National Energy Research Scientific Computing Center (NERSC)

NERSC Overview

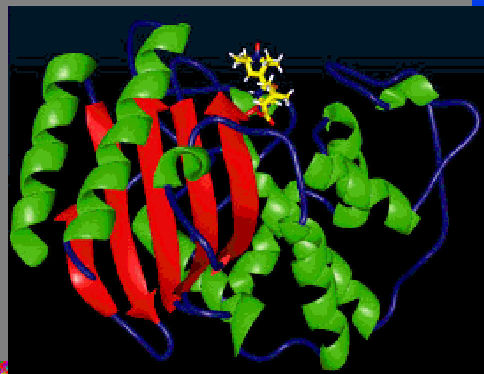
Horst D. Simon
Director, NERSC Center Division, LBNL

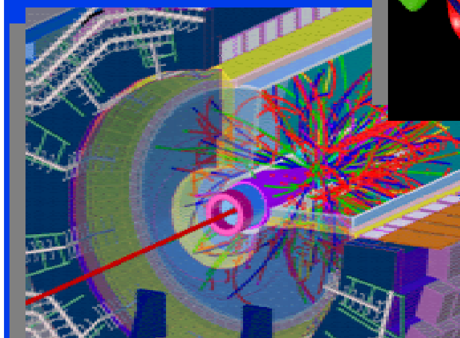
January 13, 2003

NERSC Center Overview

- Funded by DOE, annual budget $28M, about 65 staff
- Supports open, unclassified, basic research
- Located in the hills next to University of California, Berkeley campus
- close collaborations between university and NERSC in computer science and computational science

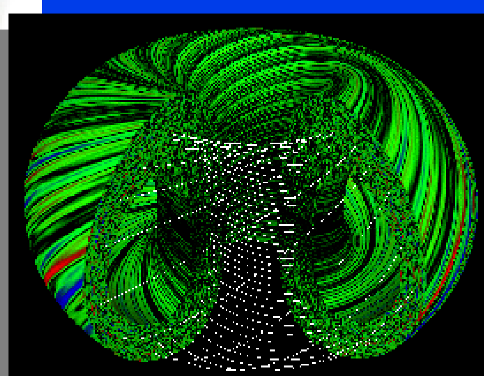
National Energy Research Scientific Computing Center

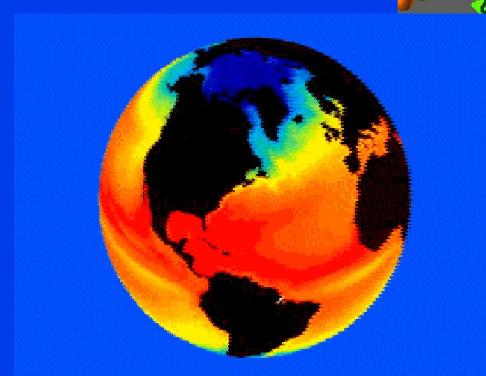




- Serves all disciplines of the DOE Office of Science
- ~2000 Users in ~400 projects

- Focus on large-scale computing





Components of the Next-Generation NERSC

Computational Resources at NERSC

NERSC-3 Vital Statistics

- 5 Teraflop/s Peak Performance – 3.05 Teraflop/s with Linpack
  - 208 nodes, 16 CPUs per node at 1.5 Gflop/s per CPU
  - "Worst case" Sustained System Performance measure .358 Tflop/s (7.2%)
  - "Best Case" Gordon Bell submission 2.46 on 134 nodes (77%)
- 4.5 TB of main memory
  - 140 nodes with 16 GB each, 64 nodes with 32 GBs, and 4 nodes with 64 GBs.
- 40 TB total disk space
  - 20 TB formatted shared, global, parallel, file space; 15 TB local disk for system usage
- Unique 512 way Double/Single switch configuration

Combined NERSC-3 Characteristics

- The combined NERSC-3/4 system (NERSC-3Base and NERSC-3Enhanced) will have
  - 416 16 way Power 3+ nodes with each CPU at 1.5 Gflop/s
    - 380 for computation
  - 6,656 CPUs – 6,080 for computation
  - Total Peak Performance of 10 Teraflop/s
  - Total Aggregate Memory is 7.8 TB
  - Total GPFS disk will be 44 TB
    - Local system disk is an additional 15 TB
  - Combined SSP-2 measure is 1.238 Tflop/s
  - NERSC-3E be in production by the end of Q1/CY03
    - Nodes arrived in the first two weeks of November
    - Acceptance end of December 2002

