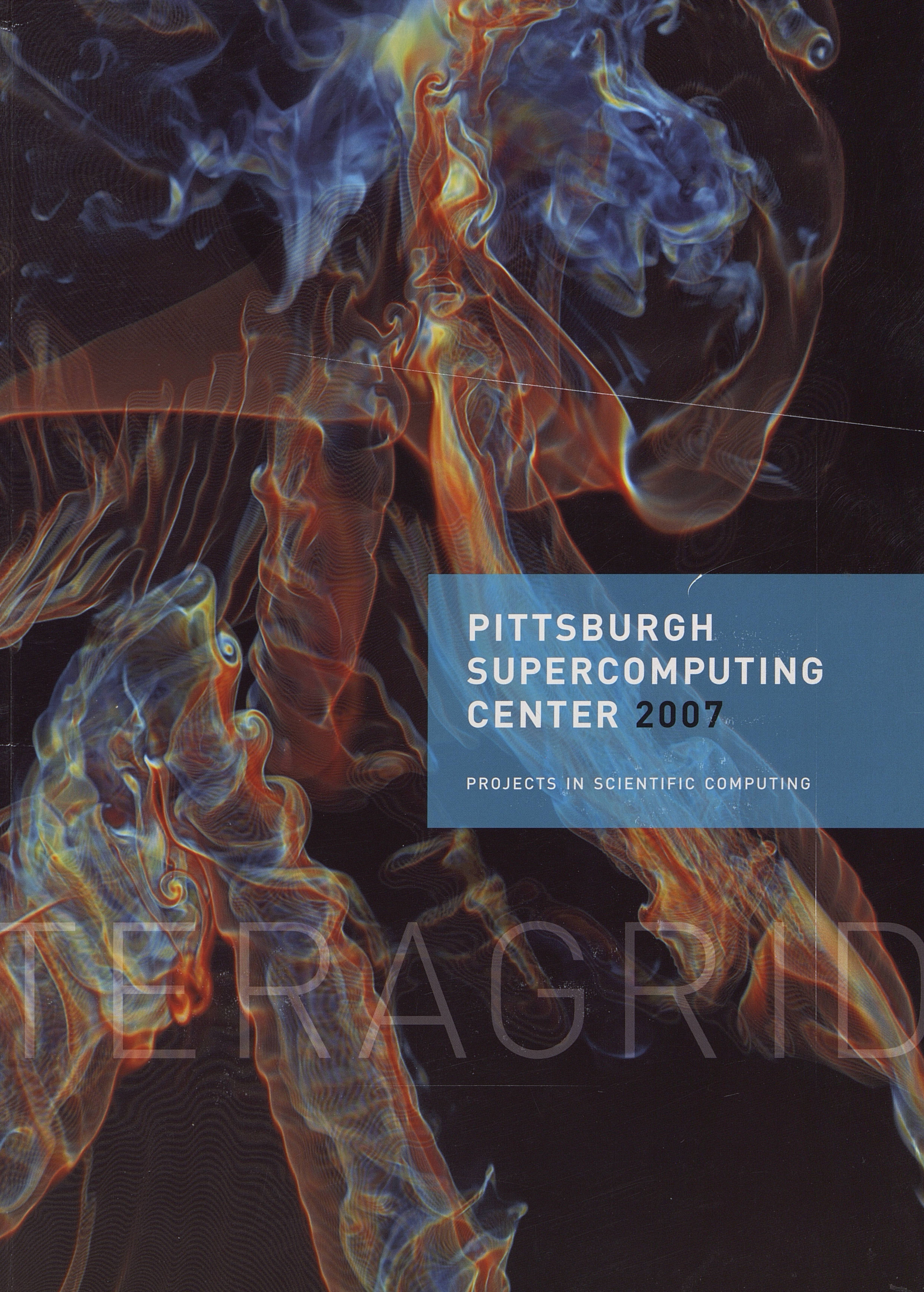


PITTSBURGH SUPERCOMPUTING CENTER 2007

PROJECTS IN SCIENTIFIC COMPUTING

TERAGRID

PSC.EDU/07

PITTSBURGH SUPERCOMPUTING CENTER 2007

The Pittsburgh Supercomputing Center provides university, government and industrial researchers with access to several of the most powerful systems for high-performance computing, communications and data-handling available to scientists and engineers nationwide for unclassified research. PSC advances the state-of-the-art in high-performance computing, communications and informatics and offers a flexible environment for solving the largest and most challenging problems in computational science. As a leading partner in the TeraGrid, the National Science Foundation's program to provide a coordinated national cyberinfrastructure for education and research, PSC works with other TeraGrid partners to harness the full range of information technologies to enable discovery in U.S. science and engineering.

WWW.PSC.EDU // 412-268-4960

FOREWORD FROM THE DIRECTORS →

Ralph Roskies and **Michael Levine**
PSC co-scientific directors

We're pleased to once again highlight research at the Pittsburgh Supercomputing Center. This year's report emphasizes the PSC staff and their critical role in enabling people outside PSC to use our technologies to conduct transformative research both efficiently and successfully.

Those of us absorbed in this work often take for granted what others may not fully appreciate — we are involved in an inherently transformative undertaking. As the microscope opened a huge unexplored realm to understanding, the technologies of supercomputing are a new lens for seeing the world and opening it to human understanding.

The results are manifest. Numerical weather prediction has since the 1960s completely transformed meteorology. When PSC set up shop twenty years ago, however, it was still taken as faith that weather at the scale of individual storms was inherently unpredictable. Our work at PSC in collaboration with the CAPS team at Oklahoma University has transformed this thinking. This spring we supported CAPS and NOAA in an unprecedented forecasting experiment (pp. 18-21) that built on our prior collaborations and took another very big step toward realizing storm-prediction technologies that will save lives.

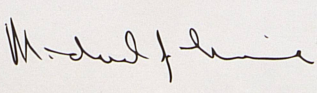
We are doing critical work in the development of "clean" power. Our efforts in founding the Supercomputing Science Consortium (p. 8) and our collaboration with the U.S. Department of Energy's National Energy Technology Laboratory have paid off in the development of coal-gasification technology (pp. 22-25) for a 285-megawatt power plant that will be, when it comes on-line in 2010, the cleanest and most efficient coal-fired power plant in the world.

PSC's National Resource for Biomedical Supercomputing (p. 12), supported by the NIH's National Center for Research Resources, is a national leader in training life-science researchers in the use of computational tools. Two projects reported here — a pioneering simulation of protein-induced membrane curvature (pp. 26-29) and a surprising new insight into the mechanism of an important enzyme (pp. 30-33) — exemplify the breathtaking progress in biological understanding made possible by information technology.

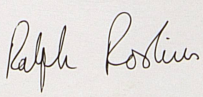
Petascale computing — tens or hundreds of thousands of processors teamed to solve the most difficult problems — will soon be a reality. PSC staff are at the forefront of the know-how, algorithms and software that will be needed to exploit petascale computing. We are involved in a number of research collaborations — in cosmology, earthquake science and others — to hasten this progress. The turbulence work of Paul Woodward and David Porter (pp. 34-37), relying on the exceptional inter-processor bandwidth of our Cray XT3, is a powerful example of how these advances will transform science. As this article explains, these exciting developments would not have been possible without PSC staff.

Cosmology and astrophysics is exploding with data from new observations, and computational simulations are essential to gain meaningful perspective within this deluge of information. Here we report on the important work of Kazantzidis and Mayer (pp. 38-41) that has forced new thought about the "missing galaxy problem" of the cold dark matter model of the universe.

This publication is testament to the creative skill, dedication and experience of the people who are PSC. We are proud to join them in the work we do and grateful for the support we receive from the National Science Foundation, the U.S. Department of Energy, the National Center for Research Resources of the National Institutes of Health, the Commonwealth of Pennsylvania and many others.



Michael Levine Scientific Director



Ralph Roskies Scientific Director

TABLE OF CONTENTS

Creating National Cyberinfrastructure: PSC & TeraGrid

CREATING CYBERINFRASTRUCTURE

The TeraGrid is the world's most comprehensive distributed cyberinfrastructure for open scientific research. As a major partner in this National Science Foundation program, PSC helps to shape the vision and progress of the TeraGrid.

