We may be able to obtain this information from the users if we do not already have it.

    iv. Tertiary tape use?

    We do not reliably keep this data per job. We may be able to obtain it from the users if we do not already have it. We can also pull an aggregate mass storage usage by user, but we cannot tie this directly to the projects (since it may have accumulated over many years with different projects) and we cannot tie it directly to jobs (since we do not keep this metric).

    v. Average wall clock time of jobs?

    Average wall clock time per processor grouping for jobs run between October 01, 2005 and April 30, 2005, where a job's wall clock time is the job's end time minus the job's start time.

Number of processors	*Allocated Projects; Average wall clock time*	*Non-Allocated Projects; Average wall clock time*	*Total*
0-50	*1.35*	*0.54*	*1.17*

51-200	*3.82*	*0.30*	*1.42*
201-400	*3.13*	*0.018*	*0.88*
401-1000	*3.54*	*0.48*	*1.01*
1001-2000	*2.15*	*0.34*	*0.57*
2001-4000	*1.76*	*0.23*	*0.41*
4001-	*3.97*	*0.16*	*0.40*

vi. Average time of jobs in the queue?

Average wait time per processor grouping for jobs run between October 01, 2005 and April 30, 2006, where a job's queue wait time is the amount of time a job spends waiting in the queue to enter a run state (the job's start time minus the job's queue submission time).

Number of processors	*Allocated Projects; Average wait time*	*Non-Allocated Projects; Average wait time*	*Total*
0-50	*1.68*	*0.85*	*1.50*
51-200	*3.68*	*0.83*	*1.74*
201-400	*5.19*	*1.06*	*2.03*
401-1000	*8.16*	*1.28*	*2.47*
1001-2000	*5.09*	*0.93*	*1.45*
2001-4000	*3.43*	*1.00*	*1.29*
4001-	*13.19*	*1.26*	*2.02*

vii. How do you measure project code performance on your machines?

We do not actively measure code performance on the systems. However, each system has tools, such as CrayPat and PAPI, that allow users to profile their code to assist in tuning it for optimum performance.

We also provide each allocated project with a liaison within the Scientific Computing Group who is available to work closely with the project on issues such as code performance.

We also monitor system usage and job sizes for each allocated project as a tool to proactively support projects.

viii. Amount of project consulting support utilized?

The Project ID is not logged in support tickets. The system simply logs the user. While we can summarize the time a ticket is open, we do not have a reliable measure of the actual time worked on a ticket.

With regard to direct support of projects, the Scientific Computing Group liaisons may better be able to summarize the amount of support utilized.

5.0 Center x User Readiness for 10x Processor Expansion

The mid-term goals for each facility call for a major expansion from machines with of order 5,000 processors to machines of order 50,000 processors or more.

a. Please outline how the center will accommodate this growth over the next 3-5 years.

In readying science applications for efficient and productive use of these systems, three aspects must be taken into account for each application: (1) it must port correctly as guided by regression tests with all required libraries and system software present; (2) it must exhibit acceptable parallel performance on up to 100K execution tasks and/or threads; and (3) a full-system simulation must have great breakthrough potential, i.e., a discovery, a higher fidelity result, or new understanding. For a science application to be mission-relevant, alignment with the DOE/SC ASCR Strategic Plan is important. A multi-faceted and systematic applications plan, with each phase building upon the previous successfully-completed phase, is necessary to ensure that the aforementioned applications will execute easily and efficiently at scale on the 250TF and 1000TF systems. This "application path", which has been documented in the CCS program plan, consists at a high level of computer science (e.g., porting to new operating systems) and algorithm (e.g., scaling, tuning) tasks. It closely follows and has a mutual dependency with the hardware, system software, and infrastructure plan. The availability and deployment of hardware and software testbeds before arrival of the final 250TF and 1000TF systems is crucial to carrying out the plan successfully. The overarching theme of this plan is to ensure an efficient factor of twenty scaling from the current 5K execution tasks (cores) to the approximately 100K tasks residing on the 1000TF system.

Key science applications must be ready for efficient and productive simulation performance before the systems undergo acceptance. This state of readiness has three components:

- *Software: the application and its required libraries port to the system correctly as guided by regression tests;*
- *Algorithms: the application exhibits acceptable initial parallel performance on the system without major algorithm overhauls required; and*
- *Science: a single simulation with the application has great potential for a higher fidelity result than ever before or a bringing to light a new understanding.*

To ensure that breakthrough science simulations occur immediately after acceptance of the next two large CCS systems (250 TF and 1000 TF), a science readiness plan, complete with key milestones, has been laid out and documented for CCS in order to ensure that a set of selected applications are in a state of readiness for one or more "science at scale" simulations immediately after system acceptance. The application selection process will include consultation with the science application PIs and their sponsoring DOE Program Managers on the potential for achieving science breakthrough results on these systems. An application selection committee consisting of the CCS Director of Science, the Scientific Computing (SC) Group Leader, and selected SC Group staff will be convened to ultimately choose a small set (three to five) of science applications codes deemed to be most "ready" as defined by software portability, algorithm scalability, and science potential. The suite selection will be a methodical process, informed by applications porting activities and scaling tests on the 50 TF and 100 TF upgrades as well as associated testbeds. The applications set considered (from which the final suite will be chosen) will be broad: current LCF/INCITE projects, other Science Laboratory codes, selected NNSA Laboratory open codes, and academic codes.

In total, 24 Level Two (L2) Science Readiness Milestones have been defined through FY09 in the CCS Program Plan: 3 for the 50 TF upgrade, 3 for the 100 TF upgrade, 9 for the 250 TF system, and 9 for the 1000 TF system. Successful execution of these milestones should insure that CCS can accommodate the twenty fold increase in processor count realized when the 1000 TF system is deployed. Further details of these milestones are available in the CCS Program Plan.

b. What do you believe is your role in preparing users for this major change?

The CCS will play an important role in preparing users for this major change, specifically because of the unique role held by the project liaisons in the Scientific Computing Group. Each liaison, typically an accomplished PhD computational scientist in their own right, is assigned to 2-3 LCF or INCITE projects (the CCS has currently 22 FY06 project allocations). The liaison's responsibilities include, among others, porting, tuning, optimizing, and improving scalability of each project's codes. Many of the liaisons are integrated deeply into the project team to the point of being a true collaborator, i.e., helping with fundamental algorithmic and physical model developments (a good example is the CCS climate modeling liaisons). The liaisons are working hard to prepare users and key codes within their projects for this major change by helping to execute some of the CCS science readiness milestones related to scaling up applications:

- *Documentation of application set requirement matrix for 250TF*
- *CCS workshop to engage SciDAC2 CS & Math projects*
- *Workshop on porting to multi-core (SciDAC2 Applications)*
- *"Science at scale" proposals submitted by PIs of each application in final application suite to application selection committee*
- *Selection of final (three to five) application suite for early Science run at 250TF*
- *Selection of an initial "science day one" application*
- *Performance of "science at scale" simulation and documentation of results*
- *Tune applications for 11K execution tasks, each having four threads*
- *Tune applications for the SSE (4 flops/clock)*

Some of the milestones established in the current CCS Program Plan cannot be successfully executed without leveraged efforts from the SciDAC-2 Program, an example being the last two milestones itemized above. Leveraged funding estimates, for example, needed for completion of these milesetones are based on resource estimates of 1 FTE per application for algorithm development required to achieve scaling on 11K four-threaded tasks, and 1/3 FTE per application for SSE tuning. These estimates are not rigorous, but based on past experience.

In addition to providing liaisons for each project and the actual platforms themselves, other roles the CCS must play in preparing users for scaling up their applications include identification of application-specific non-scalable algorithms and associated scalable remedies, availability of multi-core aware compilers and operating systems, tools and libraries for hybrid parallelism paradigm exploration (e.g., threading, OpenMP), making multi-core testbed platforms available as soon as the market allows it, and conducting workshops and tutorials on fined-grained and hybrid parallelism programming and algorithm development.

c. What effort (in terms of personnel) is devoted to code development issues today, and do you view this as adequate coverage as we move to machines with more than 25,000 processors?

In the CCS today the Scientific Computing Group and the Director of Science, or approximately 13 staff (~$4M), are available for technical support of code development activities within the 22 LCF/INCITE projects currently with FY06 allocations. This translates into one-third to one-half of an FTE per project in code development efforts supplied by the CCS. Given the challenges confronting the science applications in scaling to 25-100K

processors, the CCS technical support coverage is not adequate on a per application basis and could worsen if the number of applications (project allocations) is expected to grow (i.e., up from the current 22 allocations) . As articulated in the CCS Program Plan, current conservative estimates for tackling these scaling challenges are in the 1-4 FTE range (per application) for algorithm development required to achieve scaling on 95K tasks. In addition, efforts of 3 FTE per application for the development of math libraries and tools and 3 FTE per application for the development of computer science libraries and tools are also estimated as being needed. For success to be achieved, then, programs like SciDAC-2 must fund these efforts if the CCS is unable to grow to accommodate these needs. Currently the CCS is counting on SciDAC-2 and other ASCR Programs to leverage these needed efforts, as the CCS will continue to plan according to budgets that only allow at most one-half FTE per project in technical support.

d. Are there codes in your user portfolio that will scale today to 10,000, 25,000, or 75,000 processors. What is the nature of these codes (Monte Carlo, CFD, hydro?). Are these codes running today on other systems of comparable size?

Yes, there are codes today in the CCS user portfolio that will scale today beyond the current 5K processors available on CCS systems. In some cases scaling data (beyond 5K processors) is available, but in most cases knowledge of the algorithms is used as a basis for confidence (or lack thereof) in scalability. In putting together the CCS Program Plan for achieving science at scale at 1000 TF, six codes in particular were identified as being the current most likely candidates for "science at scale" (delivering breakthrough science at 1000 TF): LSMS, DCA, VH1, POP, S3D, GTC, and AORSA:

- *LSMS has run on 5,000 processors with excellent scaling. Parallelization is achieved by assigning system atoms to different PEs, meaning more PEs allows larger system calculations. LSMS has been run on the BG system with either one task per core on one task on two cores. To be efficient for one task on two cores, an implementation of ZGEMM that takes advantage of the multiple cores is needed.*
- *DCA is a QMC code that currently has scaling challenges at initialization that are caused by having to break up the Markov chains into pieces for the initial equilibrium calculation. Once equilibrium is established, however, the Markov chain computations are independent (embarrassingly parallel), i.e., no communication between chain. Because of the equilibrium startup computation, a large fixed startup cost involved that is less of problem as the chains become larger. Perhaps domain replication or Global Arrays will help, but this must be investigated.*
- *VH1 is an explicit (nearest neighbor) Eulerian shock physics code for astrophysics that scales well on 5K processors of the CCS XT3.*
- *S3D is an explicit nearest-neighbor turbulent combustion DNS code on structured Eulerian meshes. S3D generates high-resolution solutions to compressible Euler, turbulent model, and chemical mechanisms for multi-stage ignition*
- *GTC is a mature PIC code for magnetically confined fusion simulation (specifically turbulent transport in ITER-like configurations), nearest-neighbor, good scaling to 5000 processors demonstrated on a number of platforms. GTC utilizes MPI and OpenMPI.*
- *Pop is a global ocean circulation code with nearest neighbor communication. Pop is compute-bound as long as the number of cells per PE is high enough to swamp latency-bound 2D elliptic solve. Scaling Pop to 100K processors will be limited by the ability to generate fast, scalable elliptic solves.*
- *AORSA is a fusion code used for the prediction and control of macroscopic stability of ITER plasma and in design and application of heating and current-drive systems. It uses*

a fully spectral method to solve linearized wave equation using ScaLAPACK libraries. AORSA scales well and has been used as a benchmark code in DOE Joule audits

Many other application codes (> 30) are currently executing on the CCS platforms today, all of which have varying degree of scaling issues and problems. Those with scaling issues have been identified and the CCS liaisons are currently working with those projects to help focus CCS scaling technical support.

e. As machines become more complicated, what do you see as the challenges to your success? For example, are you (or parts of your institution) actively involved in research related to fault-tolerance, memory/bandwidth contention, job scheduling, and etc. on the future machines?

Key risks have been identified in the CCS Program Plan that could stand in the way of applications being able to scale up to 100K tasks on the CCS platforms (by late 2008). Examples of these risks include:

- *A chosen application does not have SciDAC-2 support*
- *Incomplete and/or inadequate hybrid parallel programming methods and software necessary for efficient scaling on multi-core processors*
- *Incomplete and/or inadequate math and special-purpose software libraries that invoke hybrid parallelism for efficient multi-core processor use*
- *Inadequate software infrastructure to facilitate SQE, such as testing environments*
- *Needed componentization infrastructure (e.g., CCA) is not available*
- *Inadequate IDE tools, most notably for debugging*
- *Inadequate fault-tolerant communication and parallel I/O libraries*
- *A chosen application does not have the assumed maturity of mathematics, algorithms and computer science*

The risks itemized above can be mitigated with significant leveraged efforts from other programs, e.g., the DOE/SC ASCR MICS and SciDAC-2 Programs. In particular, these investments could reside within the current SciDAC-2 Program framework (applications, SAPs, CETs, and Institutes). Some CCS staff, in particular those residing in the Technology Integration Group, are engaged in work aimed at tackling current and anticipated problems in fault-tolerance and other issues. The CCS also works closely with its sister research organization, namely the Computer Science and Mathematics (CSM) Division, in collaborative research on performance modeling, analysis, and optimization, future technologies (e.g., accelerator boards), Linux kernel development, scalable parallel I/O development, etc. The model currently established at ORNL is CSM performing the fundamental research that is then deployed by the CCS when the research reaches maturity. Requirements for research in CSM are predominantly set by the CCS, based on needs of key science applications codes.

f. How do you determine the path forward for your organization?

The "path forward" for CCS, assumed here to be a detailed and regularly updated (living) implementation plan (IP) highlighted by milestones and tangible deliverables, is the outcome of regular CCS planning sessions. The CCS IP is first shaped by high-level goals and metrics set by the ASCR Program Office in the DOE/SC, then determined in more detail by meeting requirements set by key applications codes in the various DOE/SC Program Offices. A key aspect of the IP is risk management, namely identifying all technical, programmatic, and people risks associated with each major deliverable and developing mitigation plans for those risks. Typically the mitigation plans involve taking multiple, redundant paths to a

solution, in the end selecting the path most likely to lead to success at pre-defined decision points. In addition to program goals/metrics, requirements, and risks, a final consideration for the "path forward" is determined by vendor interactions, which is a formal process by which CCS is able to stay abreast of H/W and S/W vendor technologies, plans, schedules, and costs. To summarize, the CCS path forward is best articulated in a living IP document that describes "who does what when". This IP is a formal contract with stakeholders in the ASCR DOE/SC Program Office and one that should reflect the requirements of the users of the CCS platforms.

g. What do your users want to see in the largest machines now available and those which will be available in the 3 year and 5-7 year time frames? (memory per core/node, number of processors, disk space?)

The CCS is currently engaged in a requirements process with its 22 projects via an Applications Requirements Council (ARC). The ARC develops, manages, and plans the breakthrough science requirements imposed upon the CCS leadership computing systems. The principal product of the ARC is the documentation, publication, and handoff of requirements to the CCS Technology Council (TC), which is responsible for implementing and/or aligning these requirements with deployed CCS leadership computing systems. By articulating requirements, the ARC hopes to ensure that all systems designed, procured, deployed, and operated within the CCS are aligned to the maximum extent possible with the needs and goals of the breakthrough science projects using the CCS resources. The CCS requirements document, already in draft form and due for a formal release in the 4th quarter of every FY, does address in detail the most desirable attributes a system should have in order to best meet applications needs. For example, a given LCF system has many attributes that uniquely characterizes it relative to other systems, but the CCS has determined that twelve attributes in particular are useful and important to consider from the applications perspective: Peak flops per node; Mean time to interrupt (MTTI); Wide area network (WAN) bandwidth; Node memory capacity; Local storage capacity; Archival storage capacity; Memory latency; Interconnect latency; Disk latency; and interconnect bandwidth. For each of these twelve system attributes, certain behaviors and properties of a given application warrant more importance placed on a given attribute over another. The CCS ARC has defined those application behaviors and properties that serve as drivers for each system attribute, and, for each, application, prioritized the most desirable attributes. Bottom line: the CCS requirements process is the right approach for understanding what users want to see in platforms in the 3 year and 5-7 year time frame. The annual requirements document will summarize these findings.

Appendix A – LCF 2006 Budget

		FY 2006	
		Dollars	**FTE**
Systems			
Cray X1e			
	Lease	1,448,938	
	Maintenance	1,525,316	
Cray XT3			
	Lease	19,637,447	
	Maintenance	1,474,684	
Baker			
	Lease		
	Maintenance		
	Total Lease	**21,086,385**	
	Total Maintenance	**3,000,000**	
	Total Systems	**24,086,385**	
Facility Operation			
	Space and Utilities	2,768,798	
	Facility Modifications	1,446,740	
Infrastructure			
Software license			
	MOAB Batch System	80,000	
	TotalView Debugger	100,000	
	Miscellanenous	143,542	
Servers			
	Hardware	179,904	
HPSS			
	Hardware	600,000	
	Tapes	87,233	
	Maintenance	60,000	
Disk Storage			
	Hardware	600,000	
	Software	184,100	
	Maintenance	20,000	
Networks			
	Internal	500,000	
	External	500,000	

People			
Management & Planning			
	LCF Operations	615,840	2
	Project Director	307,920	1
	Long term planning	400,296	1.3
	Project reporting	615,840	2
	Platform Planning and acquisition	123,168	0.4
HPC Operations (Ann Baker)			
	Sysadmin, Cyber Sec, Ops	2,463,360	8.0
	SAIC Computer Operators (6 people)	600,000	
	Platform implementation and testing	307,920	1.0
	System Integration		
Technology Integration (Shane Canon)			
	Sys Programming	985,344	3.2
	Operating Systems		
	File System		
	Disk Storage	431,088	1.4
	HPSS storage	923,760	3
	Networking	338,712	1.1
	Programming Environment		
User Asstance & Outreach (Julia White)			
	Helpdesk	1,324,056	4.3
	User Assistance & Outreach	1,755,144	5.7
Scientific Computing (Ricky Kendall)			
	Technical Support	3,417,912	11.1
	Data Analysis & Visualization	1,108,216	2.3
	Develop Acceptance Test	92,376	0.3
Miscellaneous			
Travel		285,911	
Small Projects and Supplies		147,478	
Center Director - Special Projects		1,180,000	
Universities		5,000,000	
	Total	**53,781,043**	**48.1**

UT-BATTELLE

AMES LABORATORY

ARGONNE NATIONAL LABORATORY

Los Alamos

Lawrence Livermore National Laboratory

NASA

NCAR

Pacific Northwest National Laboratory

PPPL

Sandia National Laboratories

Information on metrics at the ORNL National Center for Computational Sciences

Douglas B. Kothe
NCCS Director of Science

Al Geist
NCCS Chief Technology Officer

THE CENTER FOR COMPUTATIONAL SCIENCES

OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY

All those Reviews and Reports Bill Kramer talked about that the Facilities go through . . .

Not repeated here but yes ORNL goes through the same oversight

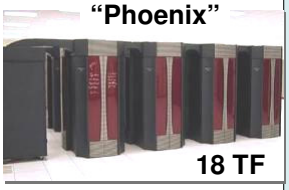
- Internal (twice year),
- External (annually),
- DOE (annually),
- Lehman reviews (twice year).

