National Energy Research Scientific Computing Center (NERSC)

Horst D. Simon
Director

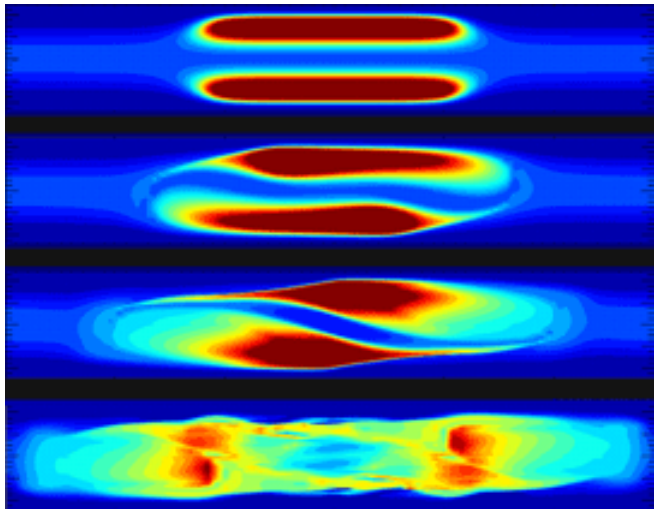
ASCAC Review of Facilites
May 2, 2001

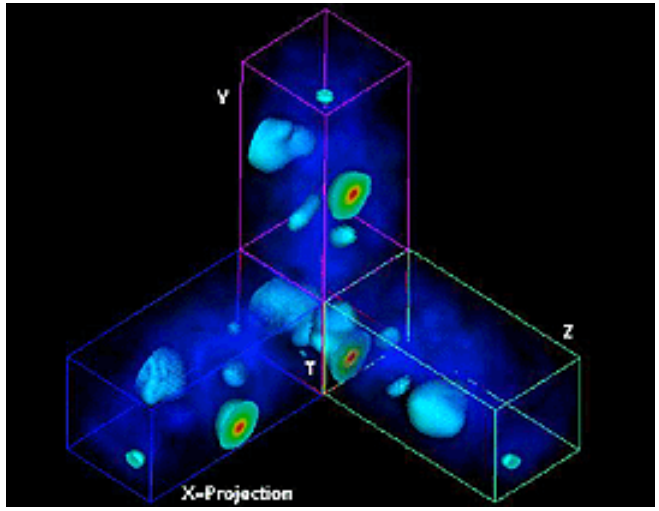
NERSC Vision

NERSC aspires to be a world leader in accelerating scientific discovery through computation. Our vision is to provide high-performance computing tools and expertise to tackle science's biggest and most challenging problems, and to play a major role in advancing large-scale computational science and computer science.

NERSC Overview

- *the* Department of Energy, Office of Science, supercomputer facility
- unclassified, open facility; serving >2000 users in all DOE mission-relevant basic science disciplines
- 25th anniversary in 1999





1996: Re-engineering Large-Scale Scientific Computing

- In 1995-96 DOE and NSF competitively re-examined the role of centers:
  - –Rapidly changing technology
  - –Better local facilities everywhere
  - –Growth of computational approaches in all disciplines

- New Model: Intellectual Services + a Major Facility
  - –New algorithms and strategies developed in medium- and long- term collaborations with scientific user community
  - –The Center is the working interface between computer science and physical science

Necessary but not sufficient!

Overview

- **NERSC resources**
  - Systems
  - Staff
  - Facilities
  - Budget
  - Collaborations
- **Who uses NERSC?**
- **Comparison to other centers**
- **Technical accomplishments**
- **Impact on DOE science**
- **Future plans**

Leading Edge Systems

Balance Newest Technology and Production Quality

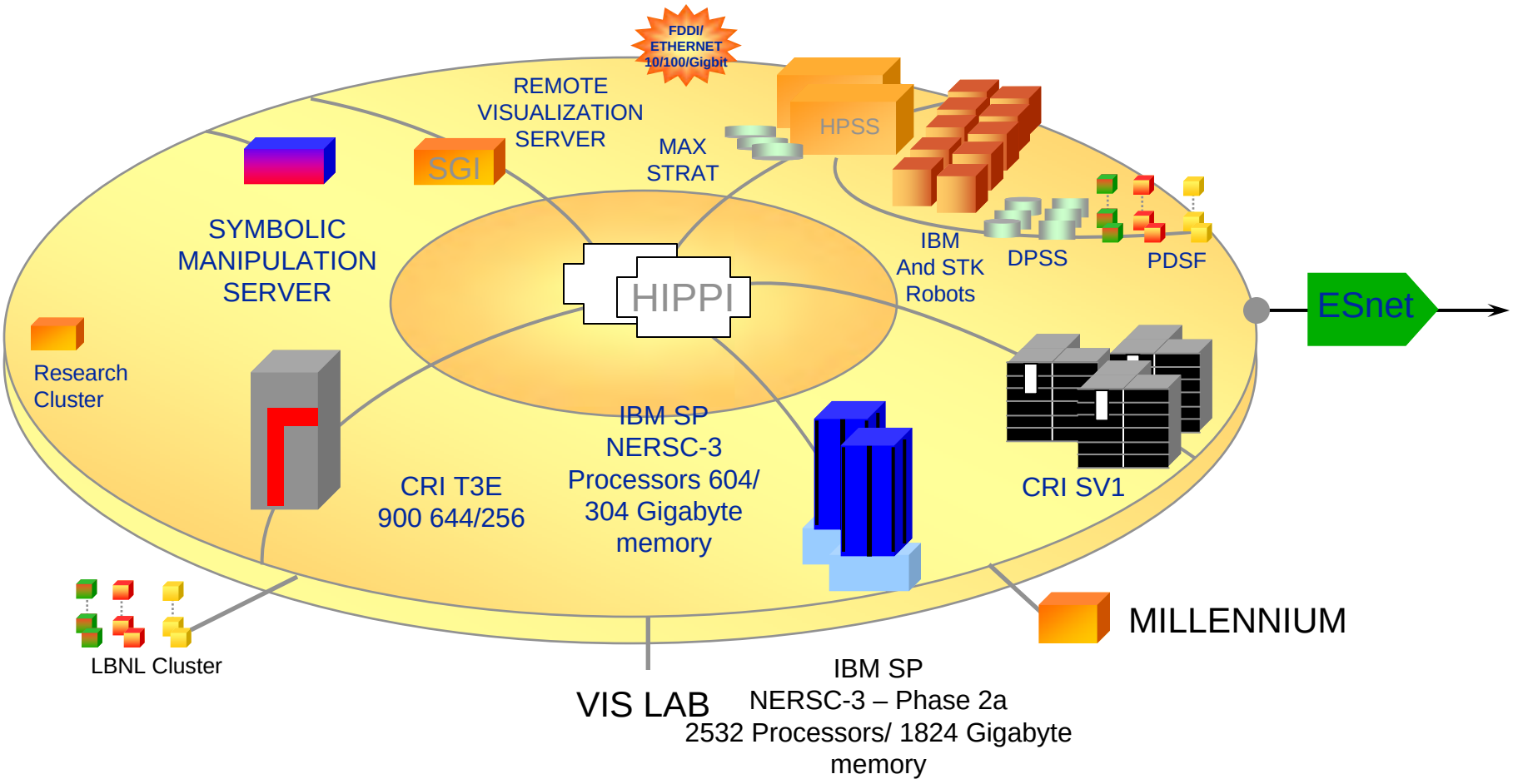
High-End Parallel Processors for Capability

Large Storage Systems

System Balance

Vector System
PC Clusters
Networks

NERSC System Architecture



Balanced Introduction of New High-End Parallel Systems

9/1996
NERSC-2 first deployment

10/1997
NERSC-2 full production

1998
NERSC-3 procurement

4/2000
NERSC-3 first deployment

2001
NERSC-3 full production

1996

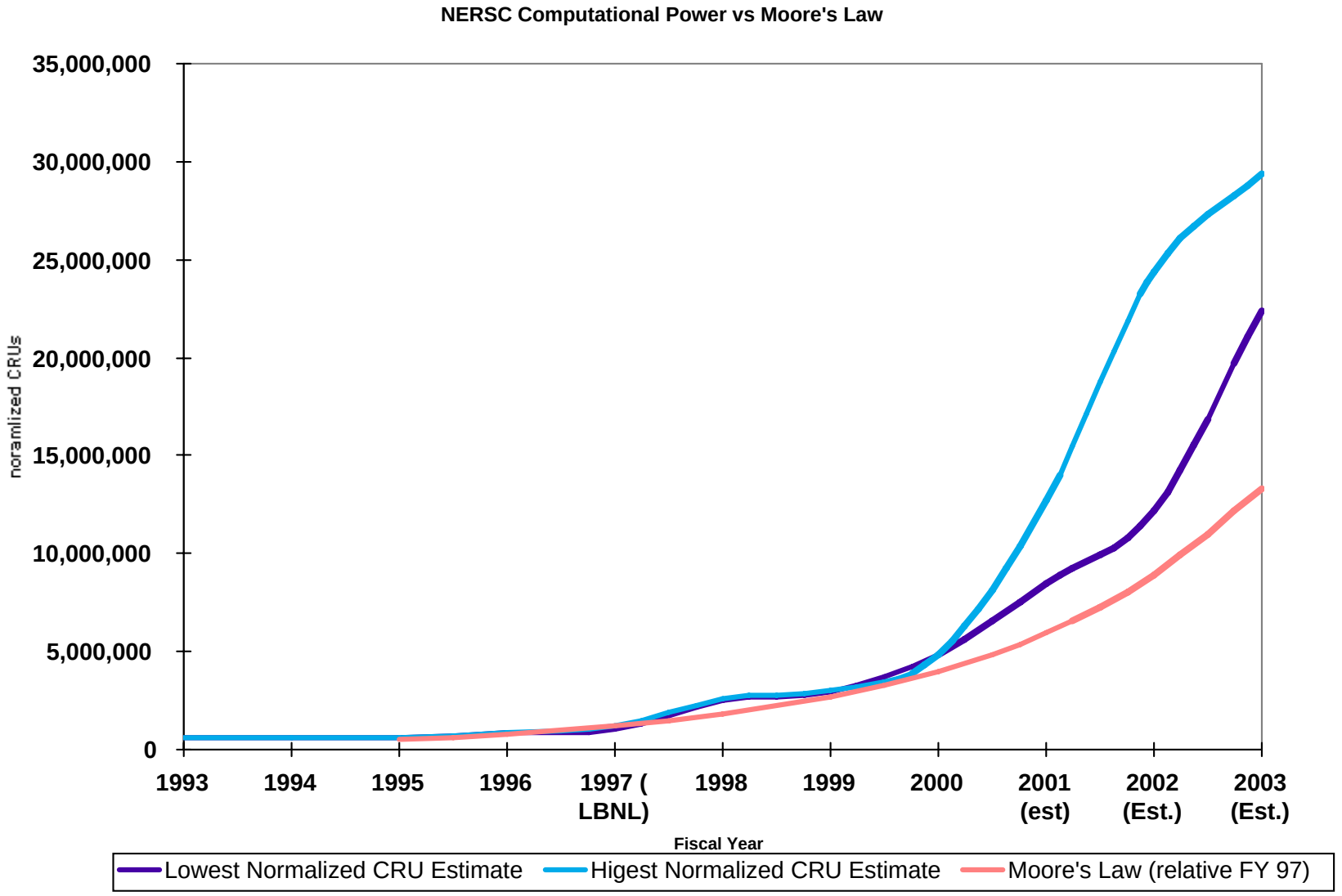
2001

High-End Parallel Systems


Cray T3E


IBM SP

NERSC Computational Power vs. Moore's Law

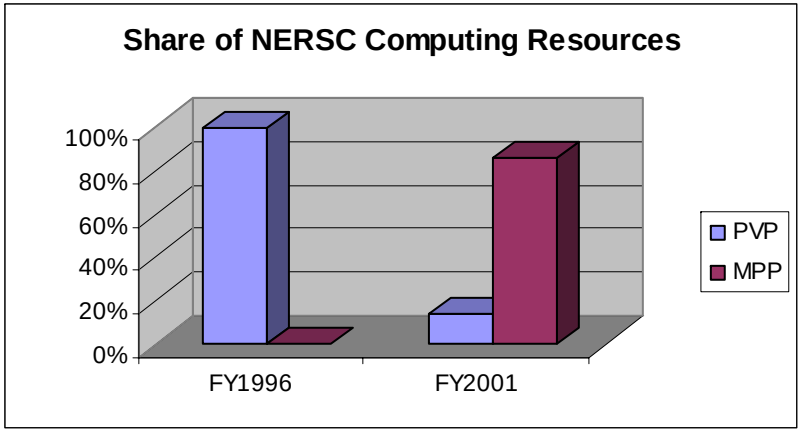


Meeting the Applications Transition Challenge in 1996

- In 1996 NERSC re-invented itself to meet the challenge of transitioning its user base from single processor vector computing to the highly parallel computing by creating the model of

  Intellectual Service + Major Facility

- This transition was highly successful
  - significant scientific results
  - smooth transition

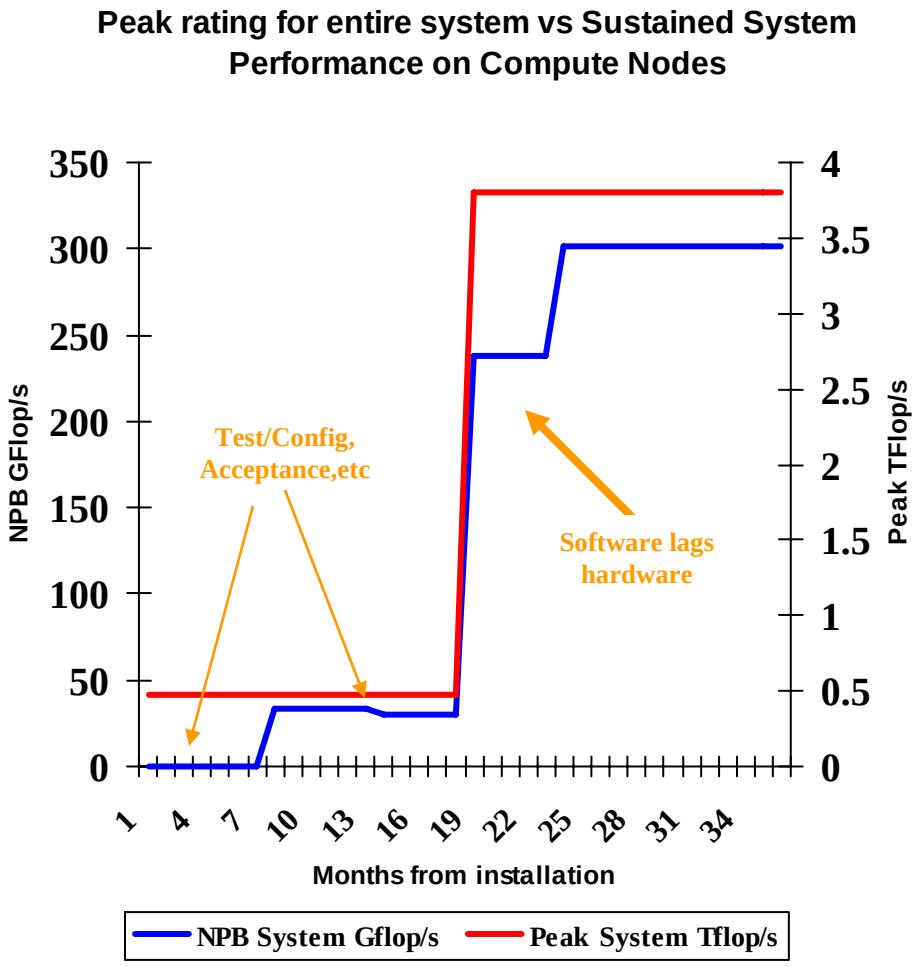


NERSC 3 – Phase 2
Delivered January 2001

- **Total system numbers: 158 Nodes of 16 CPUs each - 134 dedicated to parallel computation**
  - **–140 nodes with 12 GB of memory**
  - **– 18 nodes with 8 GB of memory**
- **Total CPUs: 2,528 at 1.5 Gflop/s peak = 3.792 Tflop/s**
- **Total memory: 1.824 TB**
- **Total user accessible shared Parallel Disk: 20 TB**
- **Total local disk: 11.4 TB of local disk (used mostly for system purposes).**