BIGBEN DOUBLED

BigBen, PSC's Cray XT3, the first Cray XT3 anywhere, is the TeraGrid's lead performer among "tightly coupled" architectures. In late 2006, PSC more than doubled BigBen's capability, replacing existing processors of the 2,090-processor system with Opteron's top-end, dual-core (2.6 GHz) chip. The upgrade doubled the processor count to 4,180 and boosted peak performance to 21.5 teraflops, while also doubling memory (from two to four terabytes).

BigBen's primary advance is its superior inter-processor bandwidth, the speed at which processors share information. This is a large advantage for projects that demand hundreds or thousands of processors working together. BigBen has demonstrated performance 10 times or more better than prior tightly coupled systems on many applications and is a champion at "scaling" — the ability to use a large quantity of processors without seriously reducing the per-processor performance, (e.g., "Bursts of Stellar Turbulence," pp. 34-37).

"The Cray XT3 has proven itself as a scientific platform of exceptional capability," said PSC scientific directors Michael Levine and Ralph Roskies. "Since becoming a production resource on the TeraGrid, this new system has made possible a number of remarkable achievements."

DAVID MOSES, PSC EXECUTIVE DIRECTOR

In September, David Moses, co-founder and former chief operating officer of Gaussian, Inc., joined PSC as executive director, managing day-to-day internal operations and overseeing a scientific and technological staff of about 75 people. His hiring culminated an extensive national search.

"We are very pleased that David will help us to carry forward our leadership in high-performance scientific computing," said PSC scientific directors Michael Levine and Ralph Roskies. "His energy, enthusiasm, solid judgment, and extensive experience in the organizational as well as the technical aspects of large-scale, collaborative computational science will catalyze our efforts and benefit the national community of scientists with whom we work."

← **Jim Kasdorf,**
PSC director of special projects

PSC AND TERAGRID

PSC is actively involved in TeraGrid leadership. Scientific director Ralph Roskies serves on the executive steering committee of the Grid Infrastructure Group that guides TeraGrid, and co-scientific director Michael Levine is PSC principal contact as one of 11 TeraGrid resource-providers.

Other PSC staff also have taken leadership roles in TeraGrid. As Area Director for User Support, PSC's Sergiu Sanielevici manages TeraGrid's user-support services and coordinates the ASTA program — Advanced Support for TeraGrid Applications. Jim Marsteller, who heads PSC's security, chairs the TeraGrid Security Working Group, responsible for risk assessments and incident response. PSC's Laura McGinnis co-leads the TeraGrid team that is developing an education, outreach and training program called "HPC University," and will chair poster sessions at the TeraGrid '08 conference. PSC director of systems and operations, J. Ray Scott, leads the TeraGrid effort in Data Movement.

Other recent PSC work in support of TeraGrid includes:

- **Securing Community Accounts:** In a major effort, PSC systems engineer Aaron Shelmire implemented a security model, adapted from NCSA-developed software, that reconciles the community-wide reach of TeraGrid Science Gateways with the secure environment of a large-scale system. TeraGrid sites are consulting with PSC for details they can emulate.
- **CTSS Compliance:** PSC grid specialist Derek Simmel identified a problem in TeraGrid's Coordinated TeraGrid Software and Services (CTSS) that obstructed user authentication. He devised a fix, solicited comment from involved working groups, and created a change plan implemented via modification of each resource provider's CTSS configuration.
- **Networking:** The TeraGrid Data Working Group recommended that all TeraGrid sites deploy HPN-SSH, a security protocol for network communications with performance enhancements — developed by PSC's Chris Rapier. Rapier coordinated to assure appropriate installations. Ben Bennett and others of PSC's network staff added "multi-threading" to OpenSSH's encryption operations, resulting in 40-percent speedup in transfer rate. PSC staff also deployed a TeraGrid version of its NPAD diagnostic service, co-developed with NCAR, which analyzes and diagnoses network path-failures, on TeraGrid's network-monitoring computers.
- **Speedpage:** PSC staff updated "speedpage," a PSC-created TeraGrid resource that measures file-transfer performance among TeraGrid sites. This includes a naming convention that PSC staff designed and shepherded into place. PSC is migrating speedpage to a new PSC information server, where it will perform better and share its database with other performance-measurement tools.
- **File Systems:** Experienced in file systems across a range of architectures, PSC staff worked with Indiana University to implement the flexible, open-source Lustre-WAN file system, deployed on PSC's TeraGrid resources, on test machines at Indiana. Performance in writing data between PSC and Indiana is only 20-percent reduced compared to similarly configured local file systems.

TERAGRID RESOURCE PROVIDERS

TeraGrid™

Indiana University

Louisiana Optical Network Initiative

National Center for Supercomputing Applications

National Center for Atmospheric Research

National Institute for Computational Sciences

Oak Ridge National Laboratory

Pittsburgh Supercomputing Center

Purdue University

San Diego Supercomputer Center

Texas Advanced Computing Center

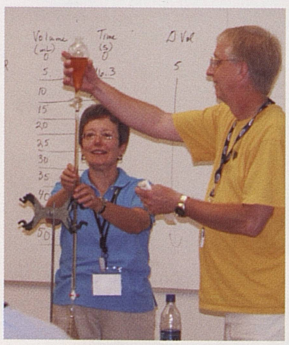
The University of Chicago/Argonne National Laboratory

↑ PSC STAFF WHOSE WORK CONTRIBUTES TO TERAGRID (l to r): **Laura McGinnis, Michael Schneider** (seated), **Nathan Stone, Josephine Palencia, Sergiu Sanielevici, Shandra Williams, Kathy Benninger, Rob Light, David O'Neal** (on floor), **R. Reddy, Derek Simmel, Rich Raymond, Jim Marsteller.**

SUPERCOMPUTING IN PENNSYLVANIA

With Commonwealth of Pennsylvania support, PSC provides education, consulting, advanced network access and computational resources to scientists and engineers across the state.

Cheryl Begandy, outreach manager, and **Pallavi Ishwad**, education outreach specialist, lead PSC's strong education and outreach programs.



K-12 SCIENCE EDUCATION

PSC's work to help prepare a more technology-ready workforce and science-literate populace thrived over the past year. In December 2006, the Heinz Endowments Education Program awarded $150,000 to support PSC's highly successful CAST (Computation and Science for Teachers) program. The award will support additional planning and a thorough program evaluation. CAST is a two-year program for teachers and school administrators that meets quarterly to introduce computational thinking and tools such as modeling and simulation into the high-school science curriculum. This year, 23 teachers from 14 school districts participated.

This year PSC initiated a second program, CMIST (Computational Modules in Science Teaching). Developed by PSC's National Resource for Biomedical Supercomputing, CMIST creates complete teaching modules in a subject area based on recent research at PSC. (More on CMIST, p. 14.) Together, CAST and CMIST offer an approach to teaching secondary science that includes both specific computational modules (CMIST) and a general framework for computational science disciplines (CAST).

PITTSBURGH STUDENTS ARE SCIENCE WINNERS AT TERAGRID '07

Pittsburgh ninth-graders Shivam Verma and Molly Joyce were winners in the TeraGrid '07 student science competition, June 4-7 in Madison, Wisconsin. From February to mid-March, a series of three "Jumpstart" seminars in computational science, spear-headed by PSC's Laura McGinnis, helped to spark the interest of these students. (l to r) Prabha Shanker Verma, Shivam's father; Shivam Verma; Pallavi Ishwad, PSC education outreach specialist; Molly Joyce; Edward Joyce, Molly's father.

