CYBER-INFRASTRUCTURE: ENABLING SCIENCE AND ENGINEERING RESEARCH

Miriam Heller, Ph.D.
Office of Cyberinfrastructure
US National Science Foundation
mheller @ nsf.gov

International Chemical Information Conference & Exhibition

Nimes, France

18 October 2005

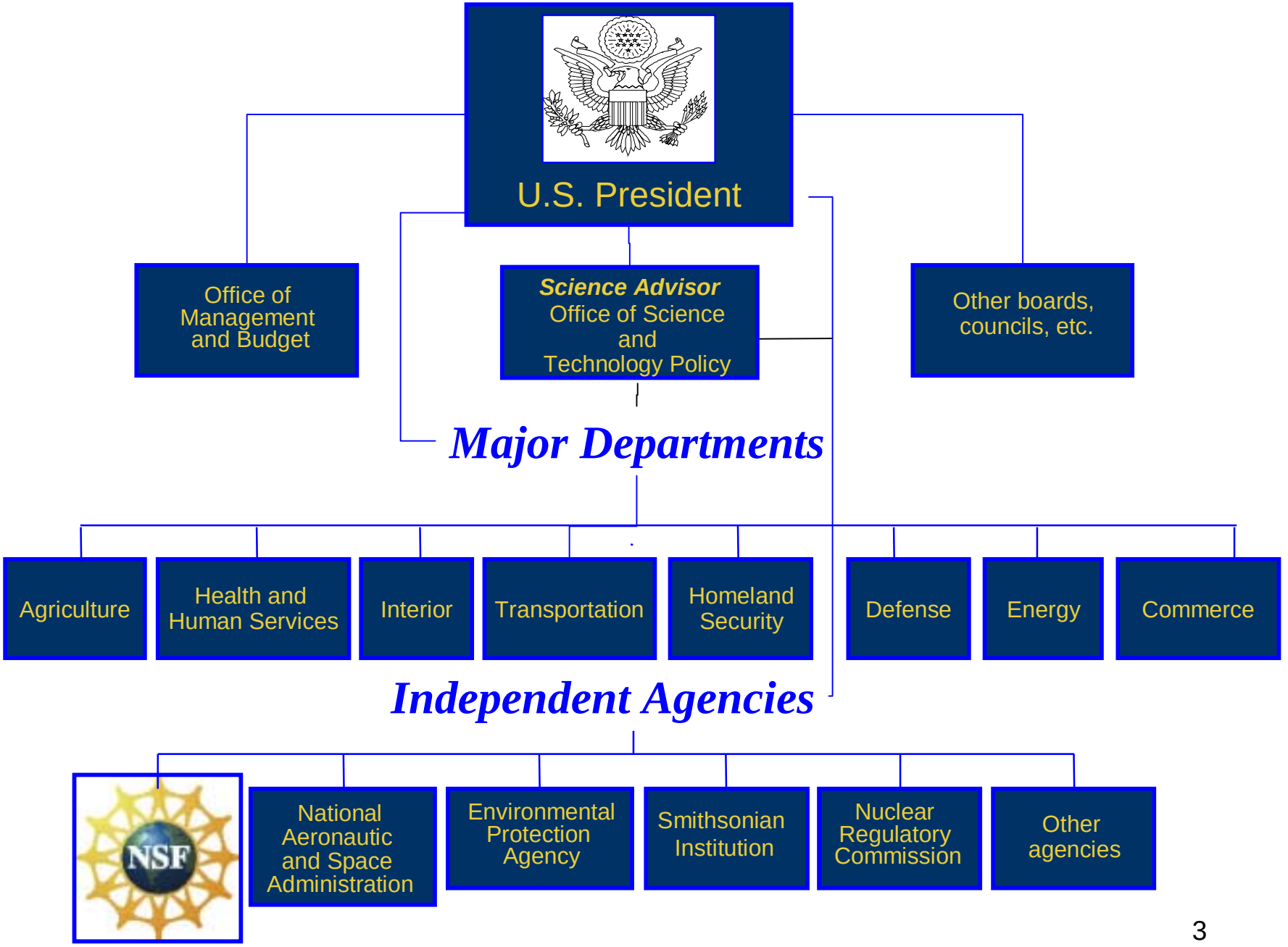
National Science Foundation

NSF Act (PL 810507) of May 10, 1950:
Establishes NSF and Its Organization

- NSB (24 members) + Director appointed by the President
- Encourage and develop a national policy for the promotion of basic research and education in the math, physical, medical, biological, engineering and other sciences
- Initiate and support basic scientific research
- Evaluate the science research programs undertaken by agencies of the Federal government
- Alan T.Waterman first Director
- 1952: first appropriation: $3.5M

National Science Foundation

NSF in the Federal Government



National Science Foundation

National Science Foundation

FY2006 Requested **$5.60 B Total**

Arden Bement

Kathie Olsen

Director

Deputy Director

National Science Board

Office of the Inspector General

Directorate for Education and Human Resources

Directorate for Biological Sciences

Directorate for Computer and Information Science and Engineering

Office of Polar and Antarctic Programs

Office of Integrative Activities (MRE, STC)

Directorate for Engineering

Directorate for Geosciences

Directorate for Mathematical and Physical Sciences

Office of Cyber-infrastructure

Office of Inter-national Science and Engineering

Directorate for Social, Behavioral, and Economic Sciences

National Science Foundation

Outline

- **Past**

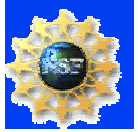
  (or how NSF changed the world …)

- **Present**

  (from compute-centric to cyberinfrastructure)