NERSC-3 Sustained System Performance

- **Estimates the amount scientific computation that can really be delivered**
  - **Depends on delivery of Phase 2b functionality**
  - **The higher the last number is, the better, since the system remains at NERSC for 4 more years**



TOP500 List 11/00

RANK	MANU-FACTURER	COMPUTER	R_{MAX} [TF/S]	INSTALLATION SITE	COUNTRY	YEAR	AREA OF INSTALLATION	# PROC
1	IBM	ASCI White SP Power3	4.93	Lawrence Livermore National Laboratory	USA	2000	Research	8192
2	IBM	NERSC-3	2.5 TF/s	NERSC	USA	2001	Research	2528
3	IBM	ASCI Blue Pacific SST, IBM SP 604E	2.14	Lawrence Livermore National Laboratory	USA	1999	Research	5808
4	SGI	ASCI Blue Mountain	1.61	Los Alamos National Laboratory	USA	1998	Research	6144
5	IBM	SP Power3 375Mhz	1.42	IBM/Naval Oceanographic Office (NAVOCEANO)	USA	2000	Research	1336
6	IBM	SPPower3 375Mhz	1.18	National Centers for Environmental Predicition	USA	2000	Research	1104
7	Hitachi	SR8000-F1	1.04	Leibniz Rechenzentrum, Munic	Germany	2000	Academic	112
8	IBM	SP Power3 375MHz 8way	0.93	San Diego Supercomputer Center	USA	2000	Academic	1152
9	Hitachi	SR8000-F1	0.92	High Energy Accelerator Research Organization/ KEK,	Japan	2000	Research	100
10	Cray Inc.	T3E 1200	0.89	Government	USA	1998	Classified	1084

Continued Improvement of Storage Capability

Moved from 70 TeraBytes to 1.3 PetaBytes

5/96: CFS and Unitree moved to Berkeley

1/97: HPSS decision

2/98: HPSS installed; Unitree conversion

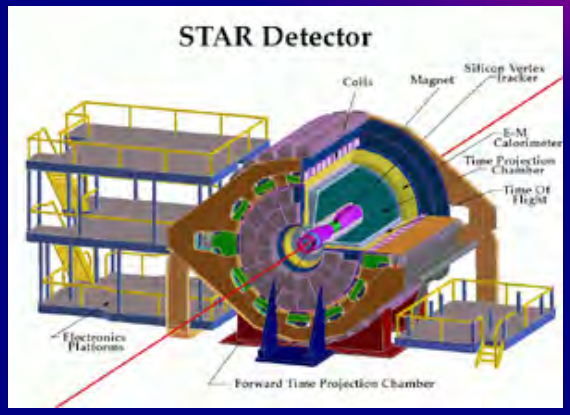
10/99: CFS conversion to HPSS

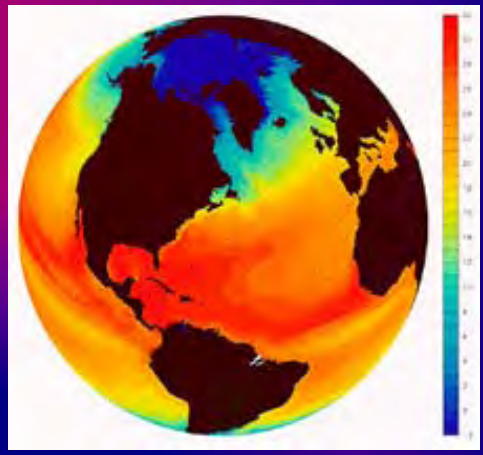
11/00: 1PetaByte

1996

2001

Storage Systems

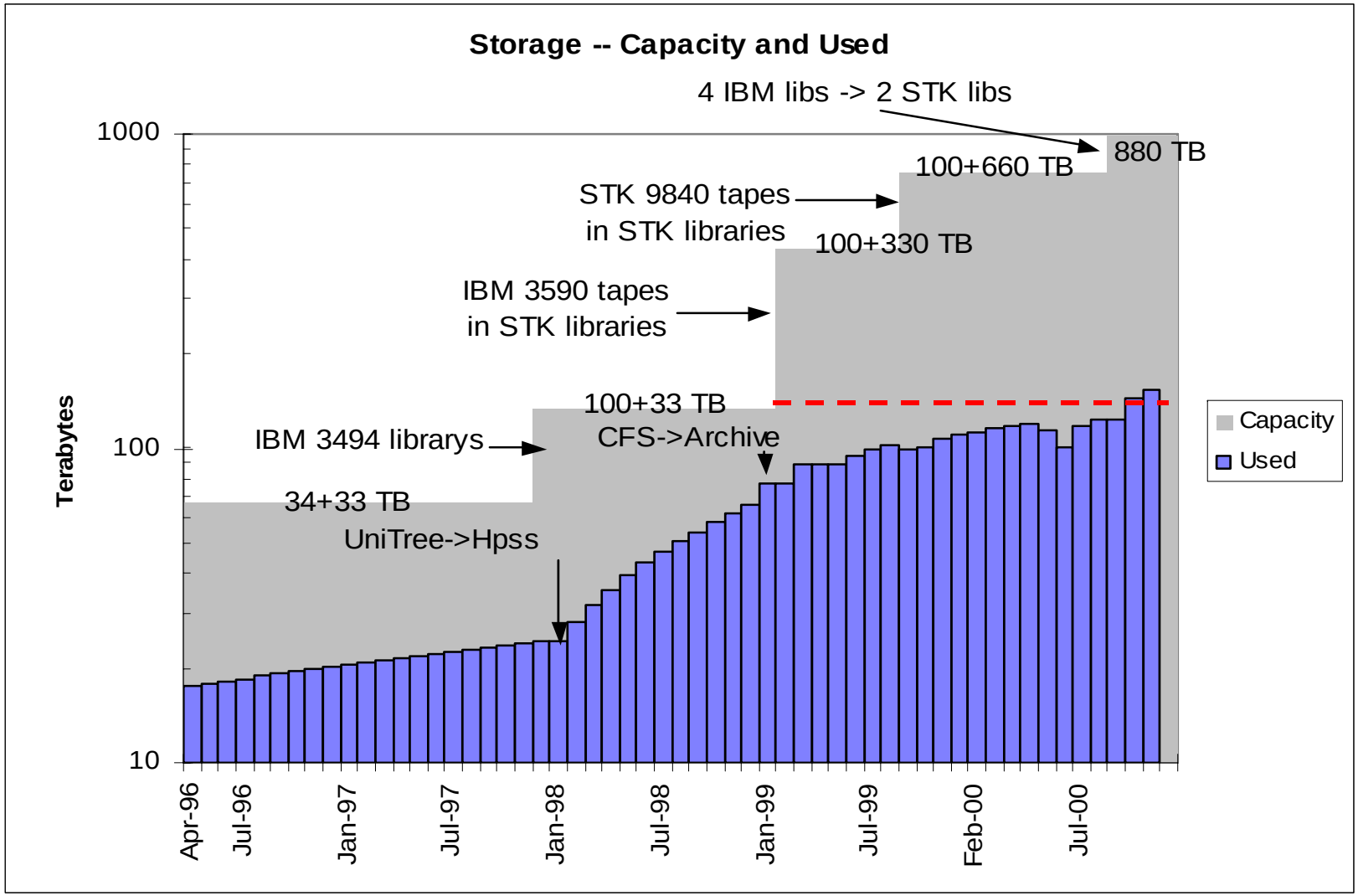




High Energy and Nuclear Physics

Climate

Storage: HPSS



System Balance: Vector Systems, PC Clusters, Visualization, Networking

1999:
Upgrade to SV-1

4/97: Upgrade to J90se

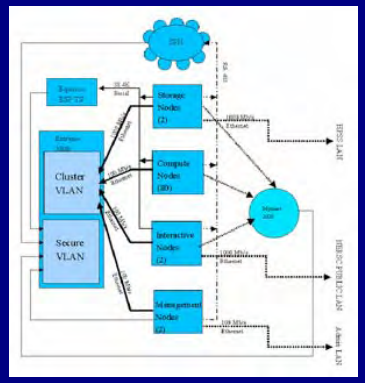
4/96: First J90 installed

1996

Other systems to balance user requirements

2001

4/96: PDSF arrives from SSC



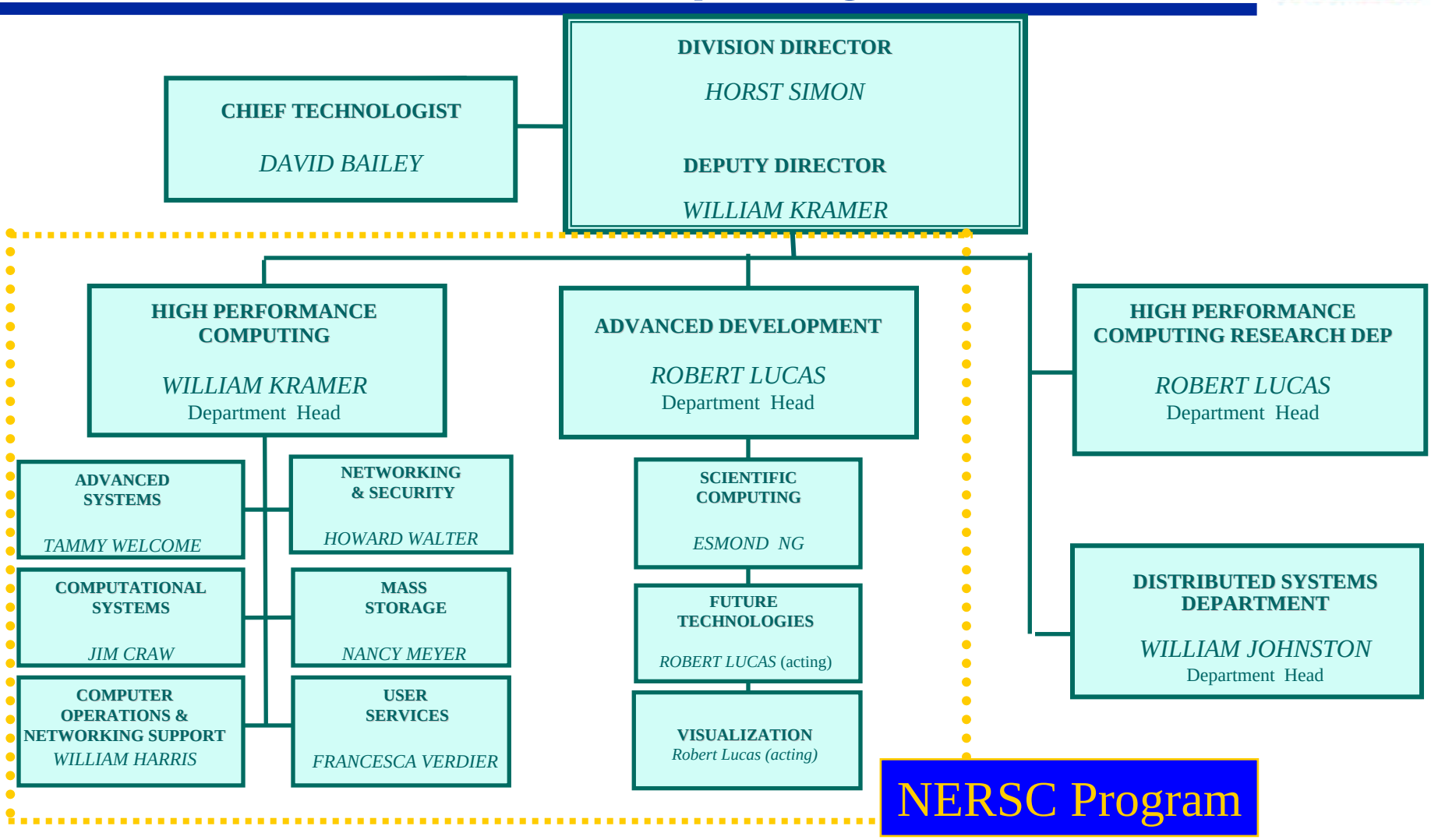
10/98: PDSF Linux PCs replace workstations

2001:
New Alvarez Cluster

Change in Staff Skills and Backgrounds

	1994	1998
Number of Technical Staff	79	59
Number of Degrees	42/53%	53/90%
Number of Advanced Degrees	21/27%	31/53%
Number of PhDs	11/14%	24/41%

National Energy Research Scientific Computing Division



FTE Profile

- 10.0 Operations
- 10.0 Consultants
- 3.0 Webmaster and webcontent
- 4.0 Account management and allocation
- 12.0 Computational Systems (IBM, T3E, PVP)
- 2.0 Computational Systems (PDSF)
- 5.0 Storage
- 4.0 Networking and Security
- 6.0 Advanced Systems
- 2.5 Future Technologies
- 2.0 Visualization
- 1.0 Management

Total: 62 FTE

Intellectual Services

New algorithms and strategies developed in medium- and long-term collaborations with scientific user community