ORNL plan of action to deliver leadership computing for DOE

- **Maintain world-class infrastructure in support of LCF**
  - Maintain state-of-the-art facilities to house LCF
  - Partner with TVA to deliver reliable, cost-effective, power
  - Deliver outstanding access and service to user community
- **Deliver leadership computers**
  - Deliver 1 PF in 2008; provide clear upgrade path of 100 TF by 2006 and 250 TF by 2007
  - Provide pathways to sustained PF computing in FY 10 and beyond
- Deliver much higher sustained performance for major scientific applications than currently achievable
  - Develop next generation models and tools in conjunction with user community
  - Engage academia and laboratories to advance scalable applications software
- Deliver science outcomes in climate, energy, fusion, biology, materials, chemistry, and other areas critical to DOE-SC and other federal agencies
  - Engage user community to enable high likelihood of breakthroughs

Two Capability Systems optimized for applications are converging as Cascade

Cray X1E "Phoenix"

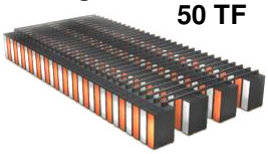


18 TF

- Proven architecture for performance and reliability
- Most powerful processors and interconnect
- Scalable, globally addressable memory and bandwidth
- Offers capability computing for key applications

Cray XT3 "Jaguar"

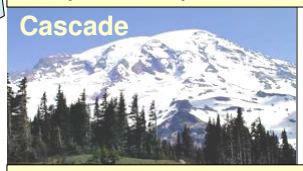
50 TF



- Extremely low latency, high bandwidth interconnect
- Efficient scalar processors, balanced interconnect
- Known system architecture – based on ASCI Red
- Linux operating system on service processors with microkernel on compute processors

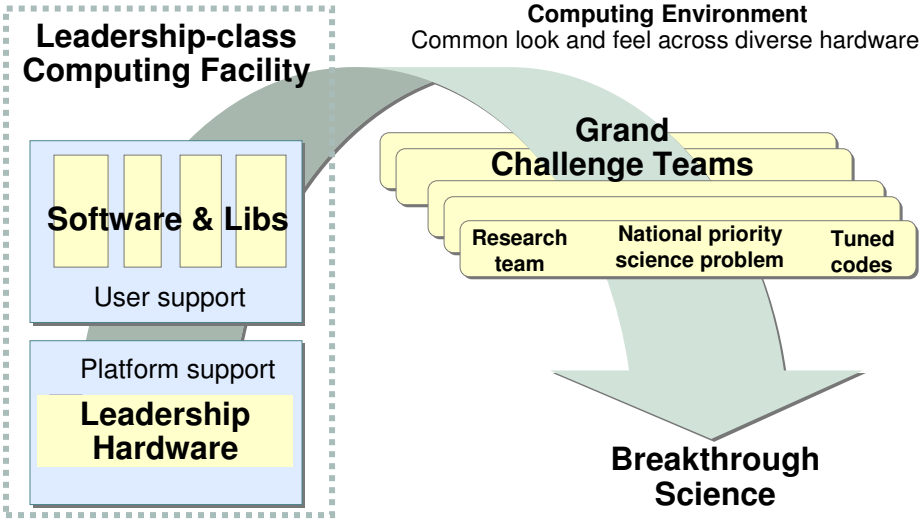
- Unified system including vector, scalar, multithreaded and potentially FPGA processors
- Scalable network and globally addressable memory
- Adaptive custom processors

Cascade

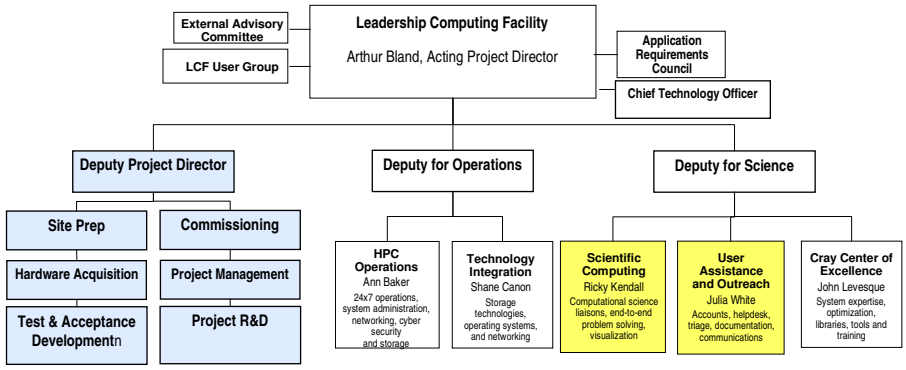


- Single Linux-based user interface and environment
- Shared global file system
- Improved performance by matching processor to job
- Single solution for diverse workloads

Facility plus hardware, software, and science teams all contribute to Science breakthroughs



NLCF organized to deliver breakthrough science



Deliver 250 TF
Deliver 1 PF

Operations
And User Support

User Assistance & Outreach Group

- **Mission**
  - Generate user satisfaction and advocacy by delivering seamless access to NCCS resources, providing swift and effective front-line support, and showcasing NCCS research in strategic communication activities.
- **User Assistance Center**
  - Phone response 24x7. User Assistance Center staffed 9-5 ET, M-F
  - *Request Tracking* system used to assign user inquiries to staff follow up and resolution
  - All email questions are triaged and assigned within one business hour
  - Functions
    - Accounts
    - General system questions
    - Batch queue assistance
    - Documentation
    - Scripts
    - Compiling/Optimization/General code help
    - Software installation

- **Additional activities**
  - S/W installation standardization; Resource usage tracking; Allocation report generation
  - Highlights of activities, research; Workshop organization; Science Themes
  - Hands-on Tutorials; End Station Meetings

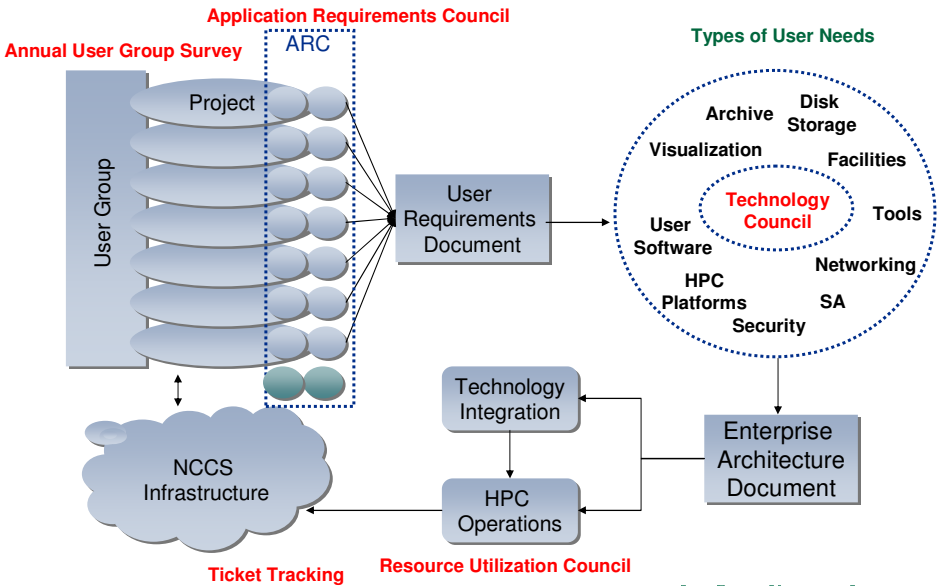
Scientific Computing Group

- **Mission**
  - Facilitate, enable & accelerate breakthrough science via targeted collaborative efforts with users
- **Metrics**
  - Effective utilization of LCF resources to provide insight and discovery
  - Elicit, analyze, and validate user requirements
  - Applications ready for the next generation systems
- **Path forward**
  - Members serve as liaisons between project teams and NCCS.
  - Collaborate directly with project teams, augmenting and extending their computational and domain-specific expertise.
  - Group members are research scientists with backgrounds in high performance computing, and various scientific domains.
  - Directly help users realize increased scientific productivity through our extensive experience in porting, tuning, and developing software on NCCS resources.
  - Reduce the total time to solution or insight for project teams by providing in-depth support for visualization, data movement and workflow needs, algorithmic development, and the choice and use of analysis tools.

SC Group must be familiar with Codes used by the 22 Projects

- **Fusion**
  - GYRO, GTC, XGC-ET, AORSA, TORIC, M3D, NIMROD, DELTA5D, GEM, AMRMHD, CQL3D
- **Accelerator Physics**
  - Omega3P, S3P, T3P
- **Computer Science**
  - Active Harmony, FPMPI, HPCTOOLKIT, IPM, KOJAK, mpiP, PAPI, PARADYN, PMaC, ROSE, SvPablo, TAU, HPCC
  - DWA/CGI
- **Nuclear Physics**
  - CCSD, HFB, SMMC
- **Climate**
  - CCSM (CAM, POP/CICE, CN, CASA', CLM), MITgcm, GEOS5, WRF
- **Combustion**
  - S3D
- **Astrophysics**
  - FLASH, SUPERNOVA, VULCAN/2D, V2D, VH1/EVH1, ZEUS-MP, BOLTZTRAN, GeNASiS
- **High Energy Physics**
  - CMS, LCG, ROOT, PYTHIA, CompHEP, MILC
- **Materials & Nano Science**
  - QMC/DCA, SPF, LSMS, VASP
- **Biology**
  - CHARMM, NAMD, AMBER, LAMMPS, GAMESS-US
- **Engineering**
  - CFL3D, OVERFLOW
- **Chemistry**
  - NWChem, VASP, PWSCF, ABinit, CPMD, ESPRESSO, OCTOPUS, MADNESS

LCF User Requirements Process



Applications Requirements Council (ARC)

- **Mission**
  - Develop, manage, and plan science requirements imposed upon the NCCS leadership computing systems *and* the science applications
  - Ensure that all systems designed, procured, deployed, and operated within the NCCS are aligned to the maximum extent possible with the needs and goals of the science projects

- **Survey projects with a detailed list of >100 requirements elicitation questions in seven different categories:**
  - Science motivation and impact
  - Science quality and productivity
  - Application models
  - Application algorithms
  - Application software
  - Application footprint on platform
  - Data management and analysis

Applications Requirements: System Attributes

System Attribute	Climate	Astrophysics	Fusion	Chemistry	Combustion	Accelerator Physics	Biology	Materials Science
Node Peak Flops								
Mean Time to Interrupt (MTTI)								
WAN Network Bandwidth								
Node Memory Capacity								
Local Storage Capacity								
Archival Storage Capacity								
Memory Latency								
Interconnect Latency								
Disk Latency								
Interconnect Bandwidth								
Memory Bandwidth								
Disk Bandwidth								

Applications Requirements: System

System Attribute	Application Behaviors and Properties That Drive a Need for this Attribute
Node Peak Flops	Scalable and required spatial resolution low; a problem domain that has strong scaling; embarrassingly parallel algorithms (e.g., SETI at home)
Mean Time to Interrupt (MTTI)	Naïve restart capability; large restart files; large restart R/W time
WAN Bandwidth	Community data/repositories; remote visualization and analysis; data analytics
Node Memory Capacity	Multi-component/multi-physics, volume visualization, data replication parallelism, restarted Krylov subspace with large bases, subgrid models (PIC),
Local Storage Capacity	High frequency/large dumps, out-of-core algorithms, debugging at scale
Archival Storage Capacity	Large data (relative to local storage) that must be preserved for future analysis, for comparison, for community data expensive to recreate;
Memory Latency	Cache-aware algorithms); random data access patterns for small data
Interconnect Latency	Global reduction of scalars; explicit algorithms using nearest-neighbor or systolic communication; interactive visualization; iterative solvers; pipelined algorithms
Disk Latency	Naïve out-of-core memory usage; many small I/O files; small record direct access files;
Interconnect Bandwidth	Big messages, global reductions of large data; implicit algorithm with large DOFs per grid point;
Memory Bandwidth	Large multi-dimensional data structures and indirect addressing; lots of data copying; lots of library calls requiring data copies; if algorithms require data retransformations; sparse matrix operations
Disk Bandwidth	Reads/writes large amounts of data at a relatively low frequency; read/writes lots of large intermediate temporary data; well-structured out-of-core memory usage

Metrics

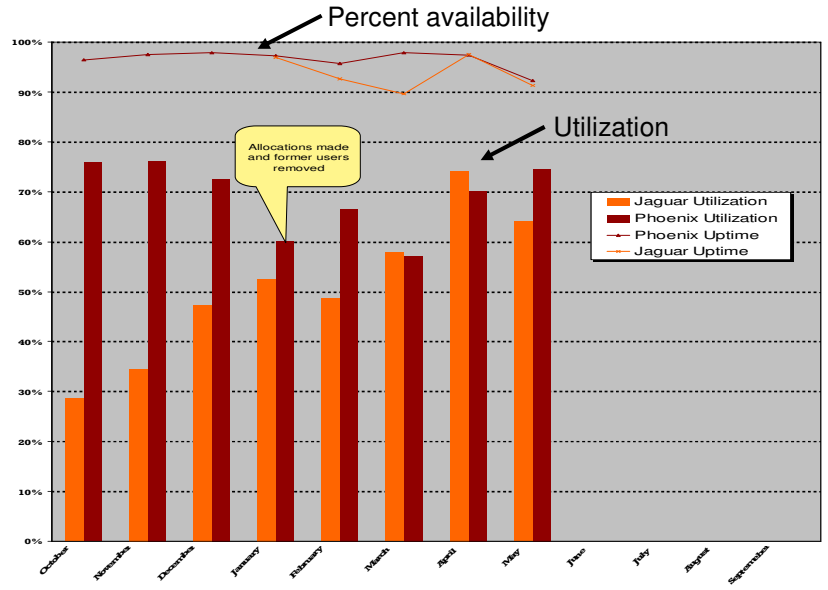
- **Old (current)**
  - Acquisitions should be no more than 10% more than planned cost and schedule
  - 40% of the computational time is used by jobs with a concurrency of 1/8 or more of the maximum usable compute CPUs
  - Every year several selected science applications are expected to increase efficiency by at least 50%.
- **New (proposed)**
  - Facility metrics
    - User satisfaction
    - Facility is ready and able to process workload
    - Facility provides timely and effective assistance
    - Facility facilitates in running capability problems
  - Computational science metrics
    - Science progress
    - Scalability

Goal #1 User Satisfaction

Results from First User Survey (Five point scale)

	Question	# Responses	Avg Rating (1 to 5)
Use Assistance Center	Solutions	12	4.50
	Speed	12	4.33
	Overall	12	4.42
Website	EaseOfNavigation	12	4.00
	StatusPageRating	12	3.92
	UserGuideRating	12	3.42
	DataXferRating	12	3.00
Jaguar (Cray XT3)	JagCrashRating	5	3.00
	JagSchedOutRating	5	3.80
	JagScratch	6	3.00
	JagHPSSInterface	4	3.75
	JagQUsability	5	4.20
	JagQTurnaround	5	4.20
	JagOverall	6	4.00
Phoenix (Cray X1E)	PhoCrashRating	6	4.33
	PhoSchedOutRating	6	3.83
	PhoScratch	6	4.17
	PhoHPSSInterface	5	3.60
	PhoQUsability	6	4.00
	PhoQTurnaround	6	3.50
	PhoOverall	7	3.57

Goal #2 System Availability & Utilization



Goal #3: Facilities provide timely and effective assistance

Helping users effectively use complex systems is a key role that leading computational facilities supply. Users desire their inquiry is heard and is being worked. Users also need to have most of their problems answered properly in a timely manner.

Metric #3.1: Problems are recorded and acknowledged

Value #3.1: **99% of user problems are acknowledged within 4 working hours.**

Metric #3.2: Most problems are solved within a reasonable time

Many problems are solved within a short time period in order to help make users effective. Some problems take longer to solve – for example if they are referred to a vendor as a bug report.

Value #3.2: **80% of user problems are addressed within 3 working days, either by resolving them to the user[1]s satisfaction within 3 working days, or for problems that will take longer, by informing the user how the problem will be handled within 3 working days (and providing periodic updates on the expected resolution).**

Goal #4: Facility facilitates running capability problems

Metric #4.2: Capability jobs are provided excellent turnaround

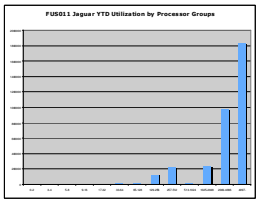
Job turnaround is an important metric for the user community and is commonly associated with user productivity. Job turnaround is typically determined as the ratio of the total amount of elapsed time a job that is eligible to run requested divided by the time the job waited to run. This is called the expansion factor.

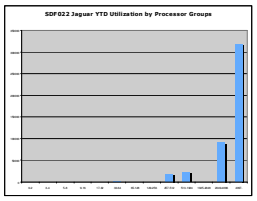
Value #4.2: **For jobs defined as capability jobs, the expansion factor is X or more. x = 10 is a potential value that may be appropriate. We are studying past data to assess this value.**

Goal #4 Majority of Time Goes to Capability

Job Size Distribution Versus Usage: Jan – Jul 06

- **Cray XT3 (Jaguar – 5212 nodes)**
  - 18.6% usage on >78.6% of processors
  - 33.1% usage on >39.3% of processors
  - 46.6% usage on >19.6% of processors
  - 68.2% usage on >5% of processors
- **Cray X1E (Phoenix – 1024 nodes)**
  - 3.4% usage on >50% of processors
  - 12.0% usage on >25% of processors
  - 60.4% usage on >12.5% of processors
  - 72.4% usage on >6.3% of processors
- **Observations**
  - Jaguar
    - A large # of 128-PE jobs in Jun have skewed results
    - Need to implement more aggressive queuing rules
  - Phoenix
    - Current queuing structures need to be revisited





Metric 4.2: Capability Job Turn-Around

Cray XT3 Expansion Factors: Jan – Jul 06

Project	Avg Job Hours Used	Avg Job Wait Time	X
AST003	17.75043011	6.05327957	1.341021571
AST004	6.009890591	5.131115974	1.853778600
AST005	3.175507246	2.090555556	1.658337517
BIO014	3.580553746	2.411856678	1.673598792
BIO015	3.279715808	2.508351687	1.764807634
CHM022	5.340755155	4.529981959	1.848191281
CLI017	3.513494624	2.325860215	1.661979159
CLI018	2.897777778	3.024444444	2.043711656
CSC023	1.169964539	1.541205674	2.317309732
CSC024	0.396666667	0.004166667	1.010504202
CSC025	0.437419355	0.132580645	1.303097345
EEF049	5.81848659	4.536949234	1.779747304
EEF051	1.38	0.01	1.007246377
EEF053	1.282857143	1.926666667	2.501855976
EES001	6.603	37.511	6.680902620
EES014	0.136666667	0.01	1.073170732
FUS011	4.694554656	4.572591093	1.974020206
FUS013	0.784844291	1.795513264	3.287731828
NPH004	3.561886792	6.157681941	2.728769694
SDF022	3.974558824	4.088382353	2.028638029

$$X = \frac{\text{job execution time} + \text{job wait time}}{\text{job execution time}}$$

Average overall expansion factor: 1.83

Metric 4.2: Capability Job Turn-Around
Cray X1E Expansion Factors: Jan – Jul 06

Project	Avg Job Hours Used	Avg Job Wait Time	X
AST005	5.406506024	5.63186747	2.04168338
CHM022	8.085729514	6.549713524	1.810033716
CLI016	14.96926829	4.729146341	1.315923681
CLI017	2.269084781	2.981607467	2.314013250
CSC013	6.092666667	1.845666667	1.302932487
CSC023	0.486810811	1.555135135	4.194536975
CSC025	1.551818182	0.677272727	1.436438196
EEF049	2.156394756	3.688528769	2.710507206
EEF050	3.803560976	17.65902439	5.642760956
EES014	1.15088	3.8744	4.366467399
FUS011	0.278	3.663	14.176258990
FUS012	8.5596	11.05602	2.291651479
FUS013	0.575510204	0.406530612	1.706382979
FUS014	14.52863636	19.85818182	2.366830398
HEP005	2.331990521	5.606445498	3.404145920
SDF006	3.235789474	2.825	1.873048146
SDF022	10.84254902	4.308823529	1.397399497

$$X = \frac{\text{job execution time} + \text{job wait time}}{\text{job execution time}}$$

Average overall expansion factor: 2.20

Very hard to schedule a leadership computer in an "optimal" manner *without* human intervention

- **Problem: it is Too many dimensions in queue space to optimize (job size, job length, job priority, job time, etc.) and science-based**
- **Solution: Resource Utilization Council**
  - Be the decision-making body for management of allocated and unallocated (discretionary) resources
  - Be the formal hearing board for ongoing user priorities, problems, and requirements
  - Issue regular utilization directives and associated actions (who/what/when) necessary to implement all decisions
- **The sole purpose of the RUC is to ensure the NCCS Leadership platforms are being efficiently and effectively utilized to the maximum extent.**
- **RUC purview: resource usage; resource requests (new/additional/exceptions); policy decisions (resource allocations, queue configurations, platform availability)**
- **Meets weekly (chaired by Director Science)**
  - Charter, minutes, action items and decisions documented & posted on web

Goal#5 Computational Science Metrics for Application Scientists (part1 of 2)

These are metrics for the science projects run at the DOE/SC facilities.

CS Goal #1: Science Progress

While there are many laudable science goals, it is vital that significant computational progress is made against the Nation's science challenges and questions.

Metric #CS1.1**: Progress is demonstrated toward the scientific milestones in the top projects at each facility based on the computational results planned and promised in their project proposals.**

Value #CS1.1**: For x% of projects at each facility, an assessment is made by the related program office regarding how well scientific milestones were met or exceeded relative to plans determined during the review period. For government funded projects, the funding office will conduct the review. Otherwise, the review will be conducted by a peer review panel selected by the DOE office of Advanced Scientific Computing Research.**

Simulation Milestones

- **All FY06 LCF projects listed simulation-based milestones as part of their proposals**
- **We have identified 74 simulation milestones for 17 LCF projects**
  - Many were not "SMART" (Specific, Measurable, Attainable, Relevant, Timely), but they were there
  - Why not hold projects responsible (or at least track progress toward milestone completion)?

Quarterly Project Updates

- **We are asking for quarterly updates from each project this FY and will *require* it in FY07 as part of their allocation**
  - Received ~50% response in Q2 of FY06
- **The update we requested asked for**
  - Recent science progress
  - Impact of recent progress
  - Next steps
  - Challenges, uncertainties, issues
  - Resource requirements drivers
  - Project productivity

Goal#5 Computational Science Metrics for Application Scientists (part 2 of 2)

CS Goal #2: Scalability of Computational Science Applications

The major challenge facing computational science during the next five to ten years is the increased parallelism needed to use more computational resources. Multi-core chips accelerate the need to respond to this challenge. Moore's Law will continue this trend as the number of CPUs on a chip double every 2 to 3 years. While this metric applies more to science projects than the facilities that host them, facility staff often must provide substantial help to the identified projects for them to be successful

Metric #CS2.1: Science applications should increase in capability.