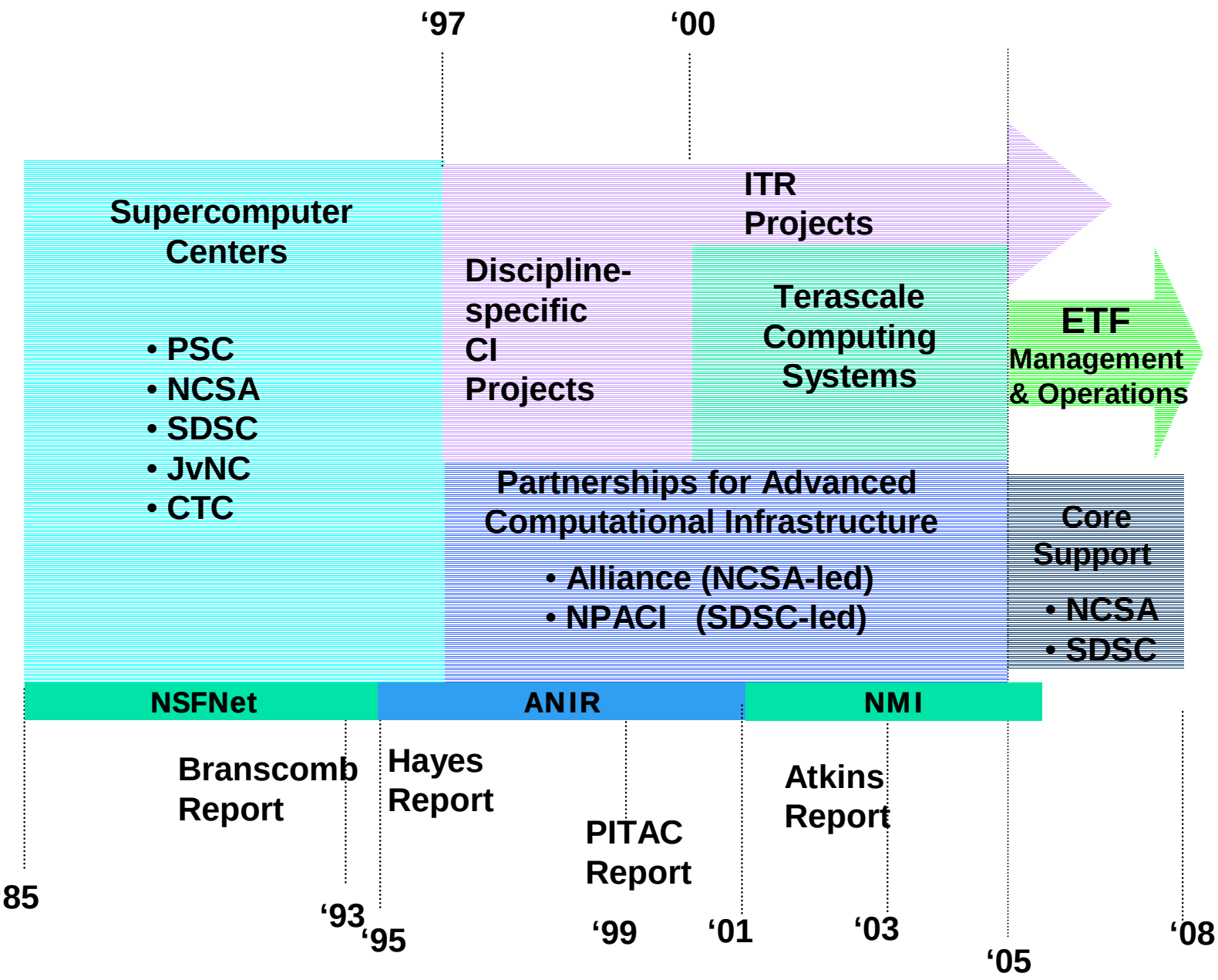
- **Future**

  (the network of things …)

