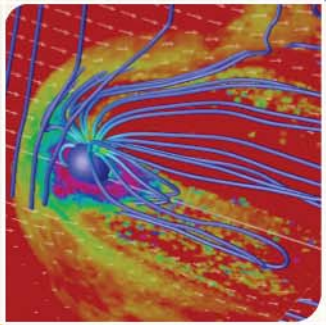
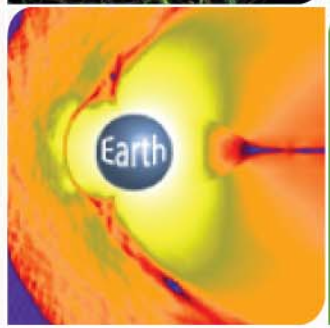
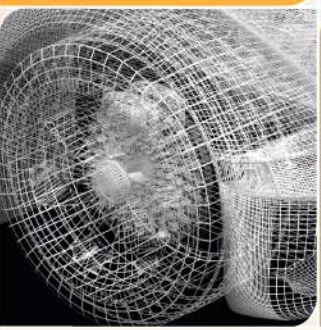


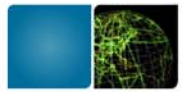


Petascale Science and Engineering:

Georgia Tech's Leadership in the Manyscale™ Transformation

David A. Bader, Executive Director of High-Performance Computing

Georgia Tech College of Computing
Computational Science and Engineering

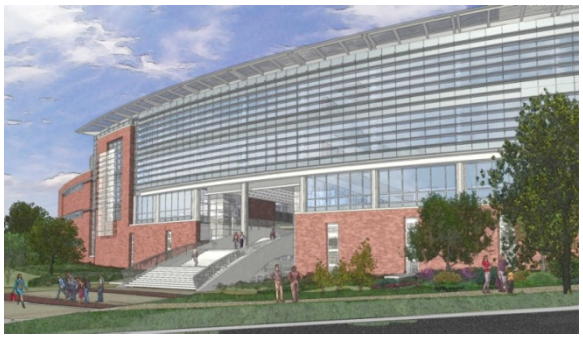




Petascale Applications

- *NSF Workshop Report on Petascale Computing in the Biological Sciences*, David A. Bader, Allan Snavely, Gwen Jacobs, August 29-30, 2006, Arlington, VA.
- *Petascale Computing: Algorithms and Applications*, David A. Bader (ed.), Chapman & Hall/CRC Computational Science Series, © 2007. (ISBN: 9781584889090)

Georgia Tech and Petascale Computing

- Computing Grad Programs Ranked 4th in Nation; and Professor Ranked 1st in World; in an article published in the June '07 issue of *Commun. of the ACM*.
- 6th ranked academic institution in the June 2006 Top100 List
- Over 20TF running on campus today by aggregate in our most capable systems
- HPC resources include approximately 7,000 processors in 35 clusters along with about 100 processors across several SMP systems. Recent HPC systems acq's:
  - IBM System Biology Center system: a 4020-processor IBM eServer BladeCenter with 1,005 blades of 2x2 Opteron cores/blade
  - Dell PowerEdge 1850 system: a 512-node supercomputing cluster with Intel Xeons and InfiniBand interconnect.
  - IBM QS20 Dual Cell/B.E. blades: 1 rack
- Klaus Advanced Computing Building (most advanced computing building in the world!) opened 26 October 2006.



- » IBM Shared University Research (SUR) grant
- » Microsoft Research Faculty Award for parallel programming of multicore processors
- » Planning underway for an HPC Center (50K ft^2 machine room, plus high-quality space for School of CSE).
- » Created a first-class Computational Science & Engineering department

Computational Science and Engineering (CSE)

Georgia Tech

Colleges

- Architecture
- **Computing**
  - Computer Science
  - Interactive Computing
  - **Computational Science and Engineering**
- Liberal Arts
- Management
- Engineering
- Sciences

- High performance computing
- Modeling and Simulation
- Data Analysis & Visualization
- Numerical methods
- Discrete algorithms

- CSE housed in a separate academic unit
- In cooperation with science and engineering

