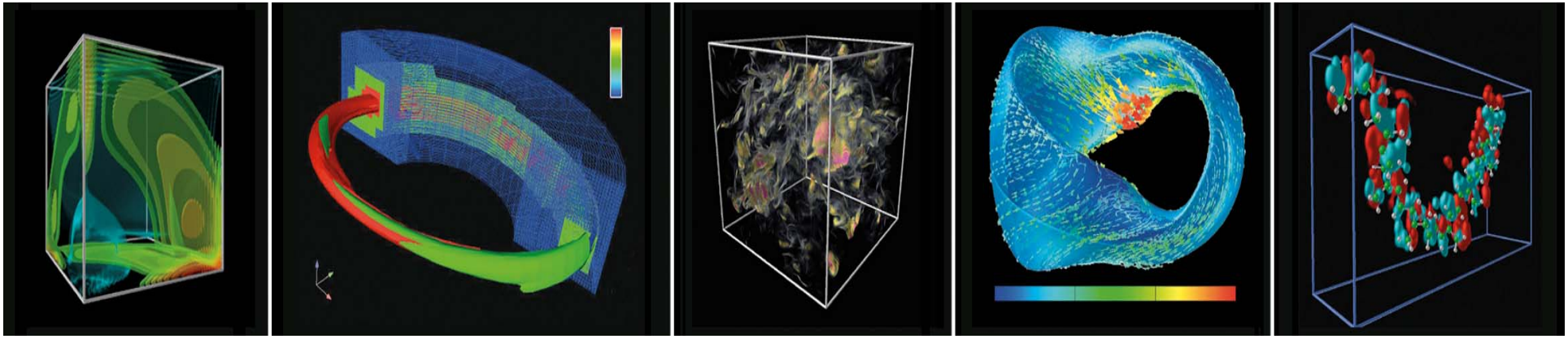


CSE @ CITRIS and LBNL

Computational Science and Engineering (CSE)

CITRIS* and LBNL Partnership

***(UC Berkeley, UC Davis, UC Merced, UC Santa Cruz)**

http://www.citris-uc.org/

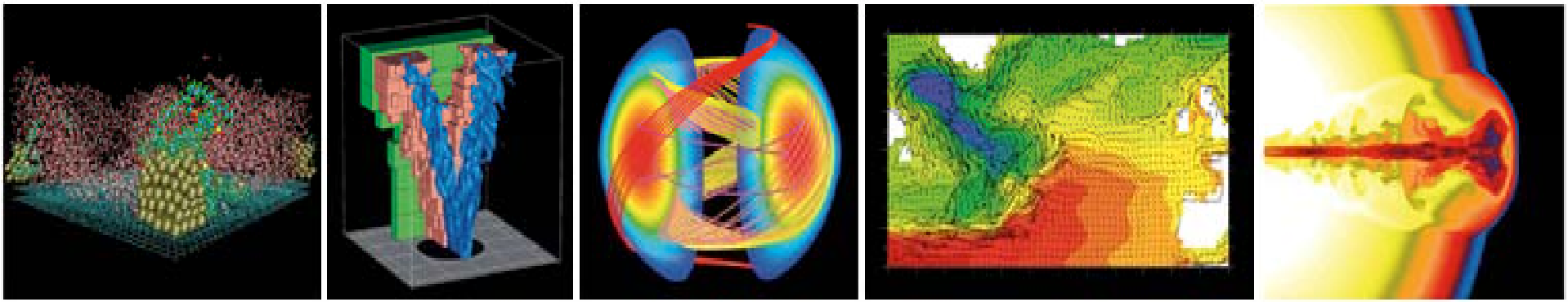
Masoud Nikravesh

CITRIS CSE Director, CITRIS and LBNL, University of California-Berkeley

Nikravesh@eecs.berkeley.edu

Leadership through Innovation

Innovation through Collaboration and Strategic Partnerships



Computation will Change Science

"An important development in sciences is occurring at the intersection of computer science and the sciences that has the potential to have a profound impact on science. It is a leap from the application of computing… to the integration of computer science concepts, tools, and theorems into the very fabric of science." -Science 2020 Report, March 2006

Nature, March 23, 2006

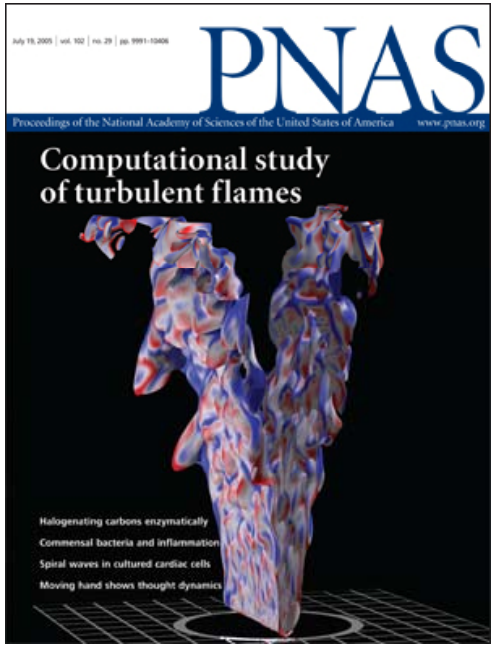
The breadth of this enterprise promises to create extensive new applications for high performance computing, and more opportunities for new products to become the tools of choice.

Computational Science and Engineering

Theory

Experiment

Simulation



High performance computing (HPC), large-scale simulations, and scientific applications all play a central role in CSE.

CITRIS

CITRIS

CITRIS is the "Center for Information Technology Research in the Interests of Society" --- in short, "Technology for Society"

In 2001, the State of California began four unique "California Institutes of Science and Innovation"

$100 million from the State of California

>$ 500 million raised in research dollars at CITRIS

Technology for Society means:

Not just the "technology-push"

Work with Business School, Law School

Public Policy, Political Science, and the

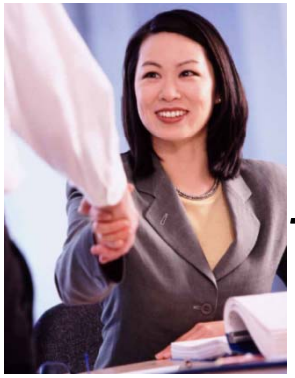
Lawrence Berkeley Laboratory and…

Create with social impact

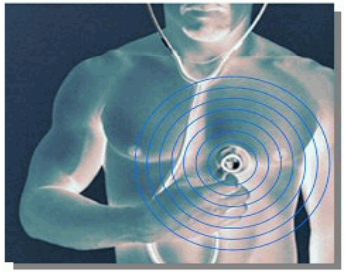
STRATEGIC DIRECTIONS

Mission: Innovation in Information Technology <u>to enable solutions to the most challenging societal problems</u> affecting the quality of life, for Californians and the world .

Best new technologies to integrate into Societal - Scale Systems:



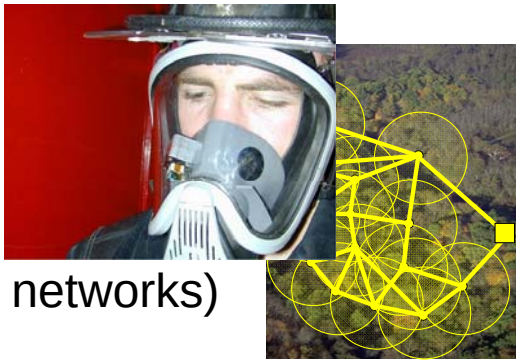
- ***Services Science & Technology***



- ***Delivery of Health Care*** (using IT)

- ***Energy & Environment*** (joint with Helios, LBNL)

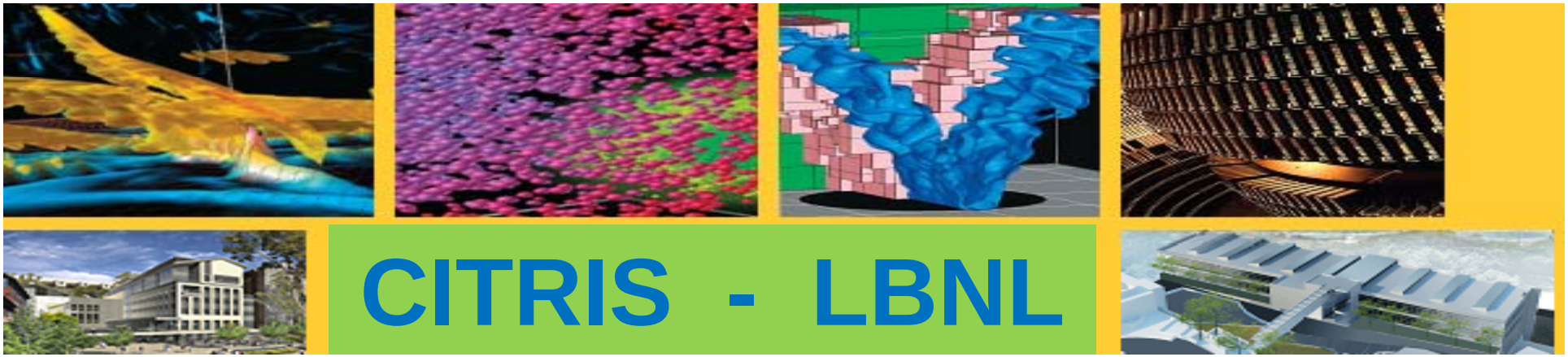
- ***Technology for Developing Economies***