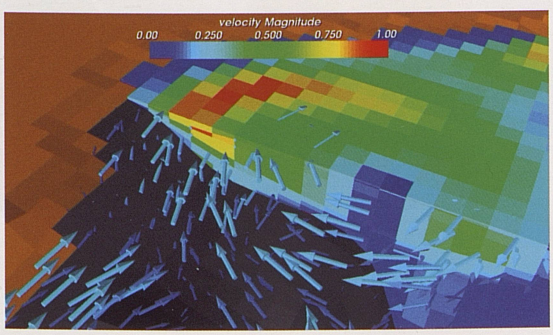
RESOURCES FOR PRIVATE-SECTOR RESEARCH

Through SC² (p. 8) and the PSC networking group, Evergreene Technology Park in Greene County provides companies with access to PSC resources, and a number of Pittsburgh-area corporations use PSC resources in product development efforts, including:

- **PPG Industries,** a global supplier of coatings, glass, fiberglass and chemicals, uses PSC systems for computational modeling of PPG compounds and in development of new products, including Transitions lenses, CeramiClear automotive clear coats and SunClean self-cleaning glass. "Our approach is to use modeling to reduce discovery and development cycle times," says Michael Makowski, who leads PPG's computational chemistry group, "and to gain fundamental understanding of our core technologies and competencies, gain a competitive advantage, and ultimately reduce costs."
- **Medrad, Inc.** of Indianola, Pennsylvania has collaborated with PSC and Carnegie Mellon to develop a novel method for safe and efficient removal of deep-vein blood clots. PSC and CMU expertise in computational fluid dynamics supported feasibility tests on the physics of Medrad's catheter device. "Using the PSC supercomputer," says John Kalafut, Medrad's senior research scientist, "we have been able to look at multiple iterations of different design parameters without building numerous, expensive prototypes."
- **NanoLambda, Inc.** of Pittsburgh, a spin-off from the University of Pittsburgh, is developing the Spectrum Sensor, a chip-scale Optical Spectrum Analyzer. Its high-resolution sensing capability enables not only mobile/wearable health monitors, such as a non-intrusive glucose monitor, but also high-standard RGB color sensors. NanoLambda has done feasibility studies with PSC systems and will use them for computationally intensive physical-optics simulations of its optically tailored devices.

CARNEGIE MUSEUM OF NATURAL HISTORY

For the three-day opening of the Carnegie Museum of Natural History's "Night of the Titanic" exhibit, PSC created a visualization of ocean currents in the North Atlantic mapped with the route of the Titanic. "Your support of Carnegie Museum of Natural History's Climate Change Weekend helped make it a success," said Kerry Handron, Earth Theater director. "The graphic depiction of the ship's path crossing the Gulf Stream emphasized one of the main points of the show." The PSC film screens regularly in CMNH's Earth Theater.



OUTREACH & TRAINING

PSC presented the sixth in a series of annual technology-briefings for **Bechtel Bettis Laboratory**, expanded this year to a two-day format. Presentations focused on managing and using a parallel distributed computing environment, including hardware, cluster architecture, visualization and grid computing. Twenty people from Bechtel Bettis participated in the two-day session with PSC staff and reported in evaluations that it was very valuable.

PSC participated in several community events to help the general public learn about supercomputing. At the **SciTech Spectacular** at the Carnegie Science Center in October 2006, which fosters understanding of science and technology among middle and high school students, PSC was one of over 30 exhibitors. PSC also participated in the **Spring Fling** held at the Pittsburgh Public School's Gifted Center in May 2007. At this interactive educational event for the district's grade two through eight gifted students, PSC presented, among other exhibits, a life-size BigBen model and a poster showing nano-scale problems.

RESEARCH AT PENNSYLVANIA COLLEGES & UNIVERSITIES, 2006-2007

From July 1, 2006 through June 30, 2007, 720 Pennsylvania researchers at these institutions used 5.5-million processor hours on PSC resources.

Bucknell University
Cabrini College
Carnegie Mellon University
Drexel University
Duquesne University
Edinboro University of Pennsylvania
Lehigh University
Pennsylvania State University, All Campuses
University of Pennsylvania
University of Pittsburgh, All Campuses
Ursinus College
Villanova University
Widener University

THE SUPER COMPUTING SCIENCE CONSORTIUM

Pennsylvania-West Virginia partners in development of clean power technologies.

Formed in 1999 and supported by the U.S. Department of Energy, the Super Computing Science Consortium is a regional partnership of research and educational institutions in Pennsylvania and West Virginia. (SC)² provides intellectual leadership and advanced computing and communications resources to solve problems in energy and the environment and to stimulate regional high-technology development and education.

Through (SC)², Evergreene Technology Park in Greene County provides a channel for companies to collaborate with local universities in southwest Pennsylvania and West Virginia and to have access to PSC computing resources.

Since the spring of 2000, a high-speed network — the first fiber-optic service to Morgantown, West Virginia — has linked the National Energy Technology Laboratory (NETL) campuses in Morgantown and Pittsburgh with PSC, facilitating NETL collaborations. Researchers at NETL and WVU have actively used this link to tap PSC computational resources. Since the founding of (SC)², 50 (SC)² researchers have used PSC systems for a range of projects, using more than 4.5-million hours of computing time, nearly 700,000 hours within the past year.

This work includes:

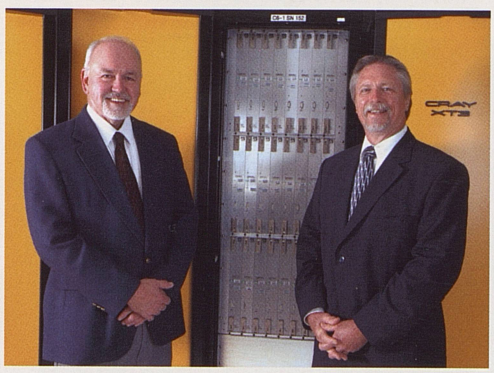
- **Clean Liquid Fuel from "Syngas"**
  http://www.psc.edu/science/2006/sc2/
- **Fuel-Quality Hydrogen from Fossil Fuels**
  http://www.psc.edu/science/2005/sc2/
- **Gas from Black Liquor**
  http://www.psc.edu/science/2004/sc2/
- **Fluidized-Bed Combustion of Silane**
  http://www.psc.edu/publicinfo/netl/
- **Lean-Fuel Mixes in Next-Generation Power-Generating Turbines**
  http://www.psc.edu/science/Richards/clean_power.html
- **A New Design for a Power-Generating Turbine**
  http://www.psc.edu/science/cizmas2002.html

(SC)² Super Computing Science Consortium

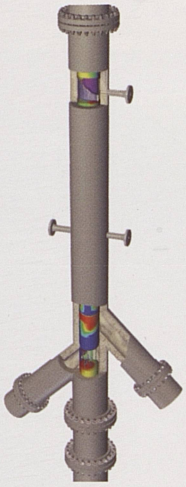
(SC)² PARTNERS

National Energy Technology Laboratory
Pittsburgh Supercomputing Center
Carnegie Mellon University
Duquesne University
University of Pittsburgh
Waynesburg University
West Virginia University
Institute for Scientific Research
NASA Independent Verification & Validation Facility
The West Virginia Governor's Office of Technology

MORE INFORMATION: http://www.sc-2.psc.edu



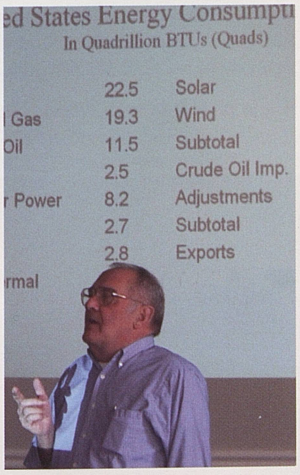
↑ **Lynn Layman**, PSC & **Bob Romanowsky**, NETL
(SC)² co-chairs



TRANSPORT GASIFICATION: CLEAN ENERGY FROM COAL

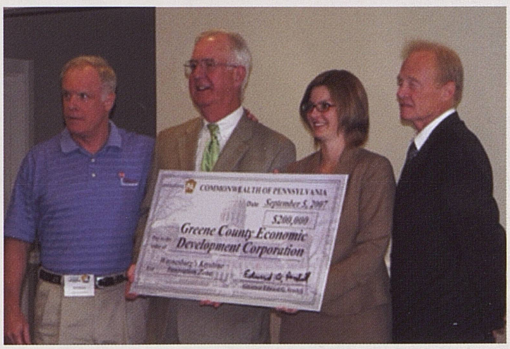
Among the projects (SC)² has made possible is an extensive series of simulations of a coal-gasification technology known as a "transport gasifier" (see pp. 22-25). This work, with NETL-developed software called MFIX (Multiphase Flow with Interphase Exchanges) and PSC's LeMieux and BigBen systems, was instrumental in scaling up the transport gasifier for a 285-megawatt power plant now under construction near Orlando, Florida. When completed, scheduled for 2010, this plant will be the cleanest, most efficient coal-fired power plant in the world.

OVERVIEW: ENERGY SOURCES AND CONSUMPTION



Retired NETL scientist Gerst Gibbon addressed the (SC)² December 2006 meeting on the topic of fossil fuel's centrality in the global and U.S. economy. In a comprehensive presentation, Gibbon, who holds a Ph.D. in chemistry from Carnegie Mellon University, discussed pros and cons of many energy sources, including ethanol, nuclear power, hydrogen and wind. The world's economies, he noted, are largely committed to petroleum and natural gas as their primary fuel sources. World reserves of these fuels are likely to be exhausted this century, which will cause significant economic disruptions in most developed countries.

