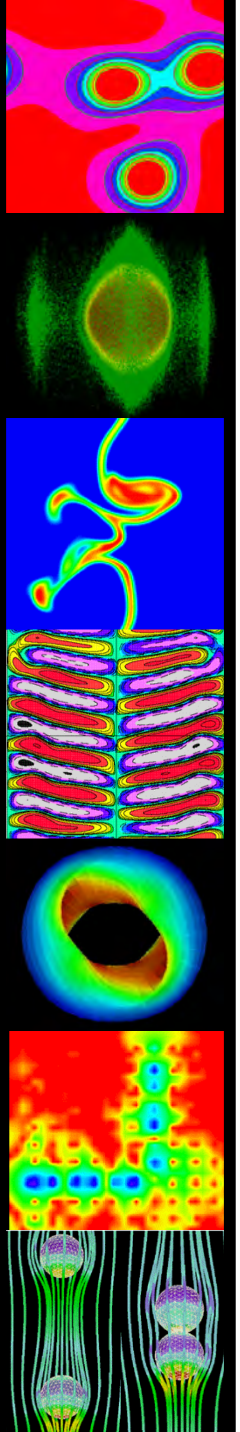
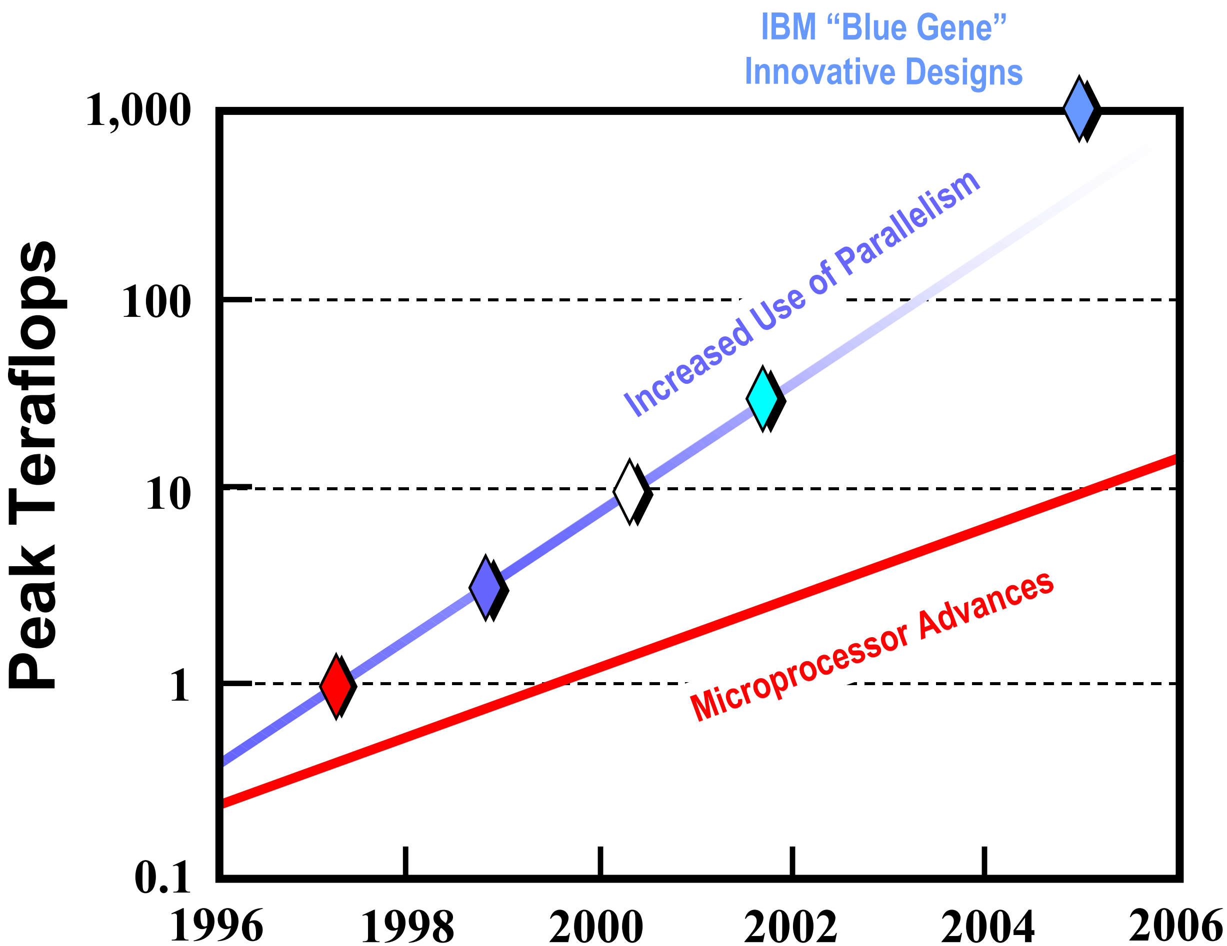


Scientific Discovery through Advanced Computing

Thom H. Dunning, Jr.

Assistant Director for Scientific Simulation
Office of Science
U.S. Department of Energy

Dramatic Advances in Computing
Terascale Today, Petascale Tomorrow



MICROPROCESSORS

2x increase in microprocessor speeds every 18-24 months ("Moore's Law")

PARALLELISM

More and more processors being used on single problem

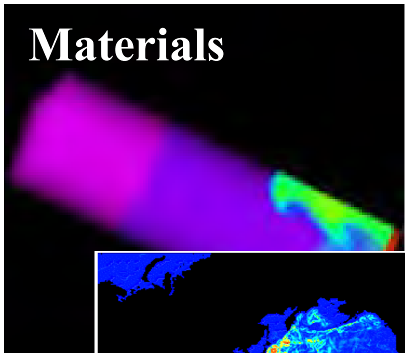
INNOVATIVE DESIGNS

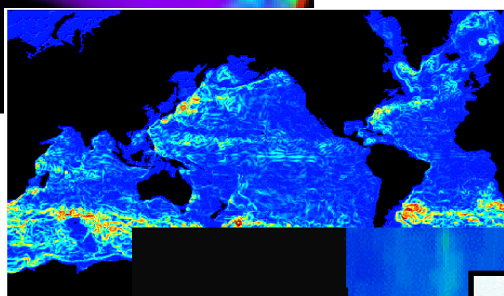
Processors-in-Memory
HTMT

Scientific Computing

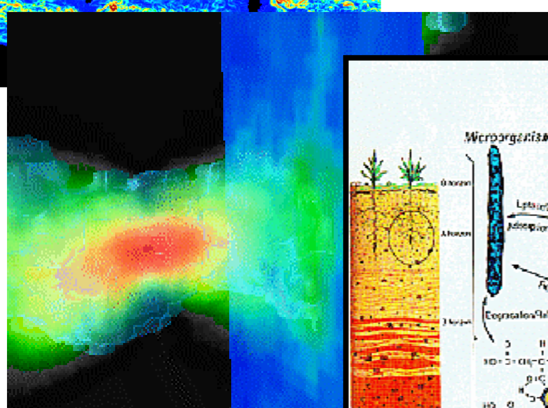
Critical to Discovery in Many Scientific Disciplines