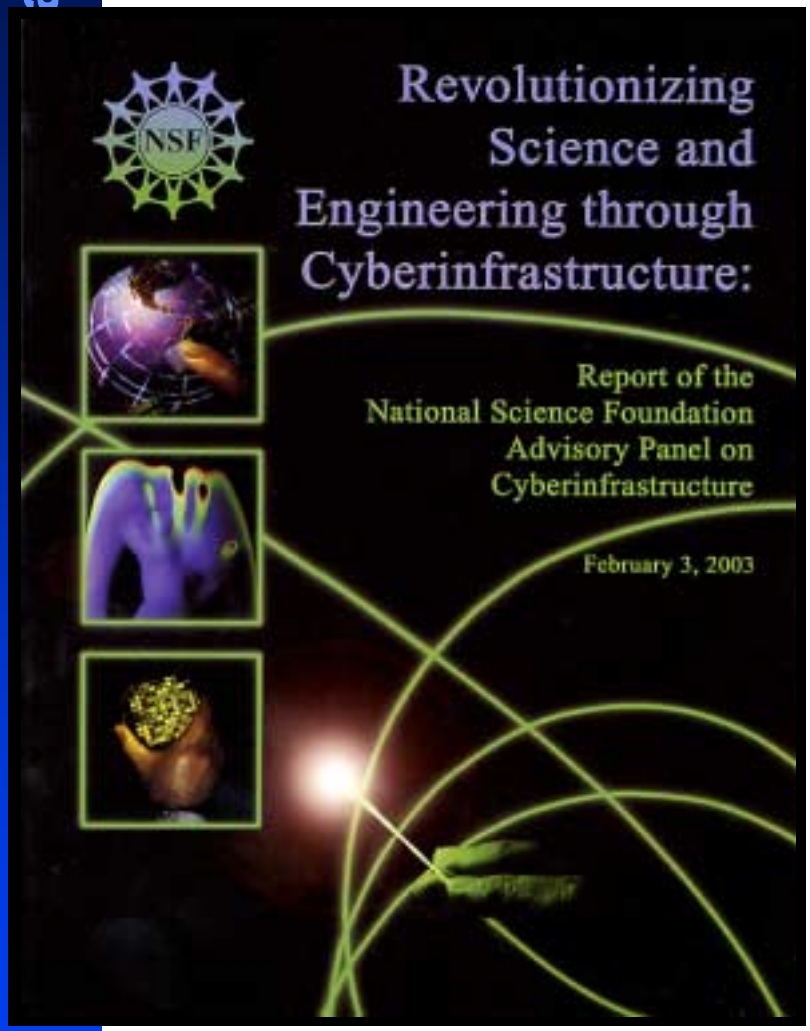


National Science Foundation

Historical Elements



Cyberinfrastructure Vision



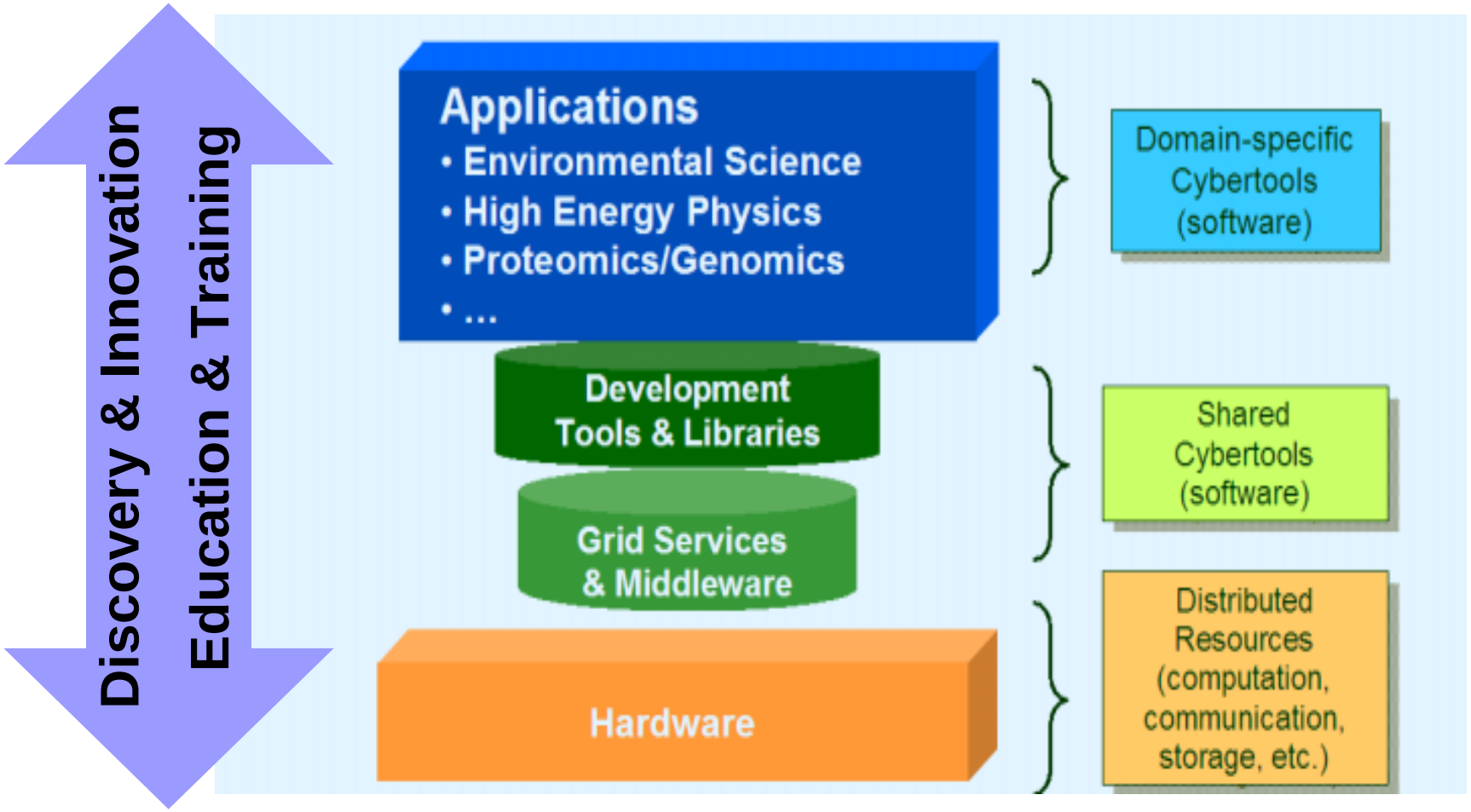
www. nsf.gov/od/oci/reports/toc.jsp

- "Atkins report" - Blue-ribbon panel, chaired by Daniel E. Atkins
- Called for a *national-level, integrated system* of hardware, software, & data resources and services
- New infrastructure to enable new paradigms of science & engineering research and education with increased efficiency

National Science Foundation

Integrated Cyberinfrastructure…

meeting the needs of a community of communities

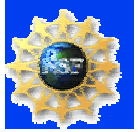


Burgeoning Number of CI Systems

National Science Foundation

Current OCI Investments

- **(High End Computing + X) HEC + X**
- **Extensible Terascale Facility (ETF)**
- **International Research Network Connections**
- **NSF Middleware Initiative**
- **Integrative Activities: Computational Science**
- **Integrative Activities: Education, Outreach & Training**
- **Social and Economic Frontiers in Cyberinfrastructure**



NCSA Scientific Computing Environment: 32 TF (HEC + Innovation; ETF)

Tungsten
Dell IA-32 Linux Cluster
16.4 TF Peak
1.2TF w/140 TB storage

TeraGrid - Mercury
IBM Itanium2 Linux Cluster
10.6 TF, 230 TB storage

Copper
IBM Power4 SMP Cluster
2 TF

Titan
IBM Itanium Linux Cluster
1 TF

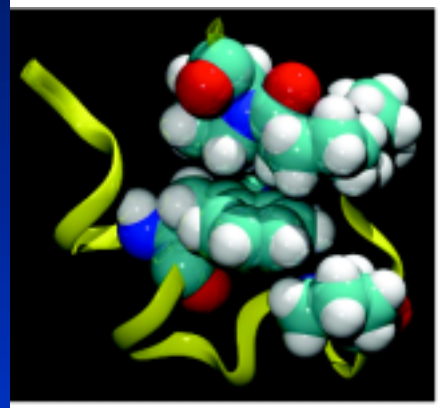
Platinum
IBM Pentium III Linux Cluster
1 TF