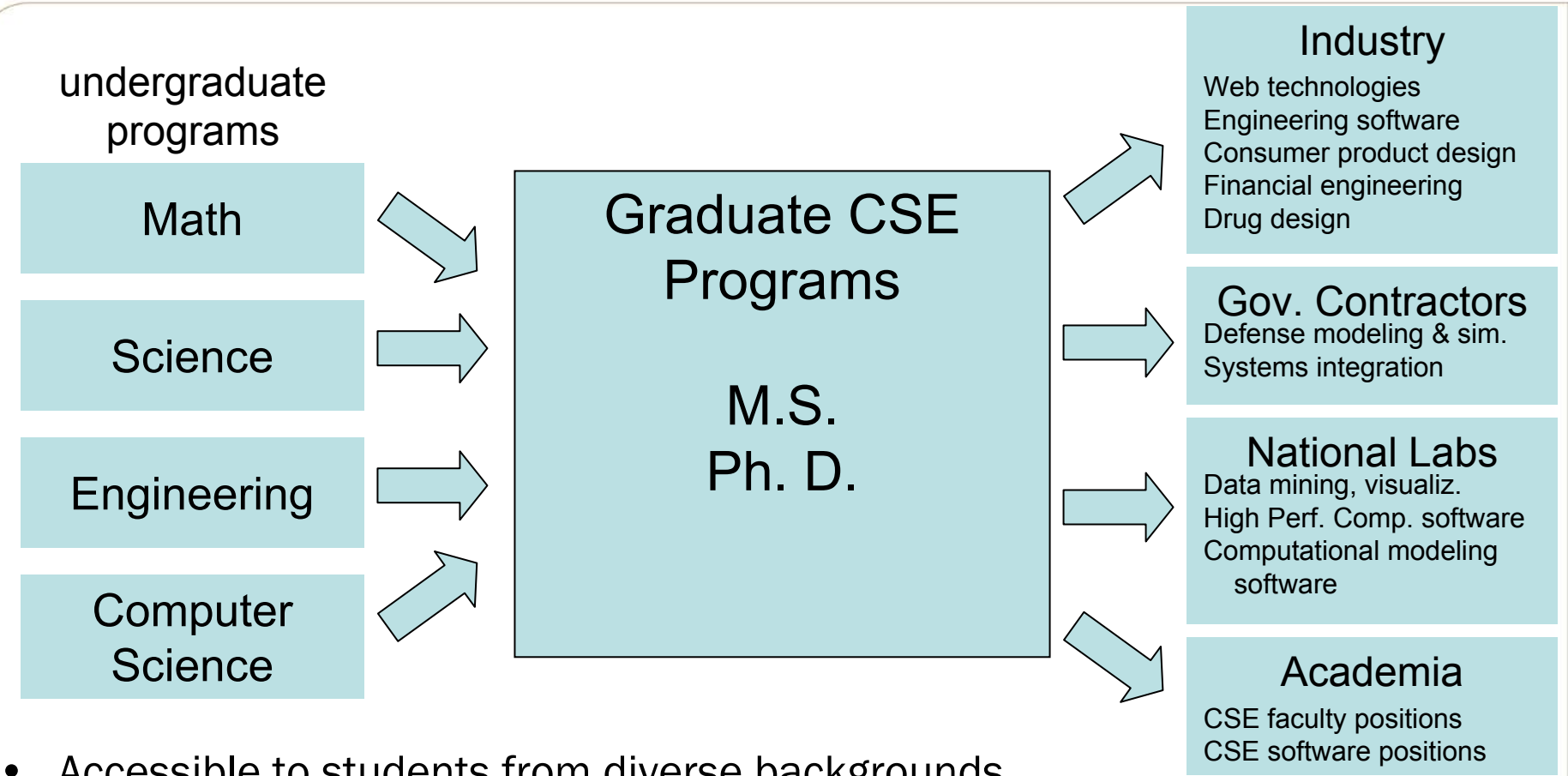
Education Initiatives

Building a pipeline of trained CSE professionals

- Multidisciplinary MS and PhD degree programs in *Computational Science and Engineering**
  - Jointly offered by Computing, Sciences, and Engineering
- Undergraduate "thread" in new CS curriculum
  - Thread in computational modeling
  - Computing core + sciences, engineering

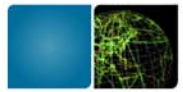
* Pending approval by Board of Regents

Graduate Programs at Georgia Tech



- Accessible to students from diverse backgrounds
  - Mathematics and computing prerequisite
- Near- and long-term demand for graduates
  - Computational models a “must have” in a growing number of fields

Sony-Toshiba-IBM Cell/B.E. Center of Competence @ Georgia Tech

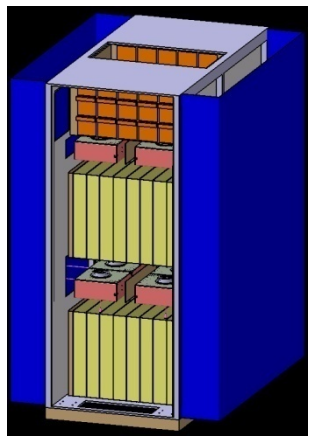

- IBM, SCEI/Sony, Toshiba Alliance formed in 2000
- Austin, TX, Design Center opened in March 2001 (~$400M investment)
- Designed for Sony PlayStation3
- Cell is an extension to IBM Power family of processors
- Sets new performance standards for computation & bandwidth
- High affinity to HPC workloads
  - Seismic processing, FFT, BLAS, etc.
- **November 2006: Georgia Tech wins competition for hosting the STI Center of Competence for the Cell**
- STI Cell/B.E. Workshop, 18-19 June 2007
  - http://sti.cc.gatech.edu

David A. Bader, Center Director

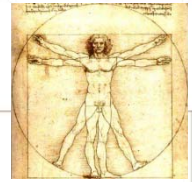
"Georgia, not Austin, gets chip center," Bob Keefe, Austin American-Statesman, November 14, 2006

HPC Systems Research
(Schwan, Wolf, Vetter, Pande, Vuduc, Bader)

- Scalable Systems Software and Runtime Systems
- I/O and Storage Virtualization
- Virtualized Platform Management
- Autonomic Computing
- IBM BlueGene/Q
- IBM Cell Broadband Engine
- SWARM: Manycore programming framework
- Automatic Tuning of Sparse Matrix Kernels



Fluid & Biochemical Transport in Human Systems: From Organs to Cells (Engineering, Sciences, Radiology)



BACKGROUND AND OVER-ARCHING GOALS

Cardiovascular and pulmonary diseases, pathogen entry and drug delivery, are all coupled to transport of species through arteries, heart, and lung.

Our goal is to develop tools to gain insight into

- transport of biochemical species through several scales from cell to the organ level
- cell response to mechanical & chemical stresses,
- *flow* of disease through human body, and
- effective methods for 'targeted' drug delivery

BREAKTHROUGHS

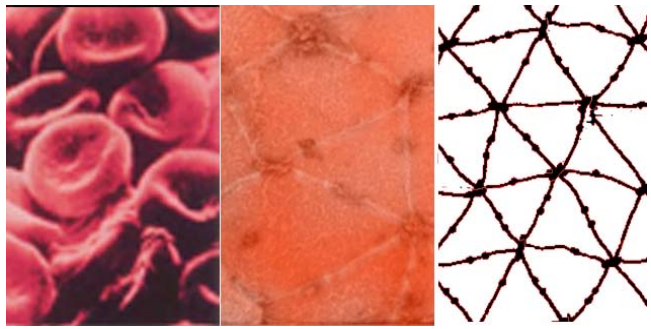
Development of a whole-body model for fluid and biochemical transport, leading to:

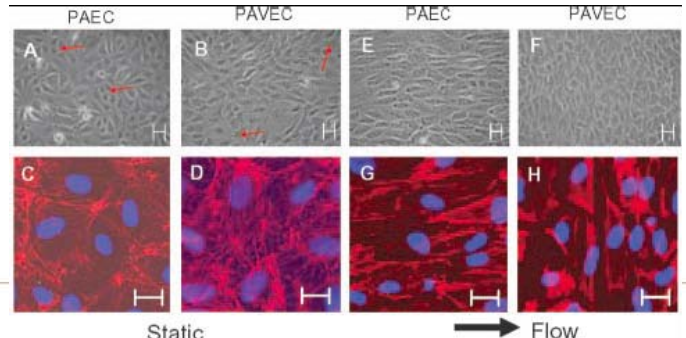
- discovery of cause-and-effect mechanisms for cardiovascular and pulmonary diseases,
- development of predictive tools to assist physicians in corrective procedure, e.g., prediction of flow patterns and consequences in bypass surgery
- insight into transport of disease in the body, such as metastasis, and potential control of spreading
- *targeted* drug delivery, such as progenitor stem cell delivery to sites of cardiovascular injury

METHODS

DNS of blood and air flow in the human system in conjunction with *in vitro* and *in vivo* experiments.