Many SC Programs Need Dramatic Advances in Simulation Capabilities To Meet Their Mission Goals

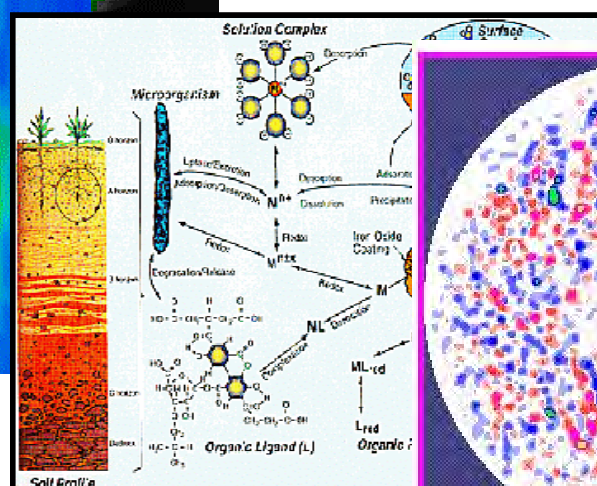




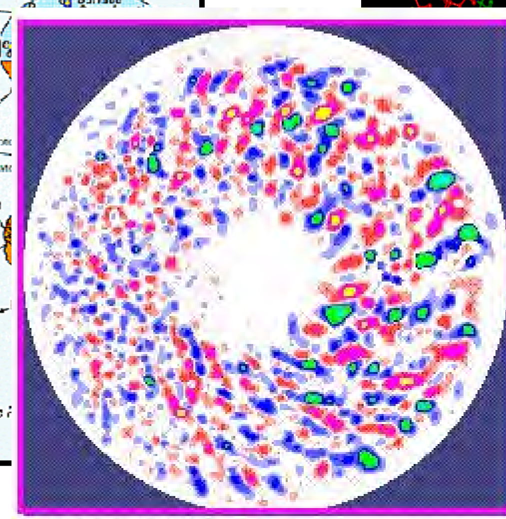
Global Climate



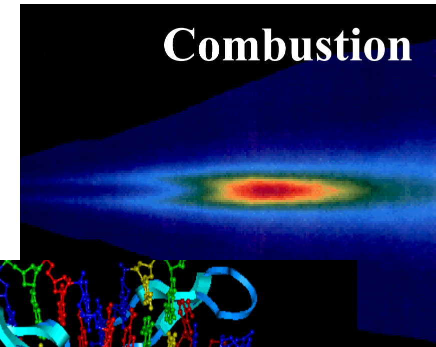
Components of Matter

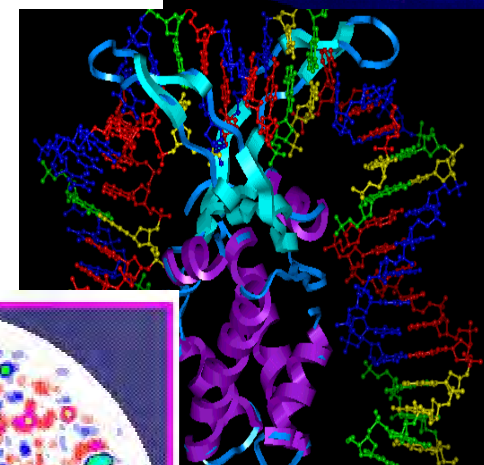


Subsurface Transport



Fusion Energy





Health Effects, Bioremediation

Goal and Strategies

"Scientific Discovery through Advanced Computing"

Goal

- Promote scientific discovery throughout the Office of Science by exploiting advances in computing technologies

Strategies

- Create *Scientific Computing Software Infrastructure* that takes full advantage of terascale computing capabilities for scientific research
- Establish *Scientific Computing Hardware Infrastructure* that supports scientific research in the most efficient, effective manner possible
- Enhance collaboration and access to facilities and data through advances in networking technologies and development of electronic collaboratories

Programmatic Elements

"Scientific Discovery through Advanced Computing"

Hardware Infrastructure

Software Infrastructure

OPERATING SYSTEM

COLLABORATORIES

DATA GRIDS

Data Analysis & Visualization

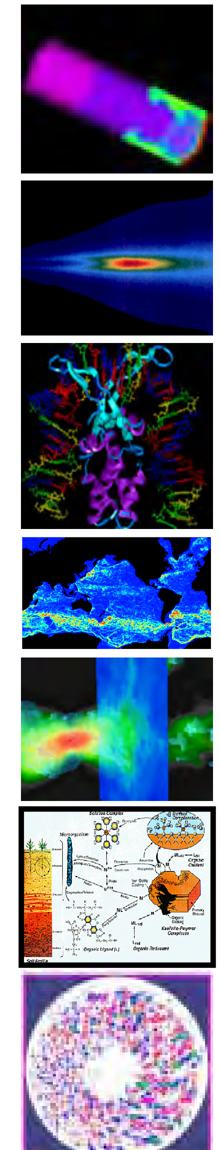
Mathematical Libraries

Programming Environments

Scientific Data Management

Problem-solving Environments

SCIENTIFIC SIMULATION CODES



BES, BER
FES, HENP

ASCR

Scientific Computing Software Infrastructure

SC Software Infrastructure
A Major Software Challenge

Peak Performance is skyrocketing

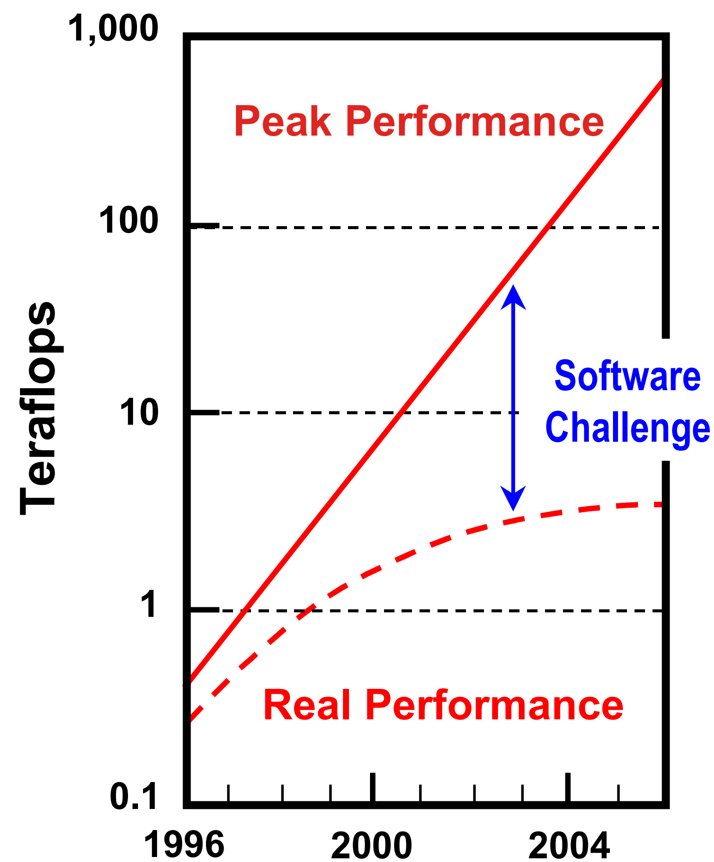
- In 90's, peak performance has increased 100x; in 00's, it will increase 1000x

But ...