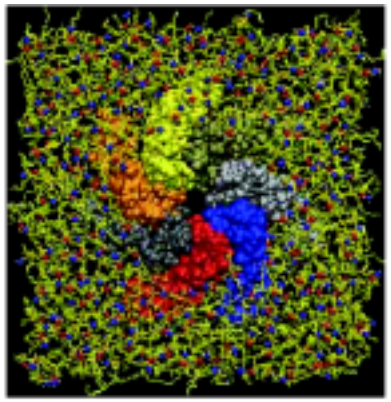
UniTree Archive
ADIC Tape Library
3 PB

National Science Foundation

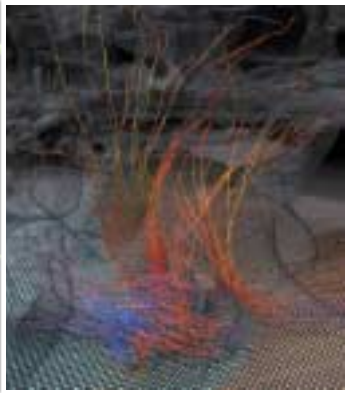
Recent NCSA Scientific Studies

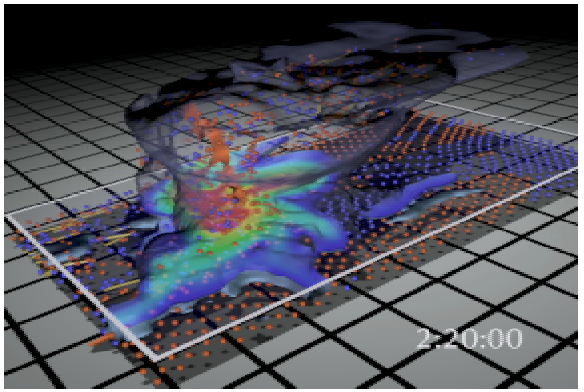
Computational Biology



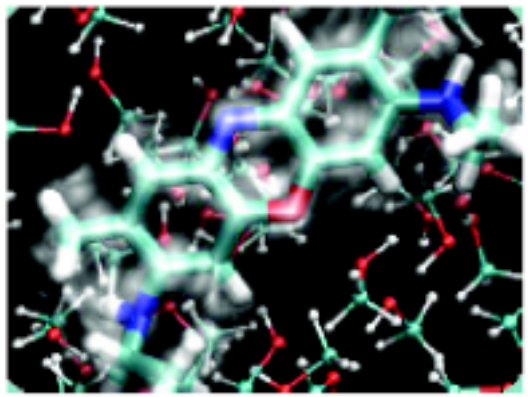


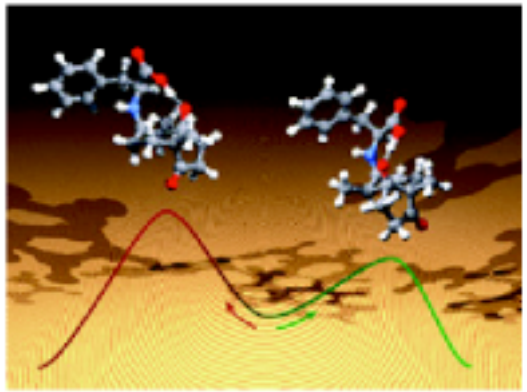
Weather Forecasting



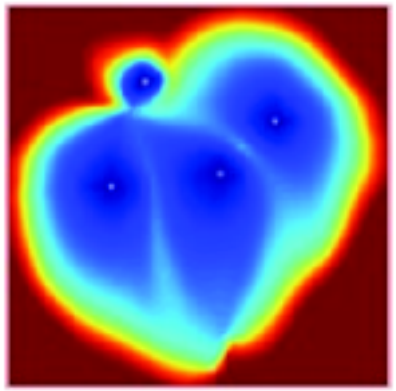


Molecular Science





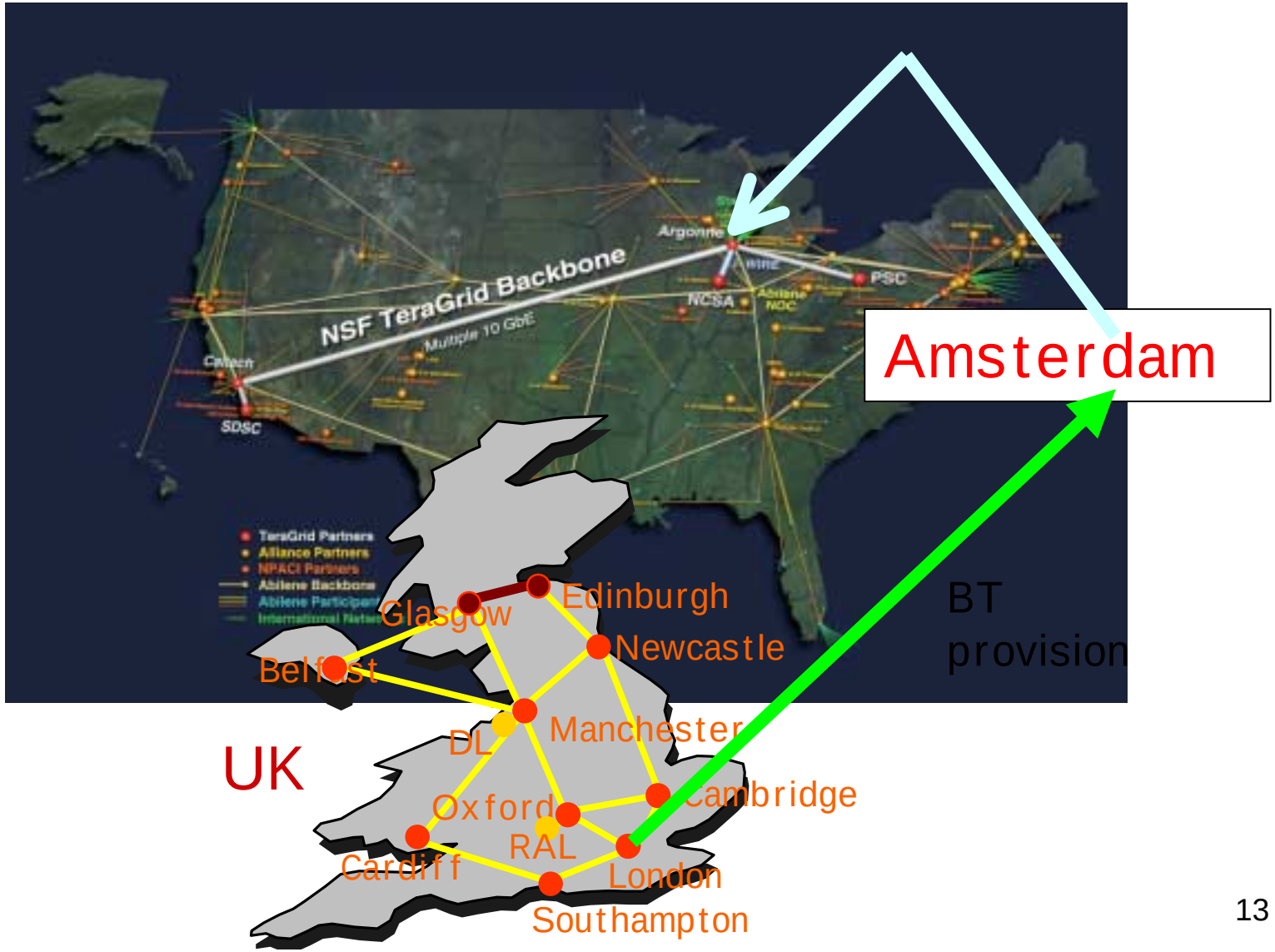
Earth Science



courtesy NCSA

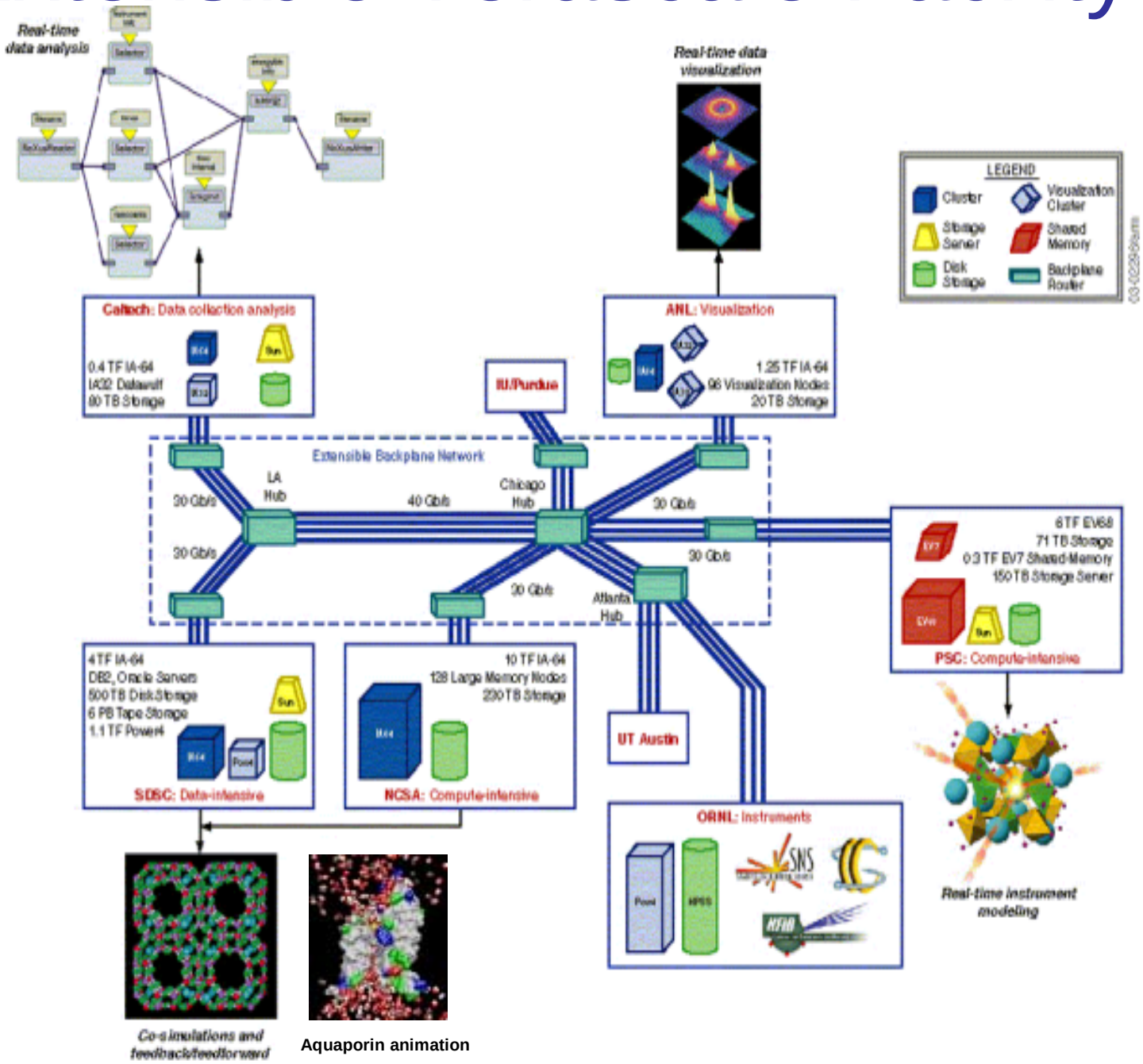
National Science Foundation

Network of Computers “the Grid”



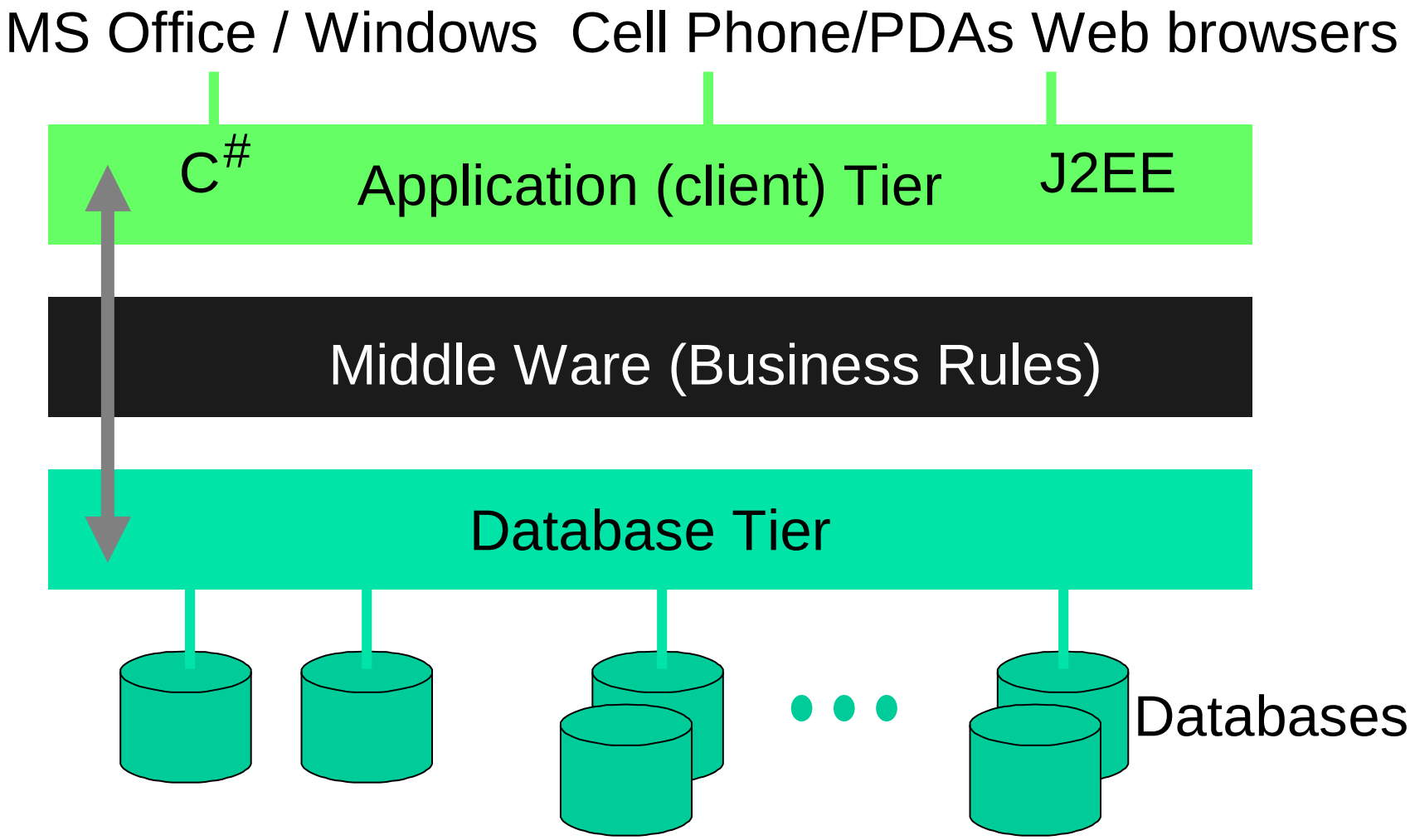
National Science Foundation

Extensible Terascale Facility (ETF)



National Science Foundation

Chemical/Pharma Informatics Architecture

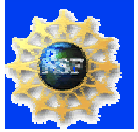


Courtesy Sangtae Kim

National Science Foundation

NSF Middleware Initiative

- **Program started in late 2001 to design, develop, test, deploy middleware**
- **Define open-source, open-architecture standards for on-line collaboration resource sharing**
- **Examples include**:
  - Community-wide access to experimental data on the Grid
  - Authorized resource access across multiple campuses using common tools
  - Web-based portals that provide a common interface to wide-ranging Grid-enabled computation resources
  - Grid access to instrumentation such as accelerators, telescopes



National Science Foundation

NMI-funded Activities in Chemistry-Related S&E Research

- **Condor** – Mature distributed computing system installed on 1000's of CPU "pools" and 10's of 1000's of CPUs.
- **GridChem** –Open source Java application launches/monitors computational chemistry calculations (Gaussian03, GAMESS, NWChem and soon Molpro, Qchem, Aces) on CCG supercomputers from remote sites.
- **NanoHub** – Extends NSF Network for Computational Nanotechnology applications, e.g., NEMO3D, nanoMOS, to a distributed environment over Teragrid, U Wisconsin, and other grid assets using InVIGO, Condor-G, etc.