Pennsylvania Representative **Bill DeWeese** (Democrat), Fayette County (part), Greene County; **Ralph Sommers**, chair, Greene County Industrial Development Corp.; **Rebecca Bagley**, Pennsylvania Department of Community and Economic Development; **Tim Thyreen**, president, Waynesburg University.



KEYSTONE INNOVATION ZONE

Through (SC)², PSC this year helped to prepare a successful grant application from Waynesburg University, the Greene County Industrial Development Authority and others for Pennsylvania's Keystone Innovation Zone (KIZ) program. Funding of $200,000 from the Ben Franklin Technology Development Authority of Pennsylvania Department of Community and Economic Development will spur new technology and economic development in Greene County. Lynn Layman of PSC, co-chair of (SC)², serves on the Waynesburg KIZ advisory board.

WAYNESBURG UNIVERSITY LINKS TO INTERNET2

Through its 3ROX high-performance network hub, PSC's networking group this year provided Internet2 connectivity for (SC)² member Waynesburg University. This advanced infrastructure opens new opportunities for research and education at Waynesburg. On May 15, PSC staff visited Waynesburg for a daylong event introducing the campus community to possibilities opened up by the Internet2 connection. A wide range of Waynesburg faculty and staff attended.

NOTES & HIGHLIGHTS

NETWORKING THE FUTURE

One of the leading resources in the world for network know-how

PSC's Advanced Networking group is one of the leading resources in the world for knowledge about networking. Through 3ROX (Three Rivers Optical Exchange), a high-speed network hub, they provide high-performance networking for research and education. Their research on network performance and analysis — in previous projects such as Web100 and the NPAD diagnostic server — has created valuable tools for improving network performance nationally.

MORE INFORMATION: http://www.psc.edu/networking/

HUNTOON ELECTED TO INTERNET2 RESEARCH ADVISORY COUNCIL

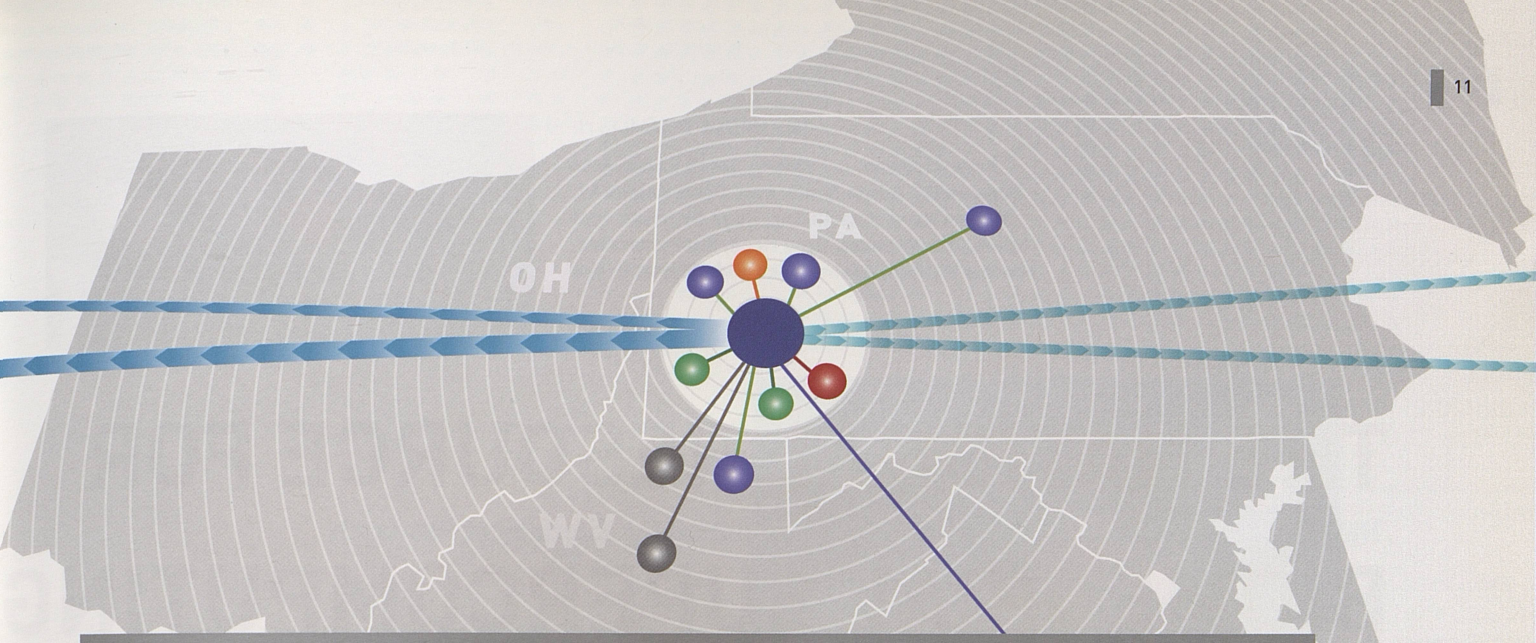
PSC director of networking Wendy Huntoon, who also directs operations for National LambdaRail (NLR), a major initiative of U.S. research universities and the private sector to provide infrastructure for research in networking technologies, this year was elected to a regional network seat on Internet2's Research Advisory Council.

Huntoon also serves on the steering committee for NSF's community-based planning activity, which gathers input on the infrastructure required by the science and engineering research and education community after the current TeraGrid project awards expire in early 2010.

PSC ADVANCED NETWORKING GROUP (seated, l to r):

Stephen Petko, Wendy Huntoon, and Michael Lambert, (standing) **Chris Rapier, Kathy Benninger, Joe Lappa, Andrew Adams, Steve Cunningham, Ken Goodwin, Shane Filus, John Heffner, Matt Mathis and Jim Miller.**

Not in photo, **Janet Brown.**



3ROX MEMBERS

Universities
Carnegie Mellon University, Norfolk State University, Pennsylvania State University, University of Pittsburgh, Waynesburg University, West Virginia University

K-12 Institutions
Pittsburgh Public Schools, Woodland Hills School District, Intermediate Units 1, 2 (Pittsburgh), 3 (Allegheny), 5 (Northwest Tri-County), 6 (Riverview), 27 (Beaver Valley) and 28 (ARIN)

Government Laboratory
The National Energy Technology Laboratory

Business
Comcast, Westinghouse Electric Co.

Other
Computer Emergency Response Team

NETWORK CONNECTIONS

National Research Networks
Internet2 — 1 Gbps, National LambdaRail PacketNet — 10 Gbps, TeraGrid Extensible Backplane Network — 30 Gbps

Other Network Connections
Southern Crossroads (SOX) — 1Gbps, TransitRail — 1Gbps

National Commodity Internet Networks
Global Crossing — 1 Gbps; Sprint — 1 Gbps.
(Pittsburgh Local Exchange Networks) Comcast

Note: Gbps: a billion (Giga) bits per second

3ROX

Through 3ROX, a high-speed network hub, PSC provides advanced network resources for education and research. 3ROX connects universities and public schools to high-performance networks, such as Internet2, which links leading U.S. universities, corporations, government research agencies, and not-for-profit networking organizations.

3ROX NEWS

3ROX contracted this year with FiberTech for a "co-location space" at Allegheny Center Mall (ACM) on Pittsburgh's near North Side. This network co-location space will provide a redundant location for 3ROX members to connect, initially only for commodity network connectivity, and eventually to include all services supported by 3ROX. ACM also provides an additional primary connection point for 3ROX members and collaborators.

Waynesburg University became a 3ROX member this year, and the University of Pittsburgh upgraded their connection to 3ROX from one to 10 Gbps. This new connection allows them to take better advantage of the NLR related services associated with 3ROX, including connectivity to PacketNet and the TransitRail peering service.

3ROX this year added several school district intermediate units, including Beaver Valley, Allegheny and a shared connection for Riverview, Northwest Tri-County and ARIN. These schools now have connectivity to Internet and NLR research and education networks as well as regional networks.

ADVANCED NETWORK RESEARCH

Through projects such as Web100 (www.web100.org), PSC's networking group has played a major role nationally in engineering the Internet. During this past year, they have authored and published several RFCs. official standards of the Internet Engineering Task Force, an open international community of network designers, operators, vendors, and researchers concerned with evolution and operation of the Internet.