- Automated coupling of large scale high-resolution imaging (e.g., MRI) with *fast and smart* image analysis and mesh generation software
- DNS of blood/air flow with cellular/particulate components and biochemical transport ,
- Modeling and analysis of cell response, and
- *in vivo* and *in vitro* detection of cell response





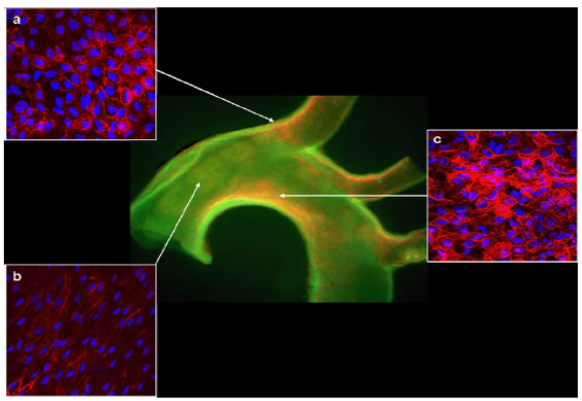
Turbulence and Chaos at Many Scales (Engineering)

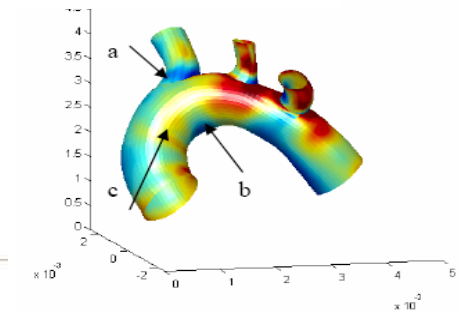
Over-arching Goals (GT-Wide)

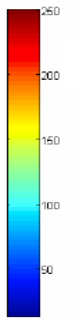
- Understand and model effects of turbulence in interdisciplinary applications (e.g. biology, oceanography, civil and mechanical engineering, aerospace)
  - Impact on environment, life and energy
- Understand Turbulence-Chemistry interactions in multi-phase (gas-liquid-solid) systems: O(8) scales
  - Future clean-burning, and "intelligent" energy production and propulsion systems

Methods

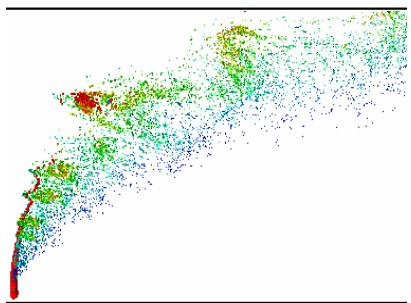
- DNS in canonical flow geometries using pseudo-spectral methods tested on BG Watson:
  - 12,288^3 by 2011 (NSF Track 1), etc.
- Hybrid DNS/LES/LBE methods for resolving physics at many scales with high fidelity
- Multi-scale LES for complex geometries
  - Towards design & construction of systems by new Cyber-enabled paradigms

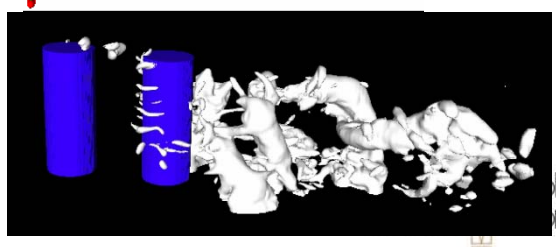


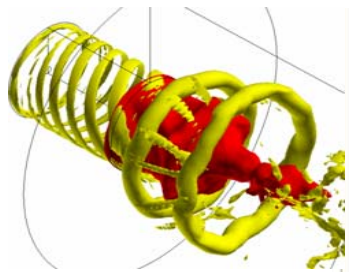


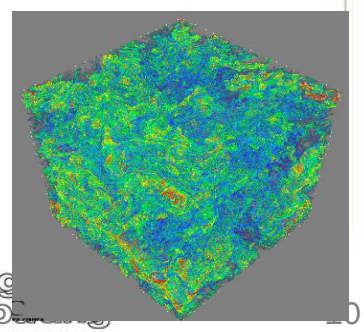


dyne/cm^2







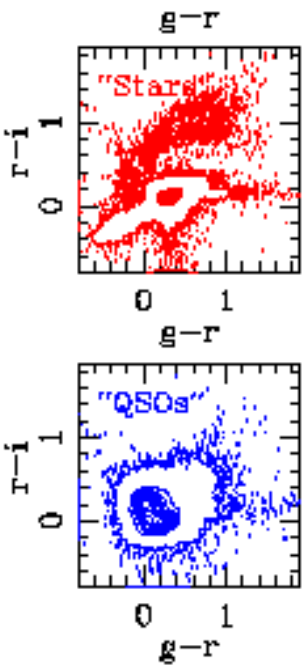


Machine Learning with Massive Data Sets

Quasar detection

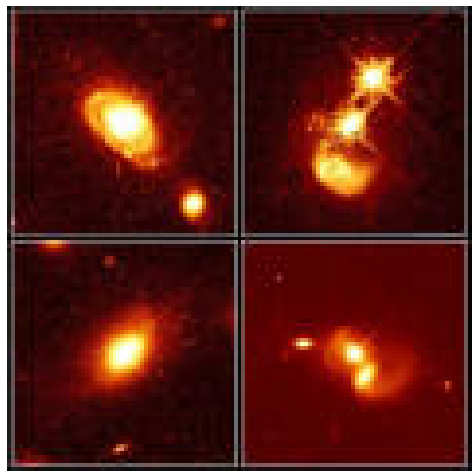
by massive-scale classification

Impact astronomy and astrophysics



Kernel discriminant analysis, SVM/DWD.
particle physics.
Generalized N-body Method