Change in staff skills and background

Innovative assistance

NERSC as working interface between science and computer science

New model of scientific computing support

Developing new user communities

Red Carpet Plan

Intellectual Services

1996

4/96: Re-invented NERSC

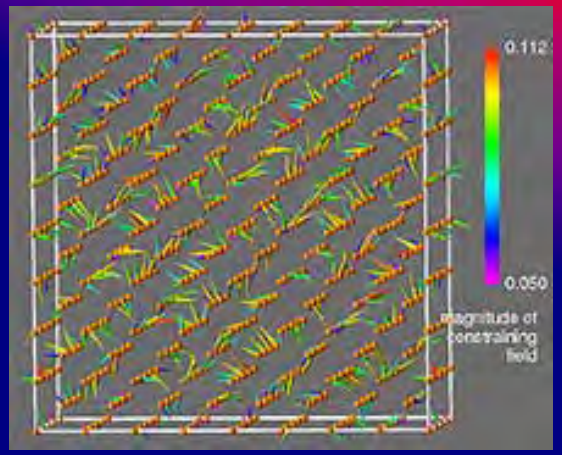
10/97: Developed Red Carpet Plan

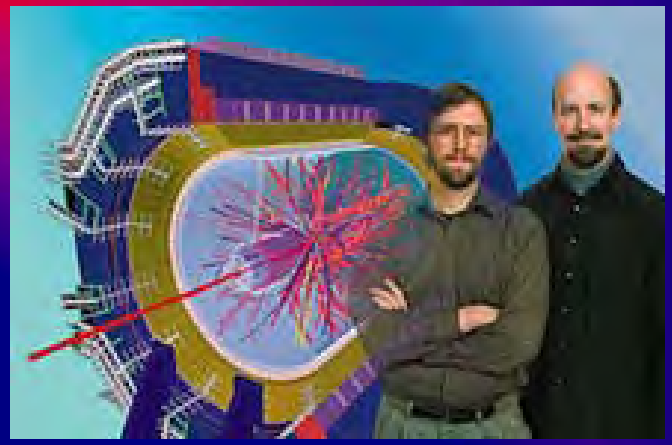
11/98: First 1 Teraflops on real application

11/99: ESP Benchmark

2001

Accomplished transition to parallel platforms



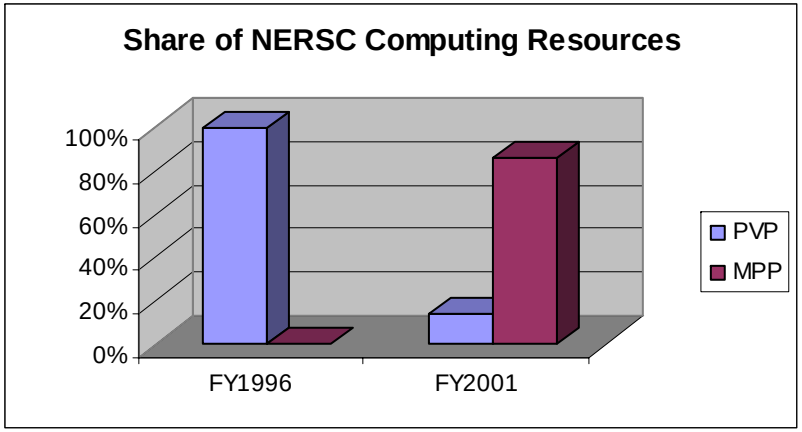


Meeting the Applications Transition Challenge in 1996

- In 1996 NERSC re-invented itself to meet the challenge of transitioning its user base from single processor vector computing to the highly parallel computing by creating the model of

  Intellectual Service + Major Facility

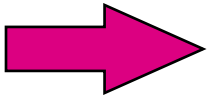
- This transition was highly successful
  - significant scientific results
  - smooth transition



Meeting the Applications Transition Challenge

New model of scientific computing support

SC	Maps to	Discipline	Maps to	Computational technology
OFE		magnetic fusion		particle in cell
BES		computational chemistry material sciences		local density functional
BER		climate research computational biology		partial diff. equations
HENP		QCD accelerator design particle detection simulation		Monte Carlo technique searching, pattern matching
OASCR		combustion applied mathematics		image processing



NERSC has or will build competency in all technological areas of relevance to SC research

Red Carpet Plan to Make Grand Challenge Projects Successful

- NERSC staff make "grand challenge" projects successful:
  - Adjusting system limits as needed (e.g., long 512-way jobs)
  - Raising priorities for jobs at critical times.
  - Software modifications to allow for larger calculations.
  - Specialized talks at conferences for specific user areas.
  - Porting software that otherwise would not have been available (e.g., Cernlib)
  - Parallelizing public domain software and optimizing it for the center's platforms (e.g., NetCDF).
- Staff researchers in the Scientific Computing Group are collaborators in multi-institution projects:
  - Material Science, Environmental Science, Astrophysics, Linear Algebra

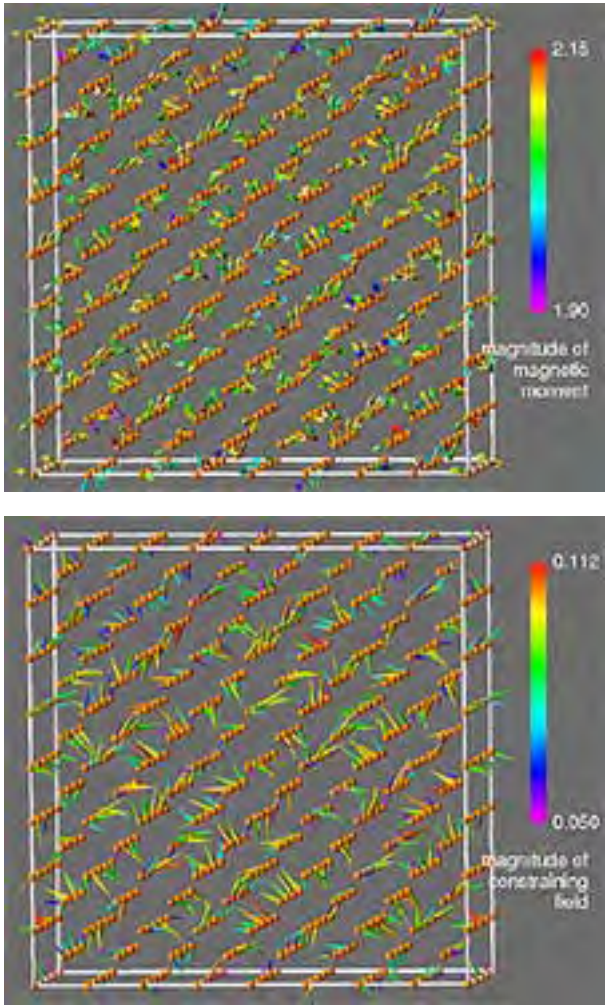
Red Carpet Plan for Grand Challenges - Example

- 1998 Gordon Bell Prize for best performance of a parallel supercomputer application for a team of collaborators from DOE's Grand Challenge on Materials, Methods, Microstructure, and Magnetism.

- Andrew Canning (NERSC) made significant algorithmic contributions to this project, and was the key force behind the large-scale simulations.

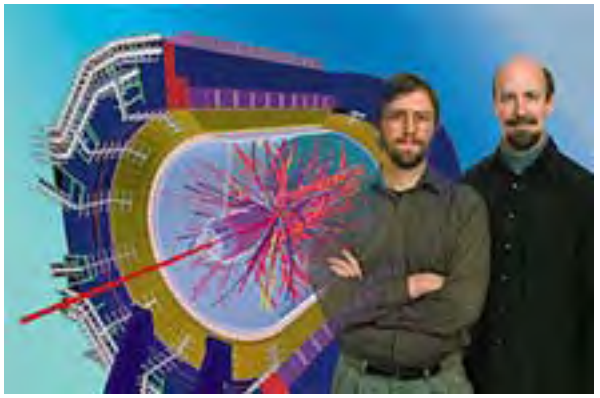
- 1024-atom first-principles simulation of metallic magnetism in iron.

- First complete application to break the 1Tflops barrier.



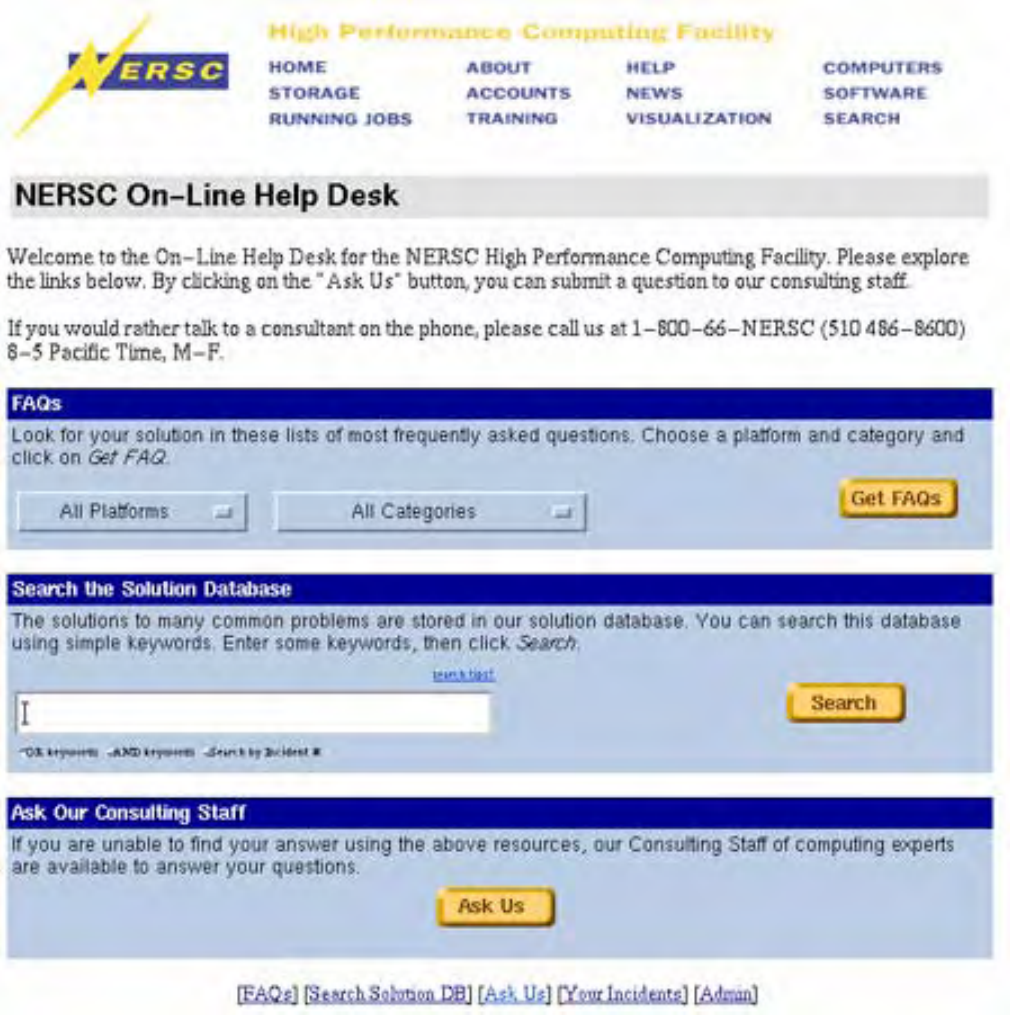
Developing New User Communities

- **Ported CERNLIB to the T3E**
  - **–first port to highly parallel platform enabled HENP community**
- **Ported NCAR CSM to the SV-1 Cluster**
  - **–only complete high performance platform installation outside NCAR for climate users**



Innovative Assistance: Transitioned Routine Questions to Web

FAQs viewable by machine & topic



Search 'Knowledge Base'

On-Line Help Request Form

Classroom Training

- The results from our 1999 User Survey showed that users still prefer face-to-face classroom training over Web-based training.

- We have experimented with Web audio/video presentations made available in Real Media format:
  - Rapid advances in technologies are continuing in this area
  - Most A/V software comes out initially on the PC platform, while more than half of our users use Unix exclusively
  - The developments of the Linux market in general and of better Windows environment support for Linux (e.g. VMWare) may improve this situation.

Innovative Assistance from NERSC Staff

- Collaborative effort - SLAC (ARDA, SCS, BaBar) & LBNL (NERSC)
- Specs: 17 Dell 410 systems, each with dual 450 MHz Pentium II processors, 256 MB memory, one 9 GB disk connected with a Cisco 5505 Fast Ethernet switch
- Operating system: RedHat Linux version 2.0.36

Omega3P running on cluster using MPI

Performance is within 25% of T3E up to 16 processors

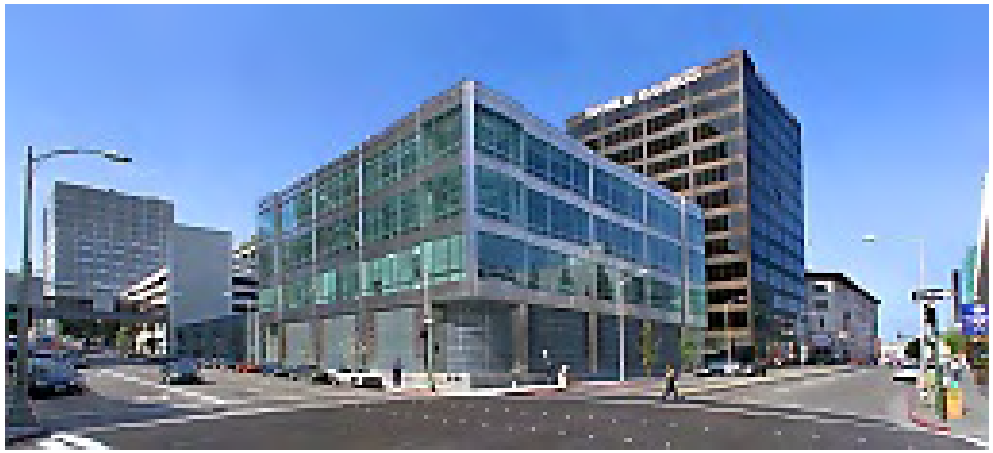
Possible low-cost, scalable alternative to supercomputers

Effective resource for high-resolution component design

Slide adapted from Kwok Ko, SLAC

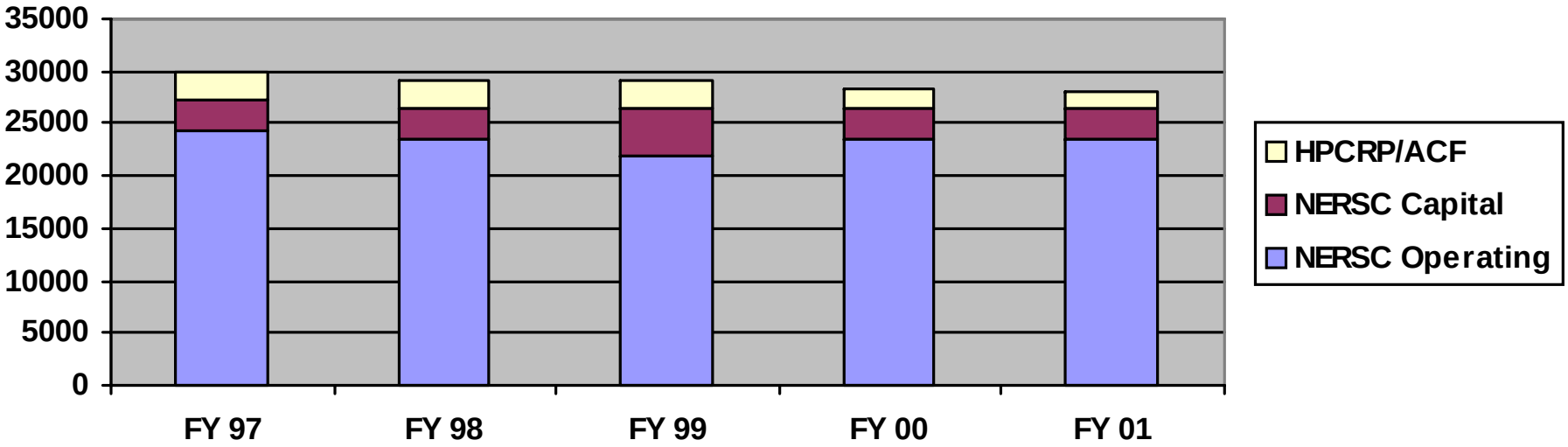
Oakland Scientific Facility

- **20,000 sf computer room; 7,000 sf office space**
  - **16,000 sf computer space built out**
  - **NERSC occupying 12,000 sf**
- **Ten-year lease with 3 five-year options**
- **$10.5M computer room construction costs**
- **Option for additional 20,000+ sf computer room**