Comparison with Other Systems

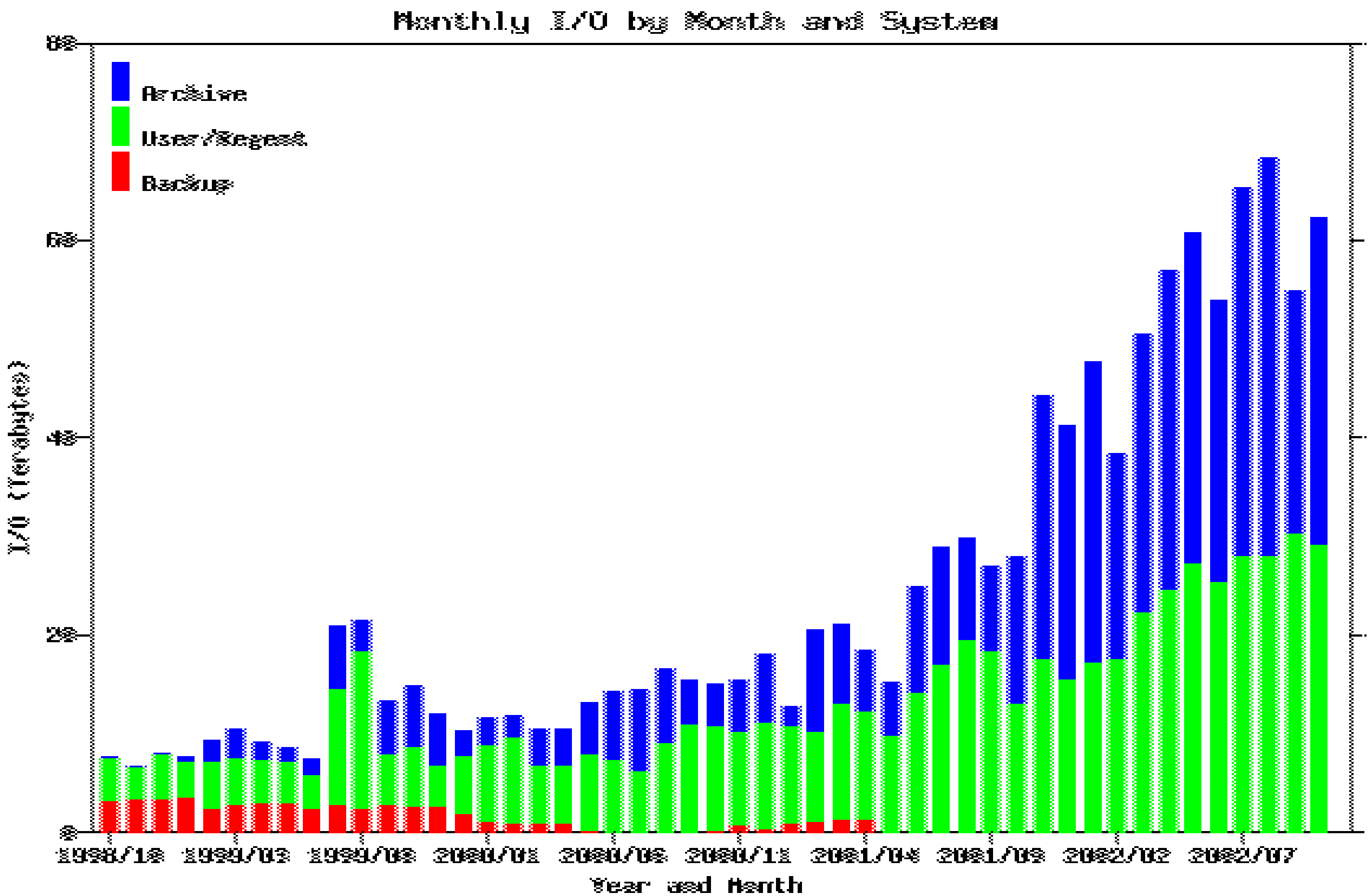
	NERSC-3 E	ASCI White	ES	Cheetah (ORNL)	PNNL Mid 2003
Nodes	416	512	640	27	960
CPUs	6,656	8,192	5,120	864	1900
Peak(Tflops)	10	12	40	4.5	11.4
Memory (TB)	7.8	4	10	1	6.8
Disk(TB)	60	150	700	9	53+234
SSP(Gflop/s)	1,238	1,652		179	