- ***Intelligent Infrastructures*** (leveraging sensor networks)

- ***Computational Science & Engineering*** (joint with CS-LBNL)

CITRIS



Computational Science and Engineering

CSE @ CITRIS

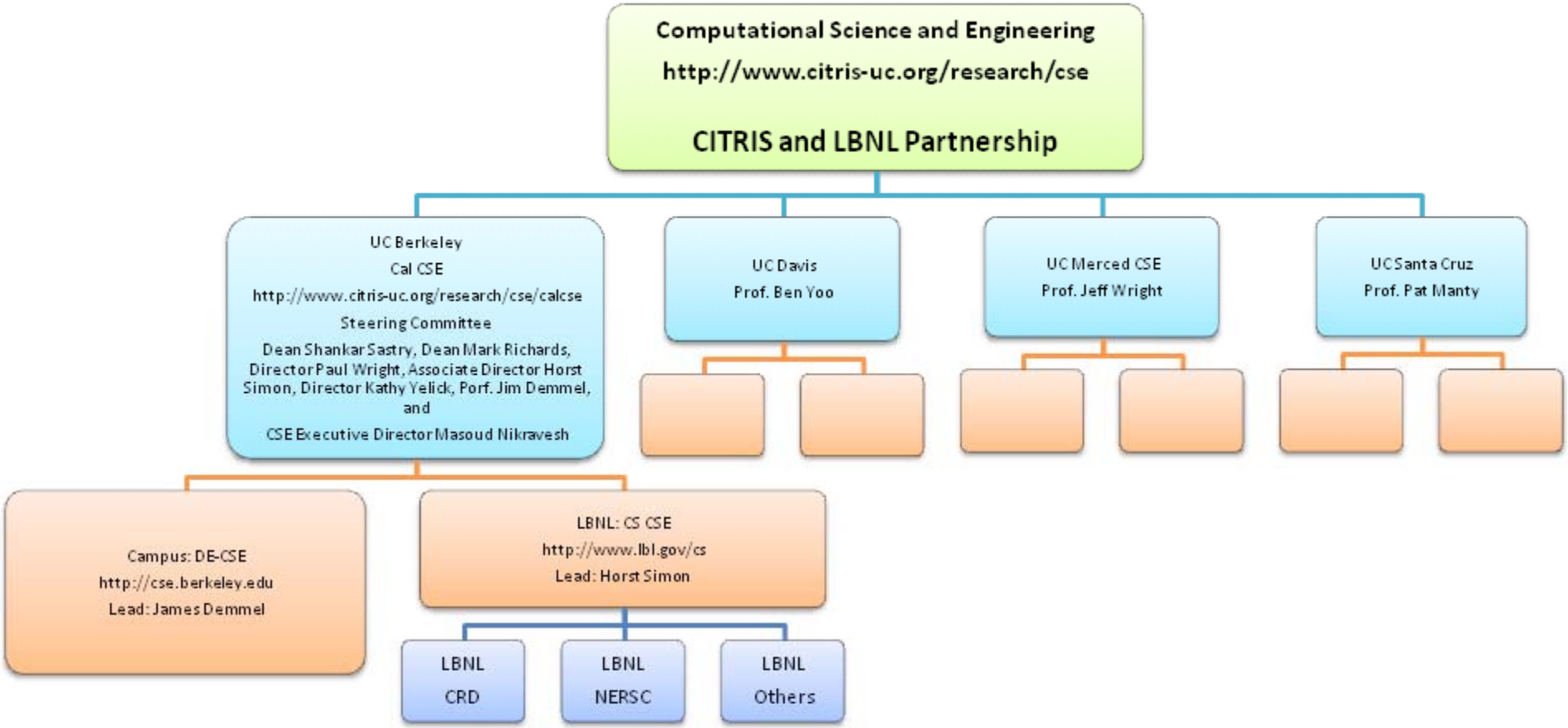
Our Mission:

To support the work of the most creative minds in science and engineering as they pursue **complex and computationally intensive** basic and applied research that will enhance national scientific, technological, and economic leadership and competitiveness while improving the **quality of life for humanity.**

The CSE Program will:

- **Conduct world-leading research** in applied mathematics and computer science to provide leadership in the area of science and engineering such as energy, environment, climate, astrophysics, nano science, engineering of mind, and cyberinfrastructure to name a few.
- Be in the forefront of the development of the **ultra-efficient largest-scale computer systems** focusing on **scientific and technological discoveries** and significant world-changing breakthrough solutions.
- **Educate and train the next generation** of CSE leaders and give students a competitive edge for the most desirable jobs in academia and industry. Critical to our mission is the education of graduate students capable of confronting challenging problems at the leading edge of CSE.

CSE: A NEW Strategic Research



‘Integration’ of CITRIS Mission

Energy and the Environment

Intelligent Infrastructure

Health care

Emerging economies

Art, technology and culture

Services

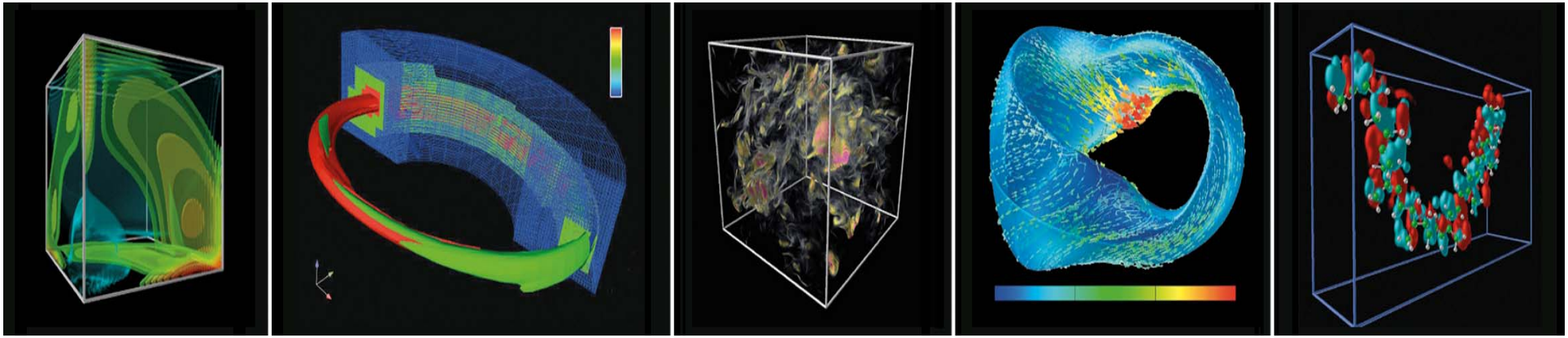
CSE

Networking

Embedded Systems

Nanotechnologies

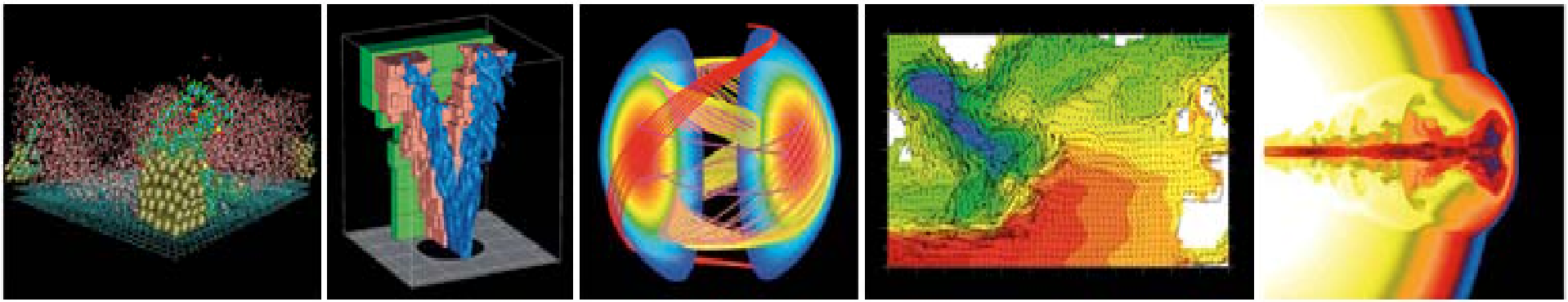
Center for IT Research in the Interest of Society