In May, PSC researchers Matt Mathis and John Heffner authored "Packetization Layer Path MTU Discovery," a new standard for improved Internet performance. This work describes a method to automatically assess the best Maximum Transfer Unit (MTU) — the biggest data packet a connection can efficiently use — for a given network connection.

In June, Mathis, Heffner and Rajiv Raghunarayan of Cisco published "TCP Extended Statistics MIB." This important standard, essentially the main deliverable from the Web100 project, describes extended performance statistics for TCP (transmission control protocol), a core software protocol that underlies the majority of Internet applications. With the performance statistics defined by this document, diagnostic tools can — in situations of poor network performance — directly ask TCP why it is slow and gather information on the nature of the problem.

"It takes the guesswork out of network debugging," says Mathis. "If a network-based application is performing poorly, TCP can determine if the bottleneck is in the sender, the receiver, or the network itself. If the bottleneck is in the network, TCP can provide specific information about its nature." Prototypes implemented to this RFC were incorporated into a number of operating systems, including Linux and Microsoft Vista.

THE NRBSC TEAM (l to r): →

Pallavi Ishwad, Art Wetzel, Jacob Czech, Alex Ropelewski, Jenda Domaracki, Boris Kaminsky, Demian Nave, Jack Chang, Hugh Nicholas, Joel Stiles, Troy Wymore, Greg Hood, Stu Pomerantz, Markus Dittrich, Adam Marko and Christal Banks

THE NATIONAL RESOURCE FOR BIOMEDICAL SUPERCOMPUTING

National Leadership in High-Performance Computing for Biomedical Research

Established in 1987, PSC's National Resource for Biomedical Supercomputing (NRBSC) was the first external biomedical supercomputing program funded by the National Institutes of Health (NIH). Along with core research at the interface of supercomputing and the life sciences, NRBSC scientists develop collaborations with biomedical researchers around the country, fostering exchange among experts in computational science and biomedicine and providing computational resources, outreach and training. In October 2006, NRBSC received $8.5 million from NIH's National Center for Research Resources (NCRR) to renew its work for five years.

"Over the past decade, computing has become essential to almost all aspects of biomedicine," says PSC's Joel Stiles, director of NRBSC, a medical doctor who also holds a doctorate in physiology. "Here at the NRBSC, we're developing and distributing computational tools in simulation, visualization, and education that are helping to transform our understanding of life and disease."

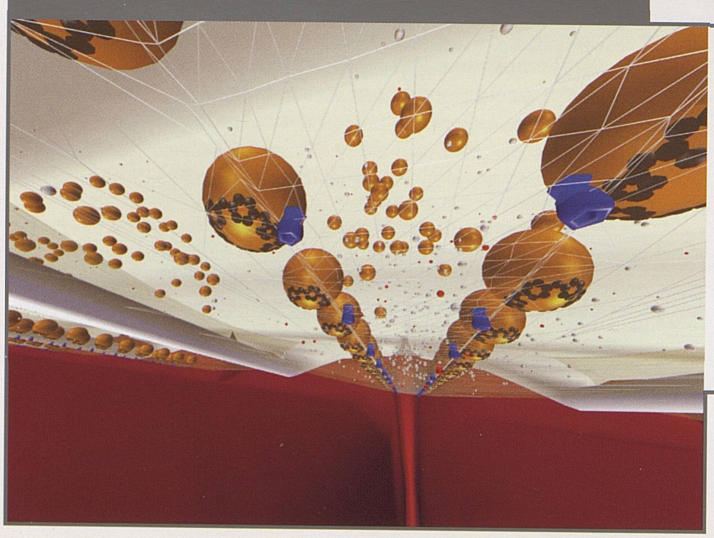
MORE INFORMATION: http://www.nrbsc.org

NRBSC Biomedical Collaborations

Albert Einstein College of Medicine
Carnegie Mellon University
Cornell University
Duke University
Howard University
Janelia Farm, Howard Hughes Medical Institute
Marine Biological Laboratory, Woods Hole
Morgan State University
North Carolina Central University
Rockefeller University
The Salk Institute
The Scripps Research Institute
University of California at Davis
University of California at San Diego
University of Kansas
University of North Carolina, Chapel Hill
University of Pittsburgh
University of Pittsburgh School of Medicine
University of Puerto Rico, Medical Sciences Campus

RESEARCH

NRBSC research focuses on three areas of biomedicine that span many scales of space and time: spatially realistic cell modeling, large-scale volumetric visualization and analysis, and computational structural biology.



Spatially realistic cell modeling centers on realistic 3-D simulations of movements and reactions of molecules within and between cells, to better understand physiological function and disease. As illustrated in this image, MCell, DReAMM and PSC_DX software is developed at the NRBSC and is used to model events like neurotransmission between a nerve and muscle cell.

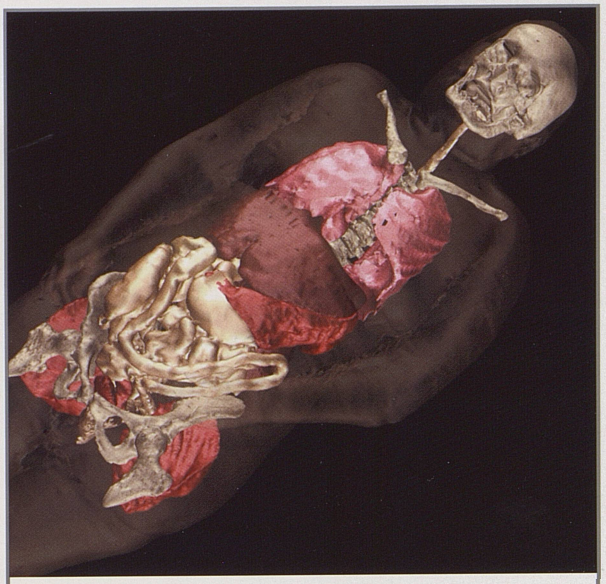
COMPUTATIONAL SERVICE & TRAINING

Since its inception, NRBSC has provided access to computing resources for more than 1,200 biomedical research projects involving more than 3,800 researchers at 255 research institutions in 45 states and two territories. Among these are several projects featured in this booklet (pp. 26 & 30).

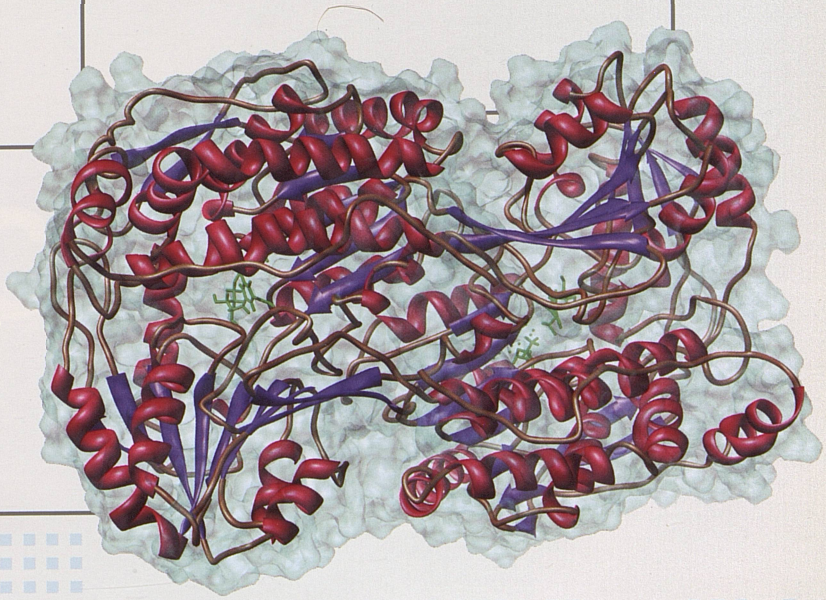
NRBSC training activities reach hundreds of scientists each year. More than 3,400 researchers have participated in NRBSC workshops in such areas as spatially realistic cell modeling, volumetric data visualization and analysis, protein and DNA structure, genome sequence analysis and biological fluid dynamics.