Value #CS2.1**: The improvement of selected applications increase by a factor of 2 every three years. The measure of improvement be it scalability, capability, fidelity will be domain and code specific.**

Tracking Science Progress

- **Tracking project progress helps us better understand project work and hence how we can best help them**
  - Identify where problems are exacerbated or caused by us
  - Increase productivity and quality of science output
- **Example of how to track project progress**
  - Liaisons in the Scientific Computing Group
    - In many cases they are "part of the team"
  - Quarterly updates (make this a requirement upon allocation award)
  - Utilize the ARC process
  - Regular communication with project teams
  - Face-to-face meetings and workshops
    - Annual User/PI Meeting, code camps, road shows, visits

Science Progress Metrics

- **Going faster is not the goal of Science projects Better, Bigger, New science is, for example:**
- **Climate project (CLI017)**
  - Stay at 5 simulation years/day, increase physics
  - Increase fidelity of models
- **Combustion project (SDF022)**
  - Progress is going to higher Reynolds, Damkohler numbers
- **Nanoscience project (EEF049)**
  - Progress is number of atoms (size of system)
  - Better physics

Four Quadrant Project View (in 7 detailed pages)

Project, team, & process	Scientific Output
Centers resources input	The "code" & code scalability

Project Name	**Scientific Output**
PIs and URL DOE Office support: DOE program manager: Scientific domain (chemistry, fusion, high energy, nuclear, other.), Support for the development of the code Degree of DOE support to develop the code? SciDAC, DOE SC program internal institutional funding sources (e.g. LDRD,..), industry, other agencies, …. What are the technical goals of the project? What problem or "grand challenge" are you trying to solve? What is the expect impact of project success? What is the project profile in total human resources including trained scientists, computational scientists and mathematicians, program development and maintenance, use(rs) of the team codes? Ext communities & sizes, that code and/or datasets support.	**The scientific accomplishments 200x to present*:** **The effect on the Office of Science programs*:** Publications/location: Citations (last 5 years): Dissertations? Prizes and other honors? Residual and supported, living datasets and/or databases that are accessed by a community? Size of the community? Change in code capabilities and quality (t) Code contributed to the centers Code contributed to the scientific community at large Company spin-offs based on code or trained people and/or CRADAs Corporation, extra-agency, etc. use Production of scientists & computational scientists during 2001-2005 Production of trained software engineers during 2001-2005 *Parts 3 & 4 of metrics approach
Centers resources Steady state production use per month; per year Processor number Processor time Disk Tertiary amount and rate of change Software provided by center Consulting Direct project support as a team member What is the size of user jobs in terms of memory, concurrency (processors), disk, and tertiary store? What is the scalability of these codes What is the wall-clock time for typical runs? What could the center provide that would enhance output?	**The Code** Problem Type Types of algorithms and computational mathematics Code size (single lines of code, function points, etc.); Code size as f (t) from the origin to the present Computer languages LOC/ language 1/LOC…n What libraries used & fraction of code Code Mix: To what extent does your team develop and use your own codes? Codes developed by others in the DOE and general scientific community? Commercial application codes provided by the center? Platforms What is the present parallelism for each of the platforms Projected or maximum scalability How is measured? Is the code massively parallel? What memory/processor ratio do your project require? (e.g. Gbytes/processor) Parallelization model (e.g. MPI, OpenMP, Threads, UPC, Co-Array Fortran, etc.) What is the "efficiency" of the code? And how is it measured? What are the major bottlenecks for code scaling? What is the split between interactive and batch use? Fraction f code development at center computer(s) versus own installation?

1.0 Project name (Background)
PI & URL

DOE Office support: DOE program manager:
Scientific domain (chemistry, fusion, high energy, nuclear, other.),
Support for the development of the code
- Degree of DOE support to develop the code?
- SciDAC, DOE SC program
- internal institutional funding sources (e.g. LDRD,..),
- industry,
- other agencies, ….

What are the technical goals of the project?
- What problem or "grand challenge" are you trying to solve?
- What is the expect impact of project success?

What is the project profile in total human resources including
- trained scientists,
- computational scientists and mathematicians,
- program development and maintenance,
- use(rs) of the team codes?

External communities & sizes that code and/or datasets support.

2.0 Project Team Resources

Team size & structure
Team institutional affiliation(s). (e.g. all the institutions involved, including universities, national labs, government agencies,..). I.e. to what extent is the team multi-institutional?
To what extent are the code team members affiliated with the computer center institution? (e.g. are the team members also members of the computer center institution?)
Team composition and experience total
- domain scientists,
- computational scientists, computer scientists, computational mathematicians, database managers
- programmers
- other

Team composition by educational level (total)
- Ph.D.,
- MS, BS, undergraduate students, graduate students, post-docs, younger faculty, senior faculty, national laboratory scientists, industrial scientists, etc.)

Team resources utilization: time spent on code and algorithm development, maintenance, problem setup, production, and results analysis

5.0 Software Engineering, Development, Verification and Validation Processes

- Software development tools used (
  - parallel development,
  - debuggers,
  - visualization,
  - production management and steering
- Software engineering practices. Please list the specific tools or processes used for
  - configuration management,
  - quality control,
  - bug reporting an tracking,
  - code reviews,
  - project planning,
  - project scheduling an tracking
- What is your verification strategy?
- What use do you make of regression tests?
- What is your validation strategy?
- What experimental facilities do you use for validation?
- Does your project have adequate resources for validation?

4.0 Project resources input from the centers

Plan with benchmarks & milestones

Steady state user of resources on a production basis per month
- Processor number
- Processor time
- Disk
- Tertiary amount and rate of change

Annual use of resources
- Processor time
- Disk
- Tertiary storage rate of change

Software provided by center

Consulting

Direct project support as a team member

What is the size of their jobs in terms of memory, concurrency (processors), disk, and tertiary store?

What is the scalability of these codes

What is the wall-clock time for typical runs?

3.Project Code

- Problem Type (data analysis, data mining, simulation, experimental design, etc.)
- Types of algorithms and computational mathematics
- What platforms does your code routinely run on?
- Code size (single lines of code, function points, etc.);
  - Code age and yearly growth.
- Computer languages employed,
  - LOC/ language 1;„LOC/ language 2 LOC/ language 3
  - Structure of the codes (e.g. 250,000 SLOC Fortran-main code, 30,000 C++-problem set-up, 30,000 SLOC Python-steering, 10,000 SLOC PERL-run scripts,…)
- What libraries are used? And What fraction of the codes does it represent?
- Code Mix:
  - To what extent does your team develop and use your own codes?
  - Codes developed by others in the DOE and general scientific community?
  - Commercial application codes provided by the center?
- What is the present parallel scalability on each of the computers the code operates on
  - Projected or maximum scalability
  - How is measured?
  - Is the code massively parallel?
- What memory/processor ratio do your project require? (e.g. Gbytes/processor)
- Parallelization model (e.g. MPI, OpenMP, Threads, UPC, Co-Array Fortran, etc.) E.g. Does your team use domain decomposition and if so what tools do you use?
- What is the "efficiency" of the code
  - how is it measured?
- What are the major bottlenecks for scaling your code?
- What is the split between interactive and batch use?
  - Why, and is interactive more productive
- What is the split between code development on the computer center computers and on computers at other institutions?

5a. Project code productivity & scalability (Project-specific measures)

- Measures of experiment productivity and performance including scalability of runs
- Scaling limits including i/o, node memory size, interconnect b/w or latency, algorithm
- History of scaling
- Projected scalability

6. Scientific | Engineering Output

The scientific accomplishments 200x to present*:
The effect on the Office of Science programs*:

Publications/location:
Citations (last 5 years):
Dissertations?
Prizes and other honors?
Residual and supported, living datasets and/or databases that are accessed by a community? Size of the community?
Change in code capabilities and quality (t)
Code and/or data contributed to the centers
Code and/or data, results, contributed to the scientific and engineering community at large
Company spin-offs based on code or trained people and/or CRADAs
Corporation, extra-agency, etc. use
Production of scientists & computational scientists during 2001-2005
Production of trained software engineers during 2001-2005

*Parts 3 & 4 of metrics approach

Argonne NATIONAL LABORATORY

Argonne Showcase Projects

Ray Bair
Project Director
Argonne Leadership Computing Facility
July 17, 2006

THE UNIVERSITY OF CHICAGO
Office of Science U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is managed by The University of Chicago for the U.S. Department of Energy

Topics

- Argonne BlueGene/L Evaluation
- Selected 2006 INCITE Projects
  - Large Scale Simulations of Fracture in Disordered Media
    - *Phani Nukala, ORNL*
  - High Resolution Protein Structure Prediction
    - *David Baker, U. Washington*
- Early Petaflops Science Candidate
  - ASC FLASH Project
    - *Don Lamb, U. Chicago*

Blue Gene/L Evaluation at Argonne

Building a BlueGene Ecosystem

- ✓ Many people are familiar with the system (240 users on BGL)
- ✓ Many applications and tools are ported (over 40 on BGL)
- ✓ Open sharing of results and know how
- ✓ Active systems software development
- ✓ Vendor involvement with the community
- ✓ Productive and stimulating research

Many thanks to DOE, IBM, LLNL and many others

System accepted 1/31/05

Argonne's 5.7 teraflops system (BGL)
1024 nodes, 2048 processors, 512 Gbytes RAM
www.bgl.mcs.anl.gov

Blue Gene Community Activities

- Blue Gene Consortium
  - Formed by Argonne and IBM, April 2004
  - Over 60 member institutions
- Blue Gene Application Workshops
  - 2 day tutorial + expert assistance for groups of 6-12 user applications
  - 4 workshops held, including one for INCITE projects
  - Most all user applications run during workshop, many on 1024 nodes
- Blue Gene System Software Workshops
  - OS, File Systems, Resource Allocation, Systems Management, Optimization
  - 3 workshops to date
- Blue Gene Consortium Days at IBM Watson
  - IBM periodically provides 2 days of access on its 114 TF system
  - Users with success on Argonne system propose large runs
  - Successful projects may apply for additional Watson time

Blue Gene/L Consortium Members (62)

DOE Laboratories

- Ames National Laboratory/Iowa State U.
- Argonne National Laboratory
- Brookhaven National Laboratory
- Fermi National Laboratory
- Jefferson Laboratory
- Lawrence Berkeley National Laboratory
- Lawrence Livermore National Laboratory
- Oak Ridge National Laboratory
- Pacific Northwest National Laboratory
- Princeton Plasma Physics Laboratory

Universities

- Boston University
- California Institute of Technology
- Columbia University
- Cornell University
- DePaul University
- Harvard University
- Illinois Institute of Technology
- Indiana University
- Iowa State University
- Louisiana State University
- Massachusetts Institute of Technology
- National Center for Atmospheric Research
- New York University/Courant Institute

Universities (continued)

- Northern Illinois University
- Northwestern University
- Ohio State University
- Pennsylvania State University
- Pittsburgh Supercomputing Center
- Princeton University
- Purdue University
- Rutgers University
- Stony Brook University (SUNY)
- Texas A&M University
- University of California (*Irvine, San Francisco)*
- University of California/*SDSC*
- University of Chicago
- University of Colorado - JILA
- University of Delaware
- University of Hawaii
- University of Illinois – Urbana Champaign
- University of Minnesota
- University of North Carolina
- University of Southern California - ISI
- University of Texas at Austin – TACC
- University of Utah
- University of Wisconsin

Industry

- Engineered Intelligence Corporation
- IBM
- Gene Network Sciences

International

- Allied Engineering Corp - Japan
- ASTRON/LOFAR, The Netherlands
- Centre of Excellence for Applied Research and Training, UAE
- Ecole Polytechnique Fédérale de Lausanne, Switzerland
- Institut de Physique du Globe de Paris
- National Institute of Advanced Industrial Science & Tech - Japan
- National University of Ireland
- Trinity College, Ireland
- John von Neumann Institute, Germany
- NIWS Co., Ltd., Japan
- University of Edinburgh, EPCC Scotland
- University of Tokyo - Japan

Argonne BG/L Joins 2006 DOE INCITE Program

- **Innovative and Novel Computational Impact on Theory and Experiment**
  - Enables high-impact scientific advances
  - Solicits large computationally intensive research projects
  - Open to all scientific researchers and organizations
  - Provides large computer time & data storage allocations
  - Small number of 1-3 year projects via peer-reviewed proposals
- IBM Partners with Argonne to provide BlueGene/L Cycles
  - 10% of 2,048 processor system at Argonne (BGL)
  - 5% of 40,960 processor system at IBM T.J. Watson Research Center (BGW)
- 10.5M BG CPU hours awarded to 6 projects in Feb. 2006

http://hpc.science.doe.gov/

Argonne NATIONAL LABORATORY

INCITE: Large Scale Simulations of Fracture in Disordered Media: Statistical Physics of Fracture

PI: Phani Nukala

Co-PI: Srdjan Simunovic

THE UNIVERSITY OF CHICAGO | Office of Science U.S. DEPARTMENT OF ENERGY

Argonne National Laboratory is managed by The University of Chicago for the U.S. Department of Energy

Large Scale Simulations of Fracture in Disordered Media: Statistical Physics of Fracture

- PI: Phani Nukala, ORNL
- Co-PI: Srdjan Simunovic, ORNL

- The main aim of the proposal is to perform large-scale 3D simulations of lattice networks in order to understand the origin of scaling laws of fracture in disordered media. In particular, the study aims at understanding the origin of universality of crack surface roughness exponents. In addition to these 3D lattice simulations, we propose to study scaling of interfacial fracture, wherein the crack front is constrained to remain on the interfacial plane.
- The authors have developed a block-circulant preconditioner that can be used in conjunction with the conjugate gradient (CG) algorithm to perform large-scale massively parallel simulation of three dimensional lattice networks. This block-circulant preconditioner has been shown to be superior to the optimal circulant preconditioner and the Fourier acceleration technique that is traditionally used in performing 3D simulation of fracture networks.
- At present, numerical simulations are limited to a lattice system size of L = 48 in three-dimensions. Using the block-circulant preconditioner, this proposal aims at performing large-scale massively parallel simulations of 3D lattice systems of sizes L = 200. …Based on these large-scale simulations with strong disorder, we propose to investigate scaling laws of fracture, avalanche precursors, universality of fracture strength distribution, size effect on the mean fracture strength, and finally the scaling and universality of crack surface roughness.

1.0 Background and 2.0 Team Resources

- Support
  - DOE ASCR/MICS (for the last year)
- Scientific Domain
  - Materials Science
- Staff Profile
  - 2 PhD Scientists (Engineering/Physics, Mathematics)
  - Not affiliated with Argonne
- External Collaborators
  - U. Rome, HUT Finland, CNRS France/Grenoble, Virginia Polytechnic
- Science Goal
  - Understand the origin of scaling laws of fracture in disordered media
- External Communities
  - Collaborators are supported as users
- Team Resources Utilization
  - 40% code development (via 50% of Phani's time)
  - Developed code over last 4 years

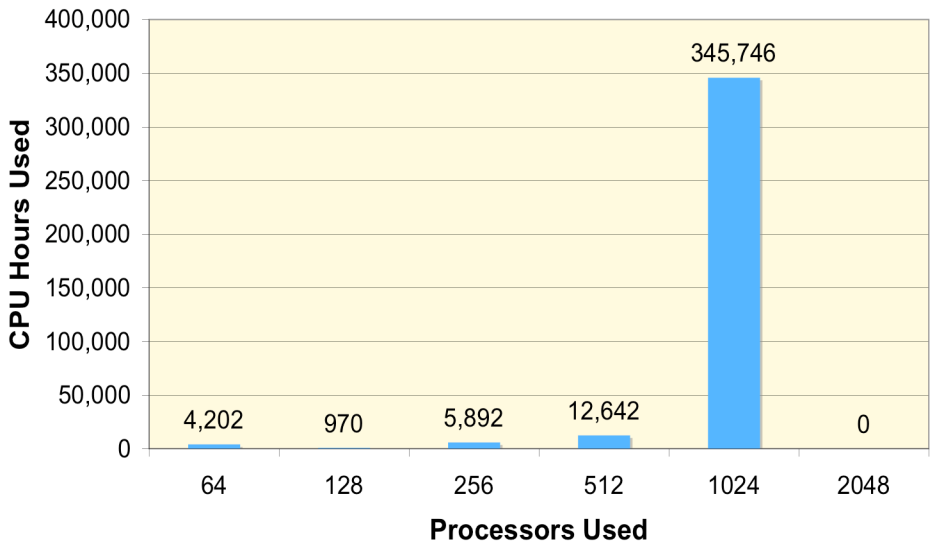
5.0 Software Engineering, Development, Verification and Validation Processes

- Software Development Tools Used
  - Parallel Solvers: PETSc
  - Debugging: IBM debuggers
  - Visualization: medit visualizer (mostly by collaborators)
- Software Engineering Practices.
  - Code Management: cvs
  - 2 person team uses 1-to-1 coordination, planning, etc.
  - No specific tools for bug tracking
- Verification Strategy
  - Suite of regression tests is checked every time a code change is made
- Validation Strategy
  - Main science intent is numerical validation against experiment

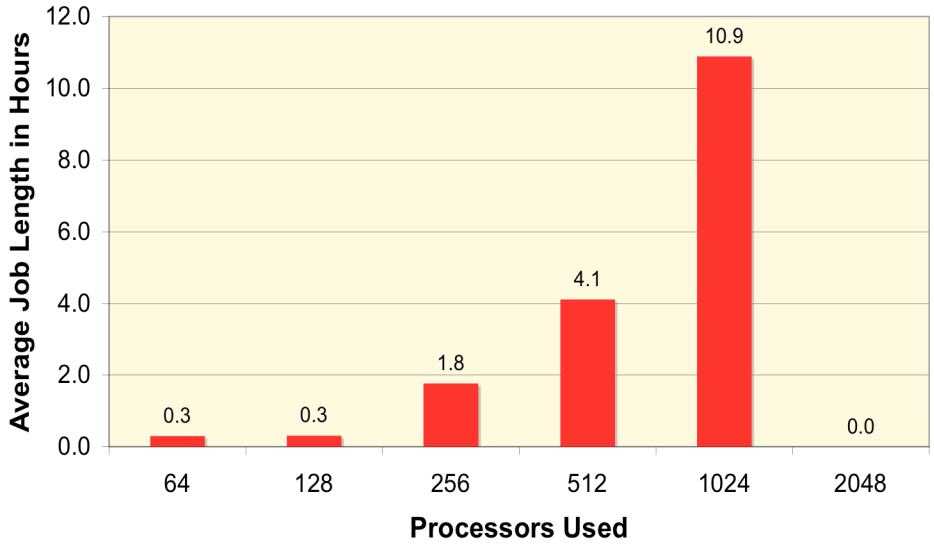
4.0 Project Resources from the Centers

- Plan with benchmarks & milestones
  - FY2006 INCITE Proposal, 3D Lattice System
    - *3D simulation at L=200 (lattice size)*
    - *10 sample ensemble simulation at L=128*
- Annual use of resources
  - Processor Time: FY2006 allocation 1.5M CPU hours
  - Disk: 300 GB per sample, then multiple samples per run
  - Tertiary Storage: N/A
- Software provided by center
  - PETSc library
- Consulting
  - BlueGene Applications Workshop Feb. 28-Mar. 2, 2006
  - Phone and e-mail support by Argonne applications and systems engineers

4.0 Project Resources Used (March-July 2006)



4.0 Project Resources Used (March-July 2006)

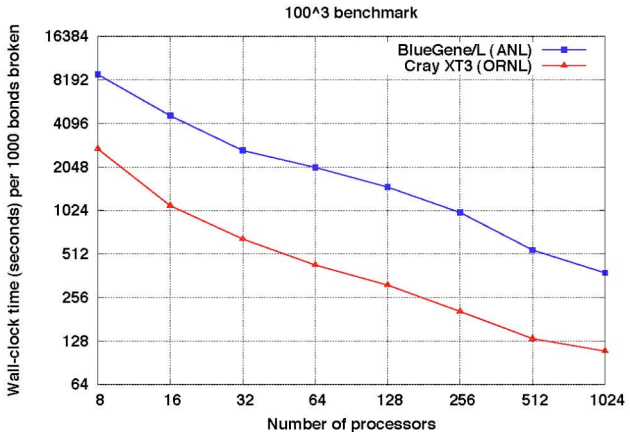


3.Project Code

- Problem Type
  - Simulation
- Algorithm
  - Random thresholds Fuse Model (RFM) with periodic boundary conditions
  - Block-circulant preconditioned conjugate gradient, iterative scheme
  - 80x improvement over pre-2003 algorithms
- Routine Platforms
  - IBM BlueGene/L, Cray XT3, Linux Clusters
- Code Size
  - Currently 3 years old, 45K lines, growing at 10-15K lines/year
- Computer Language
  - Fortran
- Libraries
  - Sparse matrix operations/solvers, including PETSc, MUMPS, SuperLU
  - Solvers are 95% of the work
- Efficiency Limitations
  - Memory intensive sparse matrix computation is limited by memory bandwidth
- Scaling Bottlenecks
  - Scaling and efficiency is dominated by the performance of the external solvers
  - Parallel code is so new that they have not done a lot of tuning

5a. Project code productivity & scalability

- Scalability on BG/L and XT3
- Science-specific measure is seconds per 1000 bonds broken (fuse model)



- Future plans for evaluation at larger scales

6. Scientific | Engineering Output

- Publications
  - 35 publications, 10-20 per year
- Citations (last 5 years)
  - 30-40 for 2003 papers (seems to be growing)
- Dissertations
  - None at ORNL, but some at collaborator universities
- Prizes/Honors
  - Phani Nukala: 2006 Science Spectrum Trailblazer Award
- Change in code capabilities and quality
  - Code growth from nothing in 3 years
  - In the past few months on INCITE specific physics options added
- Others use this code for diverse problems
  - Brittle fracture, grain boundary engineering, arctic sea-ice climate dynamics, flux through superconducting materials, and blackouts of power grid networks
- Corporation, extra-agency, etc. use
  - Contacted by industry and other labs
  - Starting collaborations with NASA and Army Research Center

Argonne NATIONAL LABORATORY

High Resolution Protein Structure Prediction

PI: David Baker, University of Washington

THE UNIVERSITY OF CHICAGO

Office of Science U.S. DEPARTMENT OF ENERGY

Argonne National Laboratory is managed by The University of Chicago for the U.S. Department of Energy

High Resolution Protein Structure Prediction

- PI: David Baker, University of Washington
- The goal of the proposed research is to compute structures for proteins of under 150 amino acids at atomic level resolution. Recent results with the Rosetta structure prediction method developed in my group suggest that the primary obstacle is adequate conformational sampling, and we will seek to overcome this bottleneck using the INCITE resources.