PNNL system available in Q3 CY2003; 53 TB SAN + 234 TB local disk

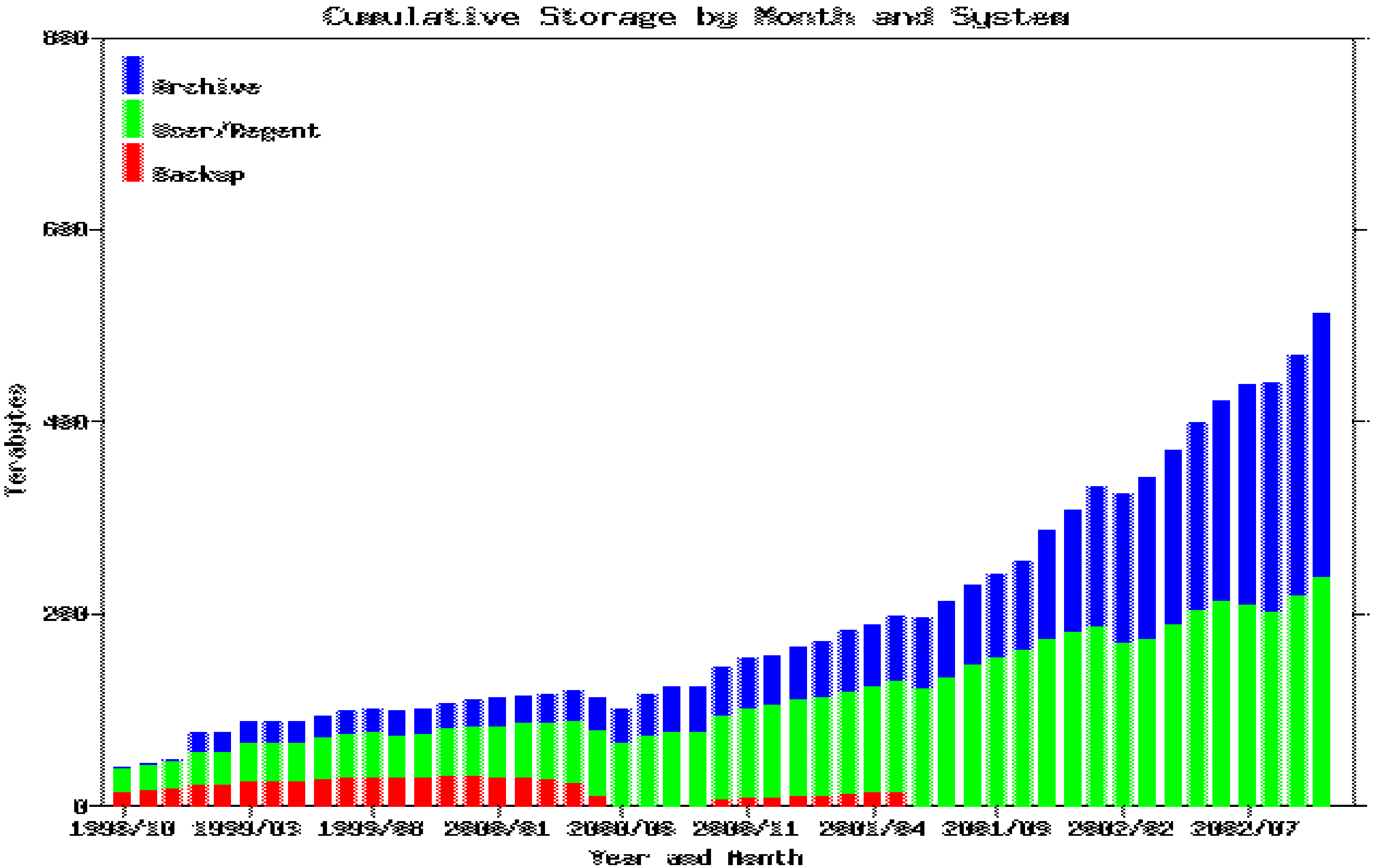
SSP = sustained systems performance (NERSC applications benchmark)