NRBSC participates in a range of undergraduate and graduate training programs. These include:

- A joint Carnegie Mellon and University of Pittsburgh Ph.D. program in computational biology (www.compbio.cmu.edu),
- The Ray and Stephanie Lane Center for Computational Biology at Carnegie Mellon (lane.compbio.cmu.edu).
- The University of Pittsburgh Department of Computational Biology (www.ccbb.pitt.edu).
- The undergraduate Bioengineering & Bioinformatics Summer Institute (www.ccbb.pitt.edu/bbsi), sponsored by NRBSC, Carnegie Mellon, the University of Pittsburgh, and Duquesne University, and funded jointly by NSF and NIH.



Volumetric visualization is based on the NRBSC's PSC_VB software, enabling multiple users to share, view and analyze extremely large datasets and time series obtained from light and electron microscopes, CAT and MRI scanners, etc. This image illustrates Visible Human data from the National Library of Medicine, analyzed with PSC_VB and rendered with DReAMM.



NRBSC structural biology focuses on computational tools used to determine the structure of proteins from their amino acid sequence and development of quantum-mechanical simulation methods for biomolecules such as enzymes. This image shows the 3-D structure of human p5cdh, an enzyme involved in metabolism of sugars.

NOTES & HIGHLIGHTS CONTINUED

K-12 SCIENCE OUTREACH

The NRBSC and PSC are developing innovative Computational Modules In Science Teaching (CMIST) for high-school biology, chemistry, physics, computer science and math. CMIST modules bring critical concepts to life in novel ways, using realistic models and simulations with visually appealing, scientifically accurate animations. NRBSC distributes the modules online and on DVDs. They include lecture slides, animations, lesson plans aligned to national and state standards, worksheets and answer keys. NRBSC developed the pilot module, "Molecular Transport in Cells," through a summer 2007 series of workshops with 37 western Pennsylvania high-school science teachers from 17 school districts, who gave enthusiastically positive feedback.

Pittsburgh Supercomputing Center Workshops (2006-2007)

Petascale Applications Symposium

Computational Methods for Spatially-Realistic Microphysiological Simulations

Summer Institute in Bioinformatics
(for minority-serving institutions)

Parallel Programming

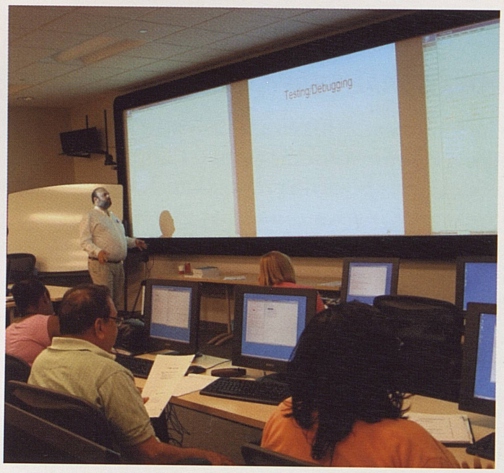
Toward Multicore Petascale Applications

Introduction to the Cray XT3

Methods and Applications of Hybrid QM/MM Simulations to Biomolecular Systems

Bioinformatics

A workshop underway in the PSC Computer Training Center, the David W. Deerfield II Training Center, equipped with 30 "dual-boot" workstations and a projector for overhead display of the instructor's desktop.



PSC'S DIRECTORS AND MANAGERS (front, l to r):

Cheryl Begandy, outreach manager; **Elvira Prologo**, manager of administration; **Bob Stock**, PSC associate director; **Rich Raymond**, manager of user support; **Wendy Huntoon**, director of networking; **Sergiu Sanielevici**, director of scientific applications and user support; (back): **David Kapcin**, director of financial affairs; **Joel Stiles**, director of NRBSC; **John Kochmar**, manager of high-performance computing facilities; **J.Ray Scott**, director of systems & operations; **Katherine Vargo**, manager of scientific computing systems.

Not in photo: **Nick Nystrom**, director of strategic applications; **Laura McGinnis**, manager of data & information resource services; **Janet Brown**, manager of networking; **Clint Perrone**, manager of systems and operations.

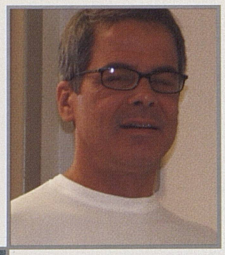
PROJECTS
IN SCIENTIFIC COMPUTING
2007

PROJECTS 2007
CONTENTS

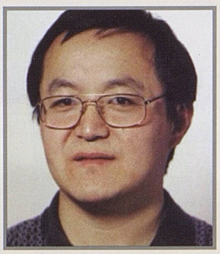
GANGING UP ON Storms

PSC staff and resources make possible an unprecedented experiment in "ensemble" forecasting of severe storms

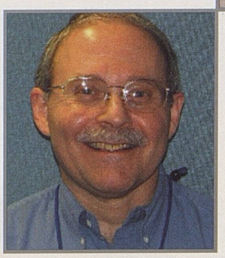
A paper by Xue, Droegemeier, Weiss and others reporting preliminary findings of the HWT experiment acknowledged that O'Neal "provided ever present technical and logistic support that was essential to the success of the forecast experiment."



David O'Neal
PSC scientist



Ming Xue
director of CAPS



Steven Weiss
science and operations officer of the NOAA Storm Prediction Center in Norman

The night of May 4, 2007 won't fade soon from the memory of people in Greensburg, Kansas. An extremely powerful tornado took only a few minutes to flatten almost every above-ground structure in this southwest Kansas town, claiming 10 lives. Catastrophic as it was, the loss would have been worse but for a very strongly worded warning from the National Weather Service office in Dodge City, about a half-hour in advance of the funnel's arrival, that residents credit with allowing most people to find safe shelter.

What if the warning had come a half-day in advance? Thunderstorms are difficult to predict, and "supercells," the high-energy vortex systems that spin-off tornados, are notoriously difficult. Nevertheless, if it were possible six or eight hours or more in advance to say when, where and with how much force a severe storm would strike, millions of dollars annually — if not billions — and countless lives would be saved.

Scientists at the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma, Norman know that it is possible to dramatically improve severe-storm forecasting, and their ground-breaking work over the past 15 years — often in partnership with PSC — has convinced many skeptics. They took a further large step this spring, collaborating with NOAA (National Oceanic and Atmospheric Administration), PSC and others in an unprecedented experiment.

Running from April 15 to June 8, the 2007 NOAA Hazardous Weather Testbed (HWT) Spring Experiment had the ambitious goal of testing "ensemble" forecasting — multiple runs of a forecast model that make it possible to specify the amount of uncertainty involved in the overall forecast. CAPS relied on PSC resources to run a 10-member ensemble, 10 runs each day of a high-resolution model (four-kilometer grid spacing) that extended from the Rocky Mountains to the East Coast, two-thirds of the continental USA. In addition, CAPS and PSC ran a single higher resolution run (two-kilometer spacing) over the same domain. The HWT experiment marked the first time ensembles were used at the storm scale, and it also marked their first use in real time in a simulated operational forecast environment.

To make this happen, PSC brought to bear BigBen, its Cray XT3 — a lead system of the TeraGrid — and helped to bring about a dedicated Pittsburgh to Oklahoma high-bandwidth link, contributed by Cisco Systems, Inc. Most critically, PSC staff pitched-in with know-how. They optimized performance of the forecast model, automated the daily forecast runs, coordinated the high-bandwidth link and — meeting the challenge of an enormous amount of computing and data-handling — produced forecasts each day for the eight weeks of the experiment with no serious problem.

"The experiment," says Ming Xue, director of CAPS, "was enormously successful. Scientists who took part were extremely impressed by what PSC was able to pull off. Most other places would struggle to do a single forecast run per day, and we did 10 ensemble runs plus a high-resolution run."

"Ensembles have been used extensively in larger-scale models," says Steven Weiss, science and operations officer of the NOAA Storm Prediction Center (SPC) in Norman. "But they have never before been used at the scale of storms. This was unique — both in terms of the forecast methodology and the enormous amount of computing. The technological logistics to make this happen were nothing short of amazing."

SCIENTIFIC COLLABORATORS IN STORM FORECASTING

Kelvin Droegemeier, director of LEAD and former director of CAPS. "We've been doing this with Pittsburgh since the mid-90s, and there's no question that had we not been developing this relationship over time, we wouldn't be where we're at today.

"PSC staff have been scientific collaborators in the deepest sense. They work hand-in-hand, not just to get our codes to run, but with networking and data-transfer, how the code is structured on the machine. You have to build-up trust, and people at Pittsburgh from Wendy Huntoon [director of networking], David O'Neal [PSC scientist], Ralph and Michael [PSC directors] — they all get a kick out of doing this work with weather."