- Efficiency has declined from 40-50% on the vector supercomputers of 1990s to as little as 5-10% on parallel supercomputers of today and may decrease further on future machines

Research challenge is software

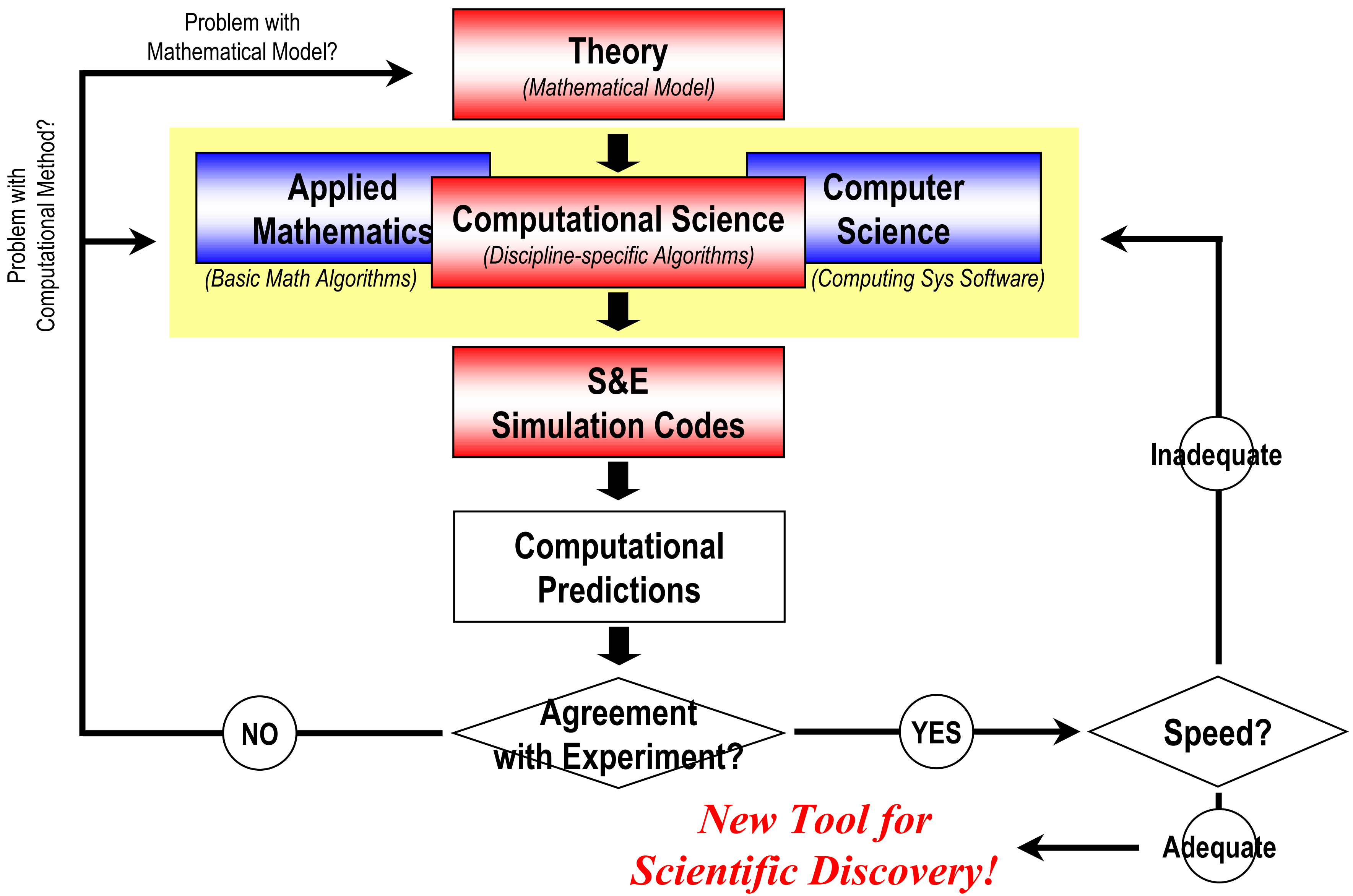
- Scientific codes to model and simulate physical processes and systems
- Computing and mathematics software to enable use of advanced computers for scientific applications
- Continuing challenge as computer architectures undergo fundamental changes



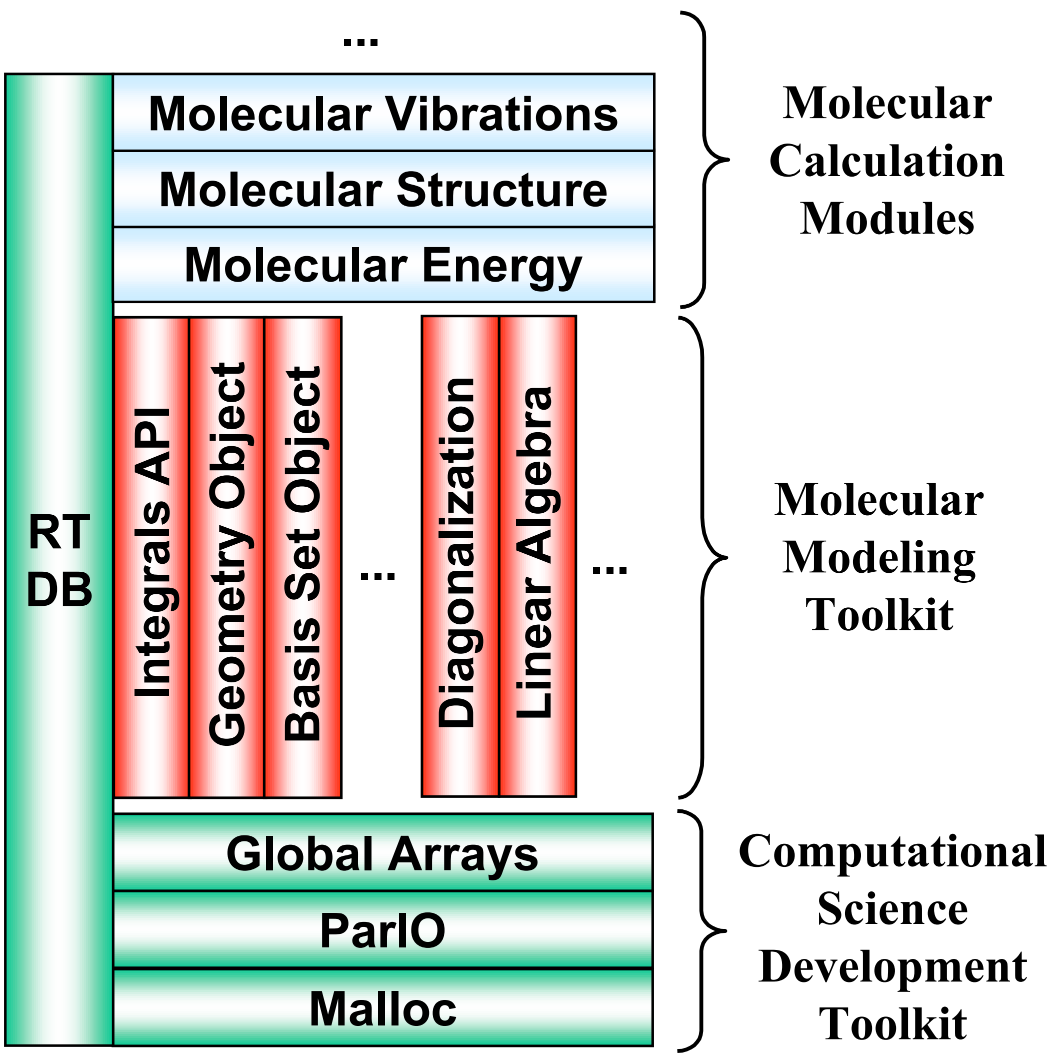
Software Challenges
Scientific Computing