1.0 Project Background

- Scientific Domain
  - Computational Biology/Bioinformatics
- Code System
  - ROSETTA - http://depts.washington.edu/bakerpg/
- Support
  - HHMI, NIH, DARPA, ROSETTA license royalties
- Technical Goal
  - Predict unknown protein structures at atomic resolution
- Science Target
  - Genome scale globular protein function prediction
- External Community
  - Very large community of users

2.0 Project Team Resources

- Project Team Profile
  - INCITE Team: PI (Baker), 3 postdocs in chem/bio, 2 grad students
  - Plus 1 code support person
  - This is the effort for one module
- Extended Development Team
  - 65 people are coming to July Developer's Meeting
    - *From UW, UCSF, Vanderbilt, Johns Hopkins, NYU, UCSC, UNC*
  - No team members are from Argonne
- Development History
  - Code is 15 years into development
  - Code is used for production and development simultaneously

5.0 Software Engineering, Development, Verification and Validation Processes

- Software Development Tools Used
  - Debugger: gdb, core files
  - Visualization: in house distributed background fill visualizer
- Software Engineering Practices
  - Code Management: cvs
  - Quality Control: 1 full time person to maintain "code etiquette"
  - Developers work in areas of their interest
- Bug Reports and Questions (by function)
  - abinitio-support, docking-support, design-support, NMR-support, DNA-support, fragments-support, general-support@rosettacommons.org
- Verification Strategy
  - Nightly regression tests from home grown scripts
  - Nightly performance checks
- Validation Strategy
  - Some validation against experiment by other lab branches and other universities
- Experimental Facilities for Validation
  - Bio Lab (bakerlab.org)
- Does your project have adequate resources for validation? Yes

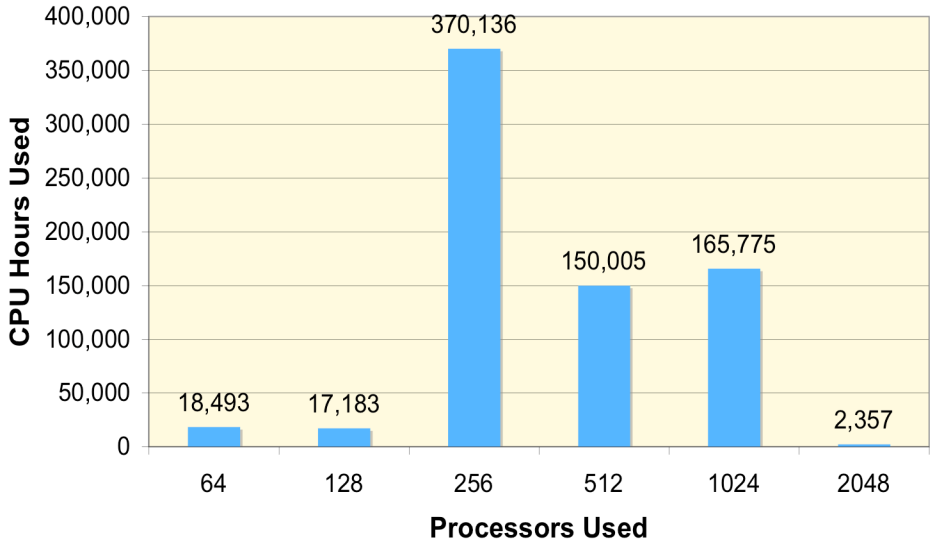
4.0 Project Plans and Milestones

- Research Plan from 2006 INCITE Proposal
  - First set of 250 proteins (year 1)
    - *2 domains (less than 150 amino acids) from each SCOP superfamily (one where there is only one structure in the superfamily)*
    - *ROSETTA low and high resolution structure prediction methodology will be used to generate models for the parent sequences*
    - *INCITE computational power will allow an order-of-magnitude improvement in the number of conformations assayed*
    - *Analyze these data by comparing to native structure*
  - Second phase 400-500 proteins (years 2-3)
    - *Functionally annotated proteins in the human genome for which no structural information is available and for which no sequence homologue has a known structure*
    - *Identify 100 proteins in the target lists of the structural genomics centers and predict structures for blind tests*
    - *Produce models for the CASP7 and CASP8 structure prediction tests to allow further independent evaluation of blind predictions*

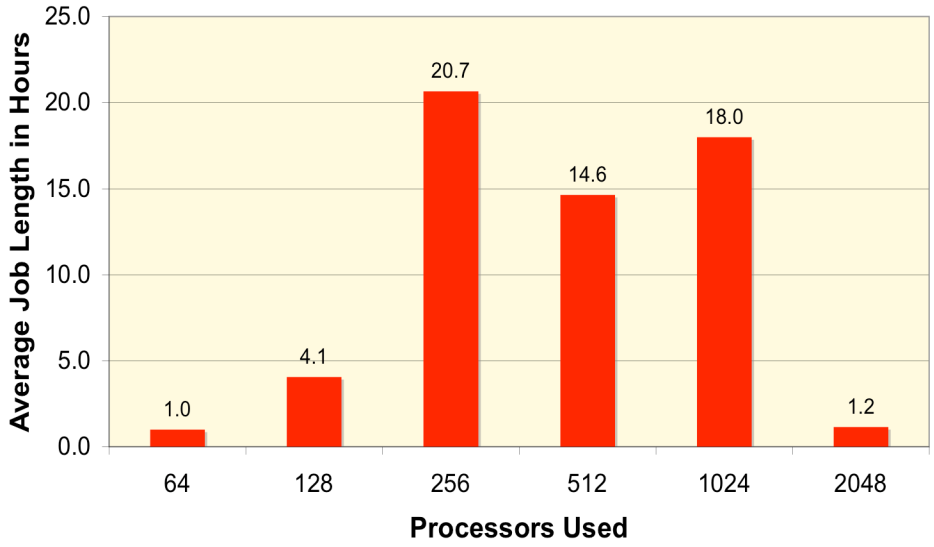
4.0 Project Resources from the Center

- Annual use of resources
  - Processor time: FY2006 allocation 5M CPU hours
  - Disk: about 1 GB of output per run, changes little over time
  - Tertiary Storage: N/A
- Software Provided by Center: MPI
- Consulting
  - BlueGene Applications Workshop Feb. 28-Mar. 2, 2006
    - *Porting and startup*
  - Phone and e-mail support by Argonne applications and systems engineers
- Scalability of the Code
  - No known limits
- Typical Runs
  - 512 processors for 4-5 days
  - CASP runs must be completed within 3 weeks of start of challenge

4.0 Project Resources Used (March-July 2006)



4.0 Project Resources Used (March-July 2006)



3. Project Code

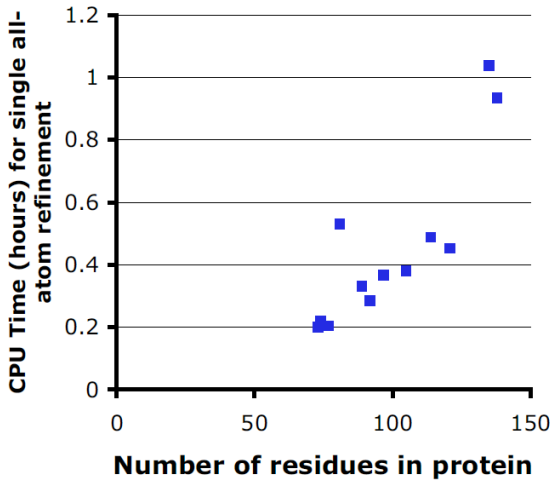
- Problem Type
  - Simulation
- Algorithm
  - Metropolis Monte Carlo minimization
    - *First stage low resolution sampling with randomly selected torsion angles*
    - *Best candidates sent to high resolution minimization*
- Platforms
  - Windows, Mac, Linux, AIX
- Code Size
  - 250,000 lines
- Computer Language Employed
  - Converted to c++ for Fortran
- Libraries Used
  - Library to allow Fortran data structure in c++ (tiny fraction of code)
  - MPI
- Code Mix
  - Code developed by Baker Group and extended by university collaborators

3. Project Code (continued)

- Parallel Scalability
  - Code distributes trials among processors
  - No known scalability limits
- Memory/Processor Ratio Required
  - Less than 256MB per processor
- Programming Model
  - MPI
- Major Bottlenecks
  - No known scaling limits, but detailed performance model has not been done
- Usage Modes
  - Batch jobs

5a. Project Code Productivity & Scalability

- Relationship of CPU time (in minutes) for all-atom refinement and scoring of a single protein conformational candidate to protein size



6. Scientific | Engineering Output

- Publications
  - 88 (entire Baker Lab)
- Citations (last 5 years):
- Dissertations
  - 20
- Prizes and Other Honors
  - David Baker has won many awards
    - *Recently: 2002 Overton Prize, 2004 AAAS Newcomb Cleveland Prize, 2004 Foresight Institute Feynman Prize in Nanotechnology*
  - Distinguished performance in Critical Assessment of Techniques for Protein Structure Prediction (CASP) challenges
- Code available to the scientific and engineering community
  - Rosetta is available under academic or corporate license terms

Argonne NATIONAL LABORATORY

ASC FLASH Project

Director: Don Lamb

1.0 Project Background

- ASC FLASH Center
  - Director: Don Lamb
  - http://flash.uchicago.edu/
- Support
  - DOE NNSA Advanced Simulation and Computing Alliances Program (ASC)
- Scientific Domain
  - Astrophysics compressible turbulence, combustion, radiation, etc.
- Support for code development
  - Entirely DOE ASC, via 10 year contract
- Technical goals of the project
  - The FLASH Center is funded to build a state-of-the-art simulator code for solving nuclear astrophysical problems related to exploding stars. Particularly, the methods of detonations in x-ray bursts, novae and type Ia supernovae.
- What is the project profile in total human resources including
  - Trained scientists: ~12
  - Computational scientists and mathematicians: ~ 12
  - Program development and maintenance: ~6
- External communities & sizes that code and/or datasets support: >200

2.0 Project Team Resources

- Astrophysics (group leader: James Truran).
  - This group is focused on the astrophysics calculations.
- Basic Physics (group leader: Todd Dupont).
  - This Basic Physics Group focuses on developing fundamental understanding of the detailed physical processes which underlie the astrophysics problems.
- Computational Physics and Validation (group leader: Todd Dupont).
  - The efforts of this group are concentrated on the development of algorithms for compressible and incompressible hydrodynamics, magnetohydrodynamics, relativistic flows, radiation transport, and methods of data analysis suitable for block-structured adaptive meshes.
- Computer Science (group leader: Rusty Lusk).
  - This group investigates and develops computer science infrastructure elements, including performance and optimization tools, tools for distributed and parallel computing, architecture standards, and data transport diagnostics.
- Flash Code (group leader: Anshu Dubey).
  - The Flash code group is focused on building and maintaining the code that carries out the core astrophysics calculations.
- Visualization (group leader: Mike Papka).

2.0 Project Team Resources

- Team Size: 30-40
- Team institutional affiliation(s)
  - University of Chicago
  - Argonne National Laboratory
- Are the code team members affiliated with the computer center? No
- Team composition by educational level (total)
  - ~55% Ph.D.
  - ~35% Graduate Students
  - ~5% Masters Degree
  - ~5% Bachelors Degree
- Team resources utilization
  - ~40% Astrophysics
  - ~40% Computational mathematics
  - ~20% Code Development

5.0 Software Engineering, Development, Verification and Validation Processes

- Software development tools used
  - Code management: svn, cvs
  - Debuggers: gdb, printf, totalview
  - Performance: TAU, Jumpshot
  - Visualization: IDL, flashviz (ANL tool)
- Software engineering practices.
  - Configuration management: automated etiquette enforcement
  - Bug reporting an tracking: bugzilla
  - Project planning and tracking
    - *weekly management meetings guide direction*
    - *weekly group meetings track progress*
- Verification strategy
  - Nightly check of correctness and performance
- Regression tests
  - Nightly regression and performance benchmarks
- Validation Strategy
  - Direct comparison with LLNL experiments lead to important results
  - Research in sensitivity analysis

4.0 Project resources input from the centers

- Plan with benchmarks & milestones
  - See http://flash.uchicago.edu/website/research/
- Steady state user of resources on a production basis per month
  - Porting and benchmarks on Argonne BlueGene/L; production elsewhere (millions of CPU hours per year)
- Software provided by center
  - mpi, hdf5/pnetcdf, fast math libraries
- Consulting
  - Phone and e-mail support by Argonne applications and systems engineers
  - Argonne applications engineer is former FLASH team member
- Size of their jobs in terms of memory, processors, disk, and tertiary store
  - Any range of processors and memory that is available
- Code scalability
  - Near perfect weak scaling to 64K nodes (and likely beyond) on BlueGene
  - Also excellent strong scaling
- Wall-clock time for typical runs
  - Big science runs take weeks on large machines

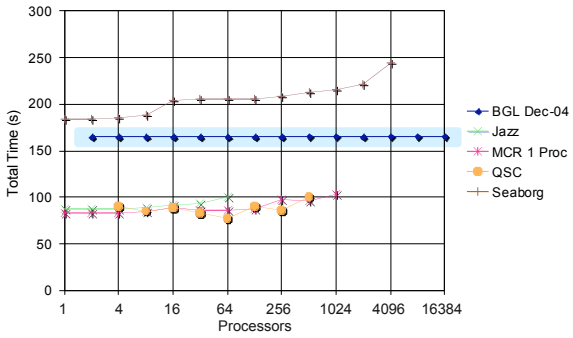
3. Project Code

- Problem Type: simulation
- Types of algorithms and computational mathematics
  - Block structured adaptive grid, explicit operator split finite volume hydro (PPM), multigrid/multipole, burning is an ODE solve
- Platforms
  - Linux (Intel, AMD), AIX, BlueGene/L, Sun, Mac, Compaq, NEC, SGI
- Code size
  - 9 years old, 500,000 lines, growing at about 10,000/year
- Computer languages employed
  - 90% Fortran
  - 5% c
  - 5% Python (at setup time)
- Code Mix
  - External: paramesh (NASA Goddard)

3.0 Project Code

- Present parallel scalability
  - See slide 4.0; to 64K nodes and beyond
- Memory/node ratio required
  - Varies; 512MB adequate for some problems
- Parallelization model
  - Adaptive, fixed, and uniform grids
- Efficiency of the code
  - Regularly gets 20-30% of peak CPU performance (via hardware FLOP counts)
  - Certain kernels are highly tuned and the entire code is written to encourage easy compiler optimizations, but it is large and complex.
  - Largest single hit, ever, was when we could no longer use IPA on any compilers because of the size and flexibility of the application.
- Split between interactive and batch use
  - All batch

Flash Hydrodynamics – Weak Scaling



Pure hydrodynamics scaling runs at NERSC (Seaborg), LANL (QSC), LLNL (MCR, BGL), ANL (Jazz)

Problem size is increased in proportion to number of processors

Recent 5-day run at LLNL

- Direct numerical simulation of 3-D, homogeneous, isotropic, weakly-compressible turbulence at one of the highest effective Reynolds numbers ever attempted.
- Gathered extensive statistics of Lagrangian tracer particles embedded within a simulated turbulent flow at effective Reynolds numbers > 500-1000.

5a. Project code productivity & scalability

- Bottlenecks for scaling and performance
  - Global gravity solve
  - FLASH is primarily limited by memory bandwidth.
- History of scaling
  - From the start, the code was designed to be scalable on large systems.
  - Largest bottleneck has always been the regridding of the adaptive grid and the global solve for gravity. But, new algorithms have partly overcome the regridding problem.
- Projected scalability
  - Currently, we expect to be able to scale to 100K + nodes
  - This is through extrapolation from real 64K data and performance models

Example petascale problem

- Whole star 3-D simulation of the gravitationally confined detonation mechanism
  1. Off-center ignition through breakout of a hot bubble produced by turbulent nuclear burning;
  2. Rapid spreading of the hot bubble material across the stellar surface, convergence of the hot bubble material at the opposite point on the surface of the star, and initiation of a detonation; and
  3. Propagation of the detonation supersonically through the entire star and the subsequent explosion of the star.

6. Scientific | Engineering Output

- Publications: 44
- Citations (last 5 years):
- Dissertations: ~20
- Prizes and other honors: 1999 Gordon Bell Award
- Community datasets: building a community dataset for turbulence data
- Change in code capabilities and quality: many new capabilities in the last 5 years
- Company spin-offs or CRADAs: None
- Production of scientists & computational scientists during 2001-2005: ~14
- Production of trained software engineers during 2001-2005: ~6

Dear Gordon:

We certainly appreciate the importance of computational research and development programs with strong science impact, and we support the efforts of your committee to examine how ASCR's production computing facilities are assessed and their impact on science. The Argonne Leadership Computing Facility is deep into its planning stages, considering many of the same questions as your committee to optimize the effectiveness of our facility for this class of science. So our responses below have been extracted from our facility plans, and reflect our current thinking about the center we will begin to physically construct in the Fall.

Sincerely,

Ray Bair
Project Director
Argonne Leadership Computing Facility

1.0 Overview of Resources Provided by the Center

a. Contact information for the project

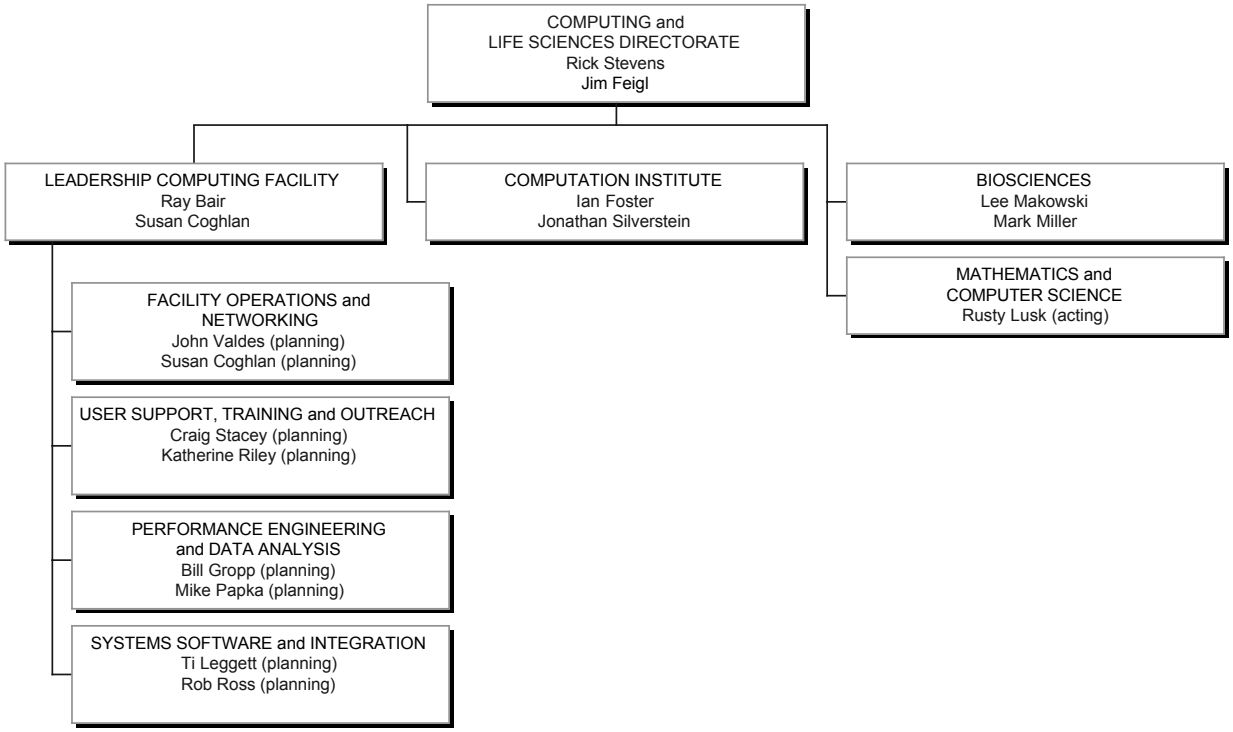
Principal Investigator
Rick L. Stevens
Associate Laboratory Director
Computing and Life Sciences
Building 221
9700 South Cass Avenue
Argonne, Illinois 60439
Email: stevens@mcs.anl.gov
Phone: 630-252-3378
Fax: 630-252-6333

Project Director
Raymond A. Bair
Argonne Leadership Computing Facility
Building 221
9700 South Cass Avenue
Argonne, Illinois 60439
Email: bair@mcs.anl.gov
Phone: 630-252-5751
Fax: 630-252-6104

We do not have an online staff directory or web site yet. The site for the pre-production BlueGene/L evaluation system is at http://www.bgl.mcs.anl.gov/

b. Organizational structure with staff sizes and functional titles (separate page)

The new ALCF organization is responsible for production high-end computing at ANL and distinct from the existing Mathematics and Computer Science Division. The planned organizational structure is:



c. FTEs

Staffing levels are under discussion with DOE. We expect to have 40-50 program-funded FTEs when fully staffed, plus some FTEs from overhead. Each of the four groups named in the organization chart above will have 7-11 people. The balance of the staff are in the Catalyst Team (science project liaisons) and facility management in the facility front office. The organization is explicitly tailored to support the cycle of activities carried out by leadership science projects, throughout their computational campaigns.

d. Physical infrastructure

ALCF will be housed in a new computer room on the Argonne site. Computer room plans include space for multiple generations of ALCF systems. The inaugural 500-1000 teraflops BlueGene/P system, its storage and support capabilities will occupy 7000-9000 sq. ft., and draw 3,000-4,000 KW (depending on the final budget). In addition, new office space will enable us to collocate facility staff with computer science, mathematics, and computational science staff from multiple disciplines, as well as visiting scientists. FY2006 electricity costs are $0.045/KW hr, and ample power is available for considerable expansion. Cooling can be scaled to meet both short and long term facility needs. ALCF will start with three 10Gbps WAN connections to ESnet, UltraScienceNet, and Internet2, and plans to increase performance to track progress in those networks.

e. Balance sheet and budget for:

Approved budgets are not available for ALCF at this time. However the FY2007 President's Budget includes $22.5M for ALCF, an amount that is expected to increase in the out-years.

f. Institutional affiliation and degree of institutional support

ALCF is hosted by Argonne National Laboratory, operated by the University of Chicago for DOE. Argonne has integrated high performance computing into its whole scientific agenda, and is providing considerable support for planning, infrastructure and conventional facilities.

g. Present and planned hardware

In the near term, ALCF is pursuing the IBM BlueGene series of computers. Based on our highly successful evaluation of BlueGene/L, we plan to deploy a large BlueGene/P system in 2007-2008. This system will come in multiple stages, starting with a 100 teraflops system, and culminating with a 500-1000 teraflops system. A 1000 teraflops system would have 72K nodes, each node with a quad core processor, for a total of 294,912 processors. ALCF plans call for the large memory configuration, 4 GB per node, which sums to 288 terabytes on a petaflops system.