CSE @ Cal

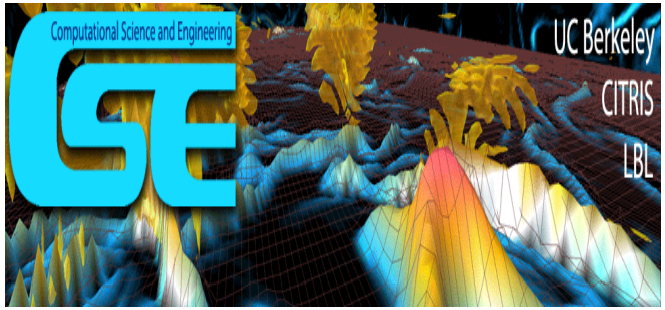
http://cse.berkeley.edu

Computational Science and Engineering

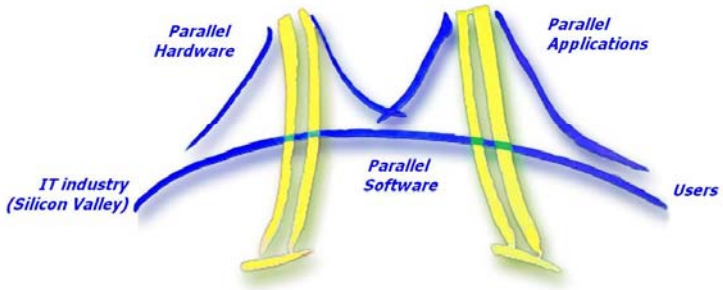


Three Major Events at Berkeley

- Berkeley's newly established Graduate Program in computational science and engineering (CSE) with over 120 faculty from 22 Departments



- ParLab/UPCRC project at Berkeley, a new research center in parallelism at UCB and RadLab and Cloud Computing Initiatives



- UC Berkeley and Lawrence Berkeley National Laboratory's (LBNL) commitment to build a new Computational Research and Theory (CRT) Building, a $113M research facility house to a PetaFlops computer

Berkeley CSE

- Combined with Berkeley's well-established strengths in computer science, applied mathematics, and scientific applications of large-scale computing and ParLab/UPCRC, we can envision Berkeley's positions at the forefront of computational science for years to come.

- This is a rare time in computing history when the needs and skills of the scientific/engineering community on the one hand, and the commercial computing community on the other, largely coincide. We should not miss this opportunity to join the skills of both sides to tackle common problems in parallel and scientific computing.

We think, long-term investment could launch an exciting new collaborative enterprise that would change the way future generations of students and researchers in CSE contribute to their fields, create new markets for new products, and help train a generation of the brightest students on these tools.

Berkeley CSE (Cal CSE)

Cal CSE is a collaborative program in multidisciplinary research and education. It was developed and is supported by the Center for Information Technology Research in the Interest of Society (CITRIS), the Computing Sciences Directorate at Lawrence Berkeley National Laboratory (Berkeley Lab), and the College of Engineering, the College of Letters and Science, the College of Chemistry, and the office of the Vice Chancellor for Research at the University of California (UC) Berkeley. Cal CSE is administered by CITRIS and the Berkeley Lab Computing Sciences Directorate.

Designated Emphasis (DE) in CSE

- New “graduate minor” – approved, starting July 1, 2008
- Motivation
  - Widespread need to train PhD students in CSE
  - Opportunities for collaboration, across campus and at LBNL
  - 18 (20) departments, 85 (107) faculty signed up (so far)
- Graduate students participate by
  - Getting accepted into existing department/program
  - Taking CSE course requirements
  - Qualifying examination with CSE component
  - Thesis with CSE component
  - Receive “PhD in X with a DE in CSE”
  - Details at cse.berkeley.edu (soon!)

Over 120 Faculty from 22 Departments

1. Astronomy, (Chair) Donald Backer
2. Bioengineering, (Chair) Dorian Liepmann
3. Biostatistics, (Chair) Sandrine Dudoit
4. Chemical Engineering, (Chair) Jeffrey Reimer
5. Chemistry, (Chair) Michael Marletta
6. Civil and Environmental Engineering, (Chair) Lisa Alvarez-Cohen
7. Earth and Planetary Science, (Chair) Rudy Wenk
8. Electrical Engineering and Computer Sciences, (Chair) Stuart Russell
9. Industrial Engineering and Operations Research, (Chair) Ilan Adler
10. School of Information, (Dean) AnnaLee Saxenian
11. Integrative Biology, (Chair) Nipam Patel
12. Materials Science and Engineering, (Chair) Robert Ritchie
13. Mathematics, (Chair) Alan Weinstein
14. Mechanical Engineering, (Chair) Al Pisano
15. Helen Wills Neuroscience Institute, (Head) John Ngai
16. Nuclear Engineering, (Chair) Jasmina Vujic
17. Physics, (Chair) Frances Hellman
18. Political Science, (Chair) Paul Pierson
19. School of Public Health, (Dean) Stephen Shortell
20. Statistics, (Chair) John Rice

Participating Departments (1/2)

(# faculty by "primary affiliation", # courses)

- Astronomy (7,3)
- Bioengineering (3,1)
- Biostatistics (2,0)
- Chemical Engineering (6,0)
- Chemistry (8,1)
- Civil and Environmental Engineering (7,8)
- Earth and Planetary Science (6,3)
- EECS (19,14)
- IEOR (5,5)
- School of Information (1,0)

Participating Departments (2/2)

(# faculty by "primary affiliation", # courses)

- Integrative Biology (1,0)
- Materials Science and Engineering (2,1)
- Mathematics (15, 4)
- Mechanical Engineering (9, 6)
- Neuroscience (7,1)
- Nuclear Engineering (2,1)
- Physics (1,1)
- Political Science (2,0)
- Statistics (5, 11)
- New: Biostatistics, Public Health

Course Structure

- 3 kinds of students, course requirements
  - CS , Math, “Applications”
- Each kind of student has 3 course requirements in other two fields
  - Goal: enforce cross-disciplinary training
  - Non-CS & Non-Math students:
    - 1 or 2 Math courses from list
    - 1 or 2 EECS courses from list
    - Other classes from Stat, IEOR
  - Math & CS students: substitute 1 or 2 courses from “applied” department for 1 or 2 inside
    - May distinguish EECS and CS students
- We have $ to support new course development

Math Courses (so far)

- Ma220 – Probabilistic Methods
- Ma221 – Numerical Linear Algebra
- Ma228AB – Numerical Solutions of Differential Equations

Stat Courses (so far)

- Stat215AB: Statistical Models
- Stat230A: Linear Models
- Stat232: Experimental Design
- Stat240: Nonparametric and Robust Methods
- Stat241AB: Statistical Learning and Decision Theory (cross-listed with CS)
- Stat244: Statistical Computing
- Stat245AC: Biostatistical Methods
- Stat246: Statistical Genetics
- Stat248: Time Series Analysis

IEOR Courses (so far)

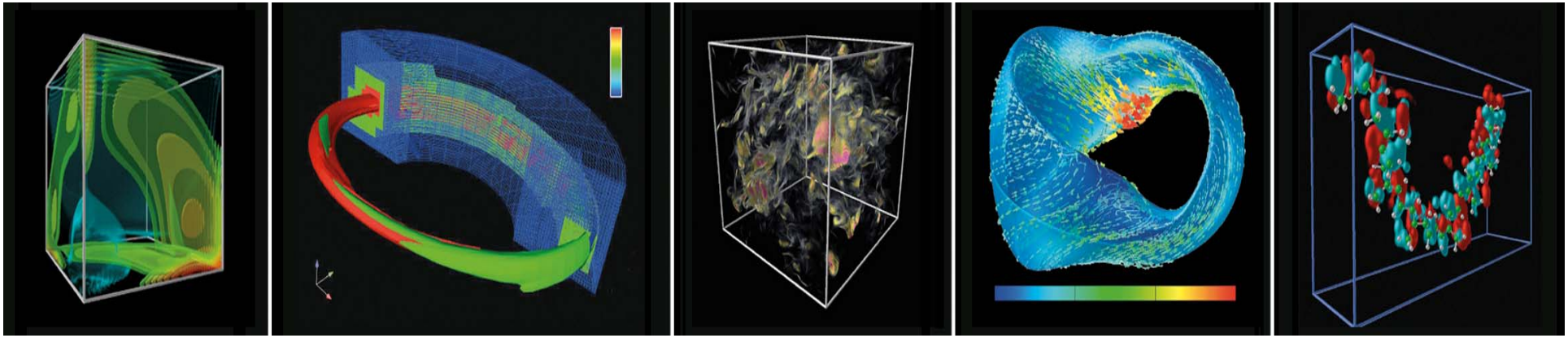
- IEOR261: Experimenting with Simulated Systems
- IEOR262AB: Mathematical Programming
- IEOR264: Computational Optimization
- IEOR269: Integer Programming and Combinatorial Optimization