- **Scientific Codes**
  - High fidelity mathematical models
  - Efficient, robust computational methods and algorithms
  - Well designed computational modeling and simulation codes
    - Readily incorporate new theoretical advances
    - Port from one computer to another with minimal changes
- **Computing Systems and Mathematical Software**
  - Increased functionality in Vendor Operating Systems
  - Computing systems software
    - Accelerate development and use of terascale scientific codes
    - Facilitate porting of software and codes among high-performance computers
    - Manage and analyze massive (petabyte) data sets, both locally and remotely
  - Algorithms that scale to thousands-millions processors

SC Software Infrastructure
Scientific Simulation Code Development Teams



An Example
Northwest Chem (NWChem)



NWChem

a major new modeling capability for molecular science

Molecular Electronic Structure

Molecular Dynamics

(Crystals)

> 600,000 lines of code and growing

Runs on ...

Cray T3D/E, IBM SP2, SGI

SMP, NOWs, Sun and other

workstations, X86 PCs (Linux)

Scales to ...

2,000+ processors

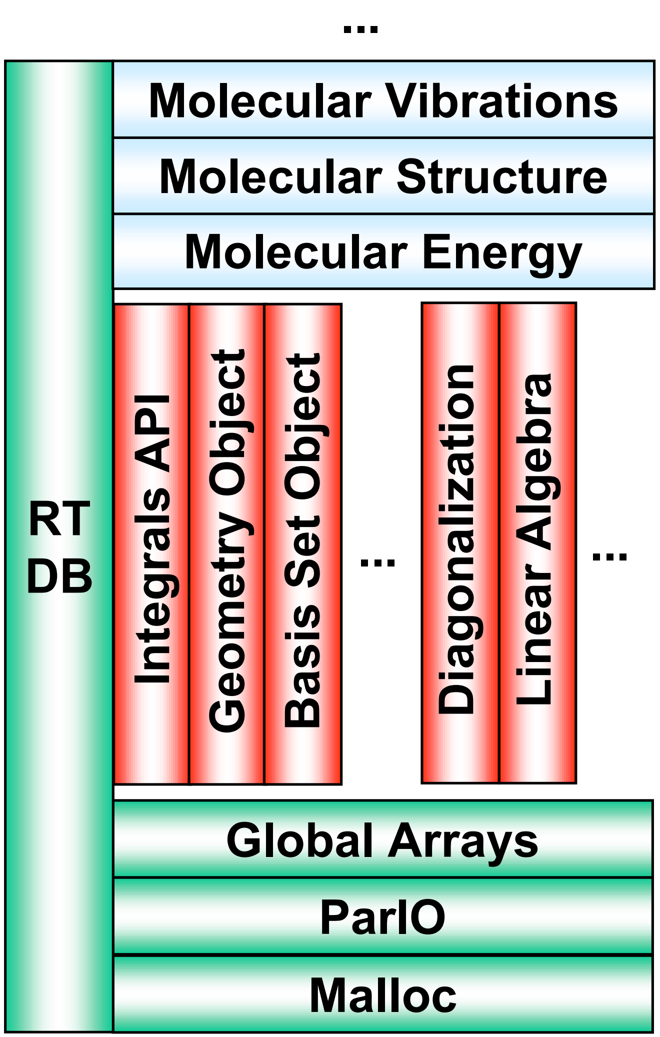
Developers ...

Core group (15) plus larger

group (20) of world-wide collaborators

>100 person-years at PNNL alone

NWChem
Chemistry-Applied Math-Computer Science



Computational Chemists

Computational Chemists & *Applied Mathematicians*

Molecular Modeling Toolkit

Basic mathematical subroutine libraries.

Computer Scientists

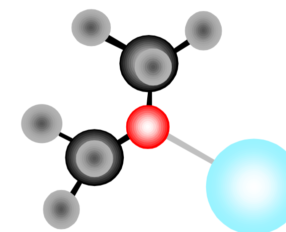
Hardware Abstraction Layer

The only layer of NWChem that must be revised to make use of new computers. Only a small fraction of the code in this layer must be changed.

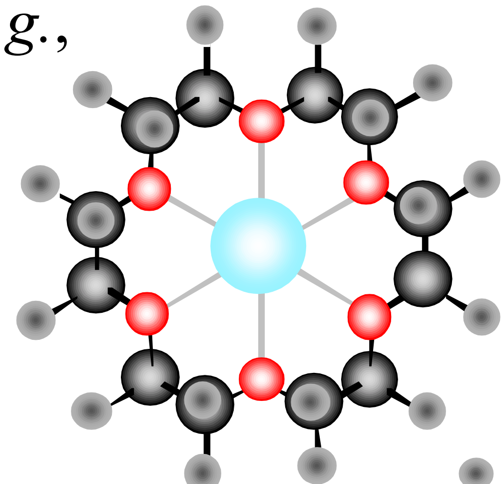
Advancements
What Have We Gained?