BlueGene also has highly scalable I/O capabilities, moving to 10 Gbps Ethernet interfaces in the coming models. Near line disk storage plans are for 10-16 petabytes of high performance SAN storage.

ALCF plans call for tertiary tape storage scaling to a potential 150 petabytes or more over the life of the BlueGene/P system, as needed.

h. Software development and production tools provided top 5 (enumerate on separate pages)

The BlueGene/P software development stack will include:

- IBM Fortran and C/C++ compilers (xlf, xlc)
- IBM Math Libraries (ESSL, MASS/V)
- Community Math Libraries (FFTW, PETSc, BLAS, LAPACK)
- Performance and Debugging Tools (IBM HPC Toolkit, TAU, Kojak, PAPI)
- Parallel I/O Libraries (HDF5, pNetCDF)
- Parallel Communications and I/O (MPICH, ROMIO, ARMCI/Global Arrays)

i. Application codes available to the users that are supported by the center (ISVs, open source, etc.) top 5 enumerate with software development tools listing

The set of supported codes will depend upon the projects that are given allocations on the system. Many community codes have already been ported to BlueGene (see 5.d below).

j. What auxiliary services do you offer your users

The ALCF is planned to have a data analytics cluster, with data reduction/analysis and rendering services. In addition large format visualization displays will be available in the facility.

2.0 User interface and communication including satisfaction monitoring and metrics

a. How do you measure the success of your facility today in being able to deliver service beyond the user surveys (e.g. the NERSC website)?

With ALCF's plans for a relatively small community (20 major projects, plus small development projects), we plan to keep abreast of the plans, campaigns, problems, and progress of each major project individually. This is one of the key roles of our Catalyst Team members who will each maintain contact with 6-7 projects.

b. Do all users-experimenter teams, team members, and any users that the team community provides utilize the survey?

ALCF plans to survey all its users.

c. Have these surveys been effective at measuring and understanding making changes in operations? (Please cite)

N/A. (ALCF is not operational yet.)

d. Describe your call center – user support function: hours of coverage, online documentation, trouble report tracking, trouble report distribution, informing the users, how do users get information regarding where their job/trouble report is in the queue?

ALCF plans to provide call-in emergency response 24x7x365, plus call-in customer support on weekdays, from 8 AM to 6 PM Central Time. Regardless of how ALCF is contacted or who is contacted within ALCF, the user's trouble ticket will be handled by the appropriate staff member(s). Users will receive prompt e-mail (or phone calls as necessary) informing them of the status of their problem, and repeated updates if the problem takes a while to address.

e. What mechanisms are provided for the user with respect to dissatisfaction with how a case is being handled?

ALCF will have a published problem escalation process.

f. What mechanisms are provided to support event-driven immediate access to your facility (e.g. Katrina or flu pandemic)

Certainly ALCF will support urgent national needs such as those mentioned in this question, rescheduling workload as needed. To provide rapid service, the applications which may be needed must be kept in a ready state (ported, validated and ready to execute), and revalidated each time the system software is updated. We have experience in this area and would be willing

to work with designated projects.

3.0 Qualitative measure of output

In a separate letter, NERSC, ORNL, and Argonne commented jointly on measures of scientific output and facility effectiveness in the context of the current PART metrics, with suggestions for new metrics.

a. Do you measure how your facility enables scientific discovery?
b. How are the results of measurement disseminated and how do they further Science and especially DOE Science Programs?
c. What impact have any of your measures had on operation of your facility?
d. What impact have the current PART measures had on your successful operations of your facility?
e. What do you view as the appropriate measures for supercomputing facilities now?
f. During the next 3-5 years?

4.0 Aggregate Projects use profiles by scale

a. How many projects does your center support?

ALCF is planning around the following project types and distribution, when we reach full production:

- ~20 Leadership Science Teams addressing the most computationally challenging science problems. These teams consume ~85% of the available cycles. We estimate ~200 users associated with these projects.
- ~5 Computer Science Testbed Teams, scaling up the next generation of systems software and numerical algorithms. These teams consume ~5% of the available cycles. We estimate ~25 users associated with these projects.
- ~60 Application Development Teams, scaling up the next generation of science codes. These teams consume ~5% of the available cycles. We estimate ~100 users associated with these projects.
- In addition 5% of the available time is reserved for projects selected by the SC Director.

b. How many users are associated with all the projects?

See estimates above.

c. How many additional users who either use project data-sets or other center resources?

This has not been determined.

d. What is the project usage profile in terms of processor count? We would like these broken down into jobs that require, or can exploit a concurrency level of (roughly) 50, 200, 400,

1,000, 2,000, and 4,000 processors to obtain the science.

N/A. The distribution will depend upon the projects assigned to ALCF.

5.0 Center x User Readiness for 10x processors expansion

The mid-term goals for each facility call for a major expansion from machines with of order 5,000 processors to machines of order 50,000 processors or more.

a. Please outline how the center will accommodate this growth over the next 3-5 years.

The petaflops BlueGene/P system planned for ALCF is quite similar in scale to the successful BlueGene/L system currently at LLNL. We will move from 64K nodes to 72K nodes, from 128K processors to 288K processors. Applications that run well on BlueGene/L are expected to do well on BlueGene/P too, so the many groups developing applications on today's BlueGene/L systems will be in a good position to scale up to BlueGene/P. We are also continuing our BlueGene Applications Workshop series (see 5.b below). In addition we will be working with a set of early science projects so their codes will be optimized and ready when the system is accepted (also see 5.c below).

b. What do you believe is your role in preparing users for this major change?

We began to prepare the community with the creation of the BlueGene Consortium in April, 2003, and the installation of our BlueGene/L evaluation system in January, 2005. The Consortium currently has over 60 member institutions and 200 individual members. We run both application porting workshops and systems software workshops at regular intervals. In addition, we began to host DOE INCITE projects in 2006. All together, the BlueGene community ecology is widespread and functioning well in dissemination of expertise.

c. What effort (in terms of personnel) is devoted to code development issues today, and do you view this as adequate coverage as we move to machines with more than 25,000 processors?

The ALCF plans include establishment of a Performance Engineering and Data Analysis Group, with the mission of assisting applications development teams with performance evaluation, performance optimization, algorithm selection, and performance validation.

d. Are there codes in your user portfolio that will scale today to 10,000, 25,000, or 75,000 processors. What is the nature of these codes (Monte Carlo, CFD, hydro?) Are these codes running today on other systems of comparable size?

Over 80 applications have been ported to BlueGene/L, spanning many domains. Many are already running at 8K processors and above, some to 128K processors. These include:

- Electronic structure (Qbox, LSMS, QMC)
- Molecular Dynamics (CPMD, NAMD, ddcMD, MDCASK, BlueMatter)

- Computational Fluid Dynamics/Multiphysics (NEK5, SAGE, Miranda)
- Nuclear Theory (GFMC)
- Quantum Chromodynamics (QCD, MILC)
- Astrophysics (FLASH, ENZO)

e. As machines become more complicated, what do you see as the challenges to your success? For example, are you (or parts of your institution) actively involved in research related to fault-tolerance, memory/bandwidth contention, job scheduling, and etc. on the future machines?

Argonne is actively involved in research related to a wide range of relevant issues, through the efforts of the Mathematics and Computer Science Division.

- Scalable, Fault Tolerant, Systems Software (to millions of processors)
- High Performance, Scalable Parallel File Systems
- High Performance Data Transport (over LAN and WAN)
- High Performance, Scalable Message Passing
- Advanced Programming Models and Languages
- Next Generation Systems Architecture

f. How do you determine the path forward for your organization?

ALCF is establishing both a Leadership Computing Science and Technology Advisory Committee and a Leadership Computing User Advisory Committee to provide input to the organization. In addition, we will incorporate lessons learned from users, feedback from Program Offices, and results of other's research in formulating our plans.

g. What do your users want to see in the largest machines now available and those which will be available in the 3 year and 5-7 year time frames? (memory per core/node, number of processors, disk space?)

Access to large systems is critical in leadership science. Our current BlueGene evaluation and INCITE users most frequently request access to larger BlueGene systems (10x or more), with larger memory, for substantial periods of time (days to weeks). Through a partnership with IBM's T.J. Watson facility, approved users are able to do scaling and science runs on IBM's 100 teraflops, 40K processor, system, and most with considerable success.

Argonne, LLNL and IBM have partnered in separate project, funded by NNSA and Office of Science (SC), to develop the next two models of BlueGene (the P and Q systems). IBM innovations will be combined with NNSA and SC applications requirements to shape these BlueGene systems for our communities.

Argonne NATIONAL LABORATORY

Argonne Leadership Computing Facility: Comments on Plans and Metrics

Ray Bair
Project Director
Argonne Leadership Computing Facility
July 18, 2006

THE UNIVERSITY OF CHICAGO
Office of Science U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is managed by The University of Chicago for the U.S. Department of Energy

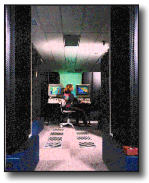
Topics

- DOE's next Leadership Computing Facility
  - ALCF systems, 2007-8
  - ALCF science opportunities
- ALCF plans
  - Surveying Users
  - Timely Assistance
  - High Availability
  - Capability Science Runs

ALCF

Over 20 years of Advanced Systems for DOE and Others

- **ACRF** period [1983-1992]
  - DOE's founding ACRF
  - Explored many parallel architectures, developed programming models and tools, trained >1000 people



- **HPCRC** period [1992-1999]
  - Production-oriented parallel computing for Grand Challenges in addition to Computer Science.
  - Fielded 1st IBM SP in DOE

- **TeraGrid** [2001-present]
  - Overall Project Lead
  - Defining, deploying and operating the integrated national cyberinfrastructure for NSF
  - 9 sites, 18 systems, 94TF
- **LCRC** [2003-present]
  - Lab-wide production supercomputer service
  - All research divisions, 74 projects, 360 users
- **BlueGene Evaluation** [2005-present]
  - Founded BlueGene Consortium with IBM
    - *60 institutions, 260 members*
    - *Applications Workshop Series*
  - 240 users, 6 INCITE Projects

ALCF

Mission and Vision for the ALCF

Our Mission

Provide the computational science research community with a world leading computing capability dedicated to breakthrough science and engineering.

Our Vision

A world center for computation-driven discovery that has

- outstandingly talented people,
- the best collaboration with computational scientists, computer scientists and applied mathematicians,
- creative, responsive and dedicated user support,
- the most capable and interesting computers and,
- a true spirit of scientific discovery.

ALCF

Desired Modes of Impact for Leadership Computing

1. Generation of significant datasets via simulation to be used by a large and important scientific community
   - Example: Providing a high-resolution first principles turbulence simulation dataset to the CFD and computational physics community
2. Demonstration of new methods or capabilities that establish feasibility of new computational approaches that are likely to have significant impact on the field
   - Example: Demonstration of the design and optimization of a new catalyst using first principles molecular dynamics and electronic structure codes
3. Analysis of large-scale datasets not possible using other methods
   - Example: Computationally screen all known microbial drug targets against the known chemical compound libraries
4. Solving a science or engineering problem at the heart of a critical DOE mission or facilities design or construction project
   - Example: Designing a passively safe reactor core for the Advanced Burner Reactor Test Facility

ALCF

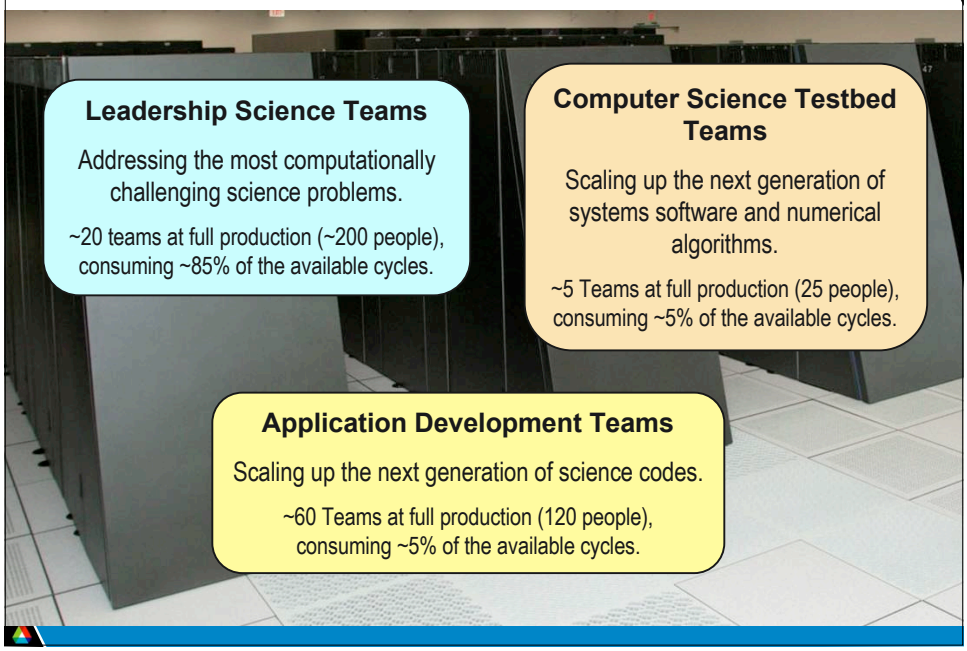
DOE Applications Drivers and Example Codes

Over 80 major applications have been ported to BG

- Computational Materials Science and Nanoscience
  - Electronic structure, First Principles ⇒ Qbox, LSMS, QMC
  - (mat) Molecular dynamics ⇒ CPMD, LJMD, ddcMD, MDCASK
  - Other materials ⇒ ParaDIS
- Nuclear Energy Systems
  - Reactor core design and analysis ⇒ NEK5, UNIC
  - Neutronics, Materials, Chemistry ⇒ QMC, Sweep3D, GAMESS
- Computational Biology/Bioinformatics
  - (bio) Molecular dynamics ⇒ NAMD, Amber7/8, BlueMatter
  - Drug Screening ⇒ DOCK5, Autodock
  - Genome-analysis ⇒ mpiBLAST, mrBayes, CLUSTALW-mpi
- Computational Physics and Hydrodynamics
  - Nuclear Theory ⇒ GFMC
  - Quantum chromo dynamics ⇒ QCD, MILC, CPS
  - Astrophysics/Cosmology ⇒ FLASH, ENZO
  - Multi-Physics/CFD ⇒ ALE3D, NEK5, Miranda, SAGE

ALCF

ALCF Science Community



Leadership Science Teams

Addressing the most computationally challenging science problems.

~20 teams at full production (~200 people), consuming ~85% of the available cycles.

Computer Science Testbed Teams

Scaling up the next generation of systems software and numerical algorithms.

~5 Teams at full production (25 people), consuming ~5% of the available cycles.

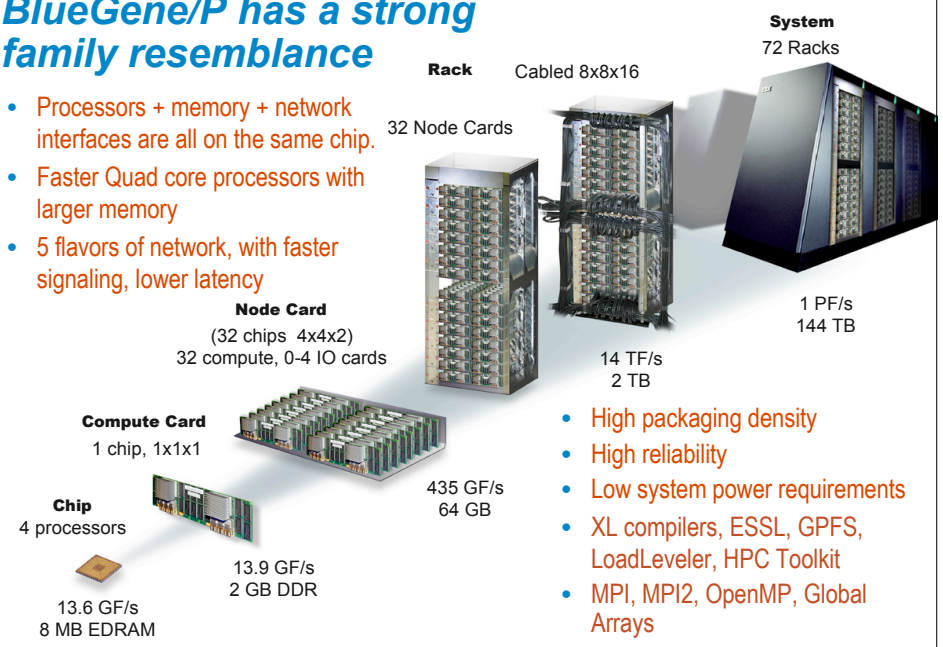
Application Development Teams

Scaling up the next generation of science codes.

~60 Teams at full production (120 people), consuming ~5% of the available cycles.

BlueGene/P has a strong family resemblance

- Processors + memory + network interfaces are all on the same chip.
- Faster Quad core processors with larger memory
- 5 flavors of network, with faster signaling, lower latency

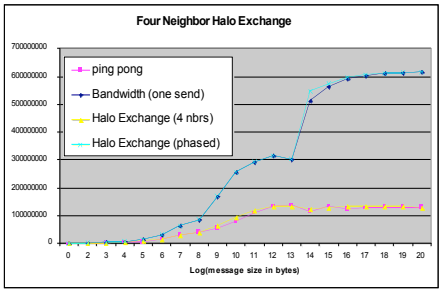


- High packaging density
- High reliability
- Low system power requirements
- XL compilers, ESSL, GPFS, LoadLeveler, HPC Toolkit
- MPI, MPI2, OpenMP, Global Arrays

BlueGene community knowledge base is preserved

Some Unique Features of Blue Gene

- Multiple links may be used concurrently
  - Bandwidth nearly 5x simple "pingpong" measurements
- Special network for collective operations such as Allreduce
  - Vital for scaling to large numbers of processors
- Low "dimensionless" message latency
- Low relative latency to memory
  - Good for unstructured calculations
- BG/P improves
  - Communication/Computation overlap
  - MPI-I/O performance



ALCF

ALCF System Deployment Plan

Compute Systems

- 100 TF Blue Gene/P
  - Arrives: Summer 2007
  - Early Science: Fall 2007
  - Leadership Projects: Winter 2007
- 500 TF Blue Gene/P
  - Arrives: Early 2008
  - Early Science: Spring-Summer 2008
  - Leadership Projects: FY2009
- Petaflops Blue Gene/P
  - Project Option

Storage Systems

- 1.5-2 PB High Performance Disk
- Data Analytics System
- 8-10 PB Tape Archive
- 11-16 PB High Performance Disk
- Data Analytics System
- 30-40 PB tape archive
  - Growing to 100-150 PB

ALCF

Goal #1: User Satisfaction

- Meeting the metric means that the users are satisfied with how well the facility provides resources and services.
- Metric #1.1: Users find the systems and services of a facility useful and helpful.
- Metric #1.2: Facility responsiveness to user feedback.

ALCF

Accessing ALCF User Satisfaction (Goal #1)

- Survey all users annually
  - Constructed carefully and drawing on most informative approaches used by other Centers
- Make key questions consistent across DOE Centers
  - Common language and rating scales in areas of DOE or OMB metrics
- Employ Best Practices
  - Opportunity to share ideas among large Centers run by DOE
  - Also TeraGrid/ETF/CIP, State and University Centers
- Issues for Leadership Centers
  - Good sample size is needed from a small user community
  - Tradeoffs between survey length and response rate

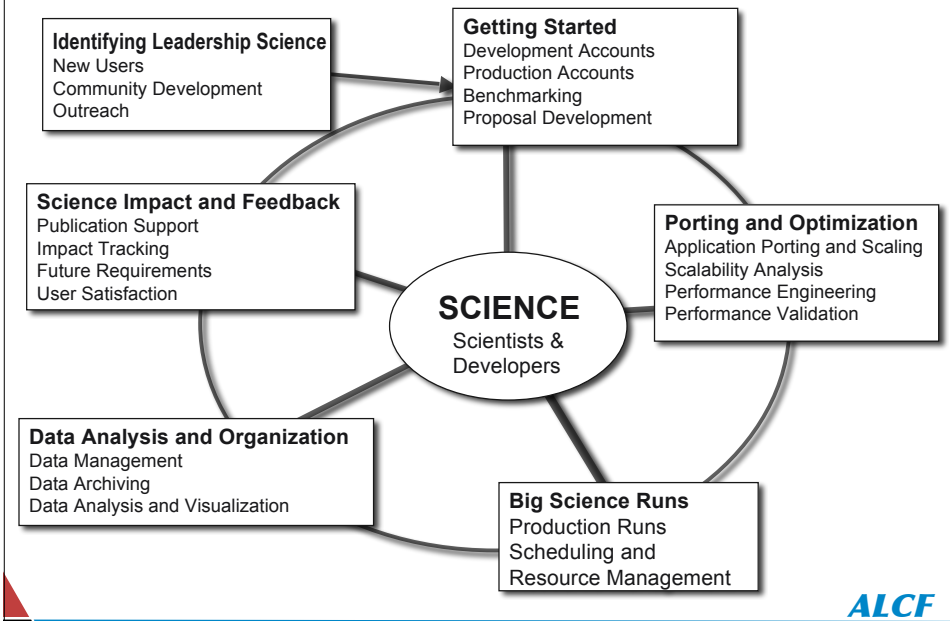
ALCF

Goal #3: Facilities provide timely and effective assistance

- Helping users effectively use complex systems is a key role that leading computational facilities supply. Users desire their inquiry is heard and is being worked. Users also need to have most of their problems answered properly in a timely manner.
- Metric #3.1: Problems are recorded and acknowledged
- Metric #3.2: Most problems are solved within a reasonable time

ALCF

Leadership Science Life-Cycle



ALCF

Big Systems Challenges

Problems & Solutions We Have Experienced

Just Getting Started….