Largest and best quasar catalog to date (by far)
Basis of recent cosmic magnification result confirming relativity (*Nature*, April 05)



Alex Gray, CSE/IIC Professor

Intermolecular Interactions

- Weak interactions govern:
  - protein folding
  - drug binding
  - crystal packing
  - supramolecular assembly of nanostructured materials
- Accurate results require very high-level quantum mechanical treatment
- First converged theoretical result for benzene dimer and crystal benzene

J. Phys. Chem. A. **110**, 10656 (2006)

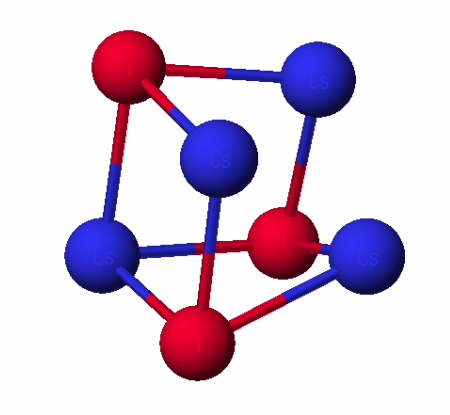
David Sherrill, CHEM/CSE Professor

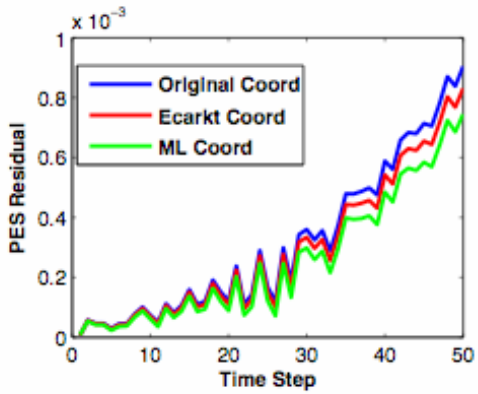
Manifold Learning and Molecular Dynamics Simulations

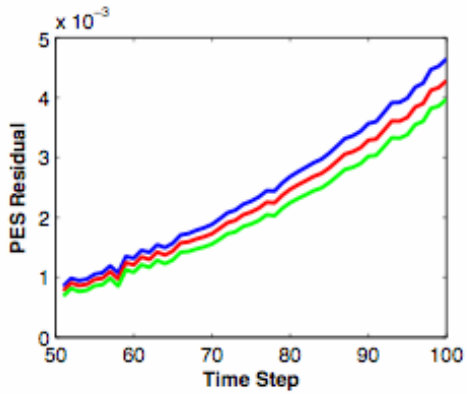
Manifold learning methods for nonlinear dimension reduction (LTSA)

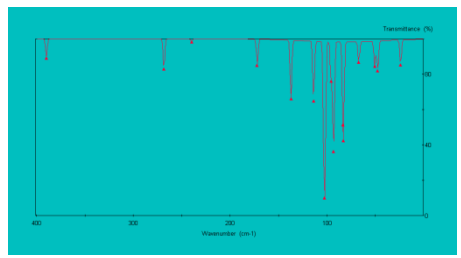
Low-dimensional structures of trajectories from molecular dynamics

Penalized Krieging methods for energy landscape interpolation









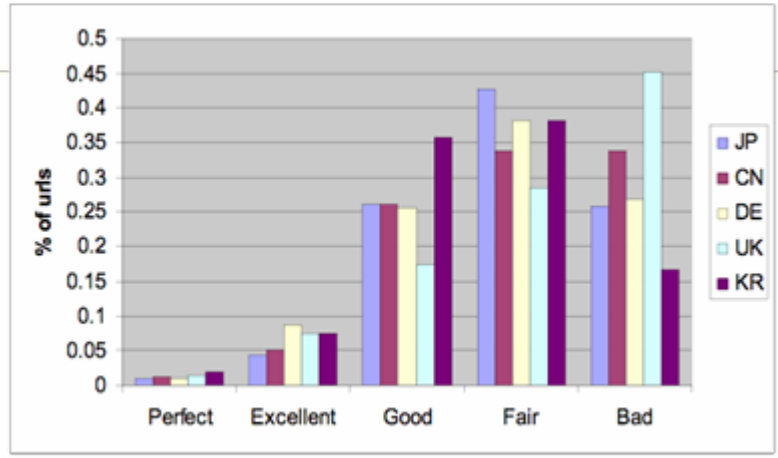
Hongyuan Zha, CSE professor

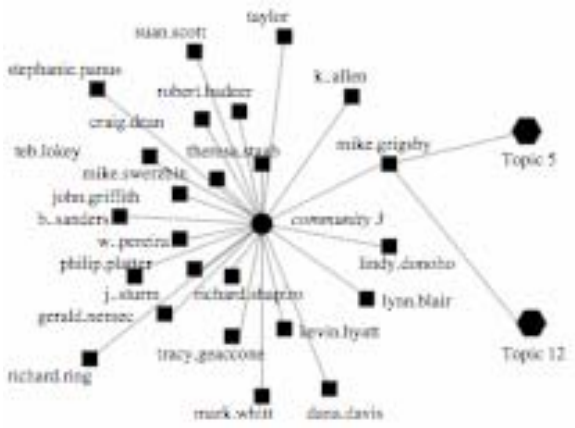
Web Search, Text Mining and Social Computing