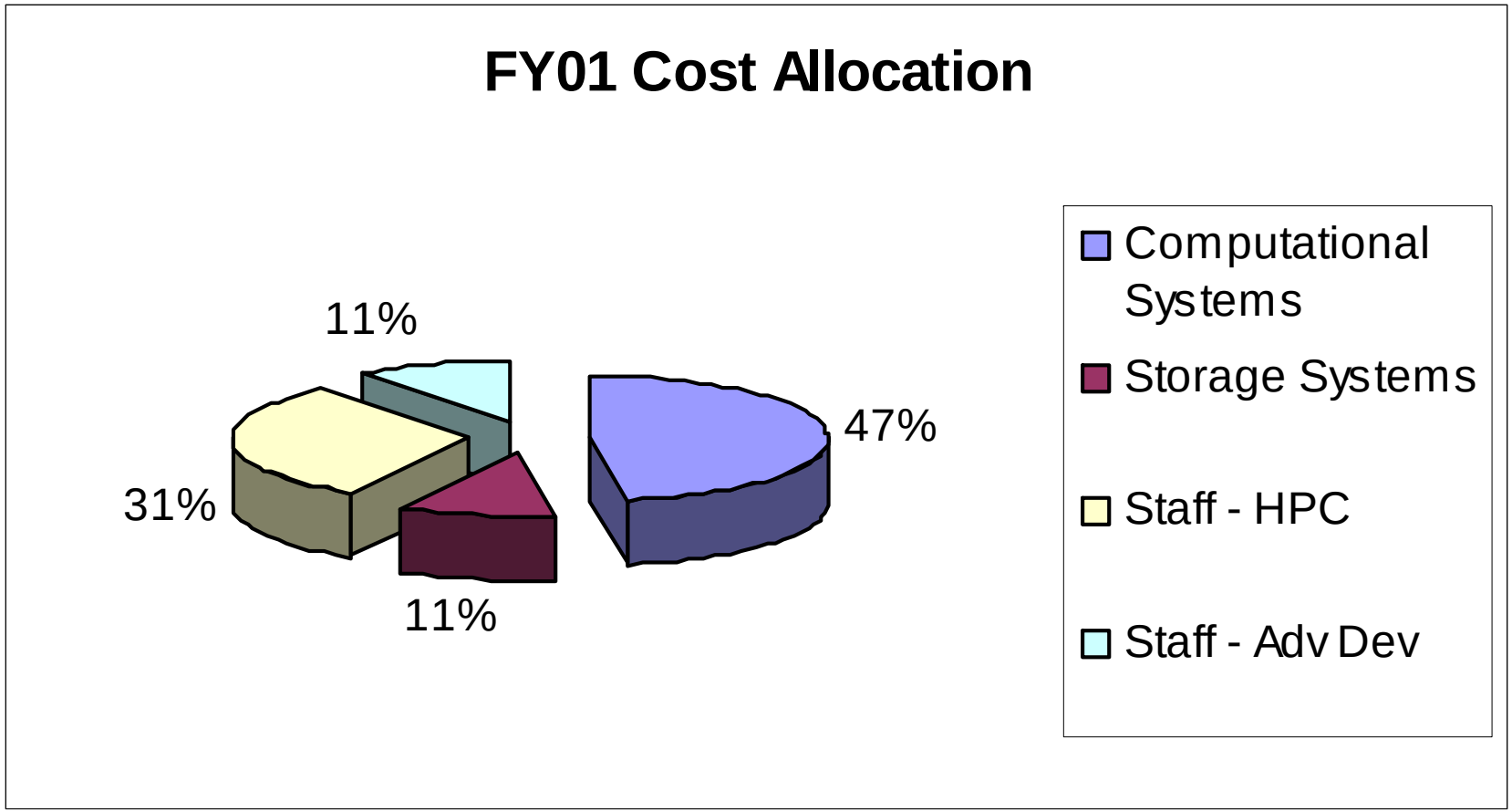


Five Year Budget History

- **Total facility budget of $28.5 M**
- **Essentially flat (actually declining) for five years**



FY01 Cost Allocation



Leveraging LBNL

- Close interaction with all other OASCR funded projects at LBNL
  - applied mathematics
  - scientific data management
  - distributed computing and collaborative technology
- LBNL is funding about $3M in computational LDRD/year
  - Initiated several new scientific projects with impact on computational science

Bay Area Community

- UC Berkeley
  - 7 joint appointments
  - NERSC staff involved in teaching
  - gradutate students
  - joint research interestes in numerical algorithms, parallel language/compilers
- UC Davis
  - 4 joint appointments
  - graduate students
  - joint research projects in visualization and numerical algorithms
- MSRI
  - 2 joint appointments
  - organized several joint workshops
  - co-hosted visitors
- ICSI
  - 1 joint appointment

Strategic Collaborations on National Scale

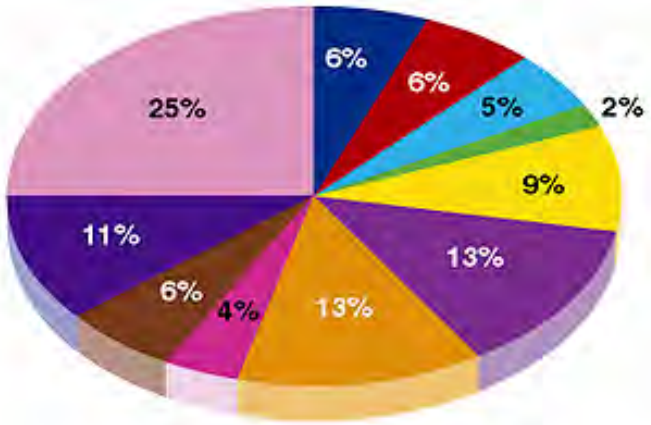
- Office of Science Labs
  - ANL (PC clusters, computational grids, visualization)
  - ORNL (PROBE - distributed storage, facilities, NIM)
  - PNNL (NWChem, collaboratory tools)
- ASCI (LLNL) – IBM platform
- NPACI – member
  - Benchmarking, Tera Evaluation

Overview

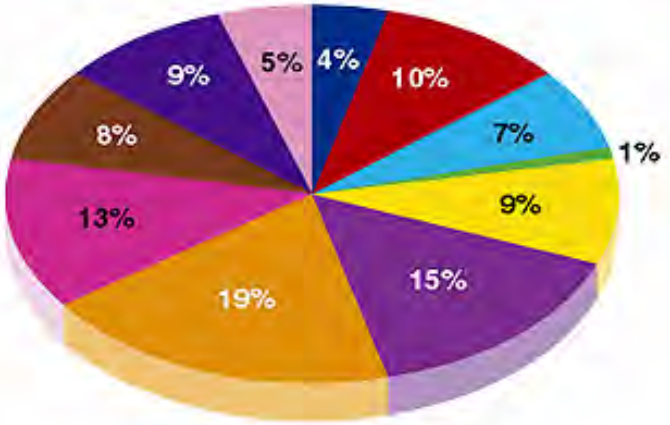
- NERSC resources
- Who uses NERSC?
  - Allocations process
- Comparison to other centers
- Technical accomplishments
- Impact on DOE science
- Future plans