EECS Courses (so far)

- CS267 – Applications of Parallel Computers
- CS270 – Combinatorial Algs. & Data Structures
- CS274 – Computational Geometry
- CS280 – Computer Vision
- CS281A – Statistical Learning Theory
- CS281B – Learning and Decision Making
- CS284 – Geometric Design and Modeling
- CS285 – Solid Modeling and Fabrication
- CS294-10 – Visualization
- EECS225AB– Digital Signal and Image Processing
- EECS227A – Convex Optimization
- EECS228B – Convex Approximation

More on possible new courses

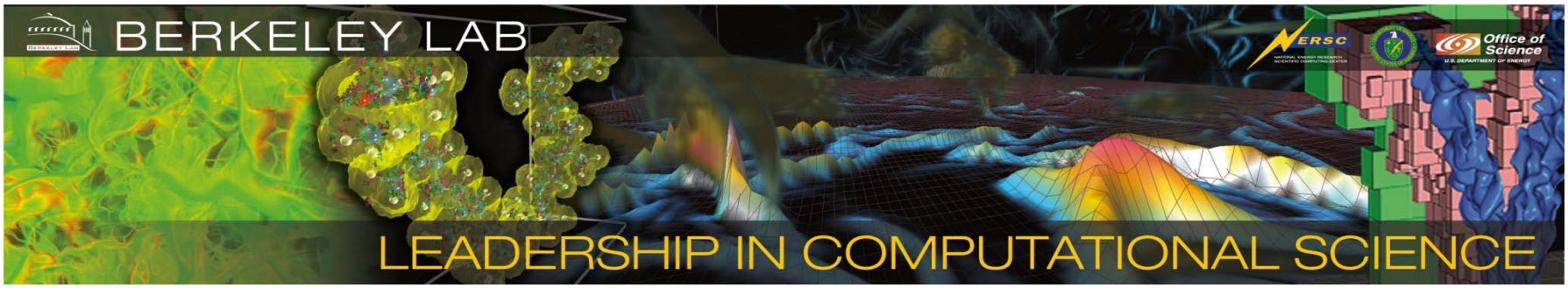
- Possible obstacles to students
  - Long prerequisite chains
  - Important material spread over multiple courses
  - Repetition of basic material in multiple courses
- 23 faculty identified following needs, opportunities:
  - New 3 unit survey course in numerical methods
  - 23 new 1-unit classes proposed with this as prerequisite
  - “Segmented courses” where one can take subset for fewer units (eg CS267)



Cloud Computing

http://cloud.citris-uc.org/

Computational Science and Engineering



Possible Research Topics for "Cloud"

- A *small* selection from among the 112 faculty
- Machine Learning (2 faculty)
  - Scale-up of Bayesian nonparametric statistics
  - Many applications (vision, speech, genetics)
- Astronomy (10 faculty)
  - Some simulations (large scale, many smaller scale), some large data sets (up to terabytes/day)
- Chemistry and Chemical Engineering (12 faculty)
  - Some large-scale simulations, some less tightly coupled
    - Ex: New materials for energy via QMC, chemical database screening

Possible Research Topics for "Cloud"

- Neuroscience and Cognitive Computing (8 faculty)
  - Some large scale simulations (of brain, auditory system)
  - Some large data set analysis (crcns.org)
- Protein Facebook (1 faculty)
  - Make large amounts of published protein data available
  - Provide easy search over many feature, collaborative filtering, alerts for relevant new results
- Natural Language and Speech Processing (1 faculty)
  - Large scale training of automatic recognition systems using much more data than possible before
- Netalyzr (1 faculty)
  - Provide network probing for anomalies and features as service

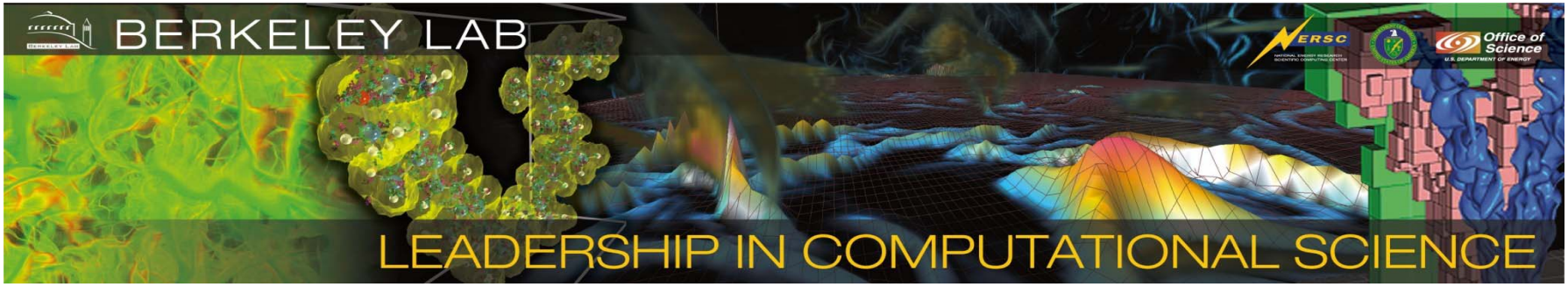
Possible Research Topics for "Cloud"

- Computational systems biology (9 faculty)
  - "Digital Human", many layers of simulation
- Econ/EECS/IEOR/Math/PoliSci/Stat (9 faculty)
  - Statistical analysis and visualization of large scale heterogeneous data bases of economic, financial, social data
  - Ex: statnews.eecs.berkeley.edu/about/project for news analysis
- Economics (8 faculty, including 1 Nobelist)
  - Econometric and social modeling
- Ultra-efficient Climate Computer (7 faculty + staff)
  - Joint with LBNL
  - 100x lower power than current supercomputers

Lawrence Berkeley National Lab
Computing Sciences Directorate

LBNL CSE

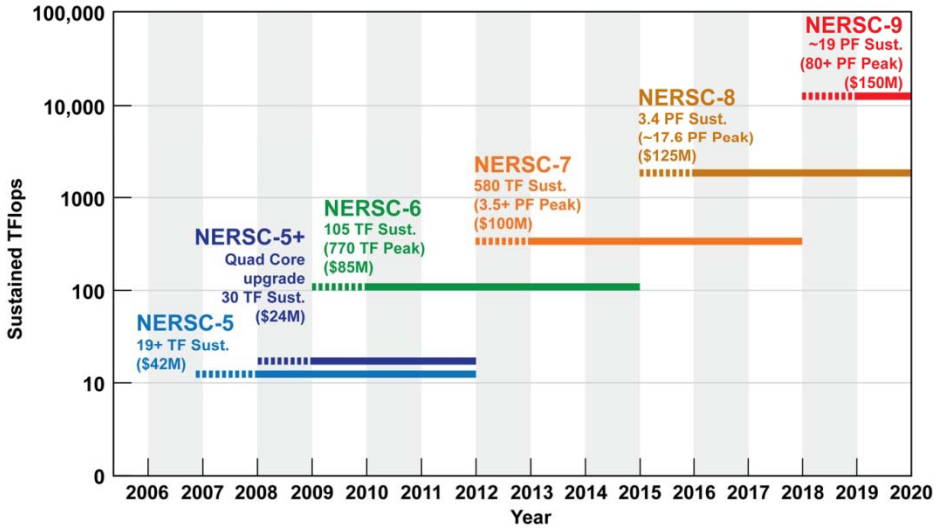
Computing Sciences consists of two divisions — the Computational Research Division (CRD) and the National Energy Research Scientific Computing (NERSC) Division. The Computing Sciences organization was created in 1996 and currently has about 200 employees. Most of the computational research and resources provided by Computing Sciences are funded by the Advanced Scientific Computing Research (ASCR) program in the U.S. Department of Energy (DOE) Office of Science.