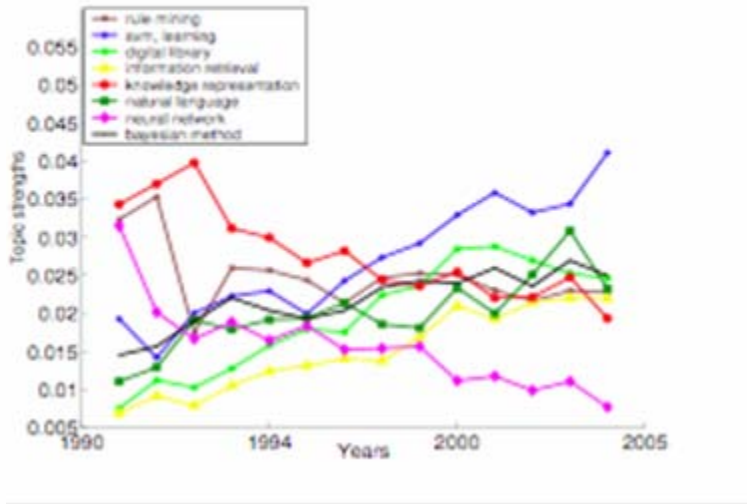
Designing retrieval functions for Web search

Probabilistic models for discovering E communities

Topic evolution and social interactions

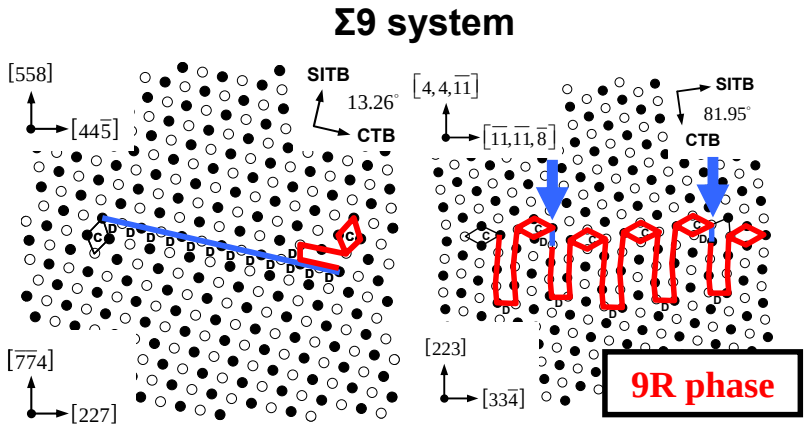




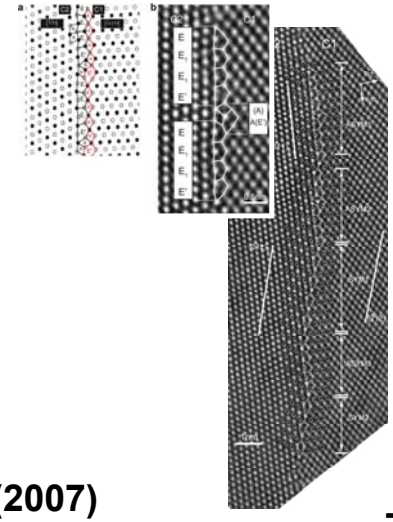


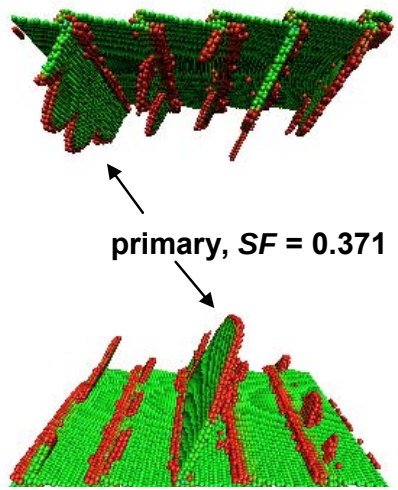
Hongyuan Zha, CSE professor

Atomistic Modeling of Grain Boundaries (McDowell Group)



Tschopp and McDowell, Phil Mag (2007)
Tschopp and McDowell, J.Mat.Sci (2007)





Tschopp and McDowell, IJP (2007)

- Nanocrystalline materials and nanoscale phenomena
- Nanocrystalline Atomistic Simulations of Inelasticity at the Nanoscale
- Bicrystal Atomistic Simulations of Inelasticity at the Nanoscale
- Bicrystal Atomistic Simulations of Σ3 Asymmetric Tilt Grain Boundaries
- Asymmetric Tilt Grain Boundaries in Low Order CSL Systems (i.e., low Σ values)

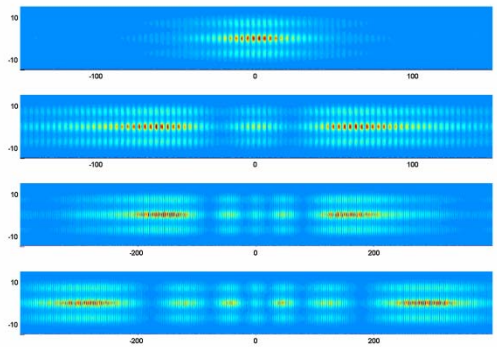
Support: US NSF, NCSA usage

First-Principles Studies of Materials Properties (Mei-Yin Chou, Physics)

State-of-the-Art Electronic-Structure Methods:

- Density Functional Theory
- Quantum Monte Carlo Methods
- Many-Body Perturbation Theory

Excitons in Si nanowire

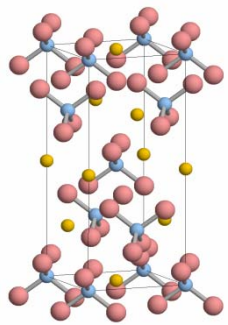


Semiconductor Nanowires (NSF)

Comprehensive simulation of the electrical, optical, vibrational, structural, and transport properties of various nanowires, with the focus on their size dependence.