National Science Foundation

Chemical & Engineering News: Government & Policy - 'Cyber-Enabled' Chemistry - Microsoft Internet Explorer provided by National

File Edit View Favorites Tools Help

Address http://pubs.acs.org/email/cen/html101705081020.html?print Go

CHEMICAL & Engineering News

Government & Policy

October 17, 2005
Volume 83, Number 42
pp. 28–30

'Cyber-Enabled' Chemistry

NSF encourages chemists to apply computers, information technology to research, education

Susan R. Morrissey

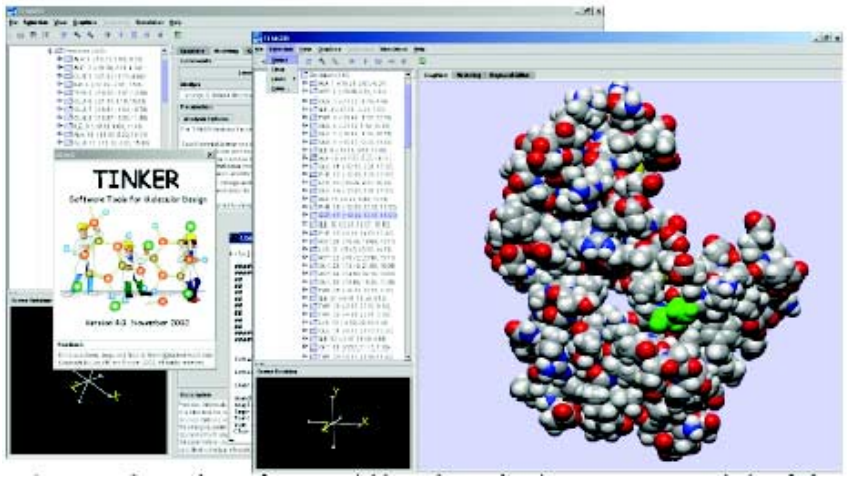
Advances in technology can be powerful drivers. Take, for instance, the progress in computer, information, and communication technologies, which have merged to produce a new, comprehensive cyber-infrastructure. It's this infrastructure that is generating excitement at the National Science Foundation: The infrastructure can provide a platform upon which scientists and engineers can build to create integrated databases and remote experimental control environments with the potential to more efficiently tackle difficult scientific problems.

At the forefront of this focus area is NSF's Chemistry Division. The division has been working to promote "cyber-enabled" chemistry-the application of this cyber-infrastructure to "enable new chemical research and education activities through grid computing, community databases, remote access to instrumentation, electronic support for geographically dispersed collaborators, and other Web- and grid-accessible services."

National Science Foundation

Integrative Activities: Computational Science – Cyber-enabled Chemistry

- **Next Generation Biomolecular Modeling** –
  - Develop and validate polarizable force fields for use in biomolecular simulations software and parallelized computer codes
  - Outcomes: parameter sets and software



Screenshots of Force Field Explorer shows energy analysis of protein calmodulin's structure. Credit: T. & M. Head-Gordon (UC Berkeley), V. Pande (Stanford), J. Ponder (Washington U)

National Science Foundation

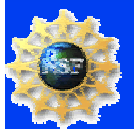
Integrative Activities: Computational Science – Cyber-enabled Chemistry

- **Collaboratory Tools for Multi-Disciplinary, Multi-Scale Research in Environmental Molecular Sciences (K.T. Mueller, Penn State)**
  - Develop a federated database
  - Link kinetic information in environmental chemistry across spatial and temporal scales
  - Integrate experimental, analytical, and simulation results from from molecular to field scales systems
  - Approximate complex interactions controlling the fate and transport of contaminants
  - Include ontological definition, search engine, and digital library services

National Science Foundation

Observations on the Present

- **Golden age of computing, but …**
- **Scalability - no longer a free ride on Moore's Law**
- **Shift from compute-centric to balance with data**
  - Emerging awareness of the importance of **middleware**



National Science Foundation

Cyberinfrastructure Vision 2005

NSF will lead the development and support of a comprehensive cyberinfrastructure essential to 21st century advances in science and engineering