FY00 MPP Users/Usage by Scientific Discipline

NERSC FY00 MPP Users by Discipline

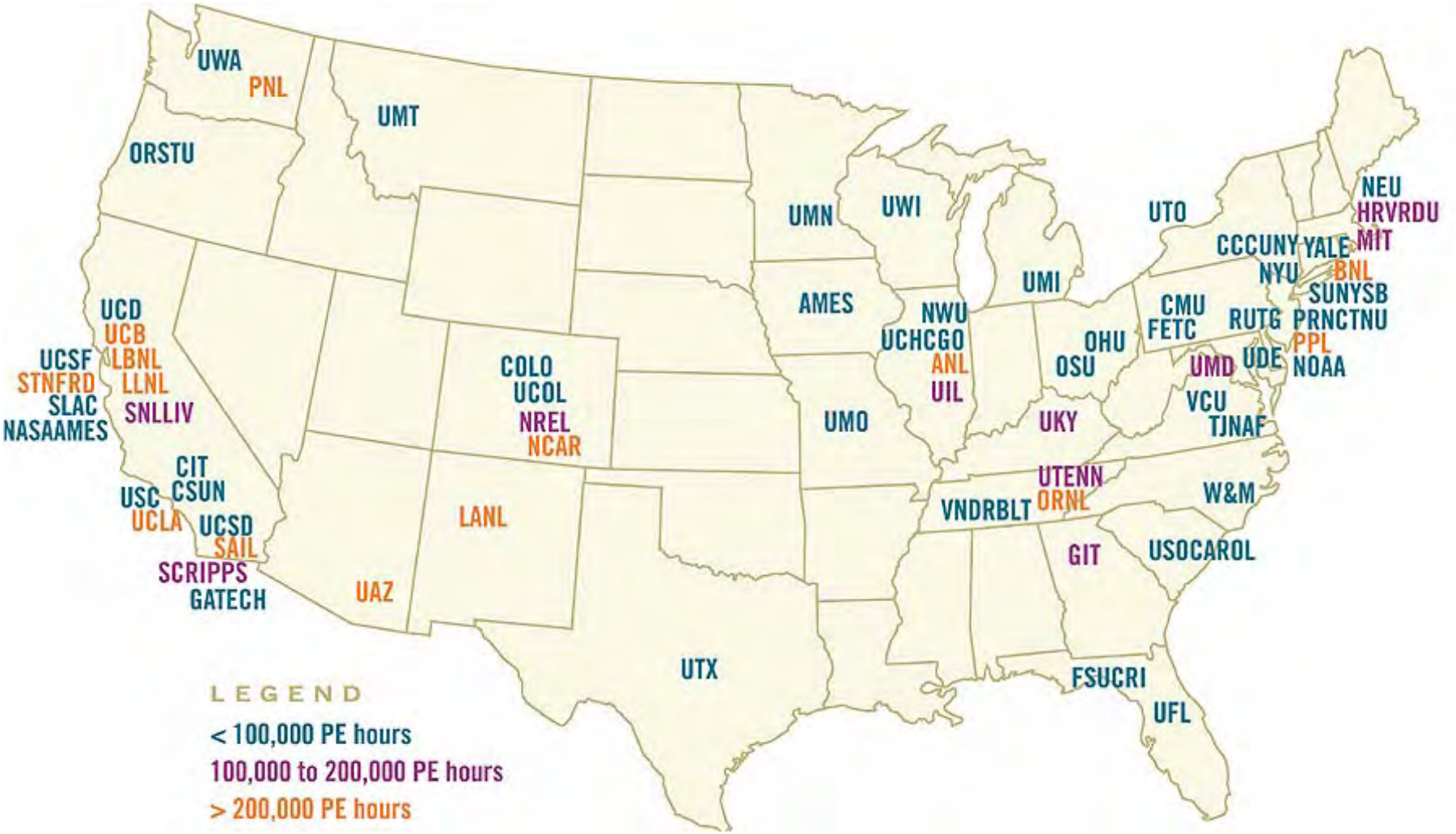


NERSC FY00 MPP Usage by Discipline

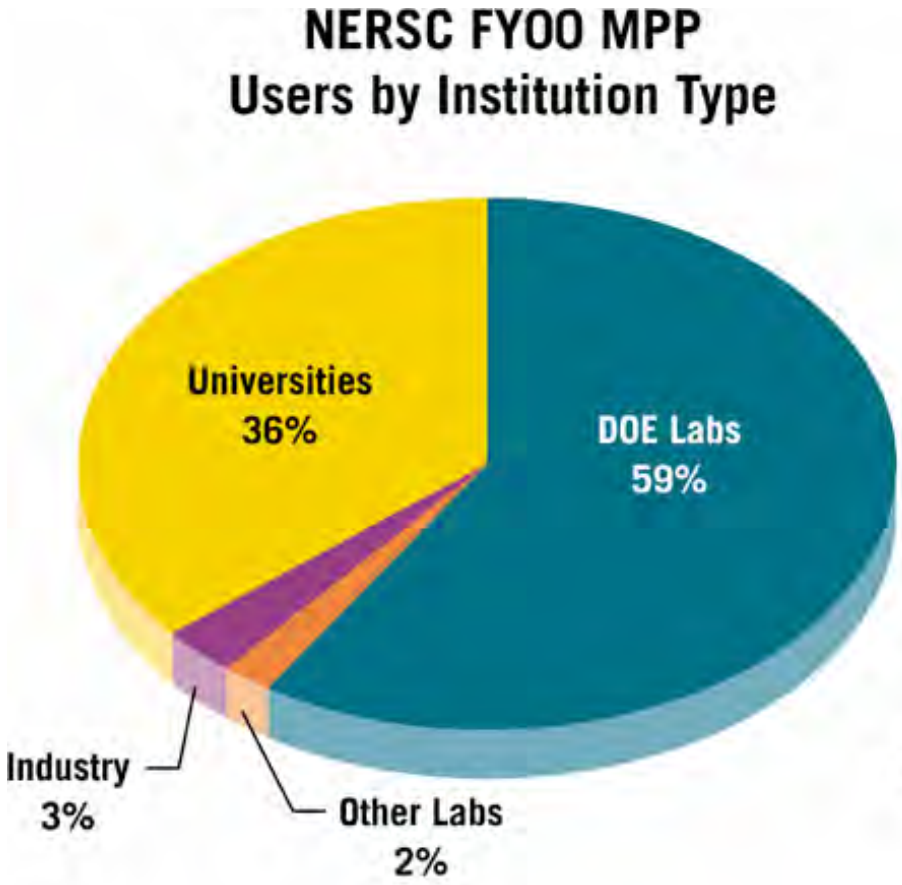


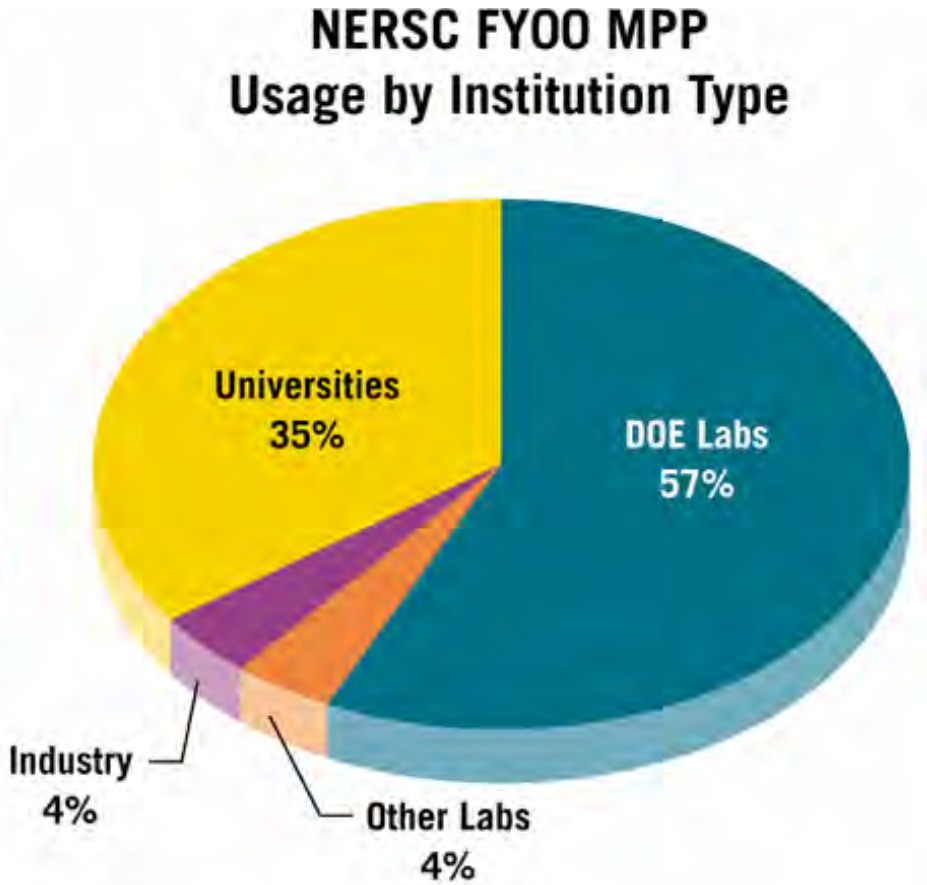
- Computational Science and Mathematics
- Environmental Sciences
- Life Sciences
- Earth and Engineering Sciences
- Chemistry
- Materials Sciences
- Fusion Energy
- High Energy Physics
- Accelerator Physics
- Nuclear Physics
- Other

NERSC FY00 MPP Usage by Site



FY00 MPP Users/Usage by Institution Type





Changes in Allocations Process in 1999

- **Two new boards to help guide NERSC:**
  - **NERSC Policy Board (NPB)**
  - **NERSC Program Advisory Committee (PAC)**

- **Peer review of large and new requests by the Program Advisory Committee**

- **Up to 10% of user resources may be allocated to projects funded by agencies other than DOE Office of Science**

NERSC Policy Board

- **Provides scientific and executive-level advice to the LBNL Director regarding the overall NERSC program and, specifically, on such issues as:**
  - **Resource utilization to maximize the present and future scientific impact of NERSC, and**
  - **Long range planning for the program, including the research and development necessary for future capabilities.**
- **Members: Al Narath (formerly Sandia), Fred Cohen (UCSF), Robert Goldstone (PPPL), Paul Messina (Caltech, DOE-DP), Larry Smarr (NCSA), Michael Witherell (Fermilab), ex officio: Steve Jardin (PPPL), Rob Ryne (LBNL)**

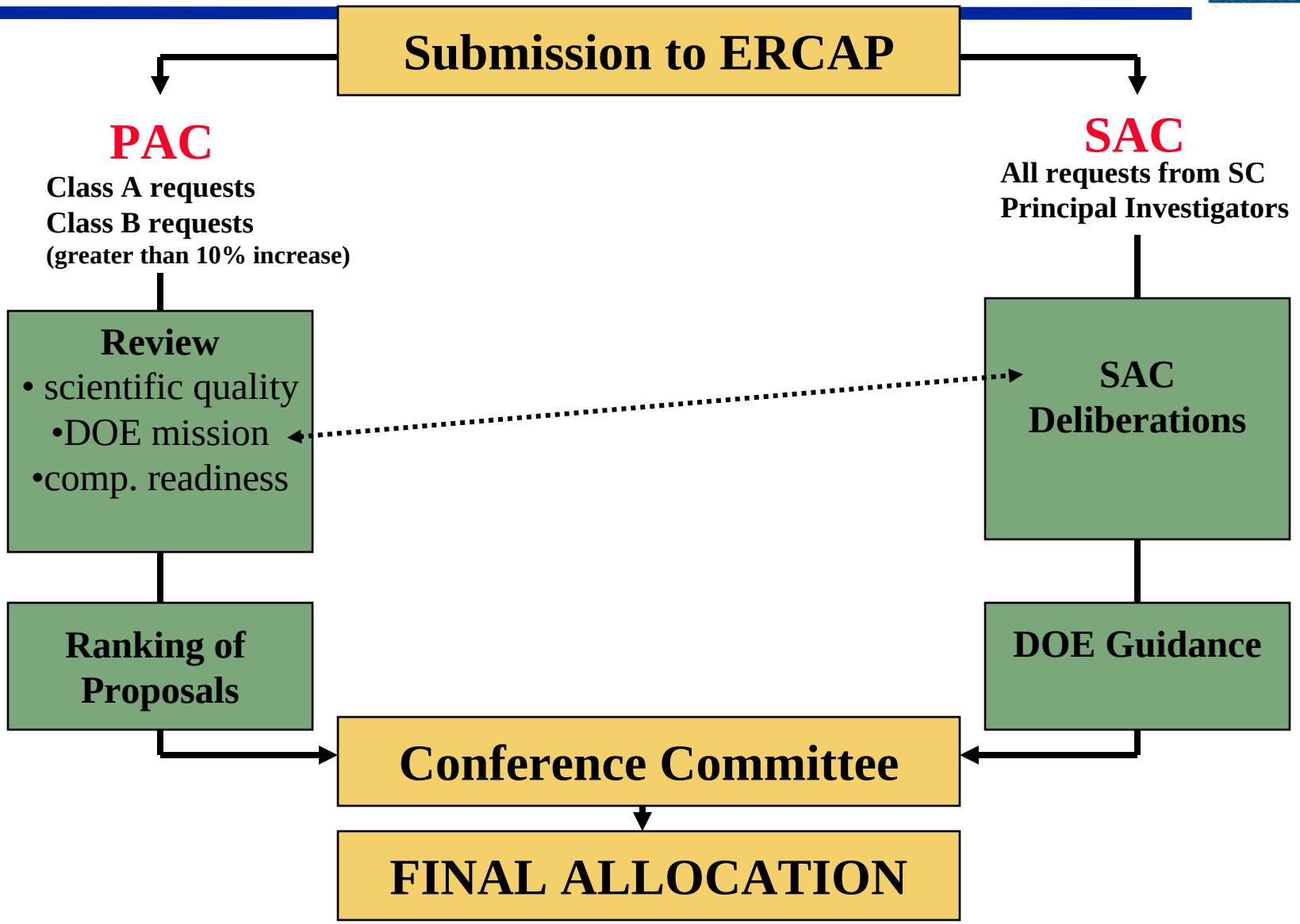
Program Advisory Committee (PAC)

- **Manages the peer review process for allocating 40% of NERSC's user resources.**
- **Advises the NERSC Director.**
- **14 members appointed by the NERSC Director:**
  - **Widely respected scientists, technically active**
  - **Knowledge of computational challenges in their fields**
  - **One representative per science discipline**
  - **Staggered three-year terms**
  - **The Berkeley Lab Director, the NERSC Policy Board, The NERSC Users Group, or the U.S. scientific community may nominate candidates for PAC**

PAC Members

Steve Jardin, PPPL	PAC Chair
David Bailey, NERSC	NERSC Liaison
Ian Foster, ANL	Computer Science
Doug Rotman, LLNL	Environmental Sciences
Bruce Harmon, Ames Lab	Material Sciences
Robert Harrison, PNNL	Chemistry
Jean-Noel Leboeuf, UCLA	Fusion Energy
Greg Newman, Sandia	Geo Sciences
Rob Ryne, LBNL	Accelerator Physics
S. Subramannian, UCSD	Life Sciences
Bob Sugar, UCSB	High Energy Physics Theory
Doug Swesty, SUNY	Nuclear/Astro Physics
Mike Weinert, BNL	Material Sciences
Mary Wheeler, UT Austin	Applied Mathematics

Allocations Process



Impact of New Allocations and Guidance Rule

- Recognition of NERSC a DOE facility

- Advisory structure consistent with DOE facilities
  - stronger advocacy role of NPB
  - user group (NUG) focus on operational aspects of NERSC

- Peer review of facility use
  - assurance of highest quality of science and balanced use of a unique facility
  - counters criticism of NERSC as “closed shop”

Overview

- NERSC resources
- Who uses NERSC?
- Comparison to other centers
- Technical accomplishments
- Impact on DOE science
- Future plans

Overview

- NERSC resources
- Who uses NERSC?
- Comparison to other centers
- Technical accomplishments
- Impact on DOE science
- Future plans

Introduction of New Technology

Cluster Computing

Algorithms and Software

Benchmarking and Performance Evaluation

Computer Science

Distributed Visualization

Scientific Data Management

Computational and Data Grids

Introduction of New Technology

4/00: Developed Grids Plan

11/99: ESP Benchmark

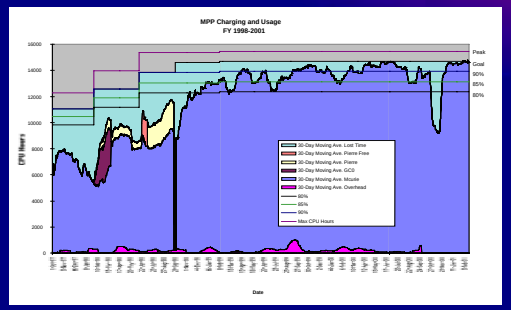
4/98: Millennium Project with UCB

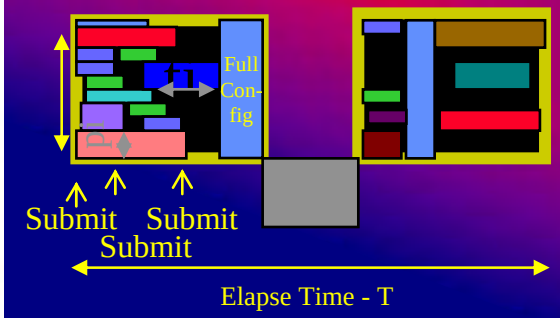
10/97: First check-point restarting on parallel platform

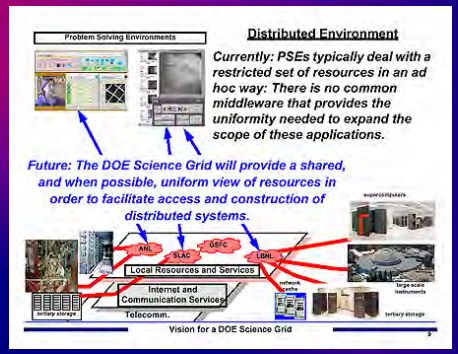
1996

2001

Balanced high quality production with rapid introduction of new technology







Cluster Computing

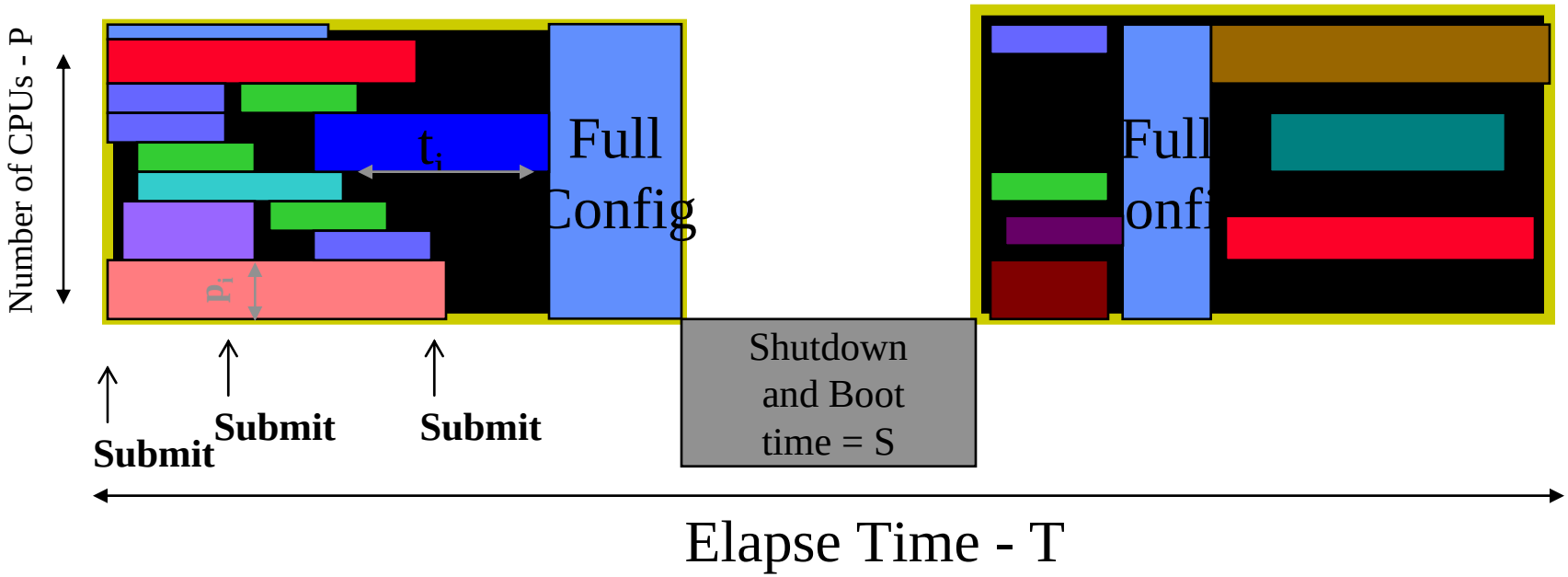
1998

- NERSC became partner in the UCB Millennium project ($6M grant from Intel for building PC clusters)
- Stimulated PC cluster project in the Future Techno Group at NERSC
- Brought UCB students to NERSC
- Leveraged into LTR funding for collaboration with Intel about VIA

2000

- NERSC Cluster team formed
- Evaluation of PC Clusters as alternate production platform
- Release of M-VIA software