- Getting accounts, credentials, time allocation, and so on
- Reproducing work
  - Porting applications
  - Finding libraries
  - Solving known problems
  - Finding existing performance data for the system
- Coordinating resources
- Porting
  - Recommendations
  - Common problems
  - Fast turn around test runs
- Demonstrating (and testing) application scaling

… Then Doing Large Runs

- Using Dedicated Time
  - Arranging for scheduling
  - Real-time resolution of problems
- Overcoming I/O Challenges
  - Commonly the largest porting barrier
  - Getting the data
- Achieving Scale
  - Scaling, performance, and debugging
  - Fast turn tests
  - Tools
- Moving data

ALCF

What LSTs Need for Success

System Solutions

- Facility-specific documentation
  - Software available
  - Experiences from previous users
    - *Performance*
    - *Issues*
    - *Code & script examples*
  - HowTo's for facility's systems and processes
  - FAQs
  - Schematics
- Storage/Data Infrastructure
  - Fast - internally and externally
  - Available
  - Easy to use
- Comprehensible scheduler

Expertise

- Focused help during startup
  - Someone to shepherd through the startup process
  - Expert-on-tap
  - Ability to visit if needed (e.g. for demo)
- Designated facilitator over project lifetime
  - Answer quick questions
  - Experience with common problems
  - Direct more detailed questions appropriately
  - Expert on system performance
  - Knowledgeable in science field
  - Arrange for reservations
  - Help solve specific problems

ALCF

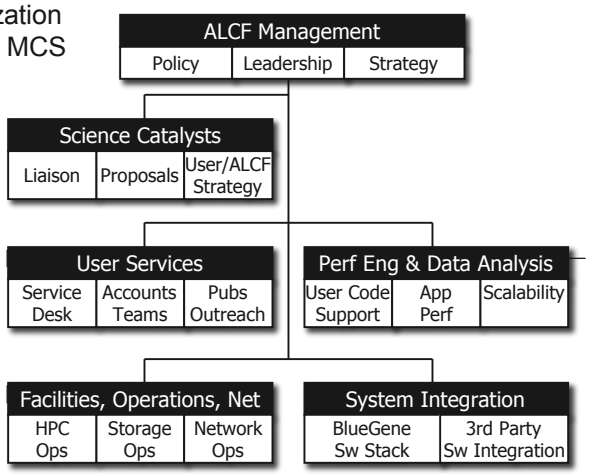
Leadership Computing Challenges Shape Response Planning

- Large application codes
  - Take time to comprehend
  - *Need to develop knowledge in anticipation of need*
- Large distributed teams
  - Current user may not be an expert in the problem area
  - *Need to develop knowledge of project organization*
- Scale-up exposes unexpected problems
  - Failures at the most inconvenient time
  - Frequently performance takes a hit
  - *Need to be able to marshal expertise quickly*
- Time allocations are used in large chunks
  - Raising the importance of each run
  - *Need to understand science campaign and be able to respond to changing plans*

ALCF

Creating a Responsive Organization (Goal #3)

- A new organization separate from MCS



- Responsiveness requires both process and culture
- Helping others is a calling

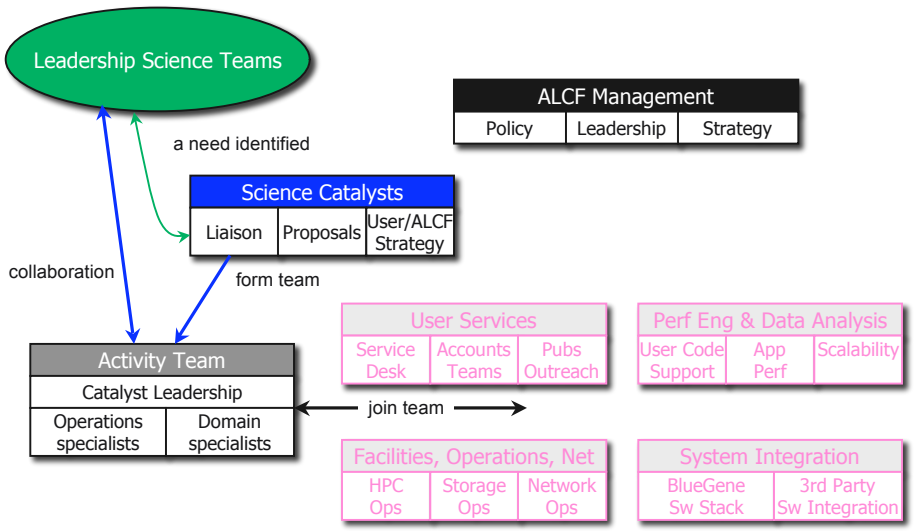
ALCF

Catalyst POC knows each Leadership Project

- Deeply understand each leadership project
  - Scientific goals
  - Development plans
  - Computational campaign plans
- Maintain a “dashboard” on each leadership project
  - Current issues and needs
  - Outstanding problems and their status
  - Hero run reservations
- Review key items on dashboards frequently
  - Assemble task forces as needed to help resolve complex issues
  - Assist in performance evaluation and optimization

ALCF

Scientific Support: Campaigns



ALCF

Goal #2: Office of Science systems are ready and able to process the user workload.

- Meeting this metric means the machines are up and available most of the time. Availability has real meaning to users.
- Metric #2.1: Scheduled (or overall) availability
- Value: High measured availability

ALCF

ALCF Availability for Science (Goal #2)

- System architecture and hardware contain features that increase reliability
  - BlueGene has few parts/node, no socketed parts, long MTBF/TF
  - Minimal compute node kernel reduces points of failure and instabilities
  - File/Tape systems robust to common failures (server, link/NIC, controller, drive)
- Early Science projects play important roles in start up
  - Contribute to acceptance tests
  - Exploit post-acceptance availability to carry out new science
- Minimizing time to repair
  - BG Reliability, Availability, Serviceability database supports trend analysis
  - Working with IBM to shorten diagnostic time
  - Trained technicians and spares on site
- Tradeoffs in availability and cost
  - Nx5 core hours + 24x7 emergency response
  - Will evaluate impact on availability and adjust accordingly

ALCF

Goal #4: Facility facilitates running capability problems

- Major computational facilities have to run capability problems. This is a complex goal that has many aspects which contribute to meeting the metric.
- Metric #4.1: The majority of computational time goes to capability jobs.
- Metric #4.2: Capability jobs are provided excellent turnaround

ALCF

Capability Jobs on ALCF BlueGene/P (Goal #4)

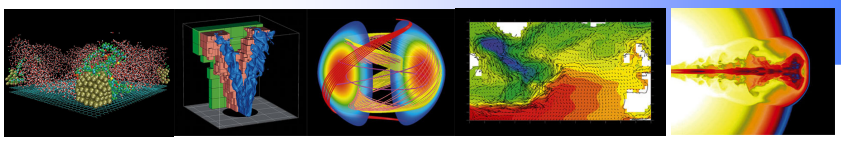
- By charter ALCF will focus on large capability runs
  - ~20 Leadership Class projects
  - Typical leadership project allocation on petaflops Blue Gene/P Annually ~100,000,000 CPU hours or ~1,000 rack-days
- BlueGene architecture is geared to large runs
  - Large partitions are 3D volumes of nodes, formed from contiguous sets of whole racks
    - *72 racks (72x32x32), or 40 racks (40x32x32) + 32 racks (32x32x32)*
  - A big partition can be run as several smaller partitions
    - *32 racks = 2 x 16 racks, or 4 x 8 racks, or 16 + 2 x 8 + 4 x 4*
- Job and queue policies shape user behavior and the scale and duration of jobs
  - Emphasis will be on jobs using many racks

ALCF

Exciting Times

- This is truly an exciting time for high performance computing
  - Many interesting computer systems and science opportunities
- We are looking forward to Blue Gene/P
  - Shaping up to be a great machine for many applications key to DOE's mission
  - Which will enable breakthrough science computations in a range of domains
- Much to be gained from collaborations among large centers
  - Both within and cross agencies
  - We share scientists, applications and data already

ALCF



NERSC Goals and Metrics for Petascale Systems and Services

William T.C. Kramer
kramer@nersc.gov
510-486-7577

National Energy Research Scientific Computing (NERSC) Facility
Ernest Orlando Lawrence
Berkeley National Laboratory

Goals and Metrics

- External, Mandated Goals/Measures
- External and Internal Reviews
- Program Plan and DME Project Progress
- Contractual Metrics
- User Survey
- Internal Goals and Metrics
- User performance Information
- Individual Performance Plans

NERSC Overview

NERSC Mission

The mission of the National Energy Research Scientific Computing Center (NERSC) is to accelerate the pace of scientific discovery by providing high performance computing, information, data, and communications services for research sponsored by the DOE Office of Science (SC).

Three Trends to Address

- The widening gap between application performance and peak performance of high-end computing systems
- The recent emergence of large, multidisciplinary computational science teams in the DOE research community
- The flood of scientific data from both simulations and experiments, and the convergence of computational simulation with experimental data collection and analysis in complex workflows

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

NERSC: A DOE Facility for the Future of Science

NERSC is the #7 priority

"…. NERSC … will … deploy a capability designed to meet the needs of an integrated science environment combining experiment, simulation, and theory by facilitating access to computing and data resources, as well as to large DOE experimental instruments. NERSC will concentrate its resources on supporting scientific challenge teams, with the goal of bridging the software gap between currently achievable and peak performance on the new terascale platforms."
(page 21)

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

NERSC in the ASCR Strategic Plan

"… NERSC provides the highest capability production resources available at any time to the SC research community. This investment is balanced by investments in supporting infrastructure as well as expert staff that provides direct support to the researchers in all of the SC program offices."

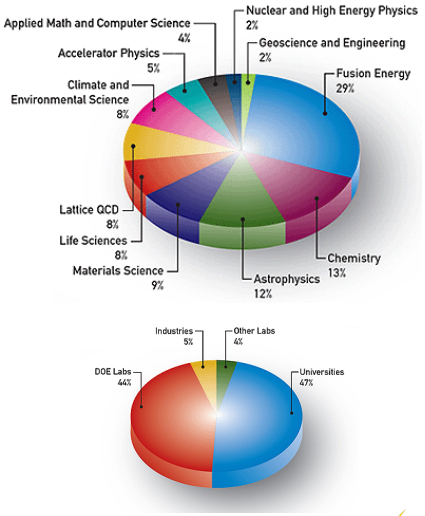
Advanced Scientific Computing Research Strategic Plan, July 30, 2004 (page 48).

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Support Different Types of Usage

- National/International User Community
- Different types of projects
  - Single PI projects
  - Large computational science collaborations
- Large variety of applications
  - All scientific applications in DOE SC
    - May not be all at once
- Range of Systems
  - Computational, storage, networking, analytics



Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Applications and Algorithms Matrix

Science areas	Multi-physics, Multi-scale	Dense linear algebra	Sparse linear algebra	FFTs	AMR	Data Intensive
Nanoscience	X	X	X	X		
Climate	X			X	X	
Chemistry	X	X	X	X		
Fusion	X	X	X		X	X
Combustion	X		X		X	X
Astrophysics	X	X	X	X	X	X
Biology	X	X				X
Nuclear		X	X			X

General purpose balanced system

High speed CPU, high Flop/s rate

High performance memory system

Bisection interconnect bandwidth

Irregular data and control flow

Storage, Network Infrastructure

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

What Scientists Want from an HPC System

- **Performance – How fast will a system process their work if everything is perfect**
- **Effectiveness – What is the likelihood they can get the system to do their work**
- **Reliability – The system is available to do work and operates correctly all the time**
- **Consistency/Variability – How often will the system process their work as fast as it can**
- **Usability – How easy is it for them to get the system to go as fast as possible**

PERCU

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Science-Driven Computing Strategy 2006 -2010

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Science-Driven Systems

- Balanced and timely introduction of best new technology for complete computational systems (computing, storage, networking, analytics)
- Engage and work directly with vendors in addressing the SC requirements in their roadmaps
- Collaborate with DOE labs and other sites in technology evaluation and introduction

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Science-Driven Services

- Provide the entire range of services from high-quality operations to direct scientific support
- Enable a broad range of scientists to effectively use NERSC in their research
- Concentrate on resources for scaling to large numbers of processors, and for supporting multidisciplinary computational science teams

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Science-Driven Analytics

- Provide architectural and systems enhancements and services to more closely integrate computational and storage resources
- Provide scientists with new tools to effectively manipulate, visualize and analyze the huge data sets from both simulations and experiments

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

What resources are required?

Task	Resource or Location
S3D (combustion)	
Problem setup and testing	Lab
Course mesh runs, test/validation	Small MPP
Deploy production code and inputs	Interactive
Submit initial production runs	Large MPP, Switch Bandwidth
Evaluation and validation of production strategy	Small MPP, Large Memory
Evaluation of performance and allocation budget	Perftools, Center staff
Production runs	Large MPP, Switch Bandwidth
Queue management	interactive
Regular snapshots saved	TB Disk, HPSS
Move data to local cluster	WAN, Sandia
Data analysis and Viz	Sandia
Publish & Share data with community	Peers, Journals, Database

Task	Resource or Location
Nearby SuperNova Factory (SNF)	
Nightly scans of the sky (NEAT)	Mt. Palomar Observatory, disk 85 GB/night , 30K images
Nightly scans of the ski (QUEST)	Mt. Palomar Observatory, disk 45 GB/night, 16K images
Transfer to HPSS via HPWREN	WAN , disk 1.5TB, tape 40TB
Morning batch processing (discover and rank SN candidates)	PDSF, serial flops, GPFS
Comparison to reference images, score by subtraction	HPSS, serial flops, GPFS
Move high scoring candidates to SCP cluster	WAN
Human evaluation of likely candidates	Custom Viz SW
Take spectrum of likely/confirmed SN	Telescope
Analyze	TB Disk, HPSS
Publish	

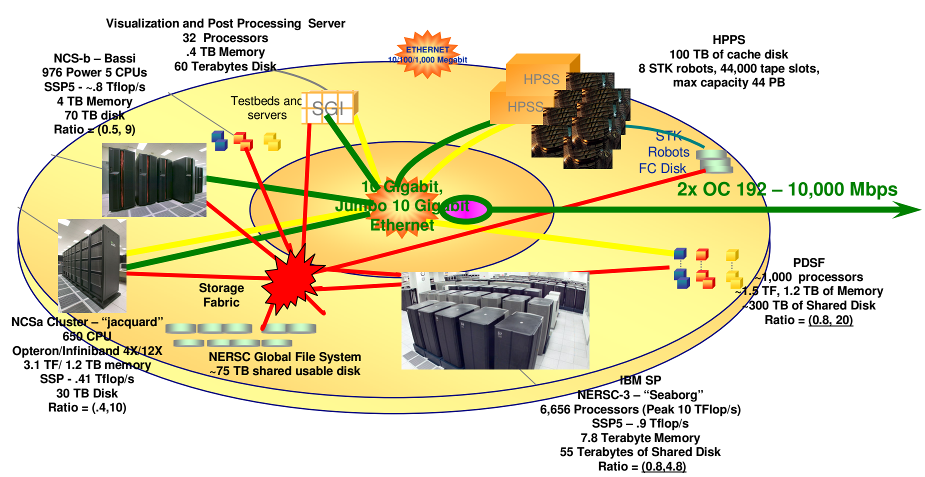
Task	Resource or Location
DNSmp (Turbulence)	
Problem choice and setup	Lab
Interactive tests	Small MPP, High throughput
Evolve initial conditions to steady state	MPP, Large Memory
Save regular snapshots (160 GB each)	Parallel FS
Test snapshots for statistical independence	Disk, serial CPU
Archive 10 independent snapshots	HPSS, queue monitoring
...	
Get 1 snapshot from tape	HPSS
Execute analysis code	Small MPP, Large Memory
Save many small analysis files (results), archive	Lab
Publish and share snapshots	WAN

Task	Resource or Location
QCD(MILC)	
Problem statement and input setup	Lab
Generate and refine quark state	MPP (2048-4096 way), low latency
Transfer quark state to MILC collaborators	WAN, Remote clusters
Analysis of reduced data, Validity testing	Small MPP
Publish and share data with community	Lab
Public archive of QCD data	WAN, http://qcd.nersc.gov

Resource
Serial
Parallel
MPP
Memory
Disk/Tape
Bandwidth
Latency
WAN

Office of Science
U.S. DEPARTMENT OF ENERGY
NERSC

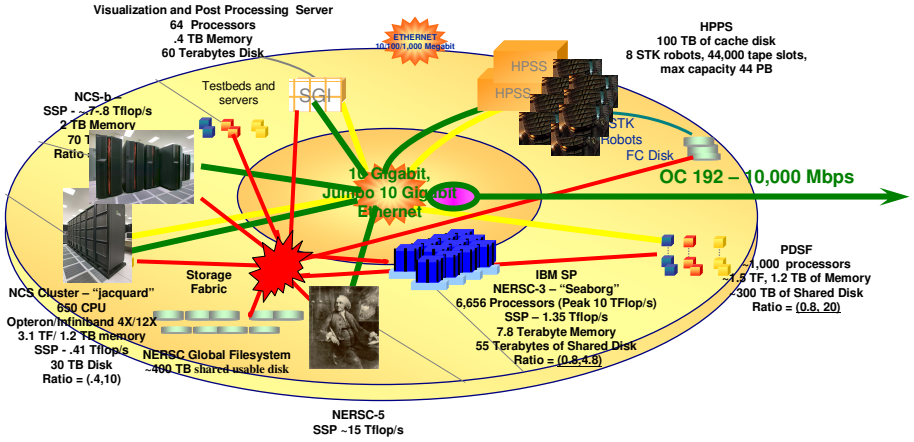
NERSC Configuration
June 2006



Ratio = (RAM Bytes per Flop, Disk Bytes per Flop)

Office of Science
U.S. DEPARTMENT OF ENERGY
NERSC

2007

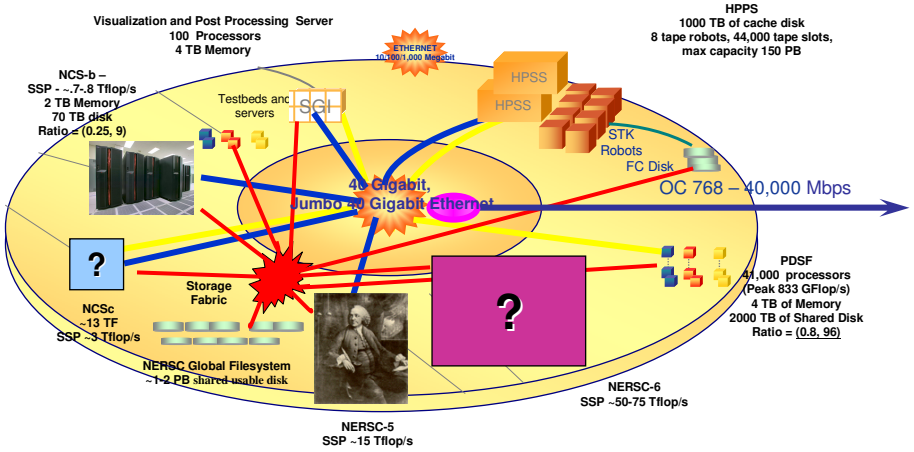


Ratio = (RAM Bytes per Flop, Disk Bytes per Flop)

Office of Science
U.S. DEPARTMENT OF ENERGY

NERSC

2009-2010



Ratio = (RAM Bytes per Flop, Disk Bytes per Flop)

Office of Science
U.S. DEPARTMENT OF ENERGY

NERSC

NERSC Impact on Science Mission

- Majority of great science in SC is done with medium- to large-scale resources
- In 2003 and 2004 NERSC users reported the publication of at least 2,206 papers that were partly based, at least partially on work done at NERSC.
- In 2005, NERSC users reported the publication of over 1,400 peer reviewed papers that were based, at least partially on work done at NERSC.

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Externally Mandated Metrics

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Current OMB PART Metrics

1. Acquisitions should be no more than 10% more than planned cost and schedule.
   This metric is reasonable.

2. 40% of the computational time is used by jobs with a concurrency of 1/8 or more of the maximum usable compute CPUs.
   Meeting this metric has positive and negative effects: motivated increased scaling of user codes; not related to the quantity, quality, or productivity of the science.

3. Every year several selected science applications are expected to increase efficiency by at least 50%.
   This metric was motivated by the desire to increase the percent of peak performance in large science applications, which now has less merit. Should be replaced by a scaling metric.