Complex Hydrides as Hydrogen-Storage Materials (DOE)

Search for new materials with favorable H contents and efficient reversible kinetics



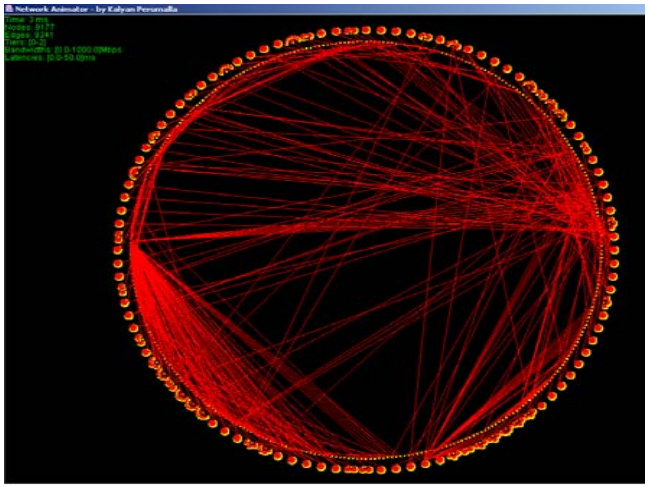
Electronic Structure and Phonons in Graphene

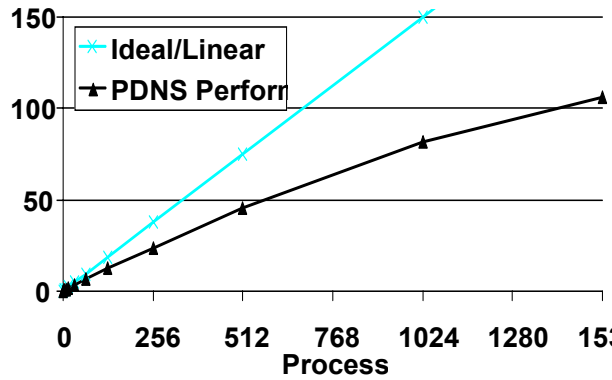
Nature of quasiparticles and magnetic field induced valley splitting

Quantum Size Effects in Metal Thin Films (DOE)

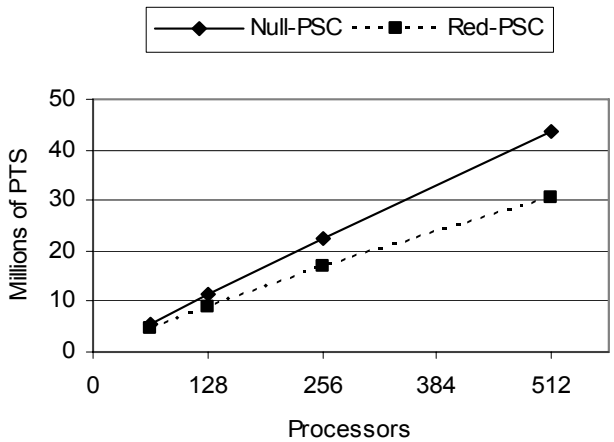
Oscillatory behavior (as a function of thickness) in stability, work function, superconductivity transition temperature, reactivity, and diffusion barrier

Large-Scale Network Simulation (Fujimoto, Perumalla, Riley)





Initial experiments: Good performance up to 1536 processors



Optimized synchronization algorithms yield improved scalability

- Faculty: Richard Fujimoto, Kalyan Perumalla (ORNL), George Riley
- Packet level network simulation
  - Detailed model of TCP/IP stack
  - Up to 4 million network nodes
- Applications
  - Denial of Service attacks
  - Internet worm analysis
  - Protocol performance
- Platform: Lemieux (Pittsburgh)

DARPA

Georgia Tech College of Computing

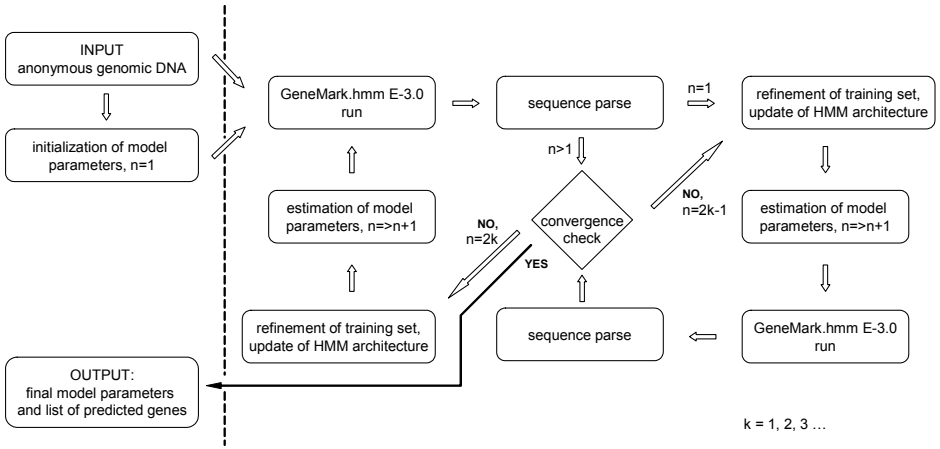
Automatic unsupervised param. est. for gene finding algorithms (Borodovsky Lab, Biology / BME)

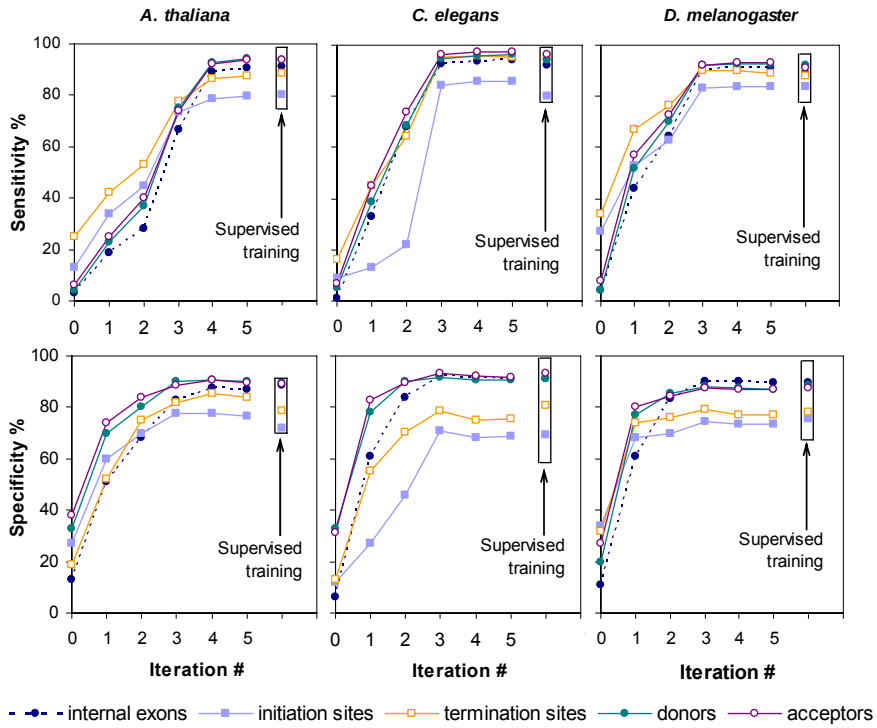
- Prokaryotic genomes

GeneMark.hmm – PS *aka - GeneMarkS*
(Besemer et al. Nucl Acids Res, 2001)