Data Intensive Computing at NERSC

Transfers= 2-3 Terabytes / day



Data in Storage = 1/2 Petabyte
Expect 1 Petabyte in 2003

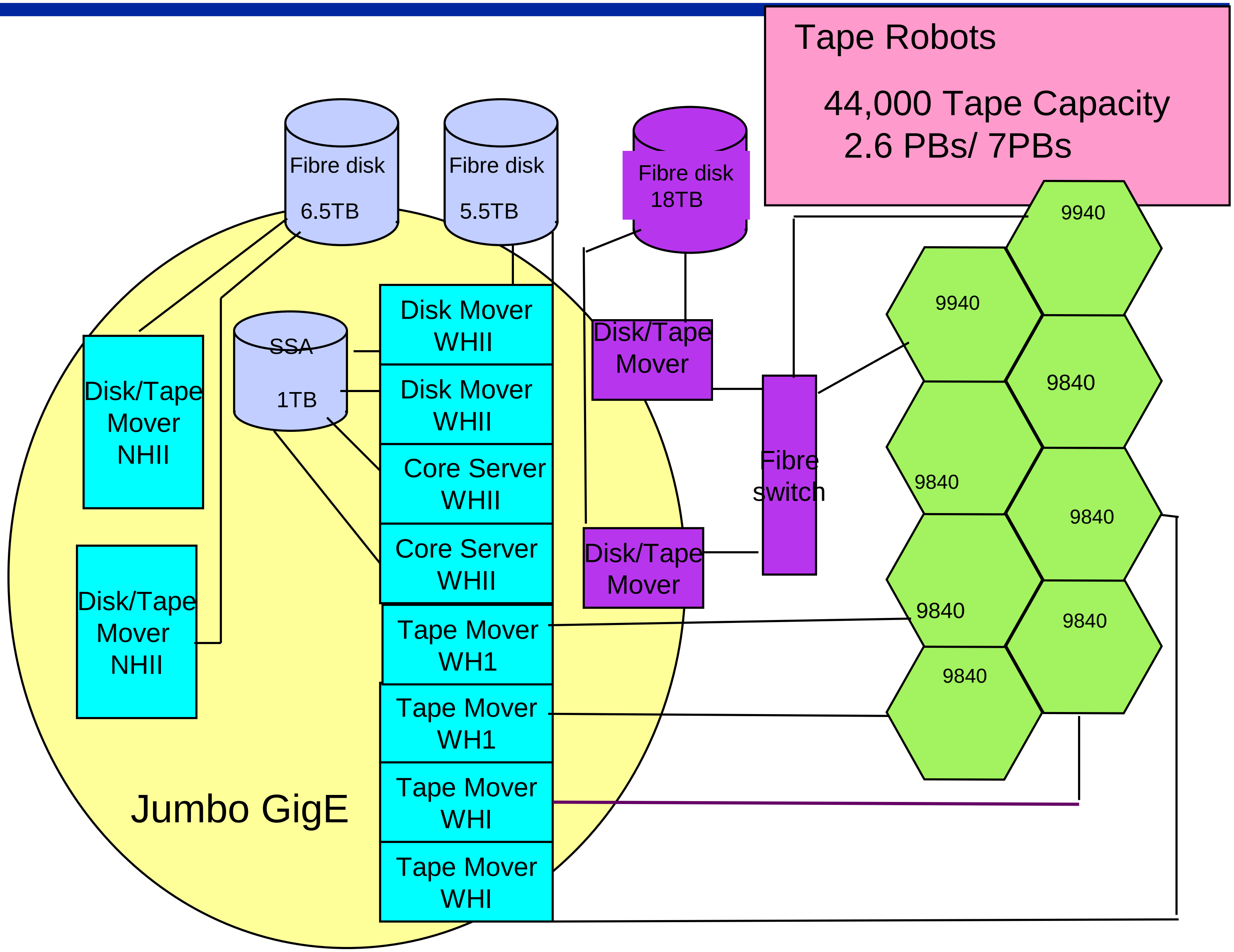


HPSS: Improvements over time

2000/2001	2001/2002	2002/2003
Fibre disk	Fibre disk	Fibre disk
7TBs 90MB/s	14TBs 90MB/s	14TB 90 MB/s 18TBs 200MB/s
GigE/Hippi 40MB/sec	GigE/Jumbo 80 MB/sec	Etherchannel 80MB/sec * N
Scsi tape	Scsi/Fibre tape	Scsi/Fibre tape
20 GB	20 /60GB	20/200 GB
1.3PB	2.6 PB:	7.5 PB
60 drives	70 drives	70 drives