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

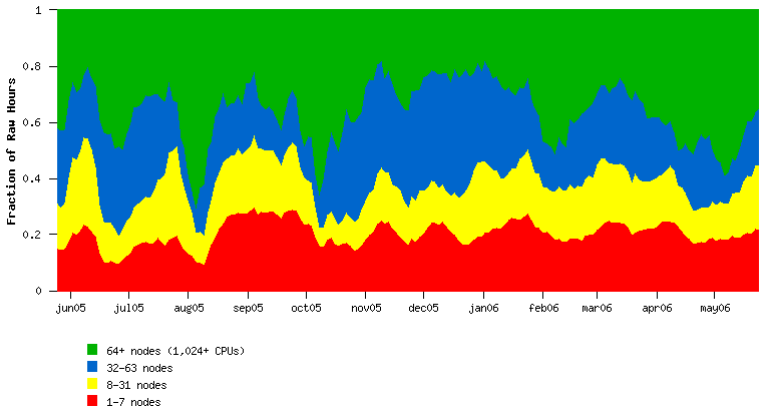
Past and Current Metrics

- 2004 Metric: >50% of all cycles to be used by jobs 1/8 of more of the max system size. NERSC interpreted this to mean ≥ 512 CPUs since there was a software limit of 4096 CPUs.
  - Achieved only in the final quarter of 2004
  - Yearly average <45% was below measure (RED)
- 2005 Metric: >40% of all cycles to be used by jobs 1/8 of more the max system size. Software limits continued.
  - Achieved throughout the year
  - Yearly average ~70%
- 2006 Metric: > same metric as in 2005, but software limits removed so this means only jobs ≥ 760 CPUs count
  - Current average >40%

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

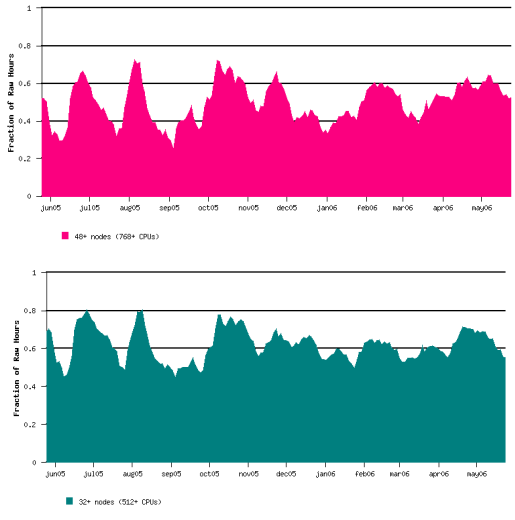
NERSC Focus is on Capability Computing



Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

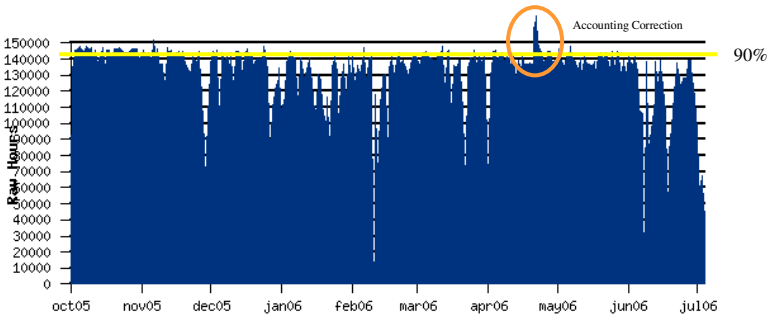
NERSC Focus is on Capability Computing



Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Time to Science Out of Theoretical Maximum Time



Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

FY 06 System Availability

System	Oct 05	Nov	Dec	Jan 06	Feb	Mar	Apr	May	Over all
Seaborg	100.00%	100.00%	100.00%	98.08%	98.27%	98.97%	100.00%	98.60%	99.24%
Jacquard	99.62%	98.85%	99.96%	95.67%	98.00%	99.87%	98.85%	100.00%	98.85%
Bassi				99.54%	100.00%	99.23%	100.00%	91.13%	97.98%
HPSS - Regent	99.31%	99.85%	99.56%	99.90%	99.84%	100.00%	100.00%	99.87%	99.79%
HPSS-Archive	100.00%	100.00%	100.00%	98.08%	98.27%	98.97%	100.00%	98.60%	99.24%
Davinci	100.00%	100.00%	98.80%	99.89%	92.83%	100.00%	99.09%	100.00%	98.83%
NGF				98.48%	99.29%	97.59%	99.46%	99.48%	98.86%

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

External and Internal Reviews

Review Organizations

- Department of Energy
  - Office of Science
  - Berkeley Site Office
  - Other Federal Oversight
- University of California
- Berkeley Lab

Program Oversight

- Strategic Proposal, Peer Review (2001)
- Programmatic and Lehman Reviews (2005)
- New Computing System Procurement Review (2004)
- Annual Allocation Review by DOE-SC
- Biannual NERSC Users Group Meeting
- Annual User Survey
- Greenbook
  - Planning document produced by NERSC Users Group

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Federal Oversight

- Office of Inspector General
  - Acquisition & Use of Supercomputers (2001)
  - Audit of User Facilities (2001)
  - Risk Assessment Guidance (2001)
  - Remote Access to Unclassified Information Systems (2002)
  - Full IT Controls Audit (2004)
- Office of Management & Budget
  - Schedules 53 & 300
    - Monthly reporting on Schedule 300
  - Quarterly Plan of Action and Milestones (POAM)
  - Earned Value Management
    - Including DME and Work Breakdown Structure
- Office of Assurance
  - Perimeter Scanning Project (2002)
  - Continuous External Scanning
  - Unannounced Red-Team scans
  - "Special Review" of cybersecurity (12/2005)
  - Cybersecurity Challenges
  - White Team cybersecurity review (5/2006)
  - Red Team cybersecurity review (pending)
- Federal Information Management Security Act (FISMA)
  - Full Authority to Operate
    - First approved in 2004
    - Must be annually renewed and includes review and approval of:
      - NERSC Enclave Security Plan
      - Risk Assessment
      - Configuration Management Plan
      - Disaster Recovery Plan

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

University of California

- **Laboratory Advisory Board**
  - Very similar to Science & Technology Panel (UCOP) under the past contract
  - Annual review of science and quality of service at UC-managed Laboratories
  - Assessment of impact and quality of User Facilities
- **Internal Audit Services**
  - Reports to UCOP/DOE/LBNL
  - Procurement oversight and audit (CFO)
  - Business Continuity (Planned)

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

BERKELEY LAB

- Annual NERSC Policy Board Meeting
  - Policy Board reports to LBNL Director
  - Addresses high level issues on the role of supercomputing in science
  - Board members are drawn from:
    - National Laboratory System
    - Universities & NSF Centers
    - International Centers
    - Industry
- Annual LBNL Director's Review of Computing Sciences
  - Every scientific division at LBNL is reviewed annually for scientific quality
  - On a rotating cycle, over three years, all of computing sciences is reviewed
  - NERSC has been part of every review except 2003 and 2005 (6/2005)
  - Computing Sciences, including NERSC, have been consistently rated 'outstanding'
- Annual Safety Self Assessment
- Annual Safety 'Walkthroughs'
- Training & Permits (ES&H)
  - Electrical Safety
  - Confined Space Procedures (raised floor)
  - Site Security
- Periodic (annual) wall to wall property inventory (>99% verified)

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Program Plan and DME Progress

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

5 Year Plan Milestones

- **2005**
  - NCS enters full service.- Completed
    - Focus is on modestly parallel and capacity computing.
    - >15–20% of Seaborg
  - WAN upgrade to 10 Gb/s .- Completed
  - Upgrade HPSS to 16 PB. Storage upgrade to support 10 GB/s for higher density and increased bandwidth. .- Completed
  - Quadruple the size of the visualization/post-processing server. .- Completed
- **2006**
  - NCSb enters full service. .- Completed
    - Focus is on modestly parallel and capacity computing
    - >30–40% of Seaborg .- Completed – Actually > 85% of Seaborg SSP

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

5 Year Plan Milestones

- **2006**
  - NERSC-5: initial delivery with possibly a phasing of delivery. – Expected – but most will be in FY 07
    - 3 to 4 times Seaborg in delivered performance – Over Achieved – more later
    - Used for entire workload and has to be balanced
  - Replace the security infrastructure for HPSS and add native Grid capability to HPSS – Completed and Underway
  - Storage and Facility-Wide File System upgrade. .- Completed and Underway
- **2007**
  - NERSC-5 enters full service. - Expected
  - Storage and Facility-Wide File System upgrade. - Expected
  - Double the size of the visualization/post processing server. – If usage dictates

Office of Science
U.S. DEPARTMENT OF ENERGY

NERSC

Competition Sensitive
CLASSIFICATION *(When filled in)*

COST PERFORMANCE REPORT
FORMAT 1 - WORK BREAKDOWN STRUCTURE
DOLLARS IN: Thousands

Page 1 of 1

1. CONTRACTOR	2. CONTRACT	3. PROGRAM	4. REPORT PERIOD
a. NAME Lawrence Berkeley National Lab	a. NAME NERSC5 W/GA	a. NAME	a. FROM *(CCYYMMDD)* 20060401
b. LOCATION *(Address and ZIP code)* 1 Cyclotron Road Berkeley, CA 94720	b. NUMBER 1		b. TO *(CCYYMMDD)* 20060430
	c. TYPE CPAF	d. SHARE RATIO 100/0 100/0	b. PHASE *(X one)* ☐ RDT&E ☐ PRODUCTION

5. CONTRACT DATA							
a. QUANTITY PROD: 0 R&D: 0	b. NEGOTIATED COST $2,122.4	c. EST COST AUTH UNPRICED WORK $0.0	d. TARGET PROFIT/ FEE $0.0 / 0.0%	e. TARGET PRICE $0.0	f. ESTIMATED PRICE $0.0	g. CONTRACT CEILING $0.0	h. ESTIMATED CONTRACT CEILING $0.0

6. ESTIMATED COST AT COMPLETION	MANAGEMENT ESTIMATE AT COMPLETION (1)	CONTRACT BUDGET BASE (2)	VARIANCE (3)
a. BEST CASE	$2,122.4		
b. WORST CASE	$2,122.4		
c. MOST LIKELY	$2,122.4	$2,122.4	$0.0

7. AUTHORIZED CONTRACTOR REPRESENTATIVE		
a. NAME *(Last, First, Middle Initial)* Bill Kramer	b. TITLE Project Manager	
c. SIGNATURE		d. DATE *(CCYYMMDD)* 20060518

8. PERFORMANCE DATA

ITEM		CURRENT PERIOD: BUDGETED COST WORK SCHEDULED (2)	BUDGETED COST WORK PERFORMED (3)	ACTUAL COST WORK PERFORMED (4)	VARIANCE SCHEDULE (5)	VARIANCE COST (6)	CUMULATIVE TO DATE: BUDGETED COST WORK SCHEDULED (7)	BUDGETED COST WORK PERFORMED (8)	ACTUAL COST WORK PERFORMED (9)	VARIANCE SCHEDULE (10)	VARIANCE COST (11)	REPROGRAM ADJUSTMENTS: COST VARIANCE (12)	REPROGRAM ADJUSTMENTS: BUDGET (13)	AT COMPLETION: BUDGETED (14)	AT COMPLETION: ESTIMATED (15)	AT COMPLETION: VARIANCE (16)
a. WBS ELEMENT																
1.1 - NERSC 5	2	102.5	136.5	7.0	34.1	129.5	1,150.5	1,201.5	685.7	51.0	515.8			2,122.4	2,122.4	0.0
1.1.1 - Planning	3	0.0	0.0	0.0	0.0	0.0	488.4	488.4	448.3	0.0	40.1			488.4	488.4	0.0
1.1.2 - Acquisition and Development (ACQ/DEV)	3	102.5	136.5	7.0	34.1	129.5	662.1	713.1	237.4	51.0	475.7			728.6	728.6	0.0
1.1.3 - Implementation/Testing	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0			905.4	905.4	0.0
b. COST OF MONEY	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0			0.0	0.0	0.0
c. GENERAL & ADMINISTRATIVE	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0			0.0	0.0	0.0
d. UNDISTRIBUTED BUDGET	2													0.0	0.0	0.0
e. SUBTOTAL *(Performance Measurement Baseline)*		102.5	136.5	7.0	34.1	129.5	1,150.5	1,201.5	685.7	51.0	515.8	0.0	0.0	2,122.4	2,122.4	0.0
f. MANAGEMENT RESERVE	2												0.0	100.0		
g. TOTAL		102.5	136.5	7.0	34.1	129.5	1,150.5	1,201.5	685.7	51.0	515.8	0.0	0.0	2,222.4		
9. RECONCILIATION TO CONTRACT BUDGET BASE																
a. VARIANCE ADJUSTMENT										0.0	0.0					
b. TOTAL CONTRACT VARIANCE										51.0	515.8			2,222.4	2,122.4	100.0

Competition Sensitive
CLASSIFICATION *(When filled in)*

DOE Quarterly Control Review Template

All sections highlighted in Yellow are required to be completed for the current review cycle for investments requiring EVM.

Investment Information

1. Date	2. Investment Name
4/30/2006	LBL NERSC
3. UPI Number/CPIC Exhibit 300	4. Program Office
019-20-01-21-01-2019-00-109-026	Office of Science
5. Project Sponsor	6. Sponsor Phone Number
Helland, Barb	301-903-3127

Project Management Certification

7. Project Manager's Name	8. Is the Project Manager for this investment certified to the level of the System?	X	Yes
Kramer, William T.			No
	If no, provide the planned date by which your project manager will be certified at the level of the investment.		

Cost, Schedule, and Performance

9. If applicable, state any actions taken to address issues that were identified in the previous quarterly Control Review cycle.

10. Recap the investment's funding profile (Funding profile should be based on the baseline in the current business case.

	PY	CY	BY	BY+1
DME	$ 6.41	$ 12.76	$ 23.34	$ 23.55
SS	$ 31.46	$ 25.11	$ 31.45	$ 31.45
Total	$ 37.87	$ 37.87	$ 54.79	$ 55.00

11. In accordance with Order 413, which Critical Decisions (CDs) have been accomplished for this investment (i.e., CD-0, 1, 2, 3, or 4)?

	Approval Date	Approver	Remarks
CD-0			NA
CD-1			NA
CD-2			NA
CD-3			NA
CD-4			NA

DME Component - Earned Value Management Data

12. For all DME activities related to the investment, enter the rolled-up EVM data from your ANSI-standard compliant EVM System and perform the required EVM calculations in to the tables below.

Project Start Baseline Date:	10/1/2004	Budget at Completion (BAC) $M:	$ 2.22
			Latest Month/Year here

Month/Year	Nov-05	Dec-05	Jan-06	Feb-06	Mar-06	Apr-06
ACWP(month)	20.1	77.2	28.7	39.7	59.2	7.0
BCWS(month)	37.3	164.7	161.3	75.4	108.3	102.5
BCWP(month)	48.4	304.6	33.9	174.4	-	138.5
ACWP(cum)	476.0	553.1	579.9	619.5	678.8	685.7
BCWS(cum)	540.2	705.0	866.2	941.6	1,049.9	1,150.5
BCWP(cum)	562.2	866.8	890.6	1,065.0	1,065.0	1,201.5

Earned Value Calculations based on cumulative data

Month/Year	Nov-05	Dec-05	Jan-06	Feb-06	Mar-06	Apr-06
BCWP(cum) - ACWP(cum)	CV	78.2	305.7	310.7	445.5	386.2	515.8
CV/BCWP(cum)	CV%	14%	35%	35%	42%	36%	43%
BCWP(cum)/ACWP(cum)	CPI	1.18	1.55	1.54	1.72	1.57	1.75
BCWP(cum) - BCWS(cum)	SV	12	161.8	24.4	123.2	17	51
SV/BCWS(cum)	SV%	2%	23%	3%	13%	2%	4%
BCWP(cum)/BCWS(cum)	SPI	1.022	1.215	1.028	1.131	1.016	1.044
BAC/CPI	EAC	2,122	2,122	2,122	2,122	2,122	2,122

13. When was, or will, the project's EVM be certified as ANSI STD 748 compliant by OECM? (Note that a site or contractor EVM certification does not necessarily indicate that certified process is being applied to the project.)

The Lab received site-wide EVMS certification in January.

14. For investments exceeding + or - 10% cost or schedule variance (CV% SV%) provide a brief description as to how and by when you plan to remediate this variance.

15. In accordance with Order 413, which Critical Decisions (CDs) have been accomplished for this investment (i.e., CD-0, 1, 2, 3, or 4)?

SC is in the process of finalizing policy with the Office of Engineering and Construction Management and the Office of the Chief Information Officer that aligns IT management with DOE Manual 413.3-1. The critical decisions accomplished for this investment will be reported when the final policy is distributed.

Show OMB Quarterly Report

Contractual Metrics
DOE/UC Contract

Office of Science
U.S. DEPARTMENT OF ENERGY

UC Contractual Metrics

- Provide for Efficient and Effective Mission Accomplishment
  - Science and Technology Results Provide Meaningful Impact on the Field
  - Provide Quality Leadership in Science and Technology
  - Provide and Sustain Science and Technology Outputs that Advanced Program Objectives and Goals
  - Provide for Effective Delivery of Science and technology
- Provide for Efficient and Effective Design, Fabrication, Construction and Operation of Research Facilities
  - Provide Effective Facility Design(s) as Required to Support Laboratory Programs
  - Provide for the Effective and Efficient Construction of Facilities and/or Fabrication of Components
  - Provide Efficient and Effective Operation of Facilities
  - Effective Utilization of Facilities to Grow and Support the Laboratory's Research Base
- Provide Effective and Efficient Science and Technology Program Management
  - Provide Effective and Efficient Stewardship of Scientific Capabilities and Program Vision
  - Provide Effective and Efficient Science and Technology Project/Program Planning and Management
  - Provide Efficient and Effective Communications and Responsiveness to Customer Needs

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

User Survey

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

FY 2005 User Survey Results

- http://www.nersc.gov/news/survey/
- 201 respondents
  - 55% of respondents are from universities; 36% from DOE labs, 9% from other labs and industry.
  - 67% of respondents are users; 11.5% Principal Investigators; 14.5% project managers;
  - 23% of respondents have NP projects; 20% BES; 19.5% HEP; 13.5% Fusion; 14% BER; 9% ASCR.
  - 41% of respondents have used NERSC over 3 years; 44% 6 months – 3 years; 15% < 6 months.
- Satisfaction rated on a 7-point scale (7 is highest score).
  - Overall average satisfaction is 6.11

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

FY 2005 Survey High Satisfaction

- Areas of highest satisfaction:
  - Account Support services – 6.73
  - HPSS reliability and uptime – 6.73
  - Consulting services– 6.73
  - NERSC security– 6.68
  - Computer and Network operations (24 by 7 control room) – 6.67
  - Network performance within NERSC– 6.45
- Largest increases in satisfaction from 2004:
  - NERSC CVS server – 6.21
  - IBM POWER3 Seaborg batch queue structure – 5.08
  - PDSF Linux cluster C/C++ compilers – 6.61
  - IBM POWER3 Seaborg up time – 6.56
  - Available computing hardware – 5.89
  - Network connectivity – 6.45

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

FY 2005 Survey
Low Satisfaction

- Areas with lowest satisfaction:
  - IBM POWER3 Seaborg batch wait time & queue structure – 5.06
  - PDSF disk configuration and I/O performance – 5.14
  - Jacquard Linux cluster batch wait time - 5.16
  - Jacquard performance and debugging tools – 5.35
- Only 3 decreases in satisfaction from 2004:
  - PDSF overall satisfaction – 6.00
  - PDSF up time – 5.89
  - Amount of time to resolve consulting issues – 6.41

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

What Does NERSC Do Well?

- 82 responses
  - 47 - NERSC provides access to powerful computing resources, without which they could not do their science
  - 32 - excellent support services, staff
  - 30 - well managed, reliable hardware
  - 11 - everything

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

What Should NERSC do Differently?

- 65 responses
  - 24 – concerns about queue turnaround time
  - 22 – concerns about job scheduling and resource allocation policies
  - 17 - need for more or different computational resources

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Some Changes Based on the 2004 Survey

- Changes in Seaborg queue scheduling:
  - we gave all premium jobs a higher scheduling priority than regular priority large-node jobs
  - we reduced the scheduling priority difference between midrange and large jobs
  - User satisfaction with Seaborg's batch queues increased by .4 points on the 2005 survey
- Hardware in support of midrange jobs:
  - In August 2005 NERSC deployed the Jacquard Linux cluster
  - In January 2006 NERSC deployed the IBM POWER5 Bassi
  - User satisfaction with NERSC's available computing hardware increased by .2 points on the 2005 survey.
- During 2005 NERSC upgraded its network infrastructure to 10 gigabits per second:
  - User satisfaction with network connectivity increased by .2 points on the 2005 survey.