- Eukaryotic genomes

GeneMark.hmm- ES
(Lomsadze et al., Nucl Acids Res, 2005)





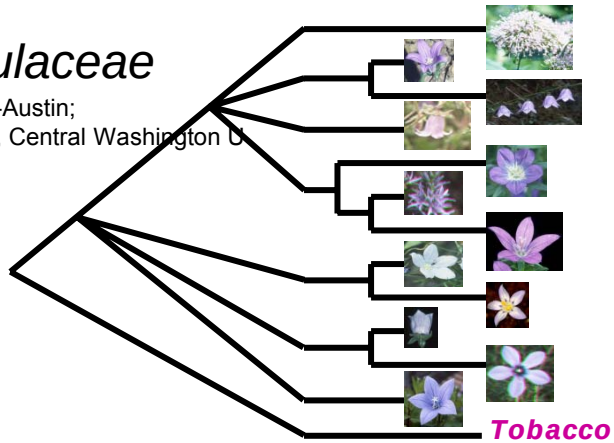
Step-wise diagram of the iterative parameterization of generalized HMM and along with iterative gene prediction (GeneMark.hmm ES)

Gene prediction Sensitivity and Specificity as functions of iteration index.

Note: one iteration requires 30 hours for a human genome (3000 MHz processor)

Lomsadze et al., Nucl Acids Res, 2005

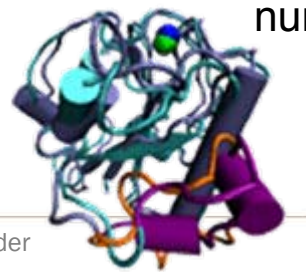
Computational Phylogeny and Genomics (Bader, CSE / Biology)



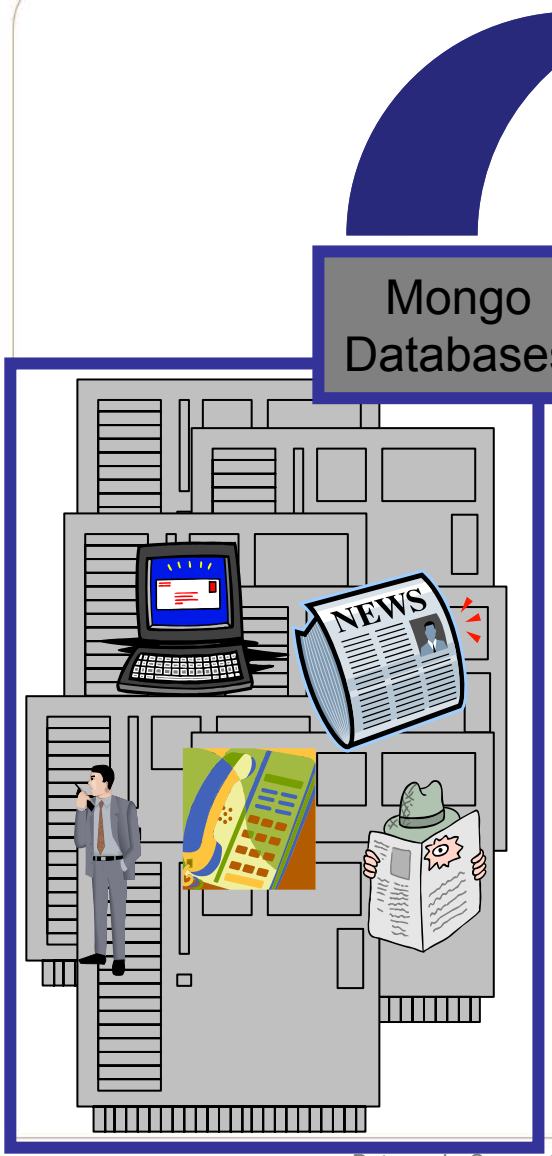
The Tree of Life

CIPRES aims to establish the cyber infrastructure (platform, software, database) required to attempt a reconstruction of the Tree of Life **(10-100M organisms)**

- GRAPPA: Genome Rearrangements Analysis under Parsimony and other Phylogenetic Algorithm
  - Freely-available, open-source, GNU GPL
  - already used by other computational phylogeny groups, Caprara, Pevzner, LANL, FBI, Smithsonian Institute, Aventis, GlaxoSmithKline, PharmCos.
- Gene-order Phylogeny Reconstruction
  - Breakpoint Median
  - Inversion Median
- over one-billion fold speedup from previous codes
- Parallelism scales linearly with the number of processors



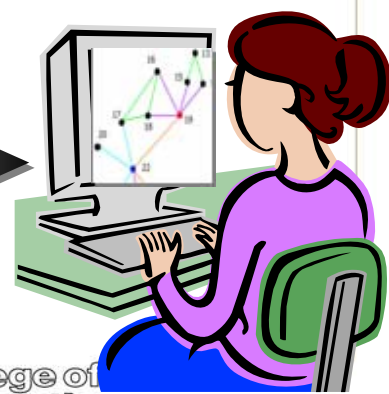
Informatics from Massive Semantic Graphs (Bader, Zha, Gray)



Graph resides in memory of supercomputer.