SETTING THE STAGE

In 1989, when CAPS started its work, the prevailing view about storm-scale weather forecasting was skepticism; thunderstorms were thought to be inherently chaotic and unpredictable. Much has changed. CAPS developed innovative techniques to gather atmospheric data from Doppler radar, and they developed a forecast model to use this data at the scale of thunderstorms. In the mid-90s, via a series of spring collaborations with PSC, CAPS proved the feasibility of forecasting thunderstorms.

In 2005, in a large-scale collaborative experiment with NOAA, again with major support from PSC, CAPS took another leap forward by showing that it was possible, in some circumstances, to predict the details of a thunderstorm as much as a full day in advance. "We provided dramatic new evidence that the predictability of organized deep convection is, in some cases, an order of magnitude longer — up to 24 hours — than suggested by prevailing theories of atmospheric predictability," says former CAPS director Kelvin Droegemeier.

An opportunity to build on this prior work, the 2007 HWT Experiment involved more than 60 researchers and forecasters from government agencies, universities and the private sector. Along with PSC, CAPS and SPC, other collaborators were the NOAA National Severe Storms Laboratory in Norman; the NOAA National Centers for Environmental Prediction Environmental Modeling Center; the National Center for Atmospheric Research; LEAD (Linked Environments for Atmospheric Discovery), an NSF Large Information Technology Research grant program and TeraGrid Science Gateway; and the National Center for Supercomputing Applications (NCSA) in Illinois, a lead TeraGrid resource provider.

To implement the CAPS daily forecast runs using the WRF (Weather Research and Forecast) model on the XT3, PSC provided technological and staff assistance at several levels:

- PSC networking staff coordinated with OneNet, a regional network of the State of Oklahoma, and National Lambda Rail (NLR), a network initiative of U.S. universities, and with Cisco Systems, who contributed a dedicated "lambda" (a 10-gigabit-per-second optical-network) for up to a 12-month period.
- PSC implemented and began testing the lambda at its end in January, using existing equipment in the Pittsburgh metro and local-area network. The backbone was provided by NLR and OneNet provided the link from Tulsa to Norman, Oklahoma.
- This dedicated link — from the Cray XT3 to OneNet in Tulsa to a supercomputer at the University of Oklahoma (which ingested and post-processed the data) — made possible the transfer of 2.6 terabytes of data per forecast day.
- PSC staff optimized the latest version of the WRF model to run on the Cray XT3, gaining a threefold speedup in input/output (I/O) of the WRF code, substantially improving overall performance.
- PSC also optimized the I/O for post-processing routines used to visualize and analyze the forecast output, achieving 100-fold speedup.
- PSC modified the reservation and job-processing logic of its job-scheduling software to automatically schedule the WRF runs and related post-processing, 760 separate jobs each day, demonstrating the TeraGrid's ability to use the Cray XT3, a very large "capability" resource, on a scheduled, real-time basis.

ANOTHER LEAP FORWARD

The daily runs of the 10-member WRF ensemble each used 66 processors of PSC's Cray XT3 for 6.5 to 9.5 hours each day. The single high-resolution WRF forecast used 600 XT3 processors for about nine hours daily, including data dumps at five-minute intervals via the dedicated network to a system in Norman, where CAPS post-processed the data for visualization and analysis.

It was an unprecedented marshaling of computing and network capability. "To do this amount of processing with the ability to transfer these huge volumes of data from Pittsburgh to Norman on a reliable, daily, timely basis was a technological success that we're still appreciating," says Weiss. "It hasn't been done before. There were people who thought 'How are they going to do that?'"

"IT HASN'T BEEN DONE BEFORE. THERE WERE PEOPLE WHO THOUGHT 'HOW ARE THEY GOING TO DO THAT?'"

Over its eight weeks, the HWT experiment produced massive amounts of data, and the analysis is ongoing. "Preliminary findings," says Xue, "show successful predictions of the overall pattern and evolution of many of the convective-scale features, sometimes out to the second day. The ensemble shows good ability in capturing storm-scale uncertainties."

The case of a large-scale north-south storm-front that developed on May 23 and continued for several days across the Great Plains exemplifies the ability of the storm-scale runs to produce a forecast that holds up for the storm system as a whole as much as 33 hours in advance. "The rather successful prediction of the overall pattern and evolution of this squall-line case," says Xue, "is very encouraging, especially considering that we are examining convective-scale predictions for up to 33 hours."

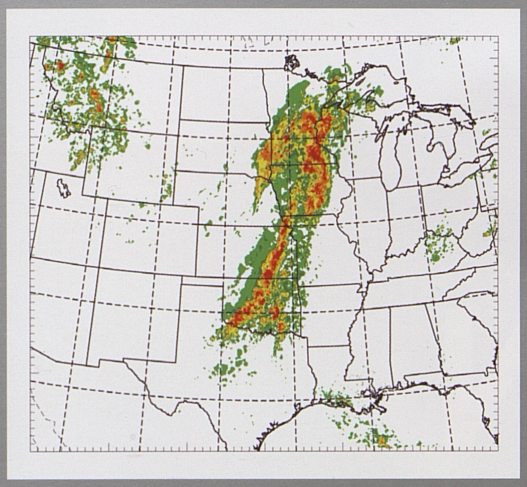
The HWT experiment also used capabilities developed by LEAD (Linked Environments for Atmospheric Discovery), a TeraGrid Science Gateway, to test "on-demand" forecasts. Triggered automatically from daily SPC forecasts indicating regions within the overall HWT domain where storms were likely to develop, these forecasts used NCSA's Ensemble Broker software and its 16-teraflop Tungsten system to run at two-kilometer resolution within the regions of high storm likelihood. More than 1,000 of these on-demand forecasts ran successfully.

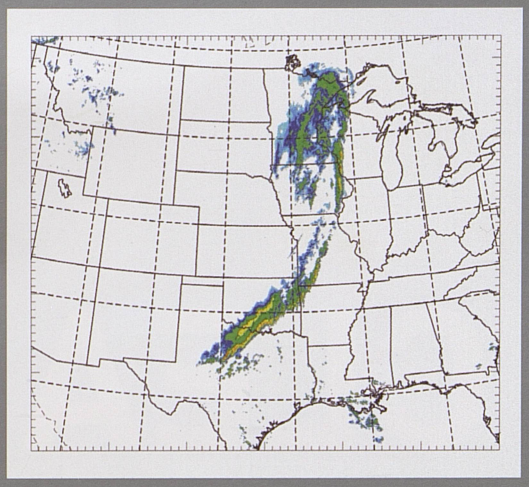
"This experiment was an enormous leap forward," says Droegemeier, who directs LEAD and as former CAPS director has led several other storm-forecast experiments over the past decade.

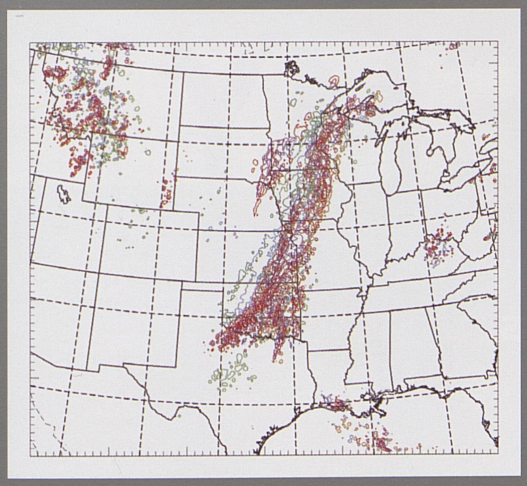
MORE INFORMATION:
www.psc.edu/science/2007/storms.html

Ensemble Predictions: Spaghetti Plots & Postage Stamps

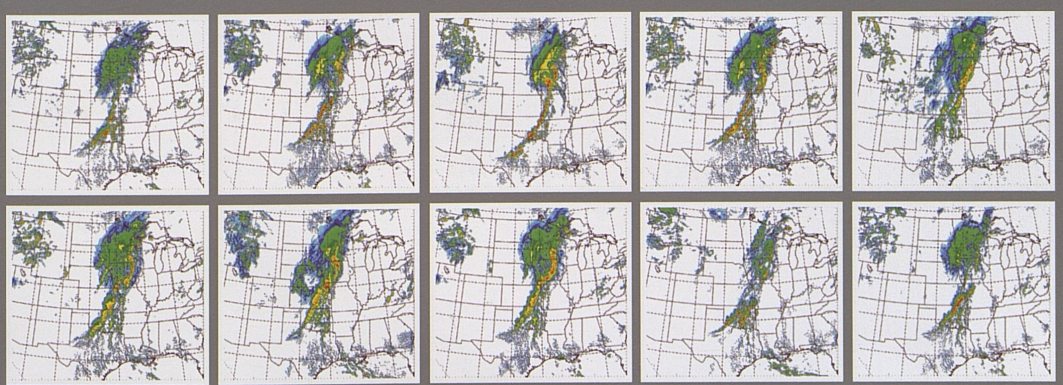
This plot (left) from the forecast for noon central standard time (18 universal time) on May 24, 2007 (from a model run on May 23) shows probability of radar reflectivity (proportional to intensityof precipitation) derived from the 10-member ensemble forecast 21 hours in advance compared to actual observed radar reflectivity on May 24 at noon CST (right).