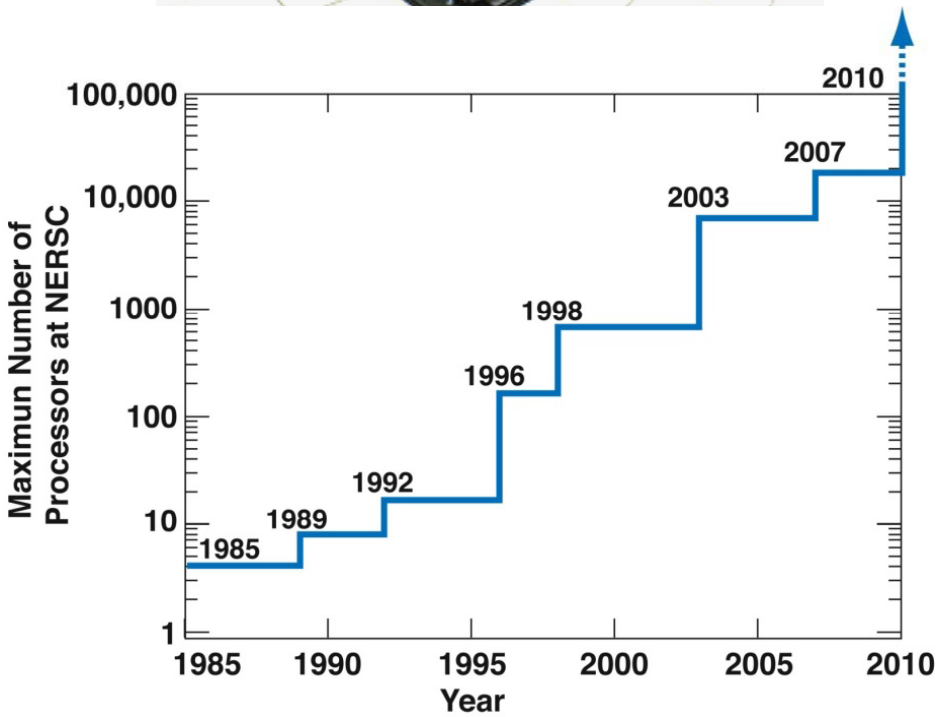


LBNL CSE

- The Computational Research Division (CRD) creates computational tools and techniques that enable scientific breakthroughs, by conducting applied research and development in computer science, computational science, and applied mathematics. CRD consists of four departments:
  - Advanced Computing for Science Department
  - Biological Data Management and Technology Center
  - Energy Sciences Network (ESnet)
  - High Performance Computing Research Department
- NERSC is the flagship scientific computing facility for the DOE Office of Science and a world leader in accelerating scientific discovery through computation.

Computing Resources





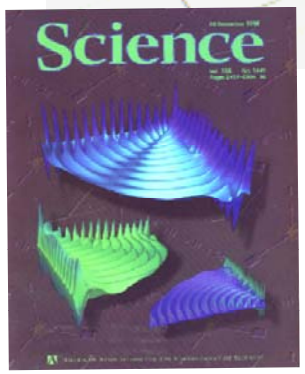
NERSC

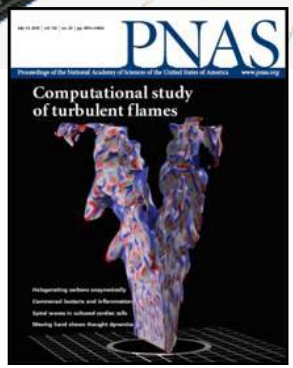
National Energy Research Scientific Computing Center

Serves the <u>entire</u> scientific community

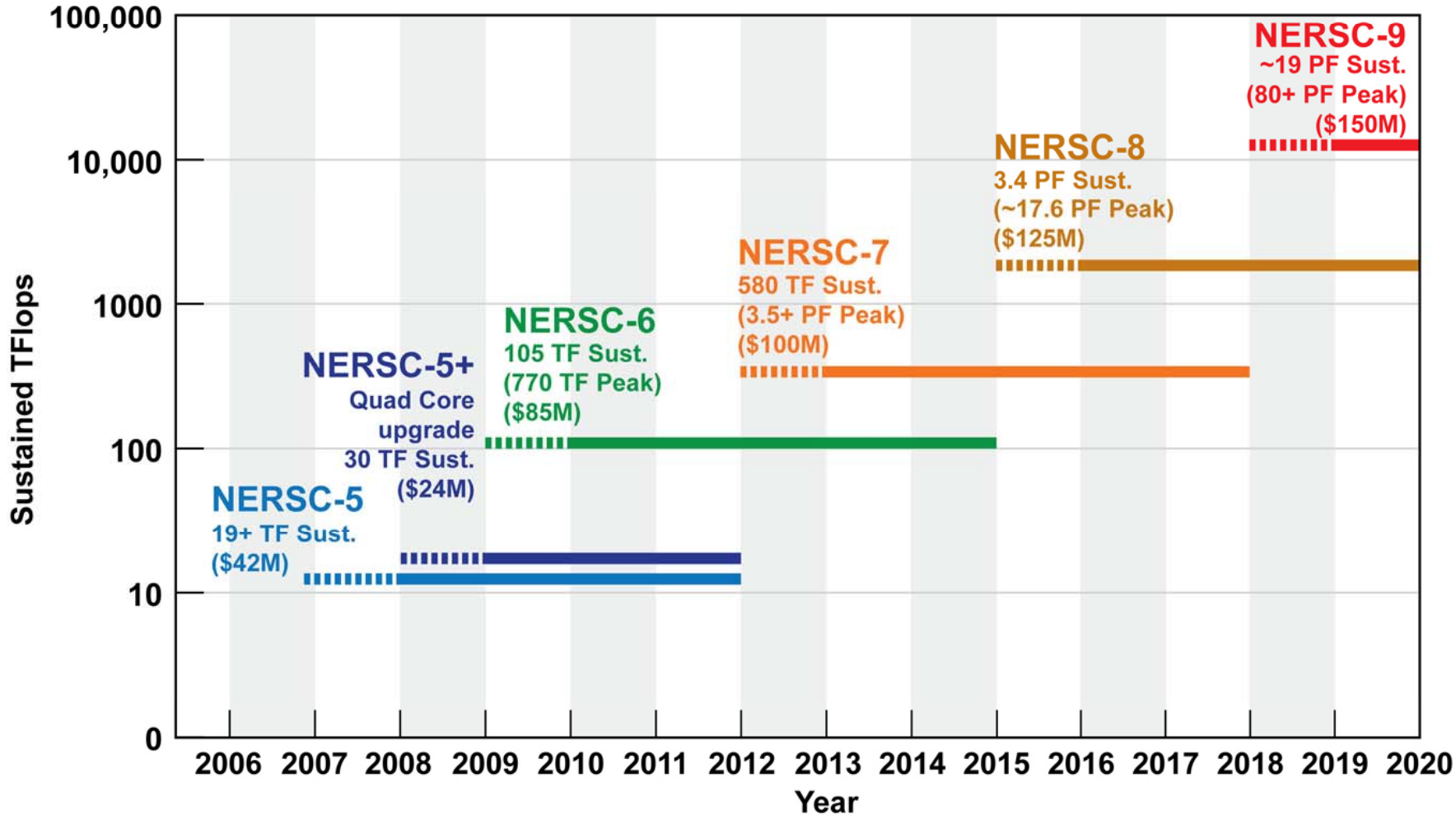
~2500 Users in ~250 projects

- **Focus on large-scale computing**





NERSC Approximate Computational System Profile



Bringing It All Together:
A New Building for Computational Science

Resources

Infrastructure

NEW CITRIS Building

The Marvell NanoTechnology Center (Floor 5)

- This is the “new Microlab” for chip-prototyping (lithography, etching, CVD etc)
- Open to all CITRIS investigators
- 20,000 sq. ft. out of the 84,000 sq.ft.
- $55M cost out of the $128M cost of the whole building

Auditorium – 150 seats fully set up for Distance Learning and high-end Conferencing (Floor 3)

- Also two smaller distance learning rooms (Floor 2)
  - Two rooms with 25 seats each and One room with 50 seats

Transmission (TEM) and Scanning (SEM) Electron Microscopes (Floor 1)

Macrolab student space – for Design and Prototyping (Floor 1)

CSE Computer Cluster / Cloud Computers

Computational Research and Theory Building (2011)

Program: Facility to accommodate scientific computing research, high performance computing centers, and the Cal CSE program.