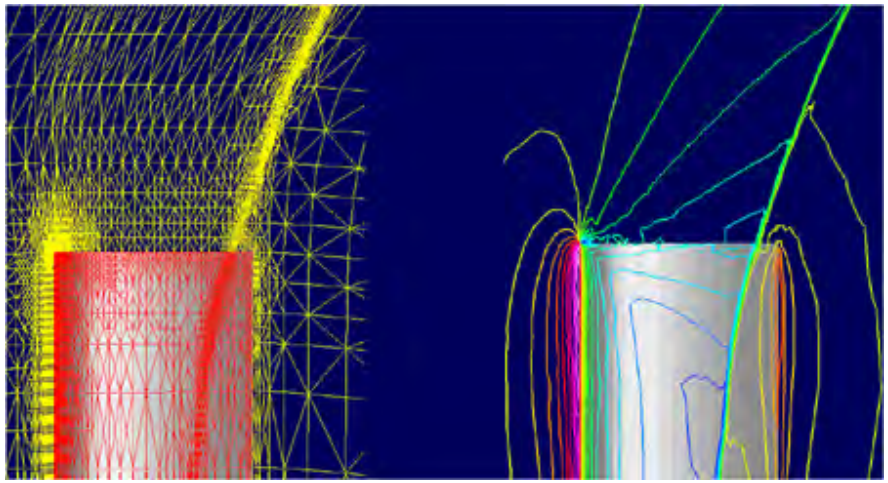
ESP Framework



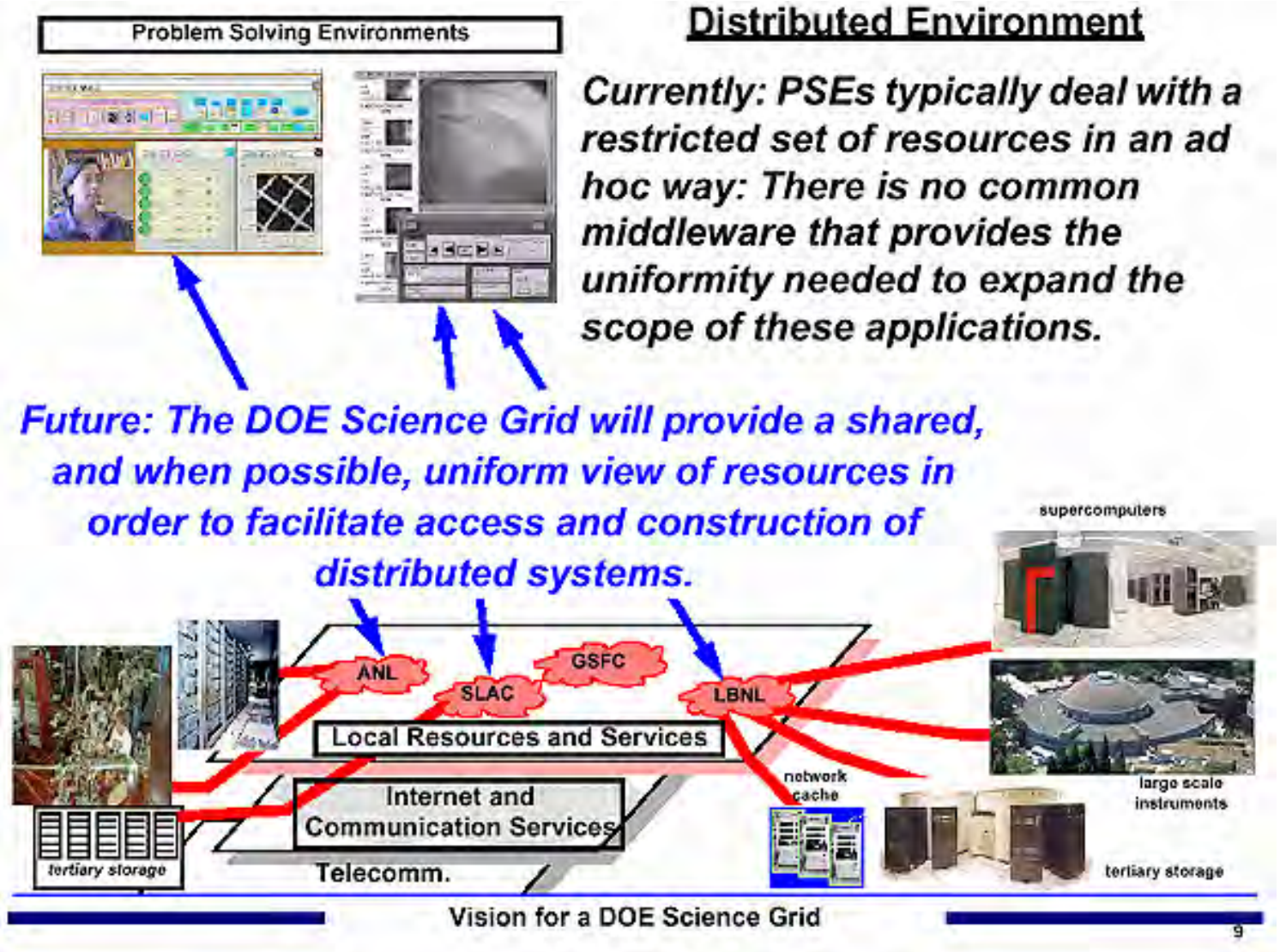
$$\text{Effectiveness} = (\sum_{i=1,N} p_i * t_i)/[P*(S+T)]$$

Technology Evaluation Example

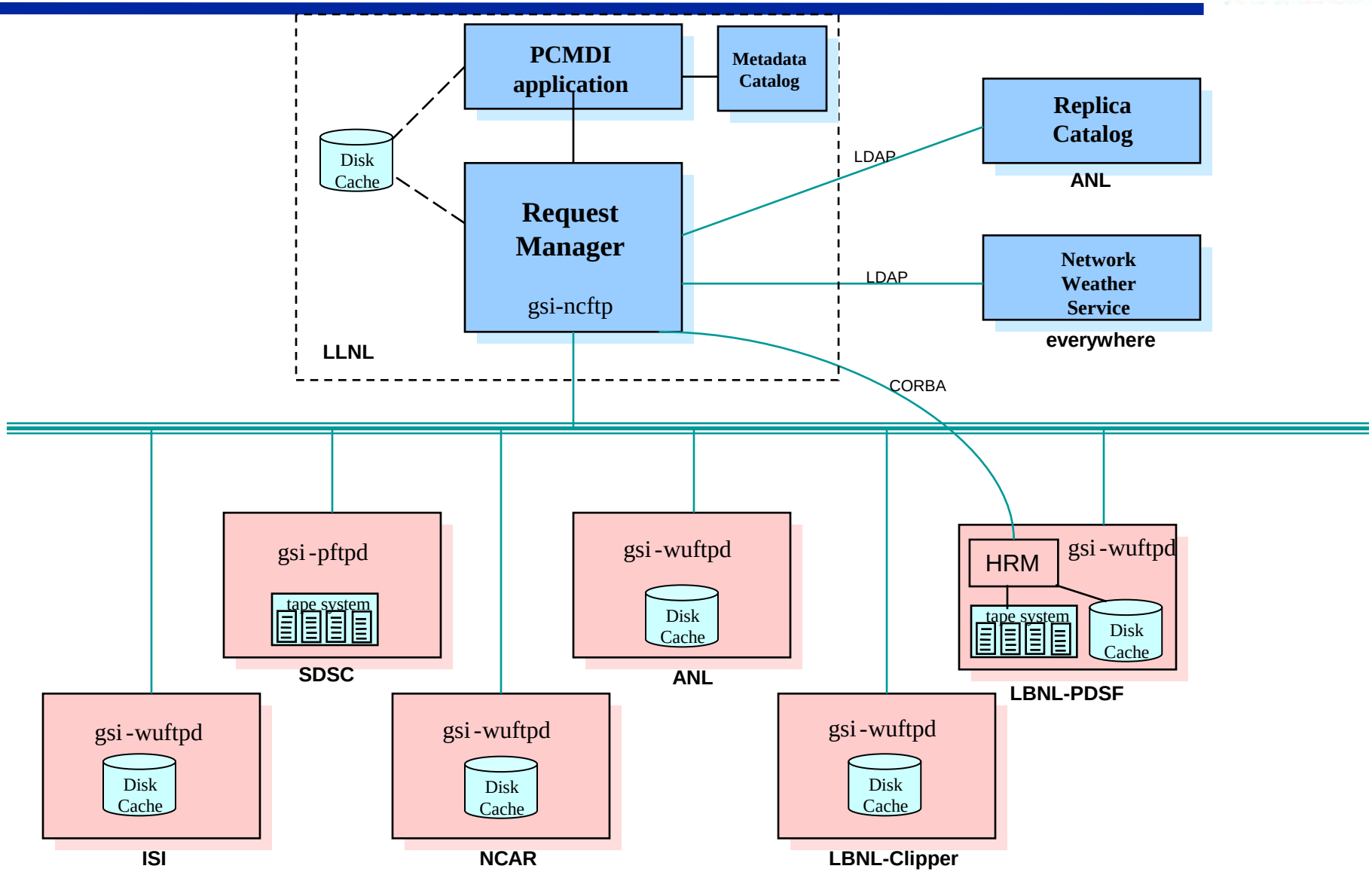
- Participated in Evaluation of Tera MTA architecture jointly with SDSC
- Best paper award for Lenny Oliker at SC99 for comparison of MPI, OpenMP, and multi-threaded implementation on T3E, Origin 2000, and Tera MTA
- Currently participating in SV-2 evaluation; study of performance trade-offs in shared memory SMP nodes



Vision for a DOE Science Grid

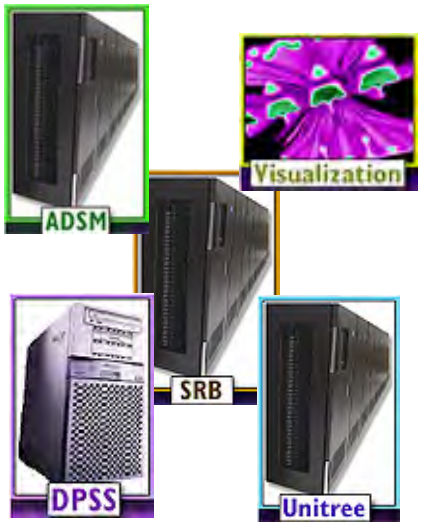


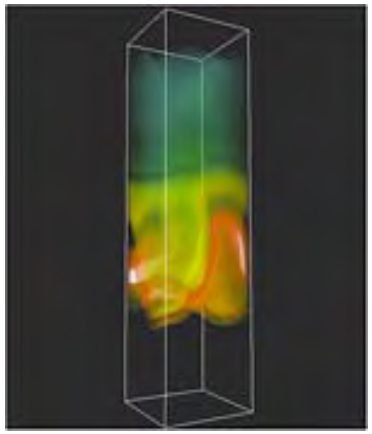
Earth Systems Grid Prototype



Visapult

Back-End Data Access/Compositing Engine



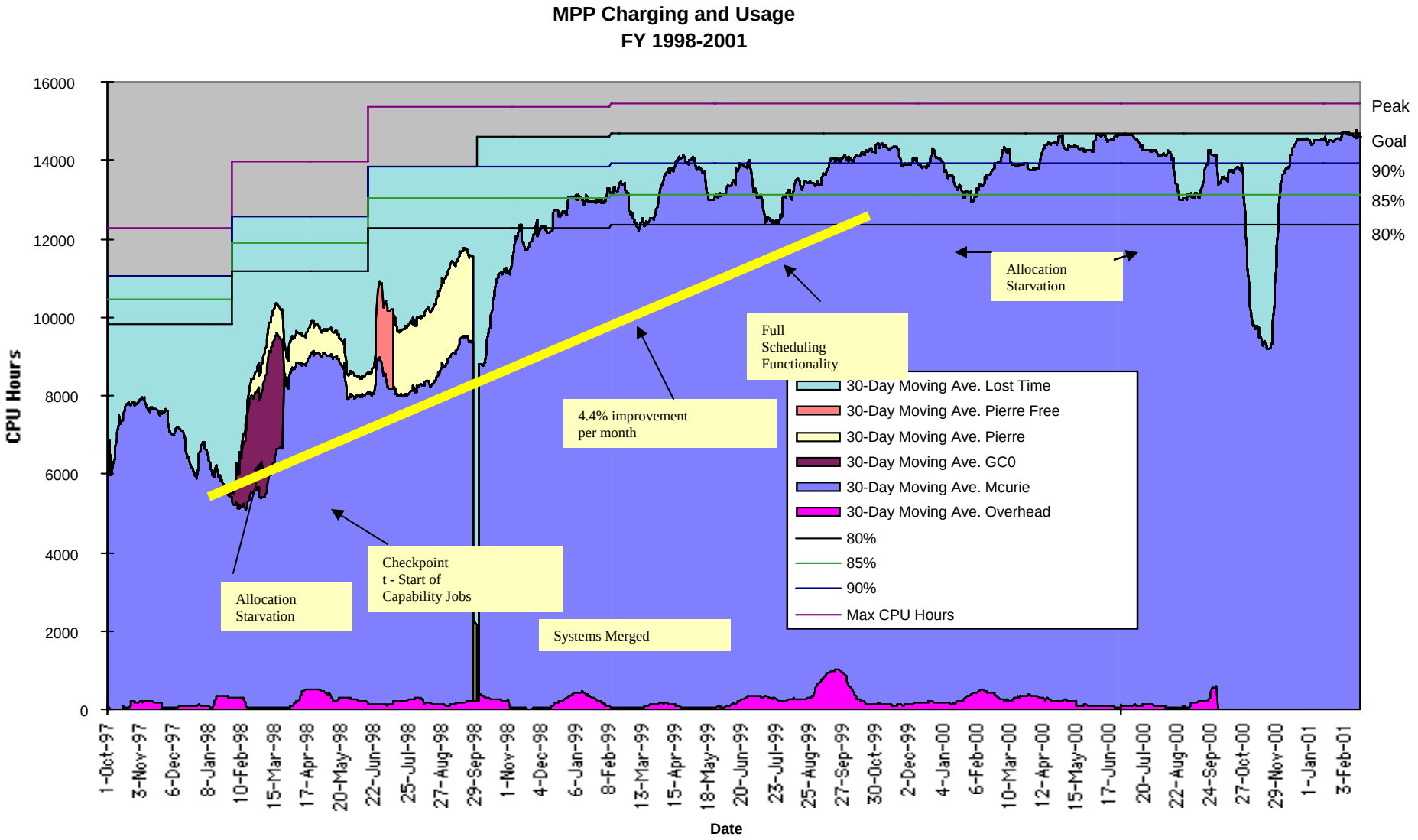


Rendering Engine

Network Interface

1.48 Gb/sec in the bandwidth challenge at SC00

Impact of Systems Software: 95% Gross Utilization on T3E



Overview

- NERSC resources
- Who uses NERSC?
- Comparison to other centers
- Technical accomplishments
- Impact on DOE science
- Future plans

Serving the DOE Office of Science Community

Allocation Process

Serving the DOE Mission

NERSC: DOE Facility and National Center

National User Community

Diverse Scientific Disciplines

NERSC Impact on DOE Mission

Computational Accelerator Physics Grand Challenge Project (Ko, SLAC and Ryne, LANL)

- **significant impact on the design of several accelerators, including the Next Linear Collider (NLC), the Accelerator Production of Tritium, and the Spallation Neutron Source.**
- **simulations of the NLC resulted in an improved linac design with a higher acceleration gradient, saving $100 million over the original design**
- **only feasible on highly parallel platform**