LHC Data Distribution Model

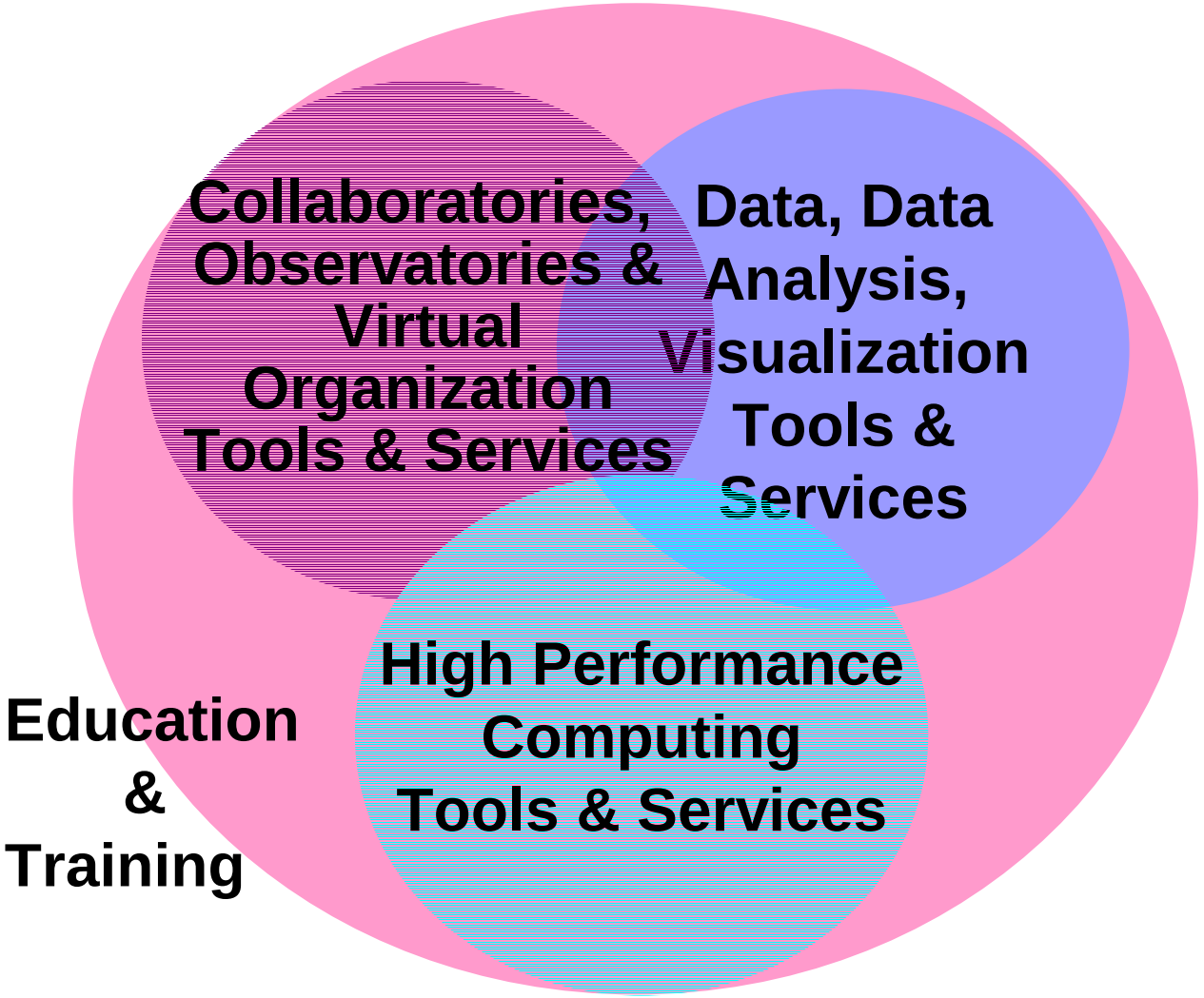
National Science Foundation

Organizational Changes

- **Office of Cyberinfrastructure**
  - Formed on 22 July 2005
  - Had been a division within CISE
  - Coordinates/supports acquisition, development, and provision of state-of-the-art cyberinfrastructure, tools, and services essential to the conduct of 21st century science and engineering research and education
- **Cyberinfrastructure Council**
  - NSF Director is Chair to Asst. Director members
- **Advisory Committee for Cyberinfrastructure**

National Science Foundation

Cyberinfrastructure Strategic Planning



National Science Foundation

Strategic Plan for High Performance Computing

- **Covers 2006-2010 period**
- **Enable petascale science and engineering by creating a world-class HPC environment**
  - Science-driven HPC Systems Architectures
  - Portable Scalable Applications Software
  - Supporting Software
- **Inter-agency and international synergies will be sought**

National Science Foundation

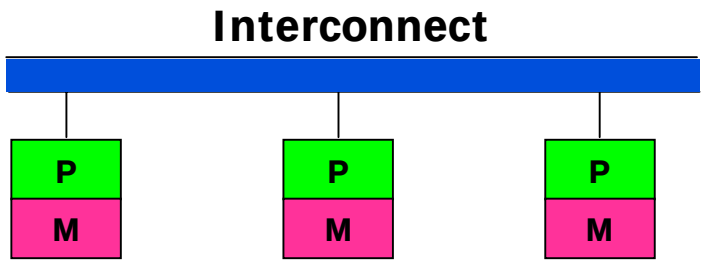
Science Driven Cyberinfrastructure

Algorithm Requirements

Science Areas	Multi-physics & multi-scale	Dense linear algebra	FFTs	Particle methods	AMR	Data parallelism	Irregular control flow
Nanoscience	X	X	X	X		X	X
Combustion	X			X	X	X	X
Fusion	X	X		X	X	X	X
Climate	X		X		X	X	X
Astrophysics	X	X	X	X	X	X	X