- **In 1992**
  - Model separations agents, *e.g.*, ether-alkali ions
- **In 2000**
  - *Serial computers*: prototype separation agents, *e.g.*, 18-crown-6
  - *Parallel computers*:* real-world separations agents, *e.g.*, Still's crown ether

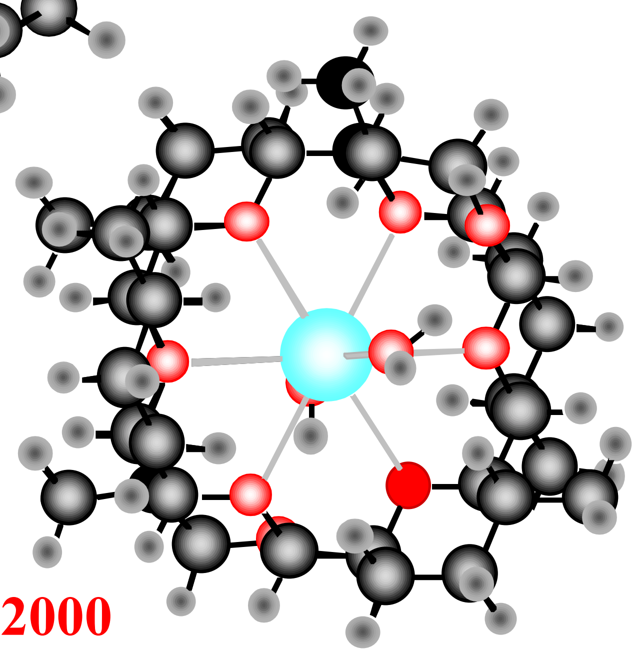
* *With **NWChem** on 0.25-tf computer at PNNL.*



1992

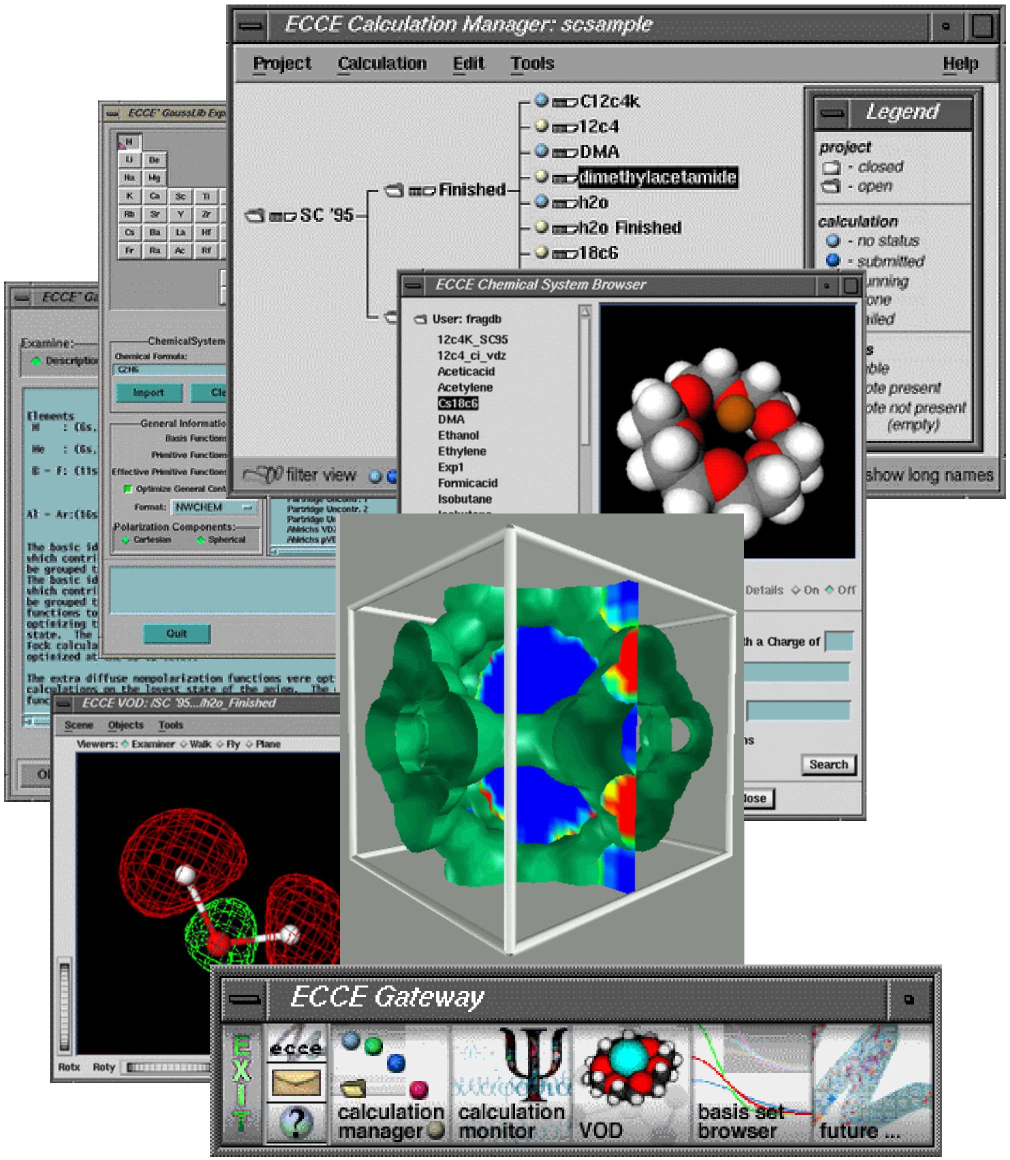


2000



2000

Another Example
Computational Chemical Environment



ECCE

extensible computational chemistry environment

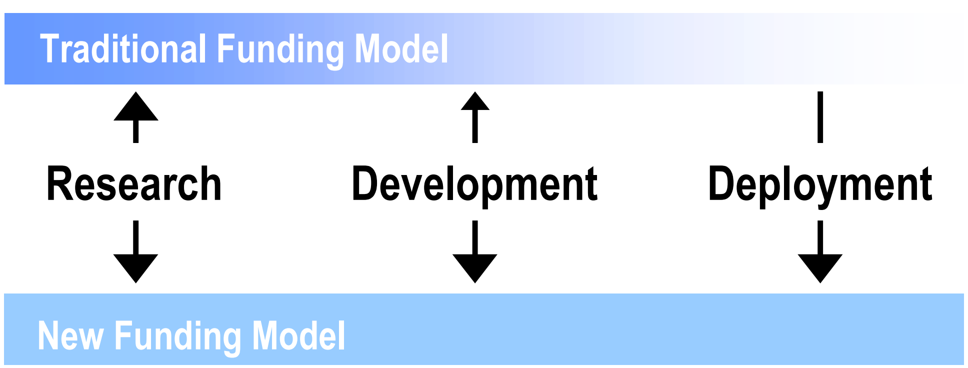
An environment to assist the researcher in using NWChem

Key Features

- Point-and-click set up of calculations
- Browser-type management of both calculations and computers
- Chemistry-specific analysis and visualization tools
- Integrated data management