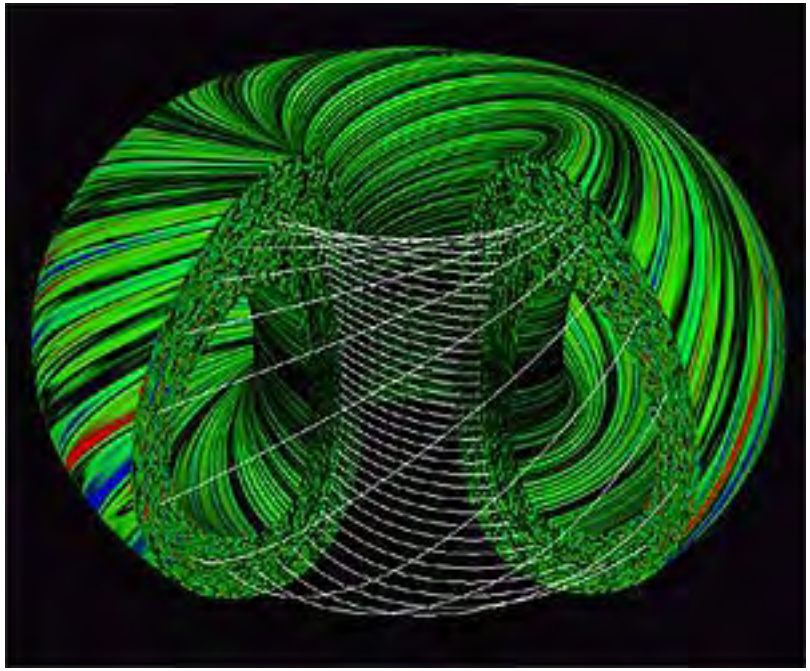
NERSC Impact on DOE Mission

NTTP Grand Challenge:

- **unprecedented progress in understanding turbulent transport in tokamak core plasmas**
- **more than 60 publications based on NERSC results**

Bruce Cohen, LLNL, about NERSC:

"… **All but a very small percentage of these results were obtained on the T3E at NERSC. Essential to obtaining these results on the T3E were the large memory (needed to support high resolution three-dimensional time-dependent calculations and good particle statistics in the kinetic calculations), the large number of processors, the high rate of availability, the excellent production environment (excellent mix of interactive and batch, excellent debugger and job scheduler, and consultants), and the big allocation that this project has enjoyed.** "



Building New Communities: HENP

- ATLAS, D0, CDF, E895, E896, GC5, PHENIX, STAR
- HENP groups which are using or have used (at a significant level) PDSF include: AMANDA, ATLAS, CDF, E871, E895, GC5, NA49, PHENIX, RHIC Theory, SNO, STAR
- Specific software/production projects include:
  - CERNLib port to T3E
  - NERSC personnel (HCG & USG) helped with port of CERNLib to T3E
  - NERSC T3E was used for port of CERNLib
  - NERSC T3E provided 1/2 of data generated by STAR GEANT for first STAR
- Mock Data Challenge
  - Pittsburgh Supercomputing Center T3E provided 1/2 of data
  - Stored on HPSS
  - Transferred using DPSS and pftp

Computing as Tool for Scientific Discovery

High Accuracy Combustion Models

Next Generation Accelerator Design

Scientific Breakthroughs Enabled by NERSC

Cosmology Data Analysis

High-Resolution Climate Models

Computing: A Tool for Discovery

1996

1997: Expanding Universe is Breakthrough of the year

1998: Fernbach and Gordon Bell Award

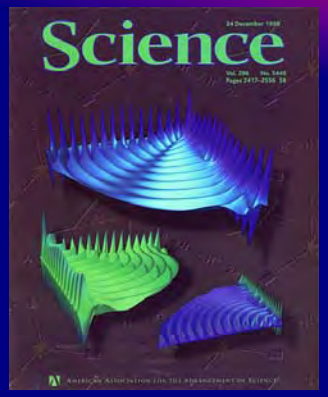
1999: Collisional breakup of quantum system

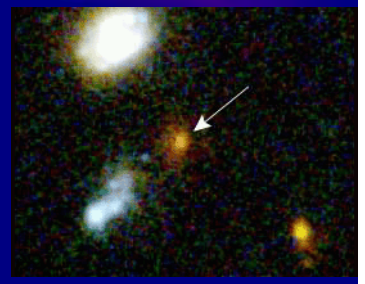
2000: BOOMERANG data analysis= flat universe

2001: Most distant supernova

2001

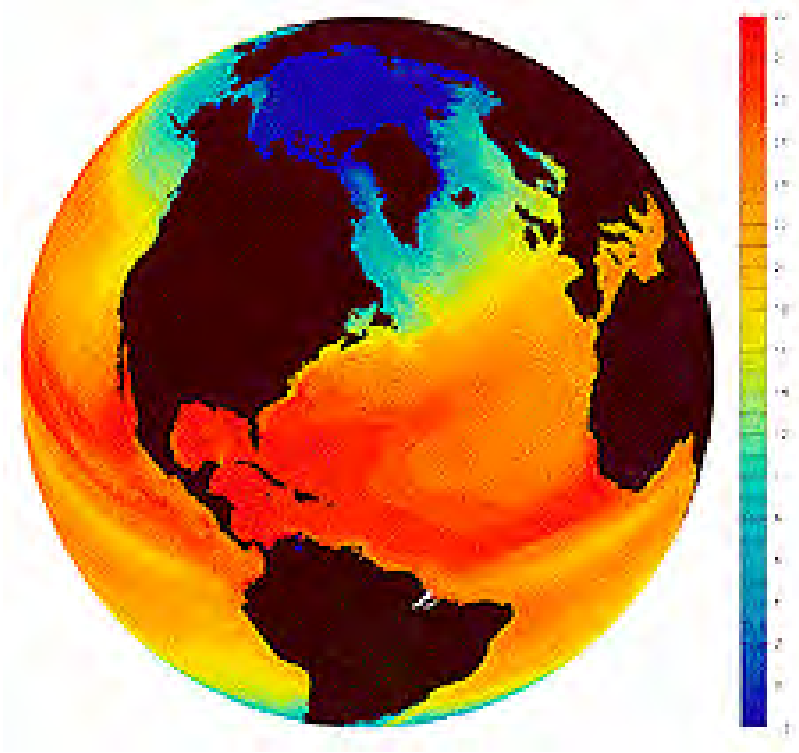
National Energy Research Scientific Computing Center





Computational Science Enabled by NERSC

- Warren Washington, NCAR
  - PCM – parallel climate model
  - Sustained 17 Gflops/s on NERSC T3E
  - Highest sustained performance on climate model in the U.S.
  - Optimized model coupler
- Ported CSM
- Collaborated with GFDL



1998 IEEE Sid Fernbach Award

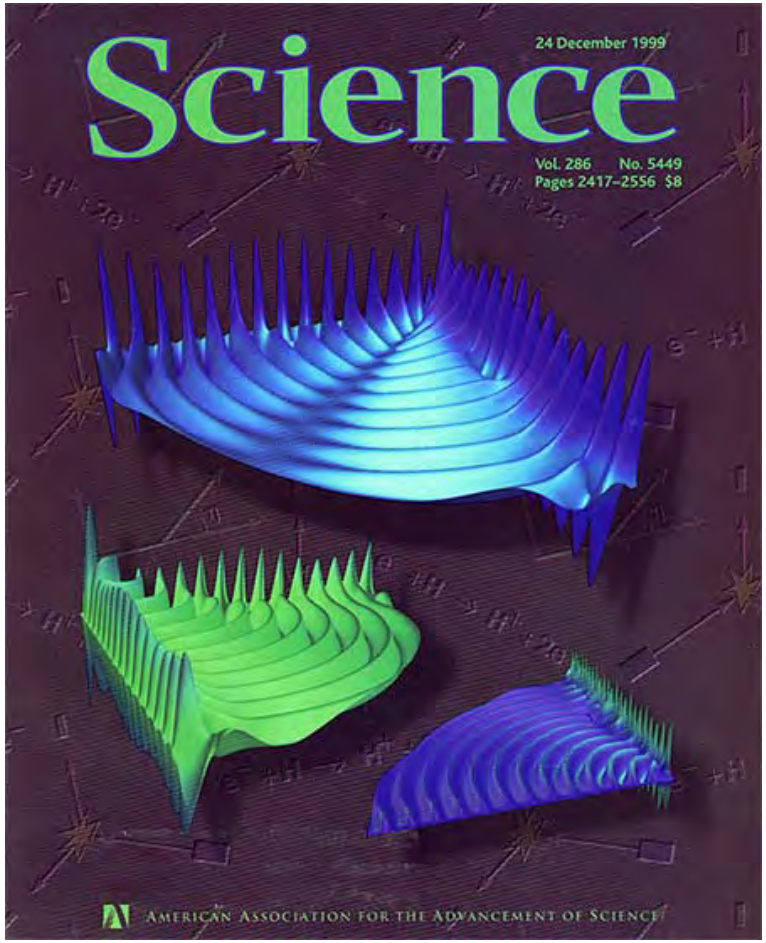
- **CCSE (Center for Computational Science and Engineering) and ANAG (Applied Numerical Algorithms Group)**
- **Combustion modeling and adaptive mesh refinement algorithms**

Phil Colella received the 1998 IEEE Sid Fernbach Award for "fundamental contributions in the development of software methodologies used to solve numerical partial differential equations, and their application to substantially expand our understanding of shock physics and other fluid dynamics problems."

Collaborations are Enabling Scientific Discoveries

- **McCurdy/Baertschy/Isaacs (LBNL)+ Rescigno (LLNL/LBNL)**

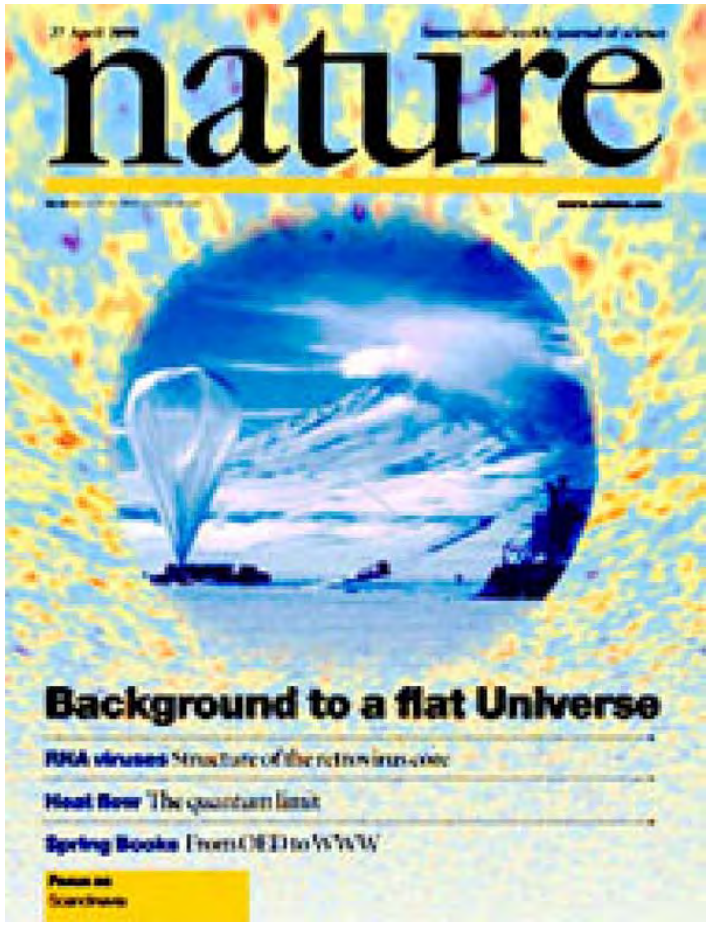
- **First complete solution to collisional breakup in a quantum system of 3 charged particles**
- **Gave rise to large sparse complex nonsymmetric linear systems; orders reach 5 million**
- **Solved with SuperLU developed by Li and Demmel at UCB/NERSC**



Collaborations are Enabling Scientific Discoveries

- **Borrill (LBNL) + CalTech + others.**

- **BOOMERANG Experiments – analyze cosmic microwave background radiation data to obtain a better understanding of the universe**
- **The data analysis provides strong evidence that the geometry of the universe is flat**
- **Developed MADCAP software and provided computational capability on NERSC platforms.**



Nature, April 27, 2000

Overview

- NERSC resources
- Who uses NERSC?
- Comparison to other centers
- Technical accomplishments
- Impact on DOE science
- Future plans

Future Strategy

- High-end systems
- Comprehensive Scientific Support
- Focus on Science Challenge Teams
- Unified Science Environment

Traditional NERSC Computational Strategy

- **Traditional strategy within existing NERSC Program funding**
  Acquire new computational capability every three years
  - **3 to 4 times capability increase of existing systems**

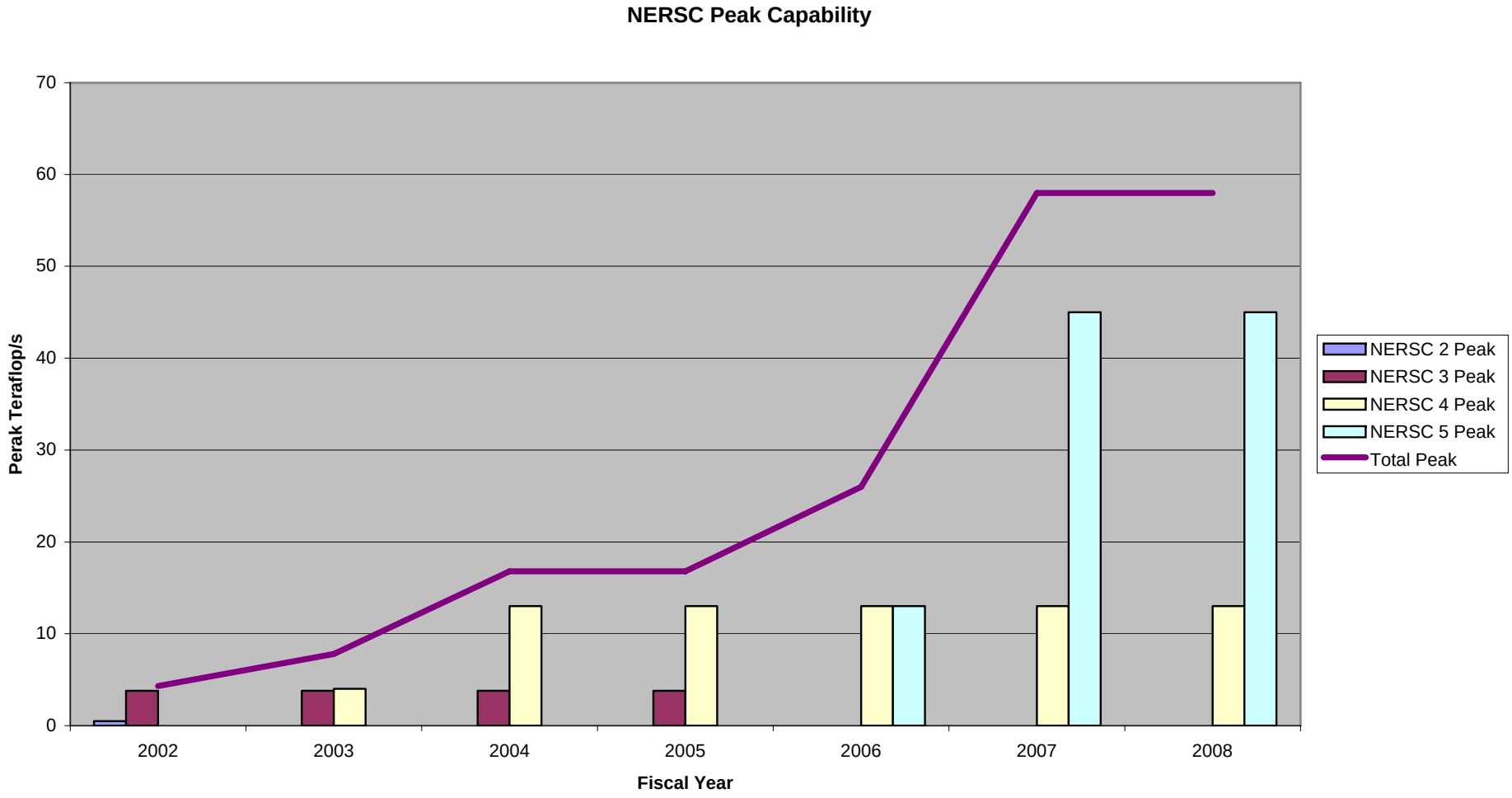
- **Early, commercial, balanced systems with focus on**
  - **stable programming environment**
  - **mature system management tools**
  - **good sustained to peak performance ratio**