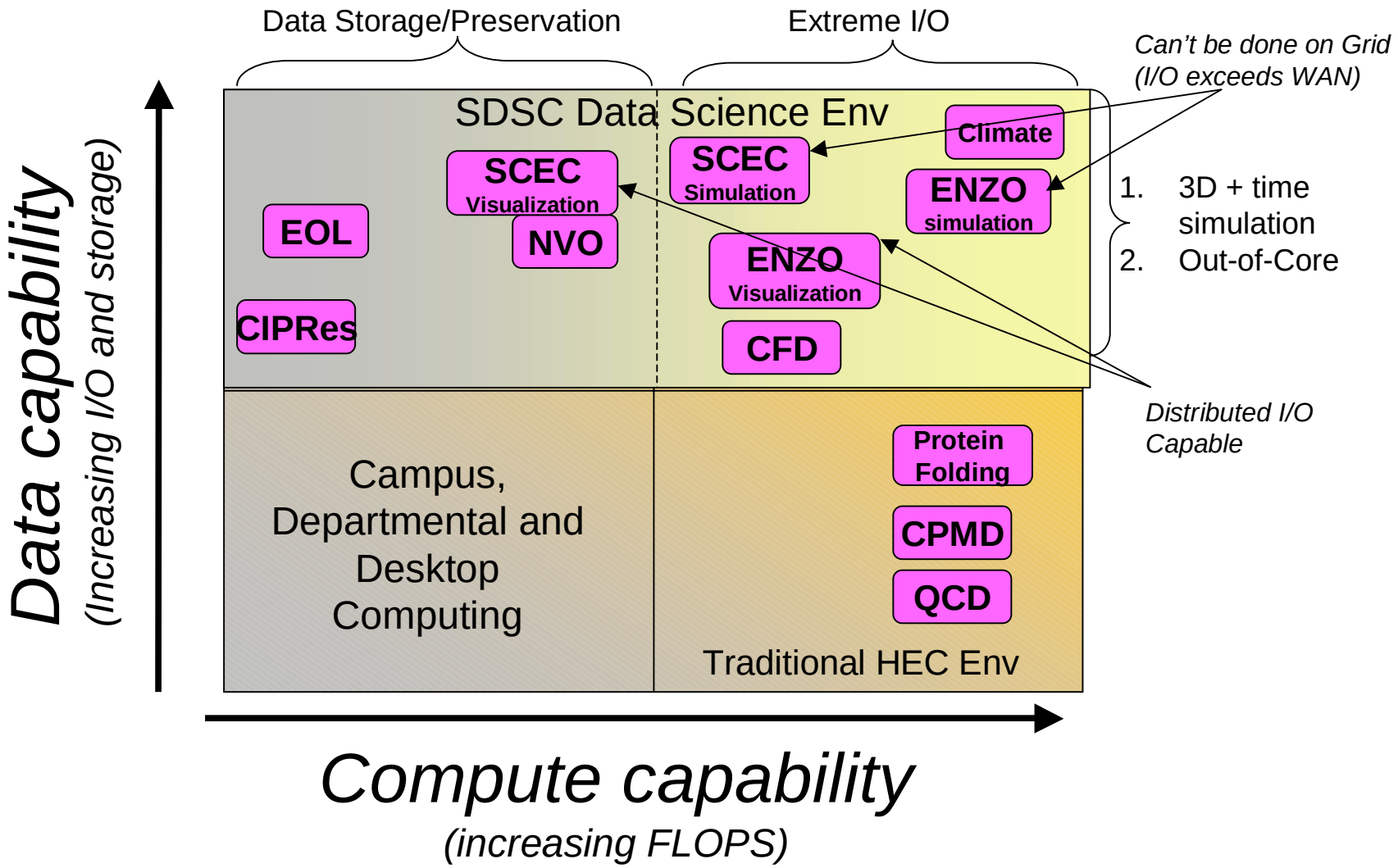
Trade-off

- Interconnect fabric
- Processing power
- Memory
- I/O



National Science Foundation

Computing: One Size Doesn't Fit All



courtesy SDSC

National Science Foundation

Strategic Plan for Collaboratoris, Observatories & Virtual Organizations

- ***Under development***
- Collaboratories, observatories and virtual organizations … allow scientists and engineers to pursue their research and education goals without regard to geographical location. In these environments, individuals will be able to access experimental and computational tools, interact with their colleagues, and share data, information and knowledge.
- This chapter will focus on the tools and services necessary to create these highly-interactive, widely-accessible environments to promote progress in science and engineering.

National Science Foundation

Strategic Plan for Data, Data Analysis & Visualization

- ***Under development***
- NSF envisions a world in which digital science and engineering data are routinely deposited in convenient repositories, can be readily discovered in well-documented form by specialists and non-specialists alike, are openly accessible, and are reliably preserved.
- This chapter will describe NSF's strategy to enable this vision through the development of appropriate data policies and the identification of robust, interoperable data services, including data analysis and visualization services.

National Science Foundation

Strategic Plan for Education & Workforce

- ***Under development***
- NSF recognizes that CI will have a profound impact on the practice of science and engineering research and education, enabling individuals, groups and organizations to advance science and engineering in ways that revolutionize *what they can do, how they do it,* and *who can participate*. To harness the full power of CI and the promise it portends for discovery, learning and innovation across and within all areas of science and engineering requires focused investments in the preparation of a science and engineering workforce with the knowledge and requisite skills to create, advance and exploit CI over the long-term. This chapter will describe NSF's approach to doing so.

National Science Foundation

Predictions for the Future

- **Pace of change will accelerate**
- **"IT Does Not Matter"**
  - **but *architecture does* … *and middleware does* – watch the response to Wal-Mart initiated RFID technology disruption**
- **Look for a new network paradigm: wireless, RFID, sensor networks, DDDAS …**
  - **Pending revolution in manufacturing and supply chain as well as the built environment**
  - **Large-scale environmental sensing, monitoring, management**

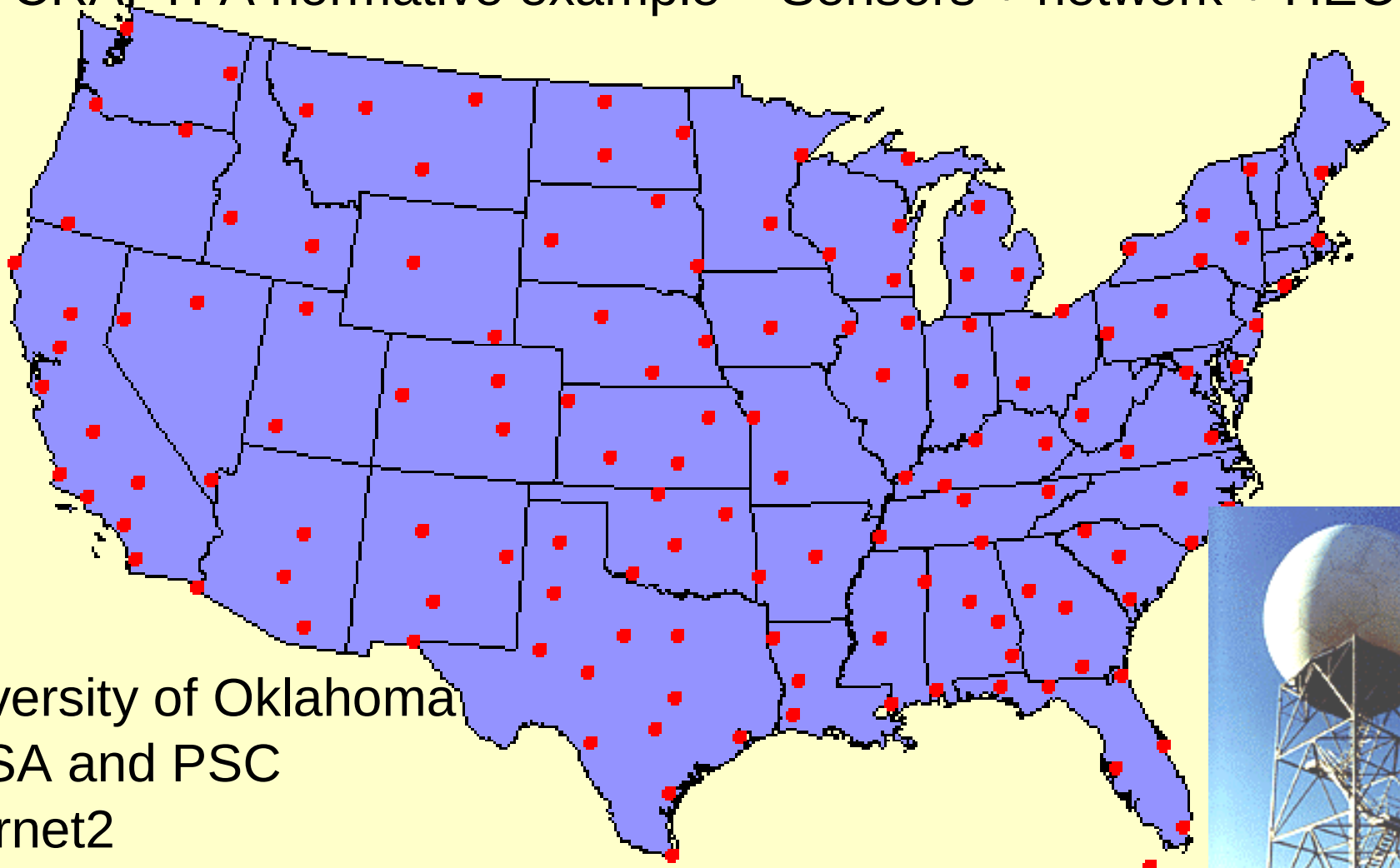
National Science Foundation

Thank You!

Miriam Heller, Ph.D.
Program Director
Office of Cyberinfrastructure
National Science Foundation
mheller@nsf.gov

Tel: +1.703.292.7025 Email: mheller@nsf.gov

CRAFT: A normative example – Sensors + network + HEC



University of Oklahoma
NCSA and PSC
Internet2
UCAR Unidata Project
National Weather Service

WSR-88D LOCATIONS