1. Assignable Sq Ft: 73,000
2. Gross Sq Ft: 126,300
3. **Total Project Budget: $112,944,000**

Leadership through Innovation

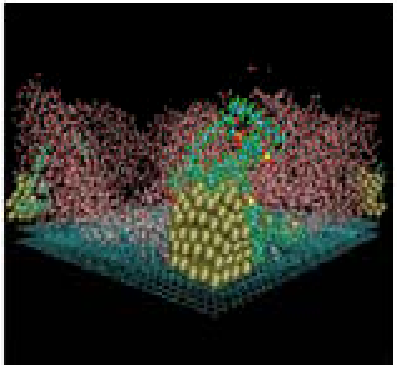
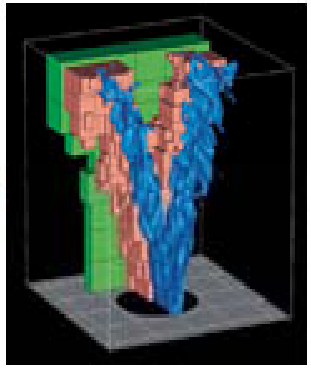
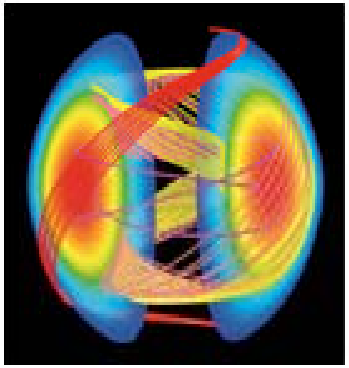
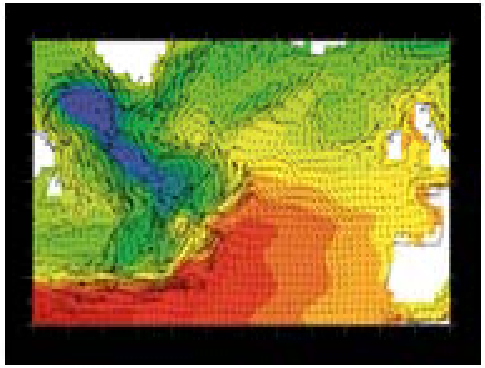
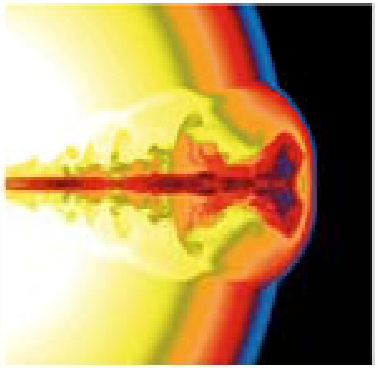
Innovation through Collaboration and Strategic Partnerships

- Joint project with UC Physical & Social Sciences and Engineering
- Collaboration and cross fertilization among the different groups
- Virtual projects of faculty, staff, students
- Persistent infrastructure for software, computing, and networking

Applications

Socio-Economic Modeling for Large-scale Quantitative Climate Change Analysis

DOE's 10 Years Vision
Exascale meeting, Group "B6"

From Terascale to Exascale

Terascale (i.e., today, almost)

- Economic models with ~10 countries & ~10 sectors
- Limited coupling with climate models
- No treatment of uncertainty and business cycle risk
- Simple impact analysis for a limited set of scenarios
- Limited ability to provide quantitative policy advice

From Terascale to Exascale

Petascale

- Economic models with more countries, sectors, income groups
- Limited treatment of uncertainty, business cycle risk
- Stronger coupling with climate models

Tera

From Terascale to Exascale

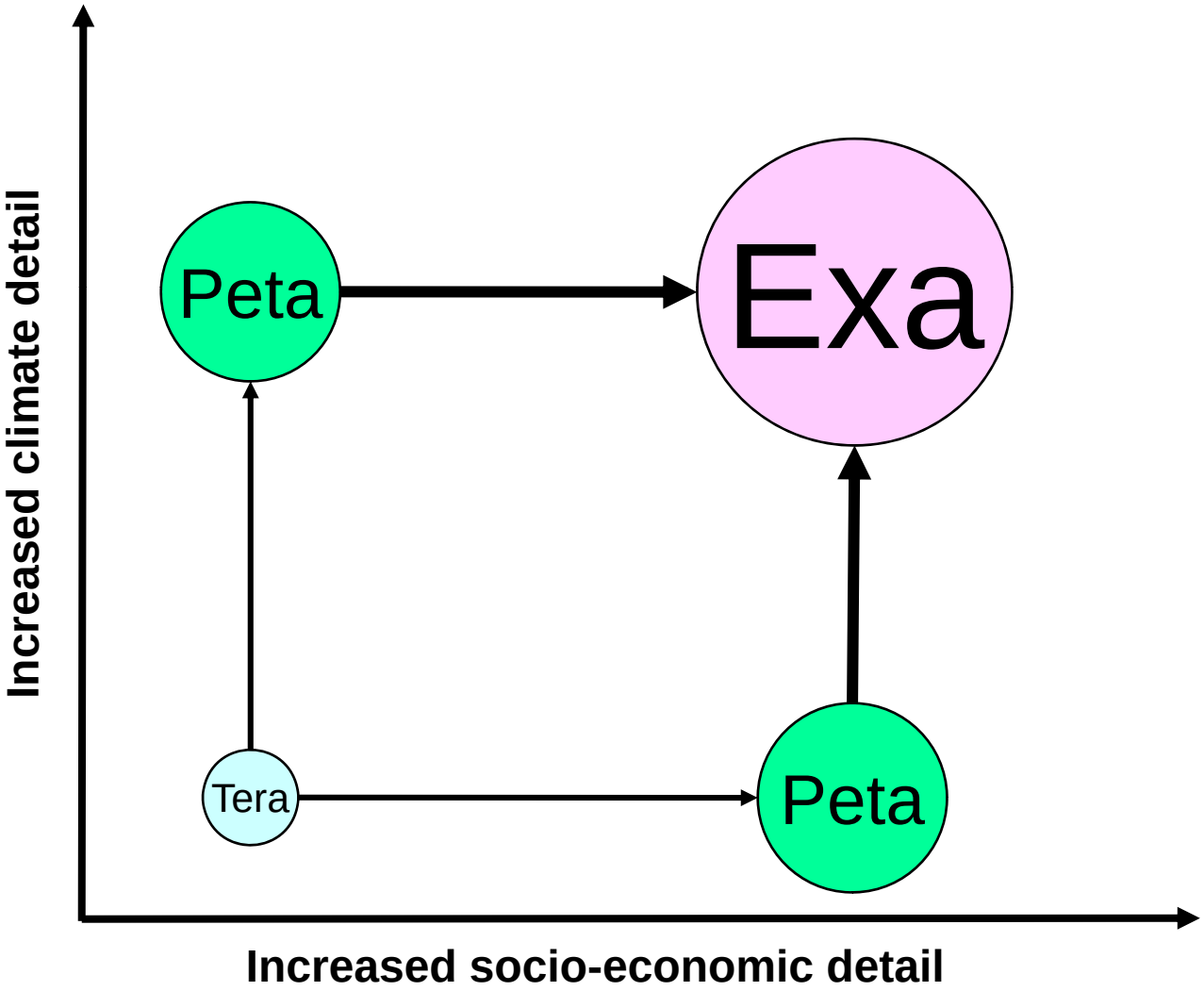
Exascale

- Economic models with all countries, many sectors, many income groups
- Many policy instruments (taxes, tariffs, quotas, CAFE, CO2 taxes), nonlinear policies, etc.
- High spatial resolution in land use, etc.
- Detailed coupling & feedbacks with climate models
- Optimization of policy instruments & technology choices over time and with respect to uncertainty
- Detailed model validation & careful data analysis
- Treatment of technological innovation, industrial competition, population changes, migration, etc.

Peta

Tera

From Terascale to Exascale



A Potential Timeline

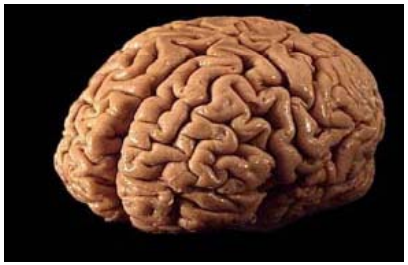
- **2007**: Socio-economic models on small clusters
- **2009**: First Petascale computer (we are told)
- **2011**: First socio-economic-environmental models on Petascale computers assuming major new program (otherwise **2016** or later?)
- **2015**: First Exascale computer (we assume)
- **2016**: Better socio-economic-environmental models running on Exascale computers assuming major new program (otherwise **2021** or later?)
- **2025**: Socio-economic-environmental Tokomak/ Stellarator

Socio-Economic Modeling Challenges

- Quantitative exploration of specific subsystems
  - E.g., energy markets, international agreements, technology innovation, human & social behaviors
  - Some are highly computationally demanding
- Data estimation, integration, analysis
  - Data access, cleaning, management, visualization
  - Empirical estimation of parameters (much computation)
- Integrated modeling of natural system & human system
  - Climate change, carbon cycle, water cycle, ecosystem
  - Multi-scale and multi-sector analysis
  - Goal: integrated natural system/human system model
- Uncertainty analysis of subsystems and systems

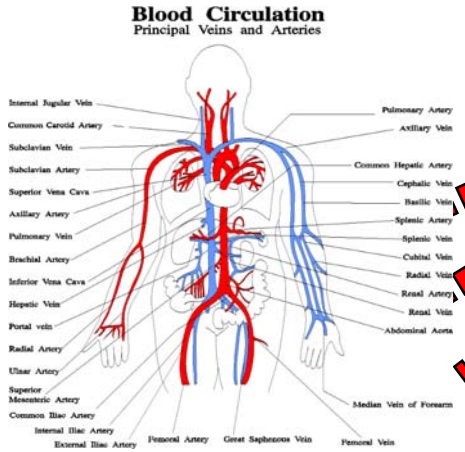
Cognitive Computing and Engineering

Cognitive Computing is a study of top-down, global, unifying theories that explain observed **cognitive phenomena ("mind")**, that are consistent with known bottom-up **neurobiological facts (the "brain"),** that are computationally feasible (for example, **implement-able on a BlueGene**), and that are mathematically principled. Cognitive Computing is a search for **computer science-type software/hardware** elements that are consistent with **known neurobiological facts about the brain** and give rise to observed mental processes of **perception, memory, language, intelligence, and, eventually, consciousness. Very simply speaking, Cognitive Computing is when computer science meets neuroscience to explain and implement psychology.** We have, in the brain and nervous system, an information processing system unrivalled by artificial means. While it trails machines in accuracy and mathematical computation, it wins on adaptability, flexibility, functionality, and parallelism. The ultimate goal is to reverse engineer enough of this system so that the design principles can be applied to building robust and adaptable computer systems.