HPSS Production Systems

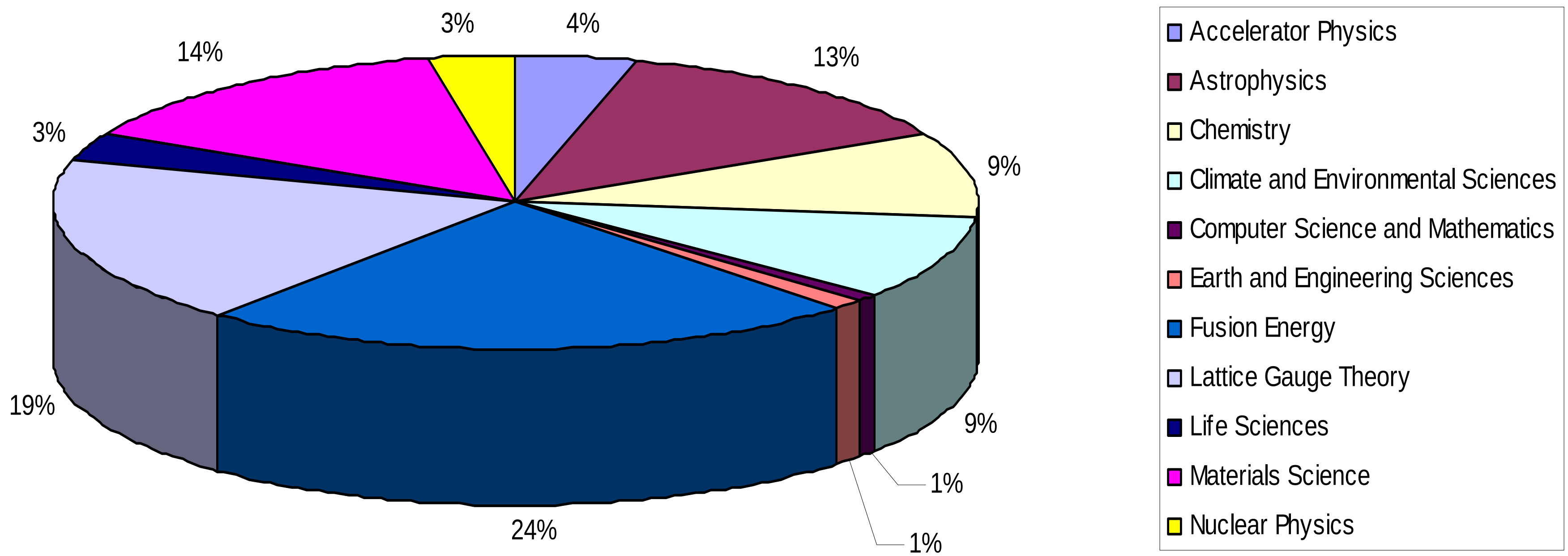
Disk Cache
11TBs 1Gbit
18TBs 2Gbit



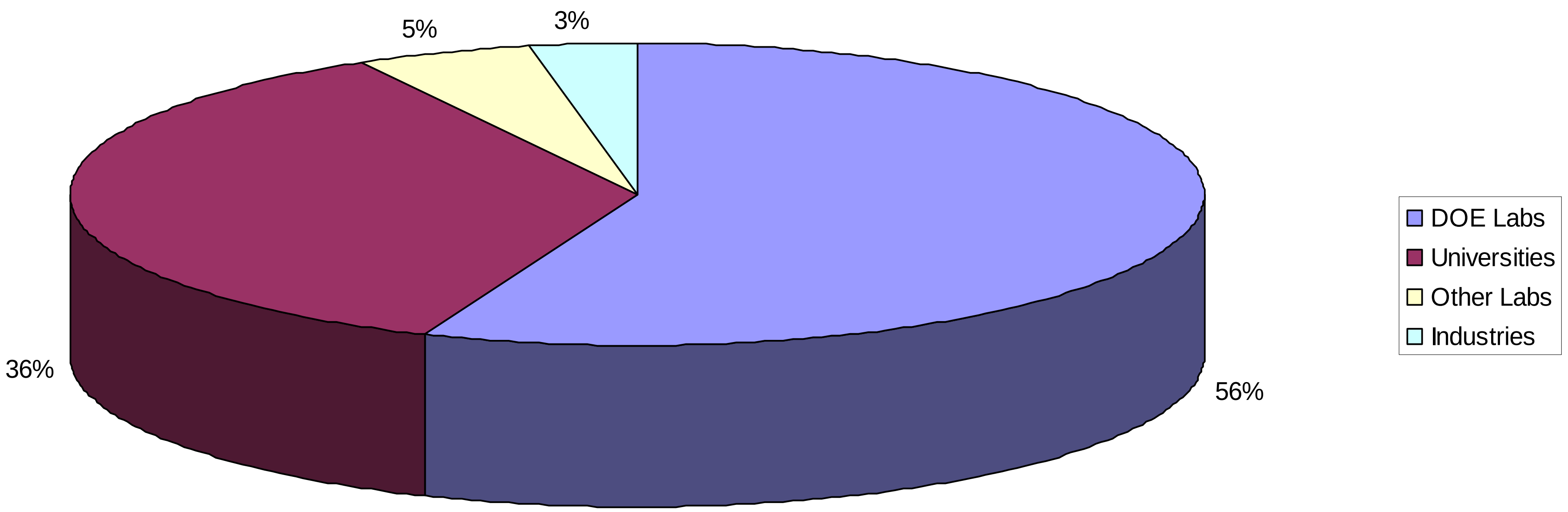
Data Movers:
GigE
Jumbo GigE
Etherchannel

Users and Allocations

NERSC Usage by Scientific Discipline, FY02



NERSC Usage by Institution Type, FY02



FY 2003 Allocations

- DOE initiated a new allocations process for FY 2003.
- Open to all DOE Office of Science mission relevant applications
- Computational Review Panel (CORP) conducts a computational review of all DOE Base requests.
- DOE Program Managers make all production (SciDAC and DOE Base) awards, considering CORP input
- NERSC makes all Startup awards
- Special selections process for “Big Splash”

FY 2003 Allocations – MPP Hours

Allocation Type	MPP K-Hours Requested	MPP K-Hours Awarded (inc. reserves)
DOE Base	78,872	36,100 (66%)
SciDAC	29,960	13,400 (24%)
Big Splash	-	5,500 (10%)
DOE Total	**105,832**	**55,000 (100%)**
Startup	1,045	2,000

Scientific Results

NERSC Goal: 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.