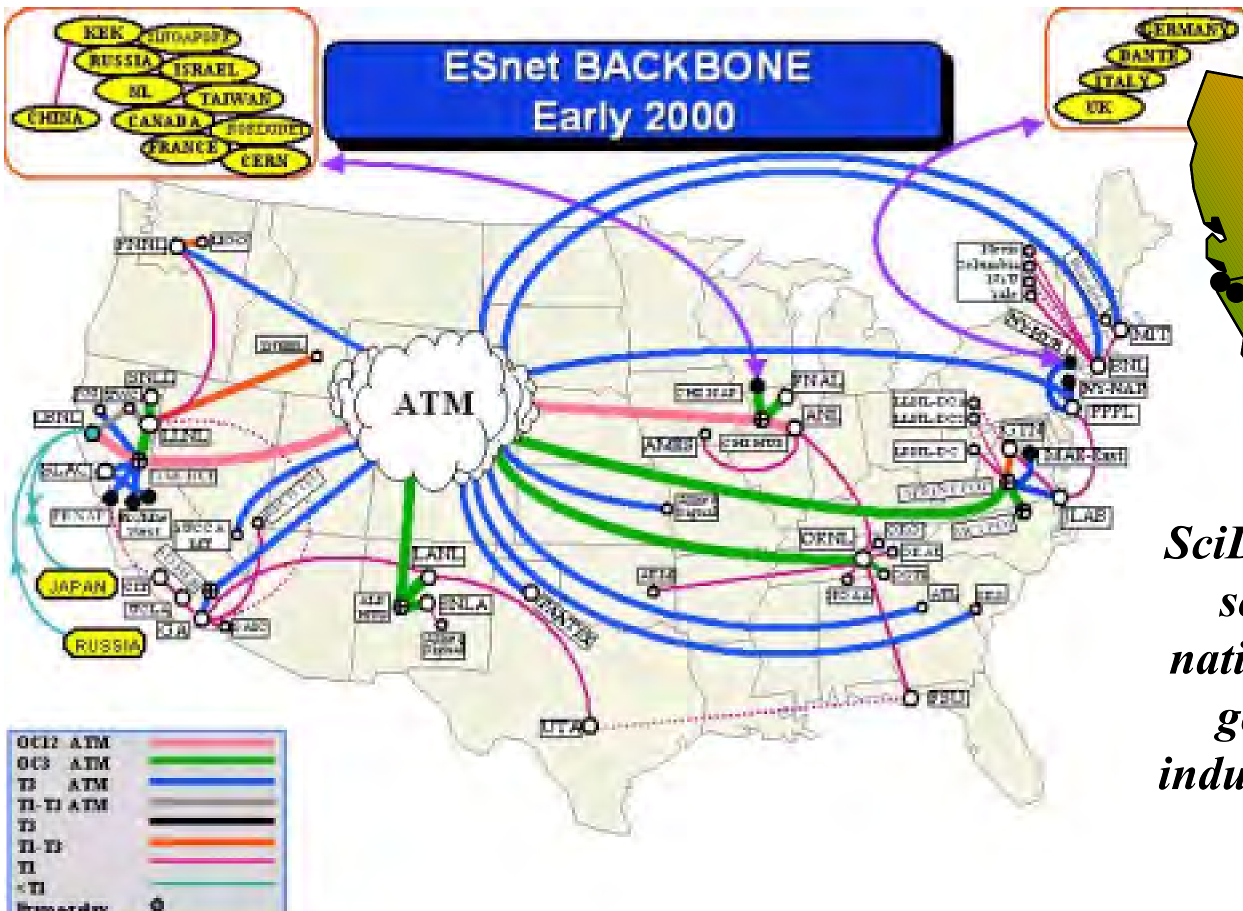
SC Software Infrastructure
Enabling Technology Centers

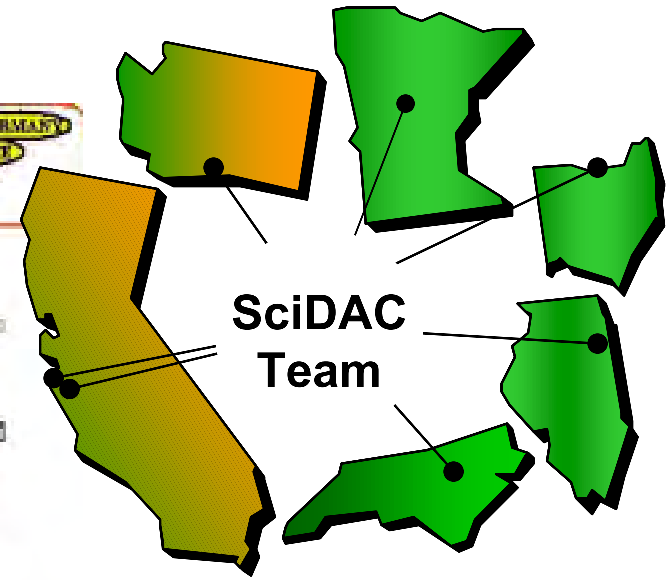
- **Teams of Mathematicians, Computer Scientists, Applications Scientists, and Software Engineers to ...**
  - Create mathematical and computing systems software to enable scientific simulation codes to take full advantage of the extraordinary capabilities of terascale computers
  - Work closely with Scientific Simulation Teams to ensure that the most critical computer science and applied mathematics issues are addressed in a timely fashion
  - Support the full software life cycle



SC Software Infrastructure
Collaboratories and Networks

Create collaboratories, software, and networks to link geographically distributed researchers, data, and tools and to enable remote access to experimental and computational facilities.





SciDAC teams will be composed of scientists and engineers from national laboratories, universities, government laboratories, and industry and will be fully supported by electronic collaboration technology.

SC Software Infrastructure Plan

- **Initiate software development efforts in FY2001**
  - **Scientific Simulation Codes**: Competitively select 4-6 teams to begin development of advanced computational modeling and simulation codes ($20.0 million = BES: $2 million; BER: $8 million; FES: $3 million; HENP: $7 million)
  - **Enabling Technologies:** Competitively select 4-6 teams to begin development of mathematical and computing systems software and fund joint efforts ($27.0 million*)
  - **Collaboratories:** Competitively select 3-4 teams to continue development of collaboratory software ($10.1 million*)
- **Strengthen and broaden the software development efforts in FY2002 and beyond (*new funds required*)**

* *OASCR total reduced by approx. $10 million for Laboratory Technology Research program.*