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Internal Goals and Metrics

Office of Science U.S. DEPARTMENT OF ENERGY

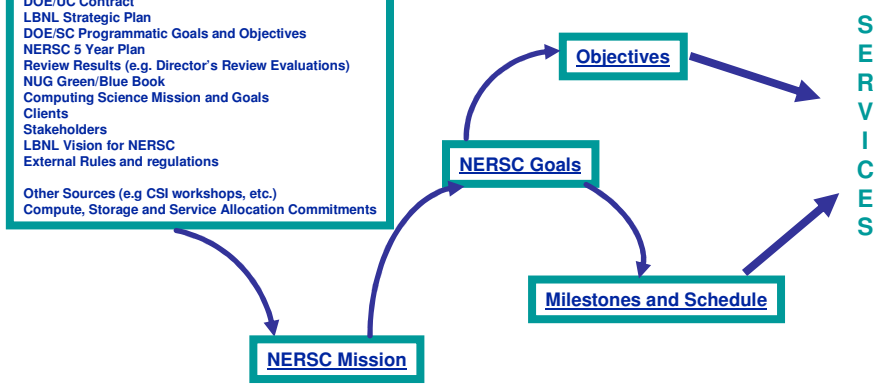
NERSC

Originating Requirements

Originating Requirements

DOE/UC Contracts
DOE Strategic Plan
DOE/UC Contract
LBNL Strategic Plan
DOE/SC Programmatic Goals and Objectives
NERSC 5 Year Plan
Review Results (e.g. Director's Review Evaluations)
NUG Green/Blue Book
Computing Science Mission and Goals
Clients
Stakeholders
LBNL Vision for NERSC
External Rules and regulations

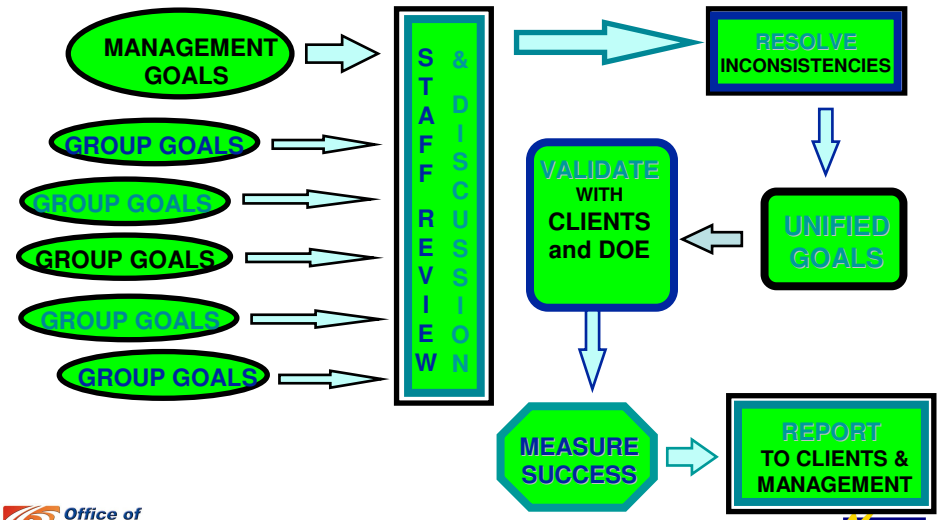
Other Sources (e.g CSI workshops, etc.)
Compute, Storage and Service Allocation Commitments



Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

NERSC Annual GOAL PROCESS



Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Example: Overall Goals FY N-1

1. Reliable and Timely Service
2. Client Support Goals
3. Never Be a Bottleneck to Moving New Technology into Service.
4. Ensure All New Technology and Changes Improve (or at Least Do Not Diminish) Service to Our Clients.
5. Develop Innovative Approaches to Help the Client Community Effectively Use NERSC Systems.
6. Develop and Implement Ways to Transfer Research Products and Knowledge into Production Systems at NERSC and Elsewhere.
7. Improve Methods of Managing Systems Within NERSC and LBNL and be a Leader in Large-Scale Systems Management and Services
8. Export Knowledge, Experience and Technology Developed at NERSC, Particularly to and Within NERSC Client Sites.
9. NERSC Will Be Able to Thrive and Improve in an Environment Where Change Is the Norm.
10. Improve the Effectiveness of NERSC Staff by Improving Infrastructure, Caring for Staff, Encouraging Professionalism and Professional Improvement

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Example Group (USG) Goals for FY N

- Provide at least one major training event per month incorporating video training, local classes and other training methods in addition to just web information. – 2, 4, 5
- Prepare and conduct the FY03 User Survey. - 1
- Provide support for strategic projects (INCITE, SciDAC, Class A). – 2, 5
- Conduct a review of the scaling characteristics of NERSC's major user codes and produce a technical report by December. – 2, 5, 8
- Manage user trouble tickets, 3rd party software support, web documentation, and user training. - 2
- Provide support for STAR, Atlas, KamLAND, and Alice software on the PDSF. Install and configure software for these experiments. Help other PDSF experiments as needed with their software installations. - 6, 2
- Provide more first-line visualization support. Learn the Ensight visualization application and help Mezzacappa's SciDAC project to make successful use of Ensight. – 2, 5

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Example Group (CSG) Goals for FY N

- Manage production systems in a manner to support "large scale" scientific research:
  - Maximizing system availability, while not impacting turnaround – 2,1
  - Create an environment that provides preferential turnaround to large scale jobs – 2,1
  - Provide timely response to customer problems
  - Provide "special" requests in a timely manner for priority scheduling, inode and disk space temporary increases, job monitoring, etc.
- Enhance production capabilities regularly and wisely.
  - Development of enhancements, that in the near-term (< 1 year in range), will improve production (computational) systems in a sufficient manner to better support "large scale" scientific research at NERSC
- Support NERSC 5 procurements on a 3 year cycle.
- Integrate, test and support division projects where needed.

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Overall Goal Support
FY N
(Primary/Secondary)

1. Reliable and Timely Service – 12
2. Client Support Goals – 15/1
3. Never Be a Bottleneck to Moving New Technology into Service. 10/7
4. Ensure All New Technology and Changes Improve (or at Least Do Not Diminish) Service to Our Clients. – 5/9
5. Develop Innovative Approaches to Help the Client Community Effectively Use NERSC Systems. 2/16
6. Develop and Implement Ways to Transfer Research Products and Knowledge into Production Systems at NERSC and Elsewhere. 2/8
7. Improve Methods of Managing Systems Within NERSC and LBNL and be a Leader in Large-Scale Systems Management and Services – 2/7
8. Export Knowledge, Experience and Technology Developed at NERSC, Particularly to and Within NERSC Client Sites. 4/7
9. NERSC Will Be Able to Thrive and Improve in an Environment Where Change Is the Norm. 0/1
10. Improve the Effectiveness of NERSC Staff by Improving Infrastructure, Caring for Staff, Encouraging Professionalism and Professional Improvement. 8/1

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Observations

- Goal 9 does not have any supporting goals. All others do
  - Three goals (5,6,7) have small numbers as primary but a number of goals listed as secondary
- All group goals relate to higher level goals
  - We need some more specifics in each area – eg the exact reliability and timeliness metrics
- Should there be any new ones or adjustments
  - E.g.
    - Do we need a goal about cost effectiveness?
    - Do we need a goal about the size of the systems or the amount we deliver.

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Overall FY 05-06 Goals

1. **Reliable and Timely Service**
   For the systems NERSC provides, service will be assessed regarding availability, mean time between interruptions and mean time to repair computational and storage systems within six months of a system going into full service.
2. **Client Support Goals**
   The end measure of a site is how much productive scientific work users accomplish. Sites must assist users in being as productive as possible by providing systems, tools, information, consulting services and training. The objective is to understand codes and how they are used, and target bottlenecks for elimination or minimization.
3. **Proactively facilitate Moving New Technology into Service.**
   NERSC is a primary vehicle for achieving the SC goal of making leading-edge technology available to its scientists. To do this, NERSC continually evaluates, tests, integrates and supports early systems and software. Therefore, NERSC must help ensure future high-performance technologies are available to Office of Science computational scientists in a timely way.
4. **Ensure All New Technology and Changes Improve (or at Least Do Not Diminish) Service to Our Clients.**
   In striving to provide users with the latest systems for computational sciences, NERSC has the responsibility to ensure system changes have a maximum benefit and minimal detrimental impact on the clients' ability to do work.

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

FY 05-06 Overall Goals

5. **Develop Innovative Approaches to Help the Client Community Effectively Use NERSC Systems.**
   NERSC must assist our clients in being as productive as possible by providing systems, enhancements, tools, information, training, consulting and other assistance. In addition to the traditional approaches that are effective, NERSC will constantly try new approaches to help make our clients effective in an ever-more-changing environment. NERSC will help design strategies and integrate and develop technology to enable our clients to improve their use of our systems and to more effectively accomplish their science.
6. **Develop and Implement Ways to Transfer Research Products and Knowledge into Production Systems at NERSC and Elsewhere.**
   NERSC is uniquely placed to establish methods and procedures that enable research products and knowledge, particularly those developed at LBNL/UC, to smoothly flow into production.
7. **Improve Methods of Managing Systems Within NERSC and LBNL and be a Leader in Large-Scale Systems Management and Services**
   As the Department of Energy's largest unclassified scientific computing facility, NERSC continually provides leadership and helps shape the field of high performance computing. As HPC technology evolves at an increasing rate, it is crucial that NERSC and LBNL remain at the forefront of getting the most out of these systems.
8. **Export Knowledge, Experience and Technology Developed at NERSC, Particularly to and Within NERSC Client Sites.**
   In order for NERSC to be a leader in large-scale computing, NERSC must export experience, knowledge, and technology. Transfer must be made to other client sites, supercomputer sites, and industry.
9. **Improve the Effectiveness of NERSC Staff by Improving Infrastructure, Caring for Staff, Encouraging Professionalism and Professional Improvement**
   Every employee has a stake in the success of NERSC and management encourages staff to contribute their ideas for helping the organization succeed. To help facilitate the professional exchange of ideas and information, NERSC staff will be strive to expand their knowledge, communication and thrive changing environments.

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

User performance Information

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

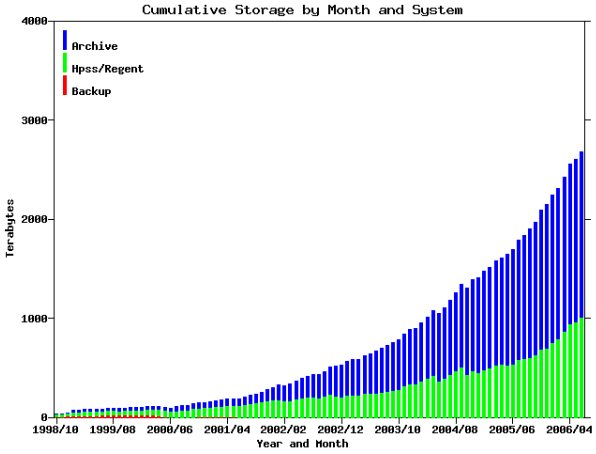
Consultation Profile

Num 2005 tickets	% of tickets	Avg hours to solve	Total hours	% of time	Topic
328	15.1	6.4	2107	43	Programming
628	28.8	1.8	1127	23	Running Jobs
393	18.1	2.0	784	16	Software
325	14.9	1.2	392	8	Data Management
131	6.0	1.9	245	5	Network Access
310	14.2	0.5	122	3	Accounts, Allocations
62	2.8	2.0	147	2	General Info
2,177		2.3	4,900		

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

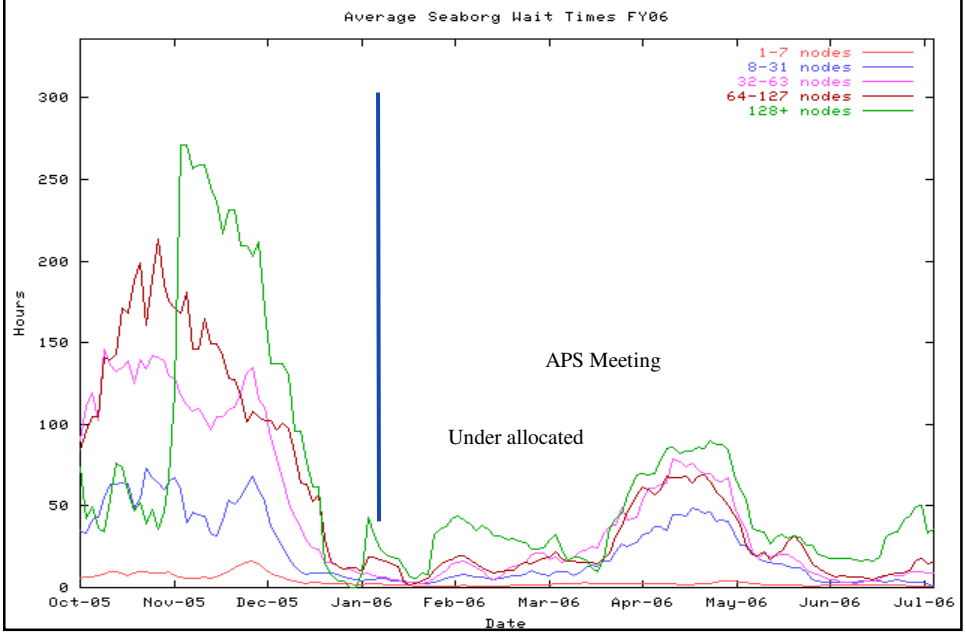
Mass Storage



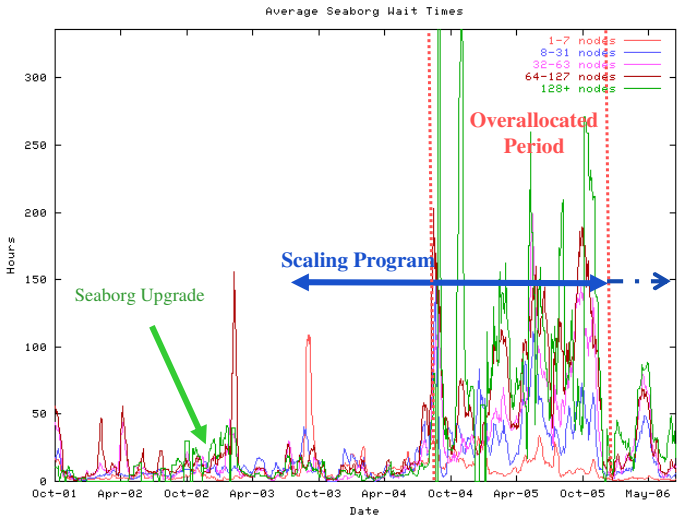
Office of Science
U.S. DEPARTMENT OF ENERGY

NERSC

AY 2005 Queue Wait Statistics



4 Year Queue Wait Statistics



Office of Science U.S. DEPARTMENT OF ENERGY

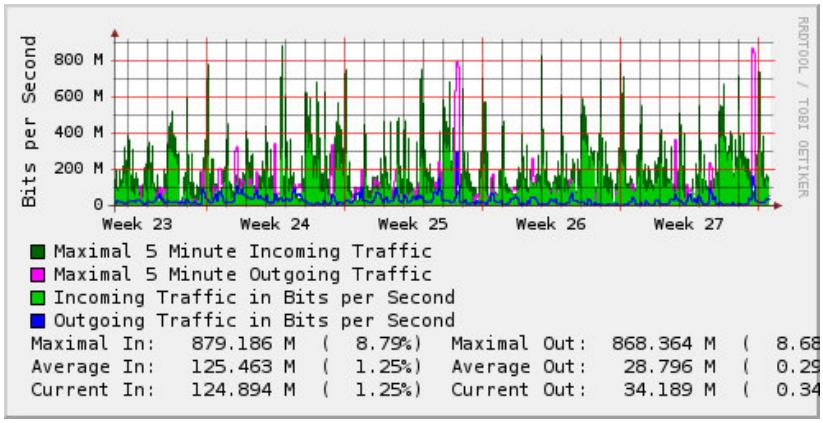
NERSC

NERSC and LBNL Border Traffic

Type	NERSC	LBNL
Bulk Data (ftp, hsi)	85%	36%
Grid	7%	<1%
Computer System Services (DNS, iperf)	4%	14%
Interactive (ssh, kshell)	3%	4%
World Wide Web	<1%	41%
Mail	<1%	1%
Database	<1%	<1%
Uncategorized	1%	3%
Total	100%	100%

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC



Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Proposed Metrics

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Goal #1: User Satisfaction

- Metric #1.1: Users find the systems and services of a facility useful and helpful.
- Value #1.1: The overall satisfaction of an annual user survey is 5.25 or better (out of 7).
  - 2005 overall rating average
    - 88 areas of ratings recorded
    - 6.11 out of 7
    - Standard Deviation -.45
    - 4 areas out of 88 categories (4.5%) with ratings below 5.25
    - 1 area was below 5 – seaborg queue wait time –
      - One lesson – do not do a survey during major system upgrade at the end of an over allocated year when people are trying to use their allocation
  - 2005 score for important topics
    - Users indicate what areas are the "most important"
    - 17 areas
    - 6.27 out of 7
    - Standard Deviation -.27
    - 0 categories with ratings below 5.25

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Goal #1: User Satisfaction

- Metric #1.2: Facility responsiveness to user feedback.
- Value #1.2: There is an improved user rating in areas where previous user ratings had fallen below 5.25 (out of 7).
  - In 2005, all of the previous year's low areas show improvement of .2 to .4
  - In 2004, all of the previous year's low areas show improvement of .3 to .5 or were works in process when survey were taken

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Goal #2: Systems ready to process the user workload.

- Value #2.1: Within 18 months of delivery and thereafter, scheduled availability is > 95%
- Value #2.1: Within 18 months of delivery and thereafter, overall availability is > 90% or another value as agreed by the program office.

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Goal #2: Systems ready to process the user workload.

System	Oct 05	Nov	Dec	Jan 06	Feb	Mar	Apr	May	Overall
Seaborg	100.00%	100.00%	100.00%	98.08%	98.27%	98.97%	100.00%	98.60%	99.24%
Jacquard	99.62%	98.85%	99.96%	95.67%	98.00%	99.87%	98.85%	100.00%	98.85%
Bassi				99.54%	100.00%	99.23%	100.00%	91.13%	97.98%
HPSS - Regent	99.31%	99.85%	99.56%	99.90%	99.84%	100.00%	100.00%	99.87%	99.79%
HPSS-Archive	100.00%	100.00%	100.00%	98.08%	98.27%	98.97%	100.00%	98.60%	99.24%
Davinci	100.00%	100.00%	98.80%	99.89%	92.83%	100.00%	99.09%	100.00%	98.83%
NGF				98.48%	99.29%	97.59%	99.46%	99.48%	98.86%

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Goal #3: Facilities provide timely and effective assistance

- Metric #3.1: Problems are recorded and acknowledged
- Value #3.1: 99% of user problems are acknowledged within 4 working hours.
  - Responded to 99.5% of tickets within 4 hrs from Oct. 1 2004 through Sept 30th 2005.
    - 13 out of 2669 tickets were longer than the metric.

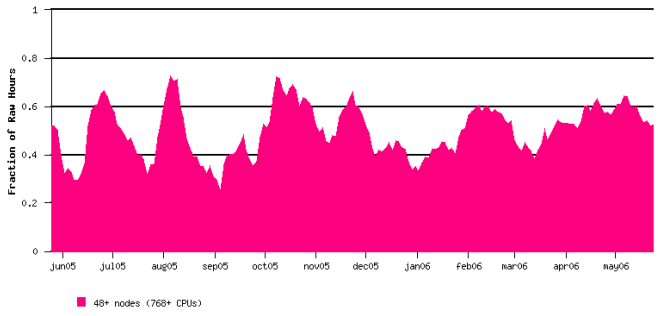
- Metric #3.2: Most problems are solved within a reasonable time
- Value #3.2: 80% of user problems are addressed within 3 working days, either by resolving them to the user's satisfaction within 3 working days, or for problems that will take longer, by informing the user how the problem will be handled within 3 working days (and providing periodic updates on the expected resolution).
  - Past survey's of our trouble ticket system indicate 80-90%

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Goal #4: Facility facilitates running capability problems

- Metric #4.1: The majority of computational time goes to capability jobs.
- Value #4.1: T% of all computational time for jobs that use more than N CPUs (or equivalently, x% of the available resources), as determined by agreement between the Program Office and the Facility.



Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Goal #4: Facility facilitates running capability problems

- Metric #4.2: Capability jobs are provided excellent turnaround
- Value #4.2: For jobs defined as capability jobs, the expansion factor is X or more. X ≤ 10 is a potential value that may be appropriate.

		Submitssion Class						
No. of Nodes	Available CPUs	Int.	Debug	Premium	Regular	Low	Other	Over all
1-7	1-112		1.16	1.09	1.63	2.56		1.67
8-15	128-240		1.13	1.08	1.85	2.48		1.69
16-31	256-496		1.18		1.81	1.75		1.77
32-63	512-1,008				2.02			2.02
64-127	1,024-2,032				1.95			1.95
128+	2,048+				6.78			6.78
Overall			1.17	1.08	2.17	2.06		2.12

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Discussion: What should the Target Expansion Factor Be?

- Traditional Expansion Factor:
  E(job) = (wait_time + run_time) / run time
- Alternative Formula (only request time can influence scheduling decisions):
  E(job) = (wait time + request time) / request time
- Weight to use in computing the Expansion Factor for a class of jobs:
  - Simple average
  - Request time
  - Request time * number of processors (this gives more weight to capability jobs)
- When to start counting wait time?
  - On Seaborg and Bassi: when the job enters Idle state
  - On Jacquard: when the job was submitted (this will change with Maui scheduler)

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Past NERSC Expansion Factors for Regular Charge Class

Quarter	Allocation Pressure	Seaborg EF	Bassi EF	Jacquard EF	NERSC EF
FY05 Q3	Over-allocated	6.72		1.39	5.14
FY05 Q4	Over-allocated	6.62		1.50	5.10
FY06 Q1	Mixed	5.69	4.61	3.62	4.89
FY06 Q2	Very Low	2.48	2.00	1.96	2.00
FY06 Q3 thru 6/5	Low	4.00	1.50	2.24	2.72

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Past Seaborg Expansion Factors for Regular Charge Class

Year	1-112 procs	128-240 procs	256-496 procs	512-1,008 procs	1,024-2,032 procs	2,048 + procs	All
FY05 Q2	3.97	7.06	9.87	5.52	7.16	17.76	6.72
FY05 Q3	4.96	10.06	13.68	5.38	7.12	8.63	6.72
FY05 Q4	4.04	5.20	10.10	5.29	7.81	9.25	6.62
FY06 Q1	2.48	5.41	7.08	5.47	8.04	6.58	5.69
FY06 Q2	1.39	1.71	2.55	1.92	2.96	4.20	2.48
FY06 Q3 thru 6/5	1.92	3.73	5.37	4.12	4.65	4.29	4.00

Office of Science U.S. DEPARTMENT OF ENERGY

NERSC

Summary

- Many people use many different metrics for NERSC
- NERSC has responded to changes in metrics and can meet almost any
- So the main message is to be very careful what you measure, because the behavior of facilities and scientists will adapt to meet the metric, even if it is disconnected science effectiveness

Office of Science
U.S. DEPARTMENT OF ENERGY

NERSC