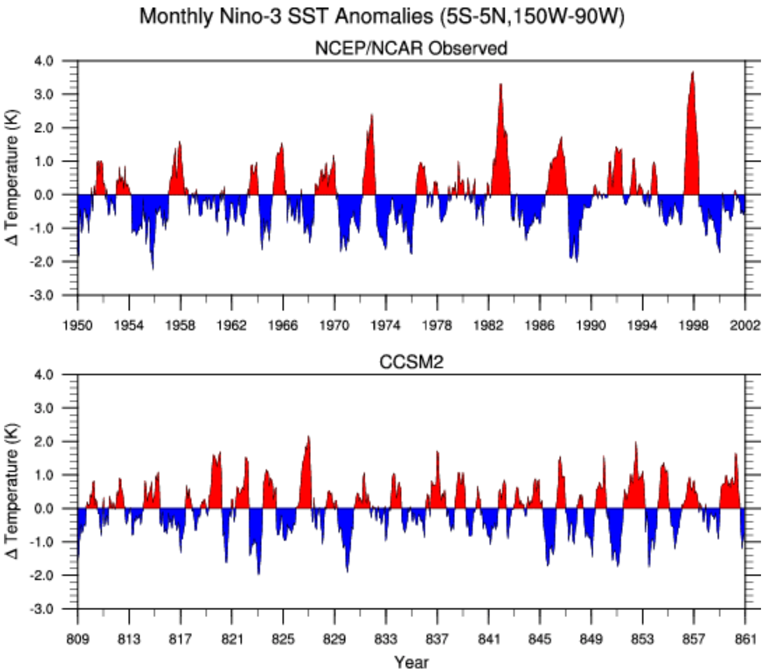
Nature, April 27, 2000

Computational Science at NERSC: A 1000 year climate simulation

- *Warren Washington and Jerry Meehl, National Center for Atmospheric Research; Bert Semtner, Naval Postgraduate School; John Weatherly, U.S. Army Cold Regions Research and Engineering Lab Laboratory; Jeff Kiehl, Jim Hack, and Peter Gent, NCAR.*

• A 1000-year simulation demonstrates the ability of the new Community Climate System Model (CCSM2) to produce a long-term, stable representation of the earth's climate.

•760,000 processor hours by July



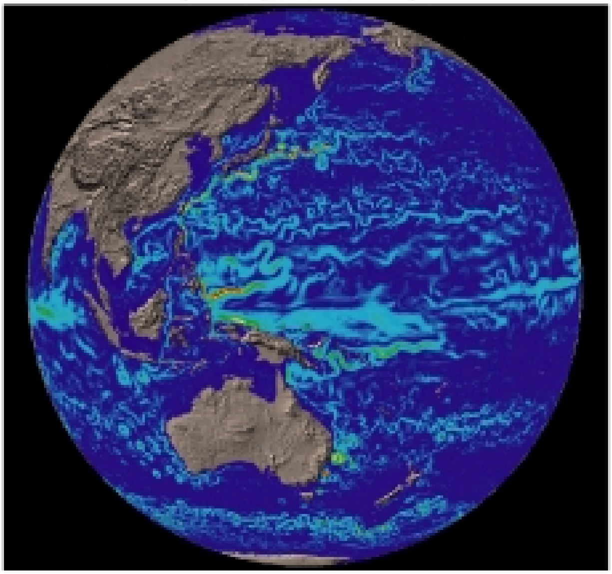
Computational Science at NERSC: High Resolution Global Coupled Ocean/Sea Ice Model

- *Mathew E. Maltrud, Los Alamos National Laboratory; Julie L. McClean, Naval Postgraduate School.*
- The objective of this project is to couple a high-resolution ocean general circulation model with a high-resolution dynamic-thermodynamic sea ice model in a global context.

•Currently, such simulations are typically performed with a horizontal grid resolution of about 1 degree. This project is running a global ocean circulation model with horizontal resolution of approximately 1/10th degree.