Scientific Computing Hardware Infrastructure

SC Hardware Infrastructure
Robust, Agile, Effective & Efficient

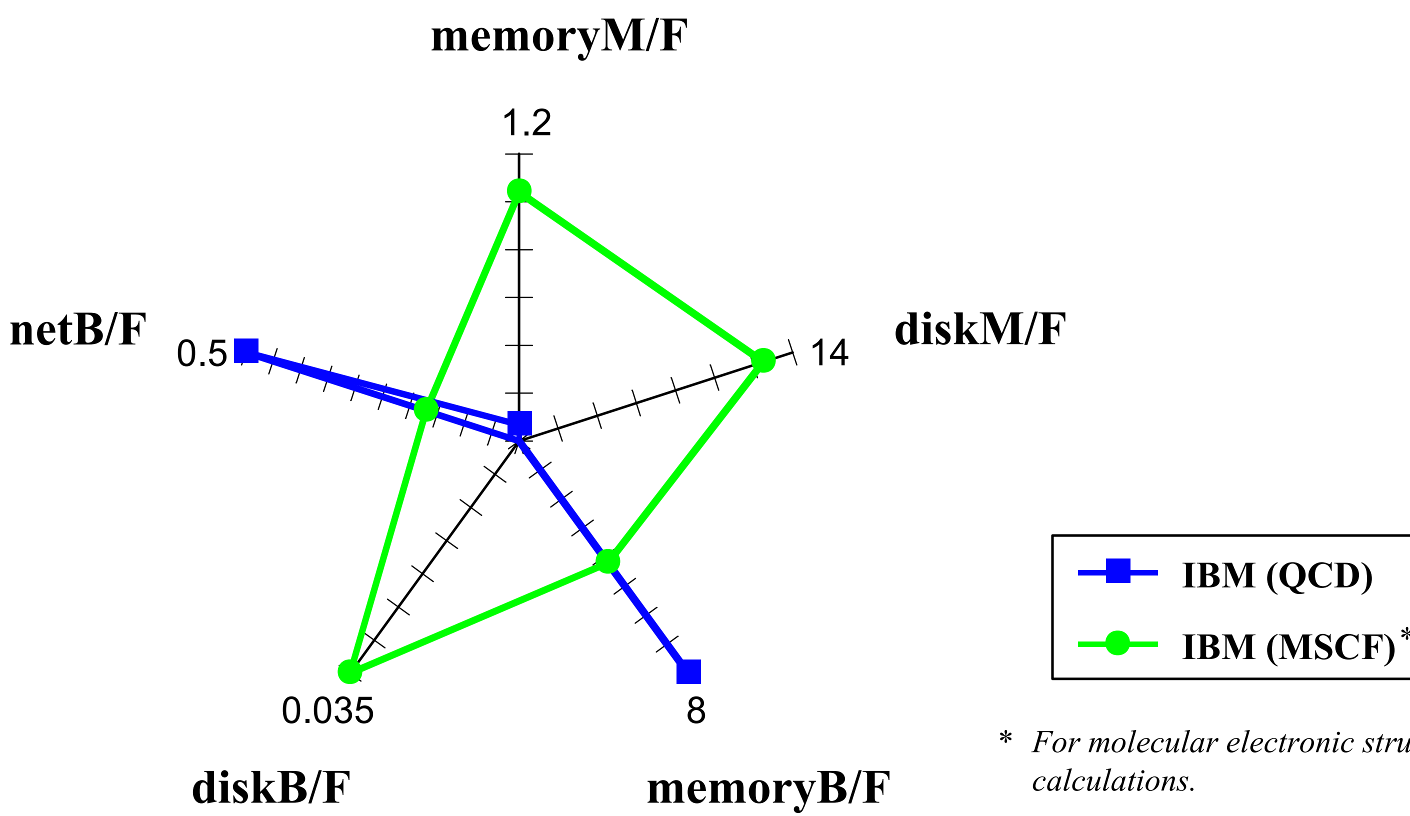
- **Flagship Computing Facility**
  - To provide robust, high-end computing resources for *all* SC research programs
- **Topical Computing Facilities**
  - To provide the most effective and efficient computing resources for a set of scientific applications
  - To serve as a focal point for a scientific research community as it adapts to new computing technologies
- **Experimental Computing Facilities**
  - To assess new computing technologies for scientific applications

Why Topical Facilities?
Variation in Scientific Application Needs*

Code	Application	Time (TFLOPS-HRS)	Memory (TBYTES)	Storage (TBYTES)	Node I/O (MBYTES/S)
Cactus	Astrophysics	300	1.8	20	5
ARPS	Weather	25	0.25	16	18
MILC	Particle Physics	10,000	0.2	1	3
PPM	Turbulent Flow	500	0.5	54	6
PUPI	Liquids	150	0.1	0.2	3
ASPCG	Fluid Dynamics	5,000	0.5	50	3
ENZO	Galaxies	1,000	0.9	10	12
	Variation	**400x**	**18x**	**100x**	**6x**

* *From "High-level Application Resource Characterization," NSF/PACI National Computational Science Alliance, May 2000. Reported by permission of Dr. D. A. Reed, Director, National Center for Supercomputing Applications.*

Topical Computing Facilities
Computer Specifications: QCD & Chemistry



* *For molecular electronic structure calculations.*

Why Topical Facilities?
A Response to the Software Challenge

- **To provide the organizational framework needed for multidisciplinary activities**
  - Addressing software challenges require strong, long term collaborations among disciplinary computational scientists, computer scientists, and applied mathematicians
- **To provide the organizational framework needed for the development of community codes**
  - Implementation of many scientific codes requires a wide range of disciplinary expertise
- **Organizational needs will continue as computers advance to petaflops scale**

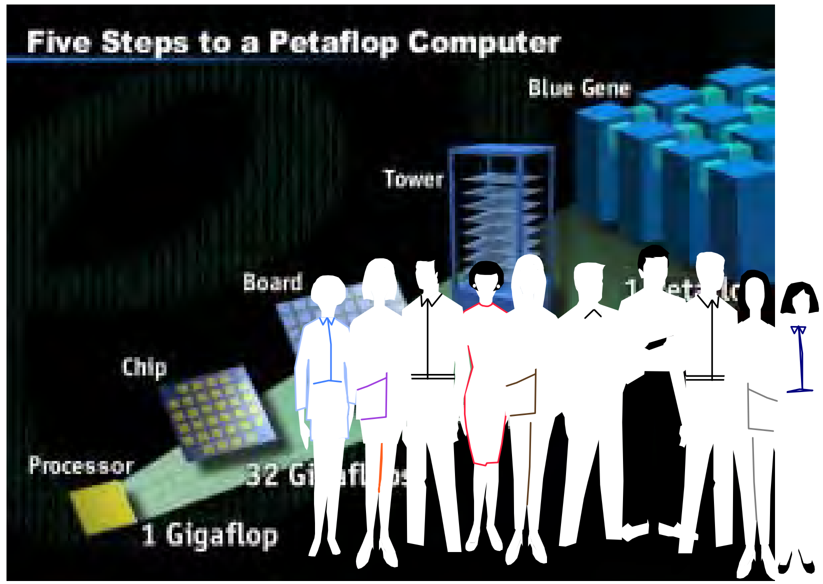
Topical Computing Facilities
Selection Criteria

- **Importance of the scientific application** to the mission of the U.S. Department of Energy
- **Need for extraordinary computing resources** to address problems of critical scientific and national importance, **ability to fully use** these resources
- **Ability to advance the state-of-the-art** in computation-al modeling and simulation in the selected application area
- **Ability to provide cost efficiencies or capability enhancements** over and above that achieved in Flagship Facility

Why Experimental Facilities?