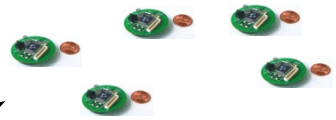


+

CITRIS Example: on-line Healthcare

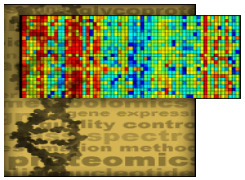


Sensors



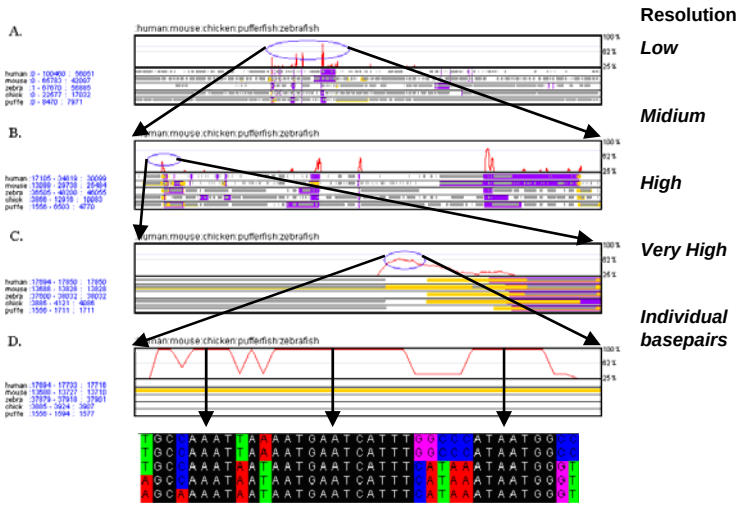
Databases

Modeling



Computational Simulation

Visualization

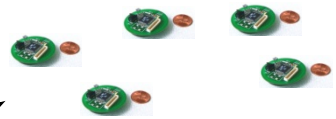


Prevention
Detection
Response
Recovery

Intelligent Networking at Works (example: Disaster Response)