•Allows resolution of geographical features critical for climate studies such as Canadian Archipelago

1/10 Degree Global POP Ocean Model Currents at 50m Depth
(blue = 0; red > 150 cm/s)



Computational Science at NERSC: Supernova Explosions and Cosmology

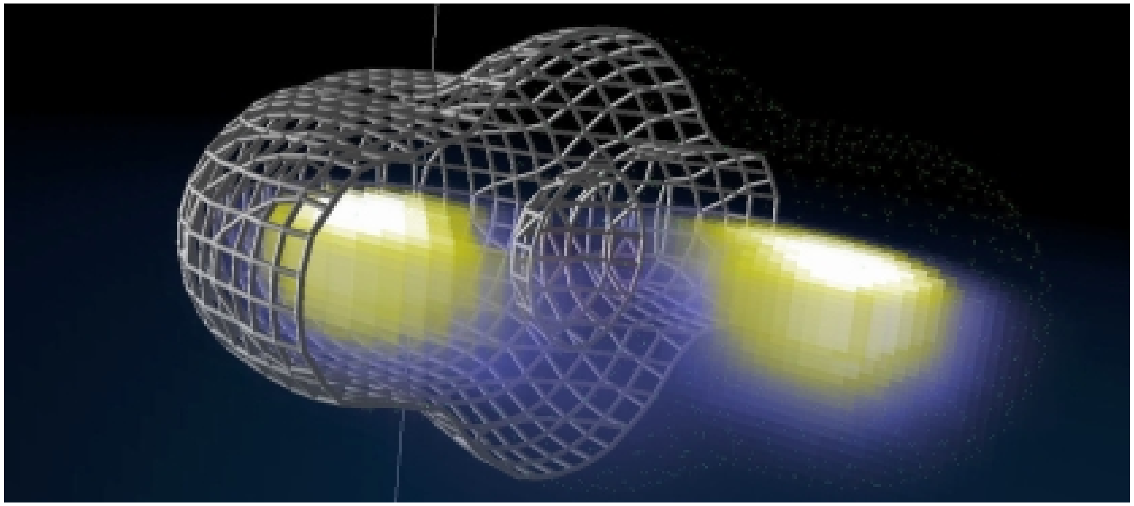
Peter Nugent and Daniel Kasen, Lawrence Berkeley National Laboratory; Peter Hauschildt, University of Georgia; Edward Baron, University of Oklahoma; Stan Woosley and Gary Glatzmaier, University of California, Santa Cruz; Tom Clune, Goddard Space Flight Center; Adam Burrows, Salim Hariri, Phil Pinto, Hessam Sarjoughian, and Bernard Ziegler, University of Arizona; Chris Fryer and Mike Warren, Los Alamos National Laboratory; Frank Dietrich and Rob Hoffman, Lawrence Livermore National Laboratory

- First 3-D supernova explosion simulation, based on computation at NERSC. This research eliminates some of the doubts about earlier 2-D modeling and paves the way for rapid advances on other questions about supernovae.

Computational Science at NERSC: Black Hole Merger Simulations

- *Ed Seidel, Gabrielle Allen, Denis Pollney, and Peter Diener, Max Planck Institute for Astrophysics; John Shalf, Lawrence Berkeley National Laboratory.*
- Simulations of the spiraling coalescence of two black holes, a problem of particular importance for interpreting the gravitational wave signatures that will soon be seen by new laser interferometric detectors around the world.

• Required 1.5 Tbytes of memory and was run on the large 64 Gbyte nodes



Future Plans

The Divergence Problem

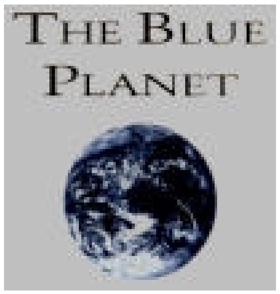
- The requirements of high performance computing for science and engineering, and the requirements of the commercial market are diverging.
- The commercial cluster of SMP approach is no longer sufficient to provide the highest level of performance
  - Lack of memory bandwidth and latency
  - High interconnect latency
  - Lack of interconnect bandwidth
  - High cost of ownership for large scale systems
- U.S. computer industry is driven by commercial applications -- not focused on scientific computing.
- The decision for NERSC-3 E can be seen as a first indication of the divergence problem: Power 4 had a low SSP number

A New Architecture Strategy: Beyond Evaluation to Cooperative Development

A proposal to establish feedback between science and computer design lasting for generations of machines

- Application teams to drive the design of new architectures
- Continued, simultaneous evaluation of multiple scientific applications replacing "rules of thumb" for computer designers
  - Example is the Performance Evaluation Research Center (PERC)
- Leveraging current components and research prototypes into new architectures
- Continual redesign and testing of prototypes in a vendor partnership to create new scientific computers
- Addressing the scientific market beyond lab and academic supercomputer centers

Cooperative Development – NERSC/ANL/IBM Workshop



- Held two joint workshops
  - Sept 2002 – defining the Blue Planet architecture
  - Nov. 2002 – IBM gathered input for Power 6
- Developed White Paper "Creating Science-Driven Computer Architecture: A New Path to Scientific Leadership,“ available at http://www.nersc.gov/news/blueplanet.html