- **Need an organized approach for evaluation of new computing technologies (processors, switches, *etc*.)**
  - Although we are currently on a plateau vis-à-vis computer architectures, this will not last through the end of the decade
  - Examples of new approaches include PIM (processors-in-memory), HTMT (hybrid technology-multithread technology)
- **Need an organized approach for interacting with computer designers as early as possible**
  - Computer designers have many variables to consider – some beneficial for scientific computing, others not
  - Earlier the scientific community can provide input, the more likely the advice will be heeded

Elements of Experimental Computing Facility



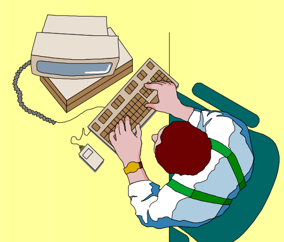
Technologies Evaluation Teams

Core Teams + Distributed Team Members

Code Development Teams

@ Flagship Computing Facility

@ Topical Computing Facilities



SC Hardware Infrastructure Plan

- **Upgrade existing Flagship Facility in FY2001**
  - NERSC-3 to 5-teraflops ($5.8 million*)
  - ESnet ($3.5 million*) + Network Testbed ($1.0 million*)
- **Upgrade Advanced Computing Research Facilities in FY2001 ($2.0 million*)**
- **Add new facilities in FY2002 and beyond**
  - Competitively establish a number of Topical Computing Facilities, beginning in FY2002 (*new funds required*)
  - Establish up to two (2) Experimental Computing Facilities, beginning in FY2002 by recompeting the Advanced Computing Research Facilities (*no new funds required*)

* *OASCR total reduced by approx. $10 million for Laboratory Technology Research program.*

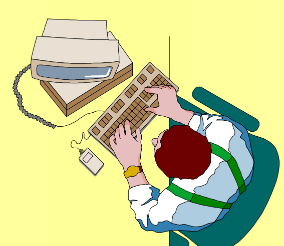
Elements of Topical Computing Facility

Computing Systems

Code Development Teams

Core Teams + Distributed Team Members

Scientific Application Users



Proposed FY2001 Investments

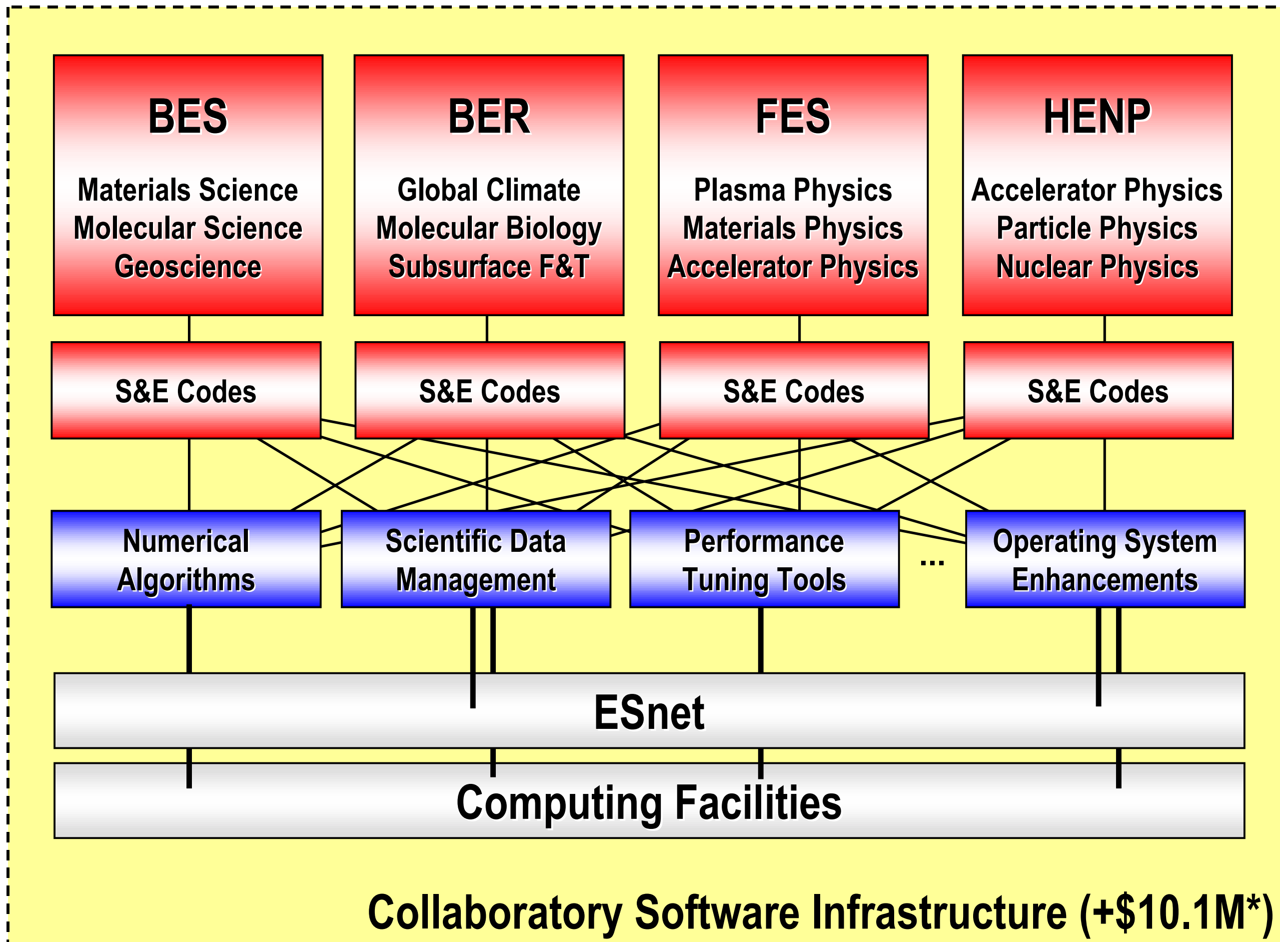
Scientific Discovery through Advanced Computing

Investments in computational modeling and simulation in the Office of Science are driven by scientific problems derived from DOE's missions.

Scientific Code Development Teams (+$20.0M)

Enabling Technology Centers (+$27.0M*)

Computing Hardware Infrastructure (+$12.3M*)



* *OASCR total reduced by approx. $10 million for Laboratory Technology Research program.*

Goal and Strategies

"Scientific Discovery through Advanced Computing"

- **Goal**
  - Promote scientific discovery throughout the Office of Science by exploiting advances in computing technologies
- **Strategies**
  - Create *Scientific Computing Software Infrastructure* that takes full advantage of terascale computing capabilities for scientific research
  - Establish *Scientific Computing Hardware Infrastructure* that supports scientific research in the most efficient, effective manner possible
  - Enhance collaboration and access to facilities and data through advances in networking technologies and development of electronic collaboratories