Fast Graph Query

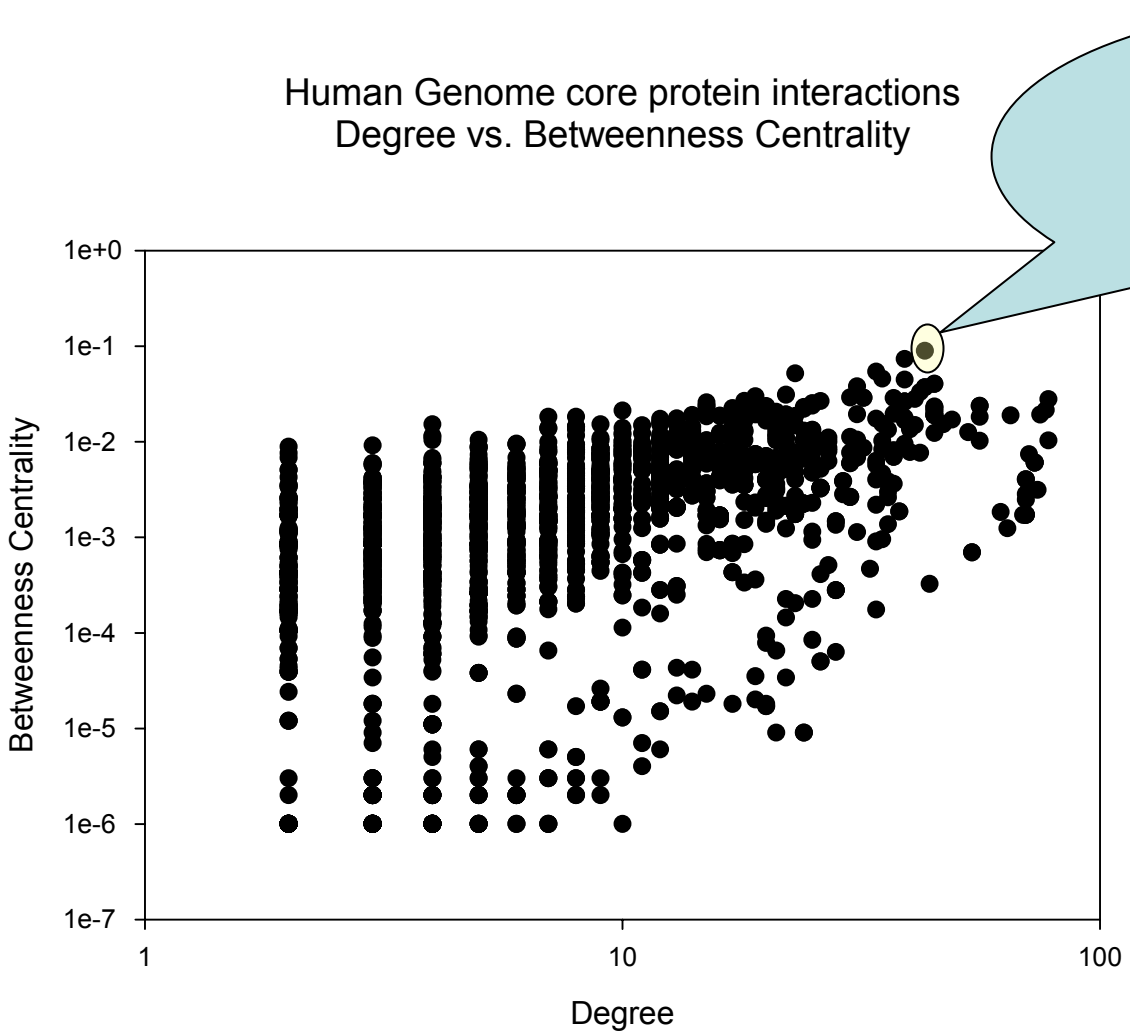
Analyst makes queries.



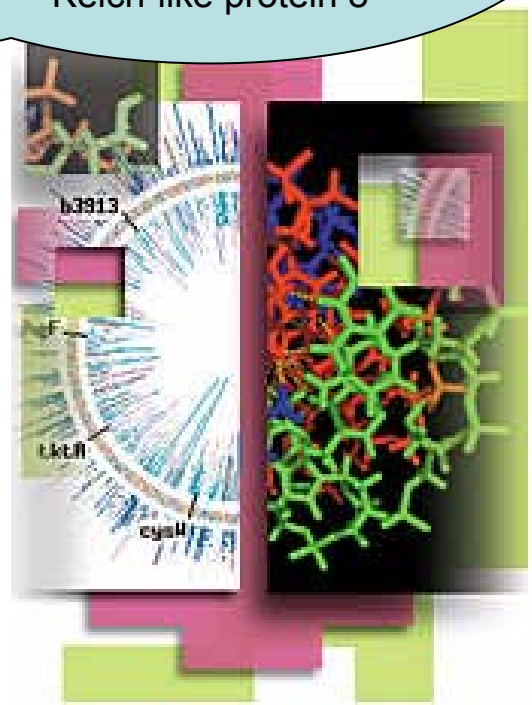
Extract "Window"

High Latency Query

Betweenness Centrality Analysis: Protein-protein interactions



43 interactions
Protein Ensembl ID
ENSG00000145332.2
Kelch-like protein 8



David A. Bader, Kamesh Madduri

Acknowledgment of Support

- National Science Foundation
  - **CSR:** A Framework for Optimizing Scientific Applications (06-14915)
  - **CAREER:** High-Performance Algorithms for Scientific Applications (06-11589; 00-93039)
  - **ITR:** Building the Tree of Life -- A National Resource for Phyloinformatics and Computational Phylogenetics (EF/BIO 03-31654)
  - **ITR/AP:** Reconstructing Complex Evolutionary Histories (01-21377)
  - **DEB** Comparative Chloroplast Genomics: Integrating Computational Methods, Molecular Evolution, and Phylogeny (01-20709)
  - **ITR/AP(DEB):** Computing Optimal Phylogenetic Trees under Genome Rearrangement Metrics (01-13095)
  - **DBI:** Acquisition of a High Performance Shared-Memory Computer for Computational Science and Engineering (04-20513).
- IBM PERCS / DARPA High Productivity Computing Systems (HPCS)
  - DARPA Contract NBCH30390004
- IBM Shared University Research (SUR) Grant
- Sony-Toshiba-IBM (STI)
- Microsoft Research
- Sun Academic Excellence Grant
- DOE: Oak Ridge, Sandia