- **Total value of $25M - $30M**
  - **About $9-10M/yr. using lease to own**

- **Have two generations in service at a time**
  - **e.g. T3E and IBM SP**

- **Phased introduction**

Projected Future Peak Capability



2000 - 2005: Technology Options

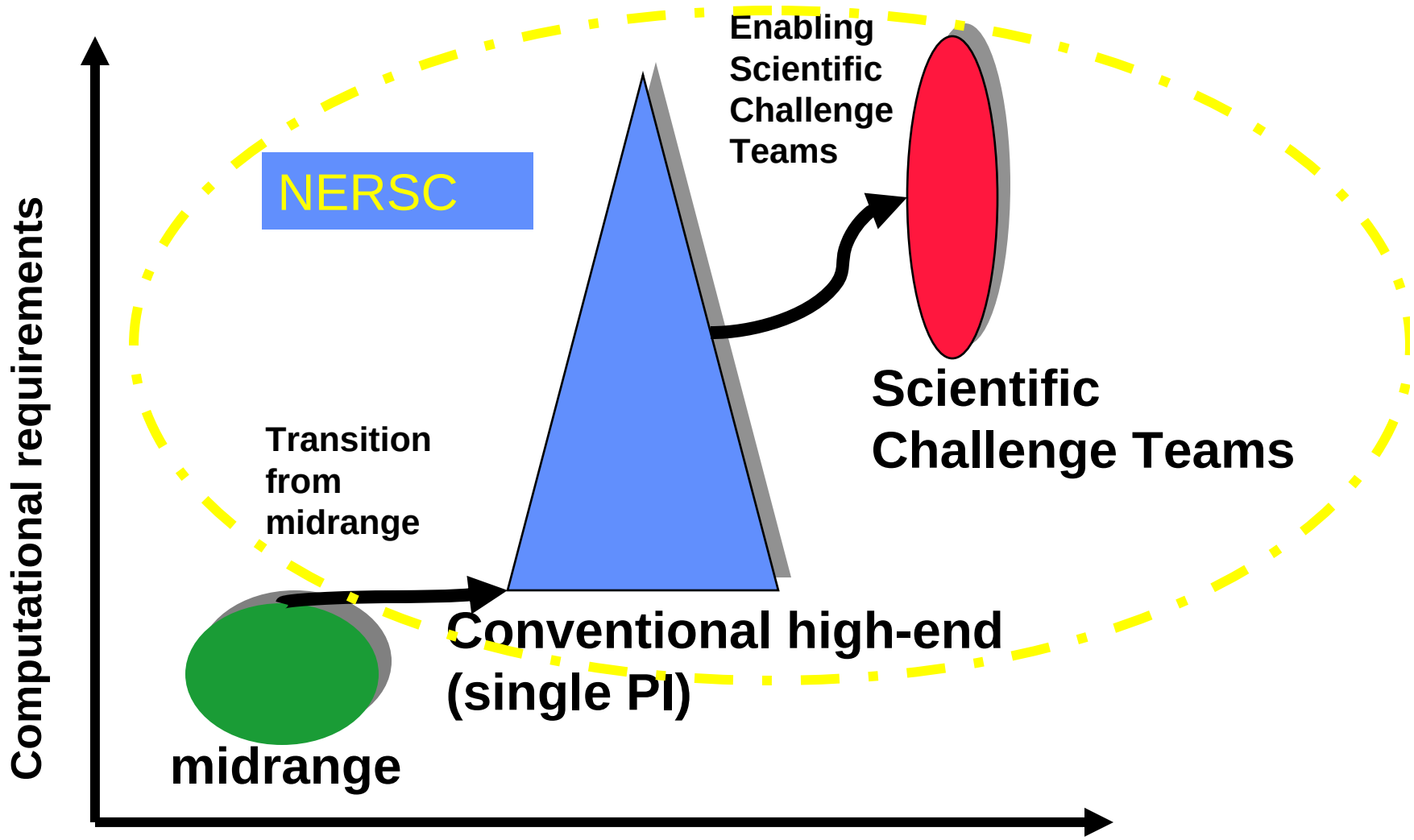
- Clusters
  - SMP nodes, with custom interconnect
  - PCs, with commodity interconnect
  - vector nodes (in Japan)
- Custom built supercomputers
  - Cray SV-2
  - IBM Blue Gene
- Other technology to influence HPC
  - IRAM/PIM
  - low power processors (Transmeta)
  - consumer electronics (Playstation 2)
  - Internet computing

Summary on Technology Assessment

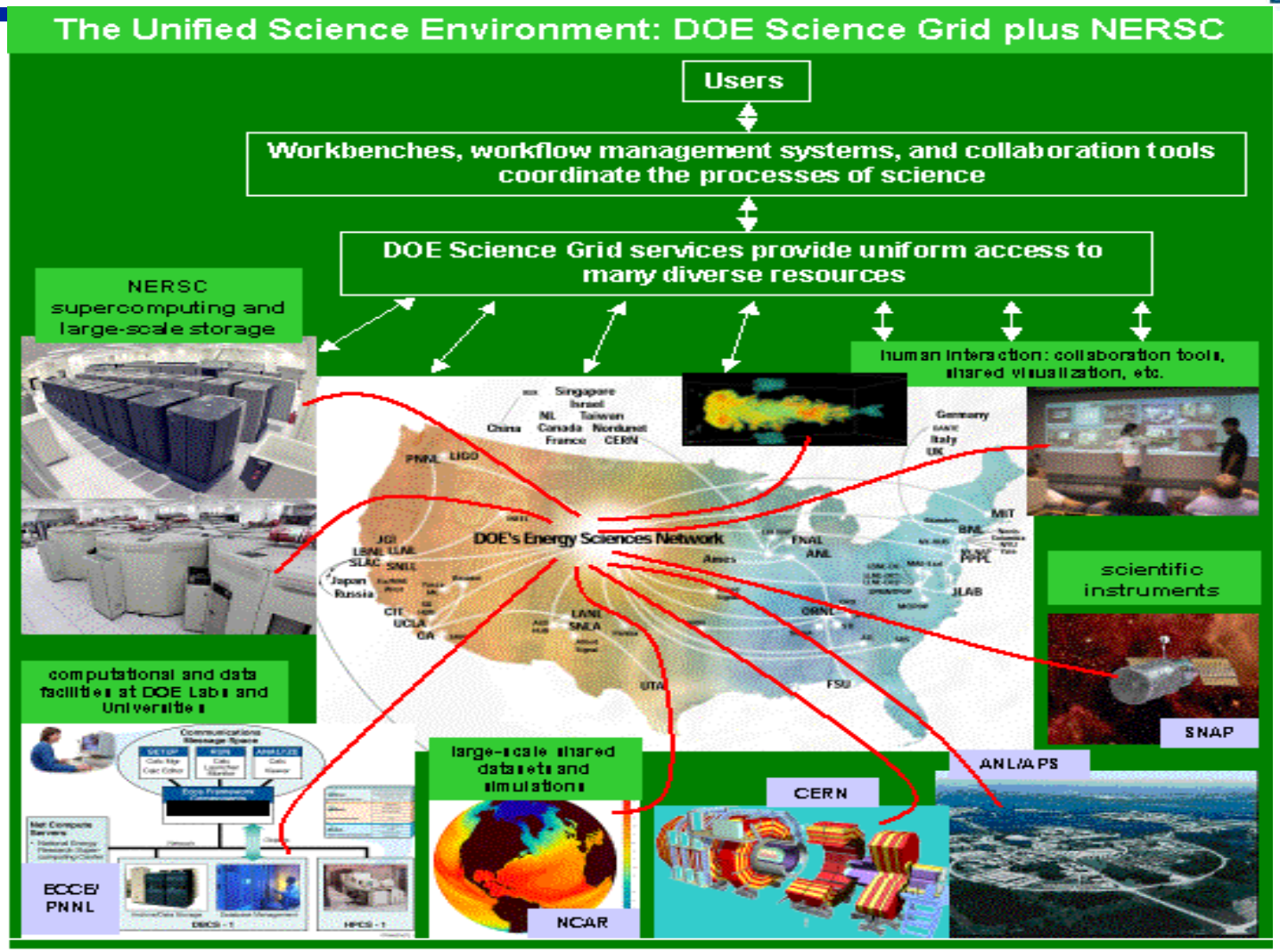
Likelihood that technology will be chosen

	NERSC-4 FY2003	NERSC-5 FY2006
Cluster of SMP	75%	40%
PC Cluster	20%	40%
Vectors (Japanese)	0.1%	0%
Custom built(SV-2)	4.9%	5% (or 0%??)
New technology	0%	15%

Future NERSC Spectrum of Usage

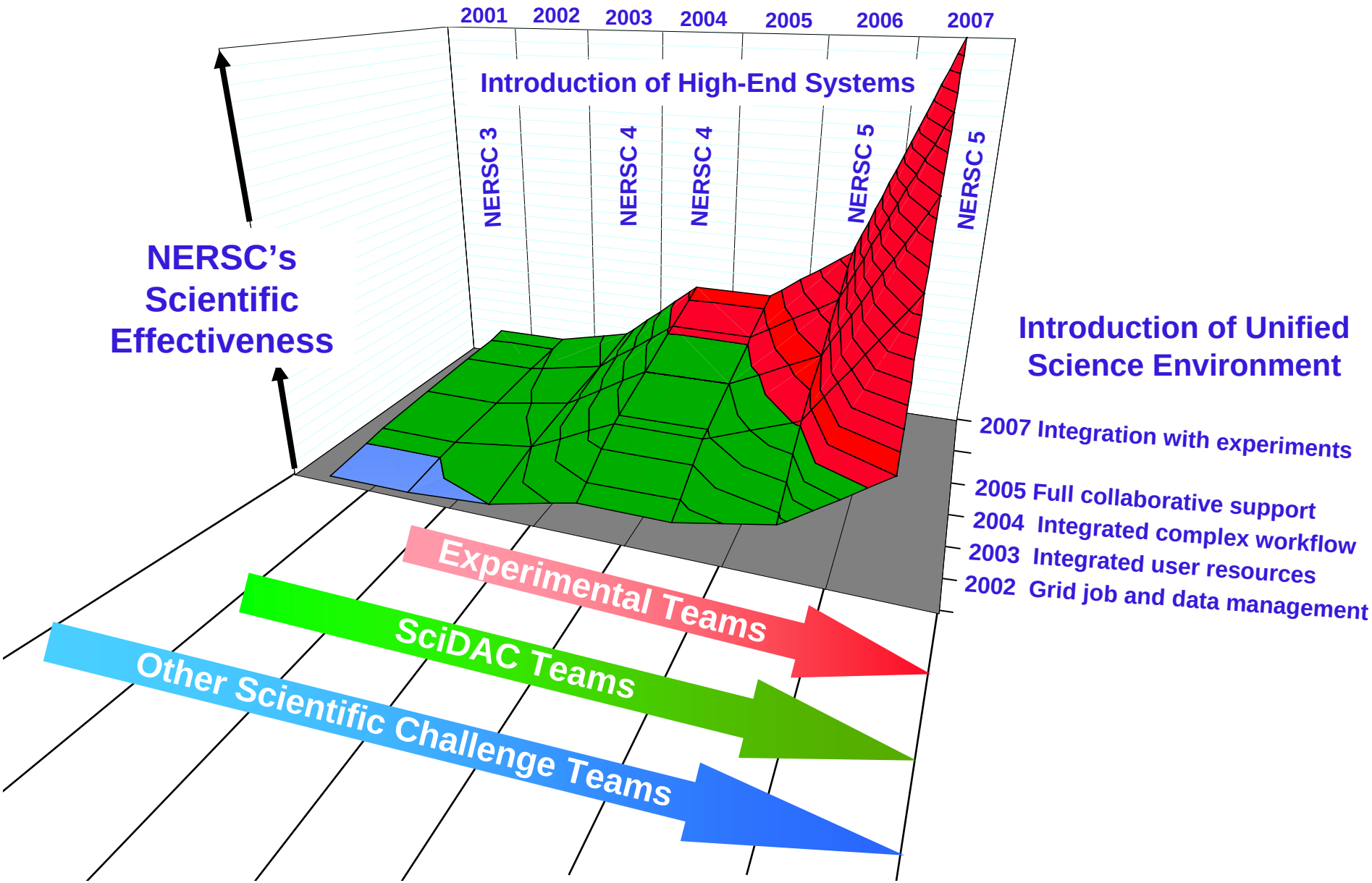


The Unified Science Environment: DOE Science Grid Plus NERSC



In the future, science will depend on the interaction and interoperation of simulation (computing), data (large-scale archives), scientific instruments, and collaborators at many different institutions. Uniform access, large-scale distributed system construction tools, security, and coupling NERSC to DOE Office of Science's other facilities, will produce a Unified Science Environment.

The next five years



Challenges

- **Flat budget means declining investments for systems**
- **NERSC must meet additional requirements: SciDAC, grids**
- **Office of Science computing (and hence NERSC) gets "squeezed" between ASCI and NSF**
  - **two new NSF facilities in FY00 and FY01**

Summary

- **NERSC has established an excellent track record in acquiring, installing, and maturing HPC technology**
- **NERSC has had a major positive impact on computational science in the last couple of years**
- **NERSC has taken maximum advantage of its intellectual resources to advance the state of the facility and to increase its value to DOE.**