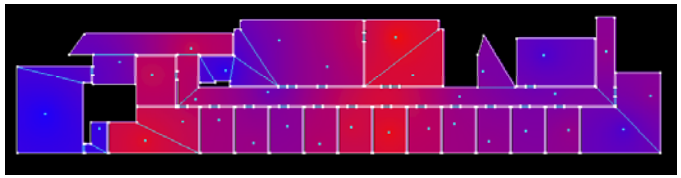
Disaster

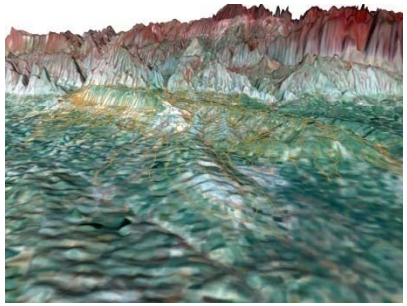
Sensors



Databases

Visualization





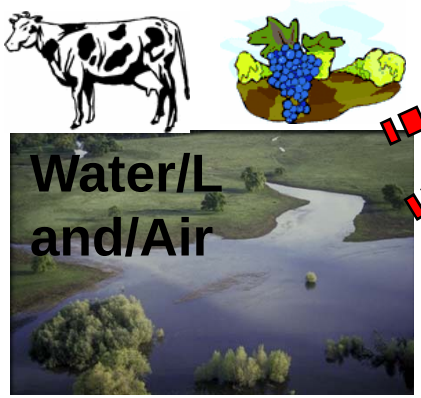
Modeling

Computational Simulation

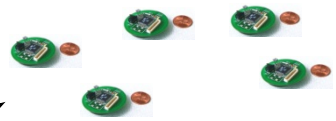
Prevention
Detection
Response
Recovery

CITRIS Example: Environment Monitoring

Environment

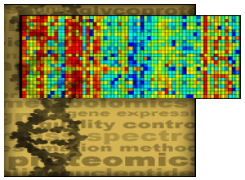


Sensors



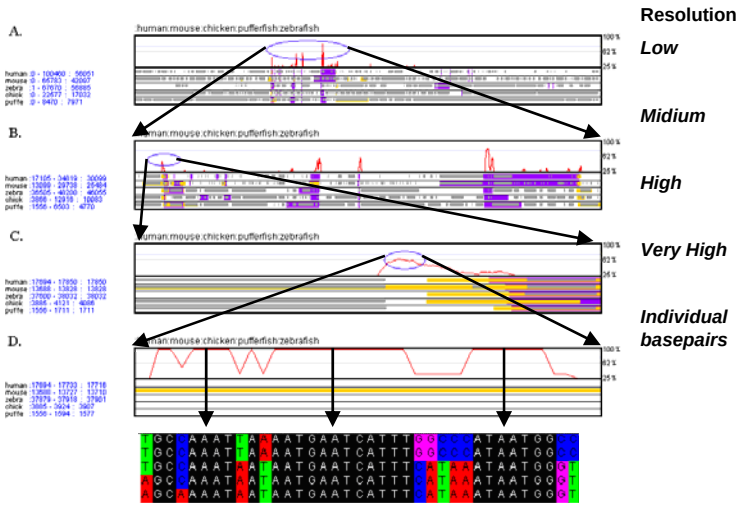
Databases

Modeling

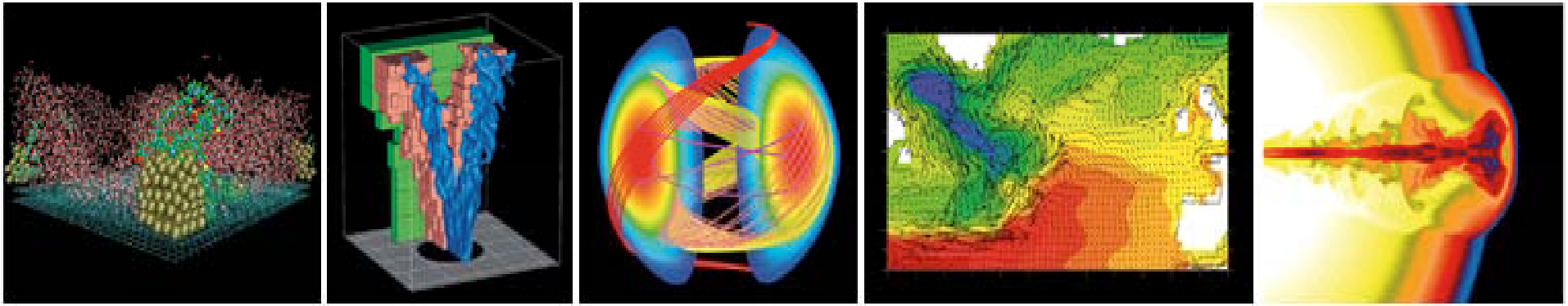


Computational Simulation

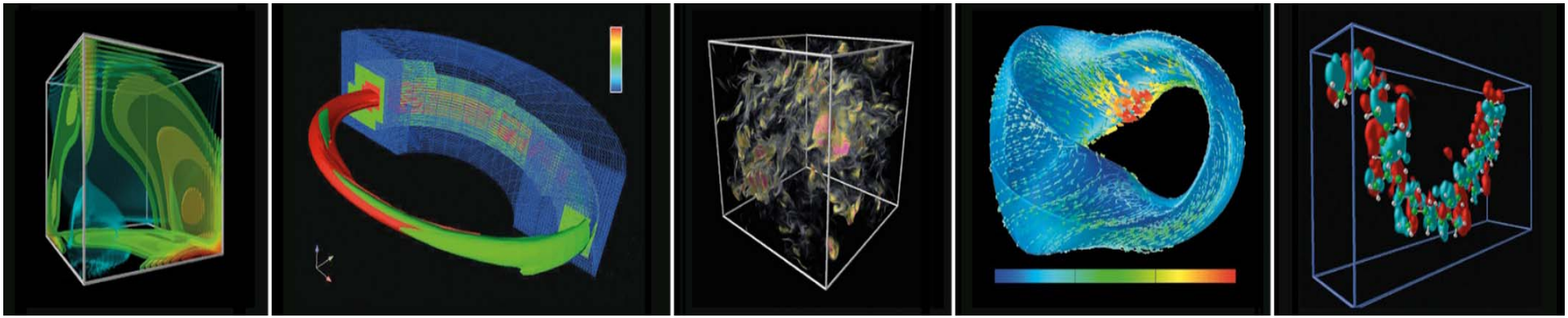
Visualization



Prevention
Detection
Response
Recovery

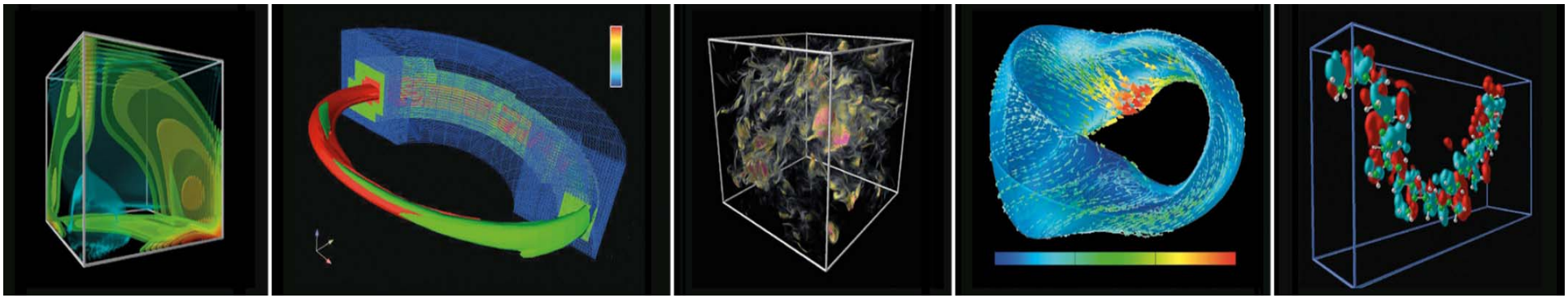


Citizen Cyber-Science



Citizen Cyber-Science

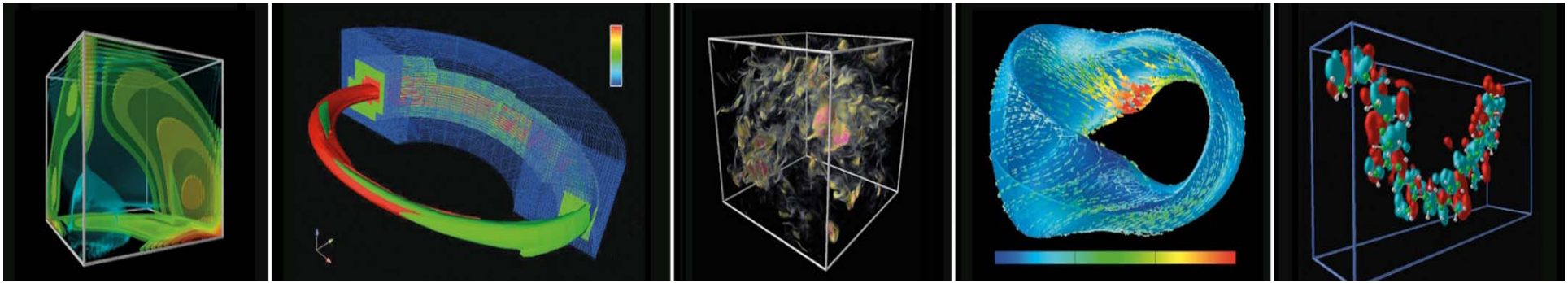
- Distributed thinking / Human computing
  - Aggregate perception, knowledge, reasoning etc.
- Physical participation coordinated via Internet



Citizen Cyber-Science

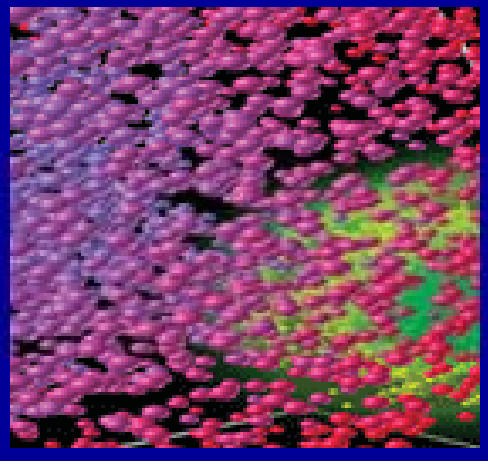
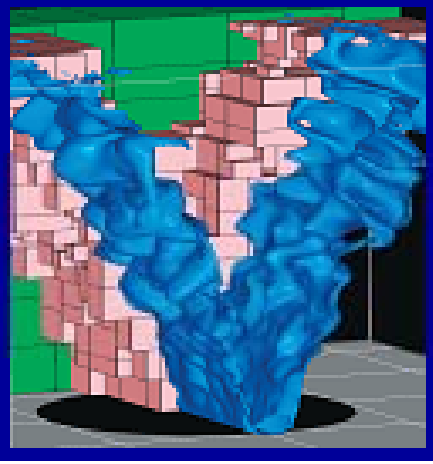
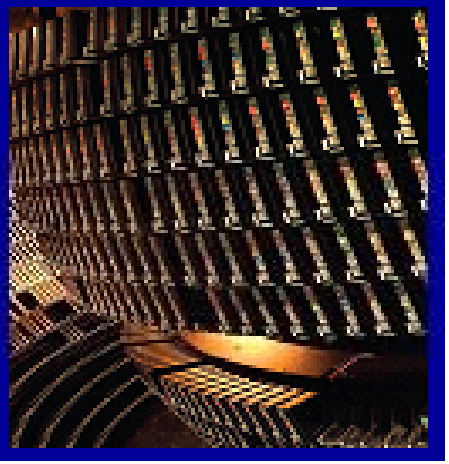
What can be aggregated?

- Visual pattern recognition
- Real-world knowledge
- 3D spatial manipulation
- Language skills



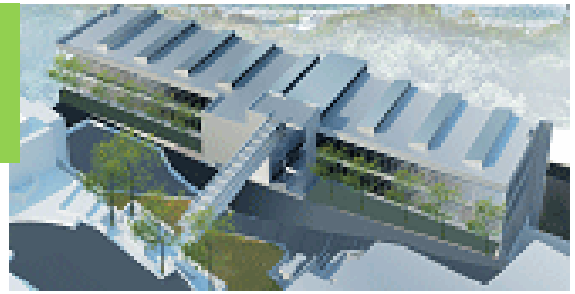
CSE @ CITRIS and LBNL

University of California, Berkeley | CITRIS | Lawrence Berkeley National Lab

CITRIS - LBNL

Multidisciplinary Research and Education
in Computational Science and Engineering



http://www.citris-uc.org

http://www.lbl.gov/cs

http://cse.berkeley.edu/
http://www.citris-uc.org/research/cse
http://www.citris-uc.org/research/cse/calcse
http://www.gov/lbl/cs/cse

CSE Contacts

- **Prof. James Demmel**
  Professor of Mathematics and Computer Science
  Dr. Richard Carl Dehmel Distinguished Professor
  Former Chief Scientist of CITRIS, the Center for Information Technology Research in the Interest of Society
  831 EVANS HALL
  University of California at Berkeley
  Berkeley CA 94720-1776
  Email: demmel@cs.berkeley.edu
  Office: (510) 643-5386
  FAX: (510) 642-3962

- **Prof. Horst Simon**
  Associate Laboratory Director for Computing Sciences
  Professor of Computer Science
  Lawrence Berkeley National Laboratory
  One Cyclotron Road, MS 50B-4230
  Berkeley, California 94720
  (510) 486-7377
  Fax: (510) 486-4300
  HDSimon@lbl.gov

- **Dr. Masoud Nikravesh**
  Executive Director of CSE
  Center for Information Technology Research in the Interest of Society
  291 Hearst Memorial Mining Building
  University of California Berkeley
  Berkeley, CA 94720
  (510) 643-4522
  Fax: (510) 642-1800
  Nikravesh@citris-uc.org