"Blue Planet": Extending IBM Power Technology and "Virtual Vector" Processing

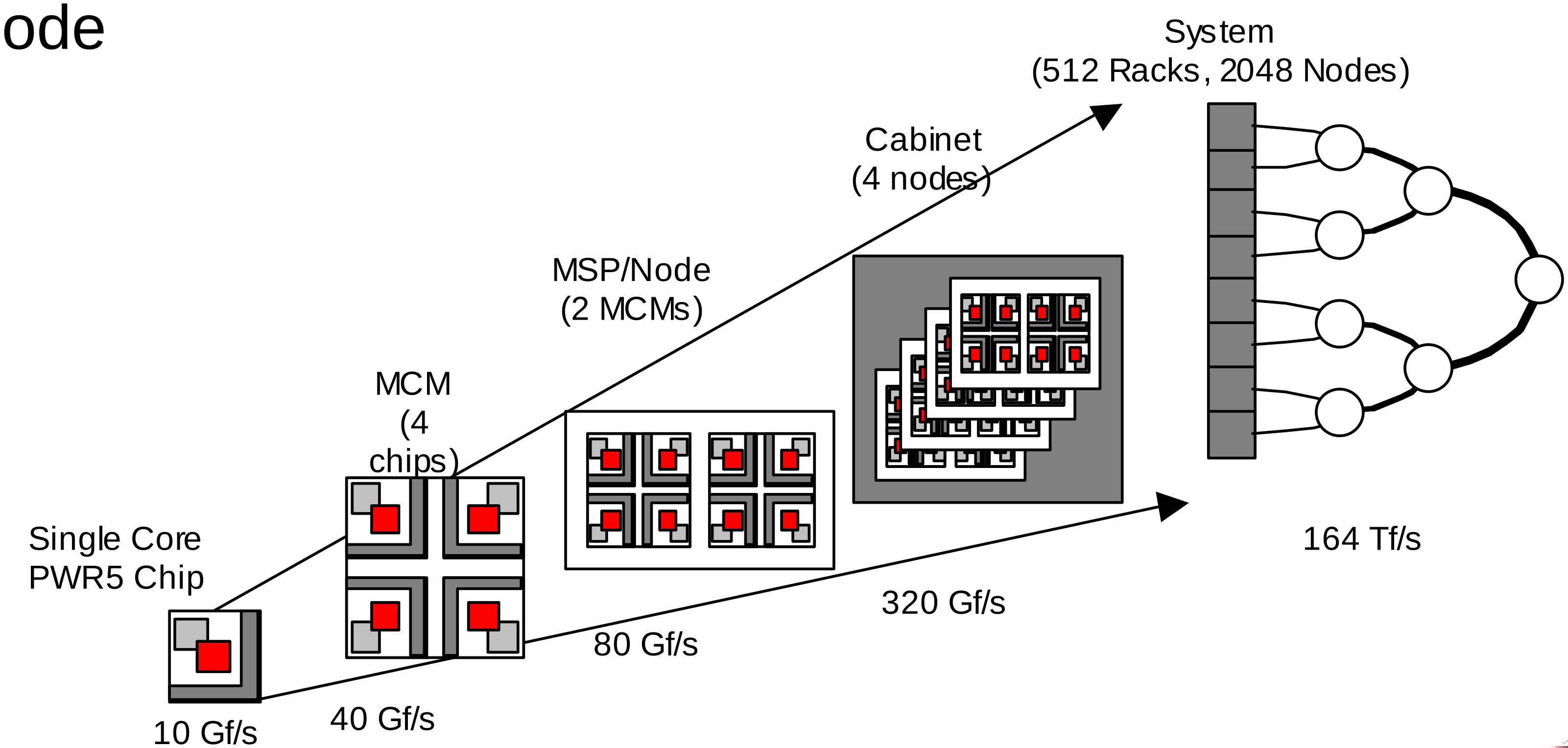
Addressing the key barriers to effective scientific computing

- Memory bandwidth and latency
- Interconnect bandwidth and latency
- Programmability for scientific applications

- The Strategy is to get back"inside the box" of commercial servers (SMPs)
  - Increasing memory and switch bandwidth using commercial parts available over the the next two years
- Exploration of new architectures with the IBM design team
- Enabling the vector programming model inside a Power 5 SMP node
- Changing the design of subsequent generations of microprocessors

Blue Planet: A Conceptual View

- Increasing memory bandwidth – single core chips with dedicated caches for 8 way nodes
- Increasing switch bandwidth and decreasing latency
- Enabling "vector" programming model inside each SMP node



NERSC Is Delivering on Its Commitment to Make the Entire DOE Scientific Computing Enterprise Successful

- NERSC provides very effective supercomputing resources
- NERSC is helping develop approaches to address the “Divergence Problem” with near term and long term computational technology for computer architectures optimized for scientific computing are critical to enable 21st Century Science.
- NERSC is providing targeted support to Large Scale Computational projects
- NERSC is a major player in SciDAC as well as supporting it’s projects and collaborations