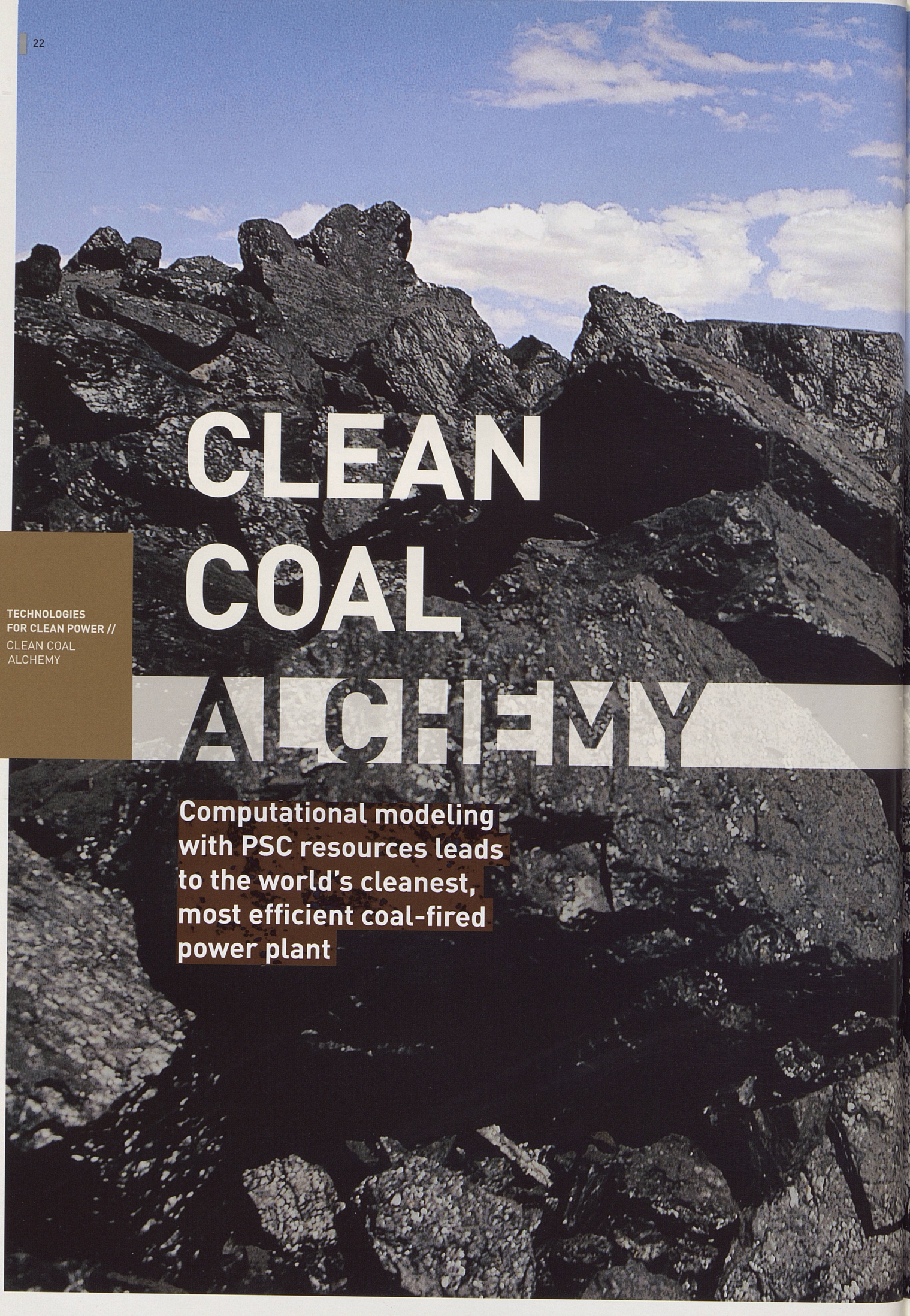






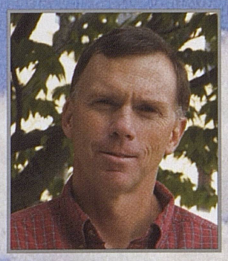
The "spaghetti plot" (left) for May 24 at noon CST represents the distribution of the radar-reflectivity forecast among the 10 ensemble runs, each represented by a different color. The corresponding "postage stamp" (below) display shows each of the 10 ensemble forecasts. Differences among them comprise a quantifiable range of the uncertainty associated with model errors and limitations on storm-scale predictability, as opposed to the traditional single-member (deterministic) forecast that produces only one prediction, good or bad.





CLEAN COAL ALCHEMY

Computational modeling with PSC resources leads to the world's cleanest, most efficient coal-fired power plant



Chris Guenther
National Energy Technology Laboratory,
Morgantown, West Virginia

While rising awareness of global warming stirs momentum for "clean" energies such as wind and solar power, the reality for now and the predictable future is that the indispensable workhorse of U.S. power generation is coal. More than half of U.S. electricity comes from coal-fired power, and none of the alternatives are without problems and high cost. Moreover, from the standpoint of geopolitics and our need for energy independence, we can count ourselves lucky to have coal as abundantly as we do — at current consumption, known U.S. coal will last another 200 years.

Are there ways to use this invaluable resource with less environmental impact? It's one of the primary questions driving research at the National Energy Technology Laboratory (NETL), a U.S. Department of Energy (DOE) national laboratory with three primary research sites: Morgantown, West Virginia; Pittsburgh, Pennsylvania; and Albany, Oregon. NETL's mission is to develop clean, affordable fossil-fuel technologies, and one of the most promising is coal gasification.

Long on the drawing boards as an idea, coal gasification is what it sounds like: a process to turn coal into a gas that in turn produces electricity much more cleanly than direct combustion of coal. With help from PSC's LeMieux and BigBen, NETL researchers have taken a big step toward making this vision real. Based on success during the past five years of a NETL program — on behalf of DOE's Office of Fossil Energy — construction has begun on a commercial coal-gasification plant near Orlando, Florida that is anticipated to be the cleanest, most efficient coal-fired power plant in the world by 2010.

"This major, unprecedented accomplishment in coal-gasification technology will allow us to continue using our most valuable energy resource in an environmentally acceptable manner," says Anthony Cugini, director of NETL's Office of Research and Development. "With sophisticated software and access to PSC supercomputing, our engineering-design team developed deep understanding of the complex processes of a coal-gasifier reactor system."

The Power Systems Development Facility
Wilsonville, Alabama

TRANSPORT GASIFICATION

DOE announced the Orlando project in October 2004 as part of its Clean Coal Power Initiative. They selected the Southern Company, an Atlanta-based energy company, along with Orlando Utilities Commission, to build the $557 million venture, for which DOE is contributing $235 million. Houston-based KBR and Southern Company are responsible for engineering.

The 285-megawatt plant is based on a technology called "transport gasification," invented and bench-scale tested by KBR in the early 1990s. For further development, the project team relied on a pre-commercial-scale plant — the Power Systems Development Facility (PSDF) — in Wilsonville, Alabama, built by Southern Company and DOE in the mid-1990s to evaluate and commercialize advanced coal-based power technologies. With the PSDF, engineers are able to adjust operating conditions of systems being evaluated, to learn what works and what doesn't. For the transport gasifier, to reduce the trial-and-error guesswork and gain improved understanding of the physics and reaction chemistry, PSDF engineers worked closely with a NETL computational modeling team, the Device Scale Modeling Group, led by Chris Guenther.

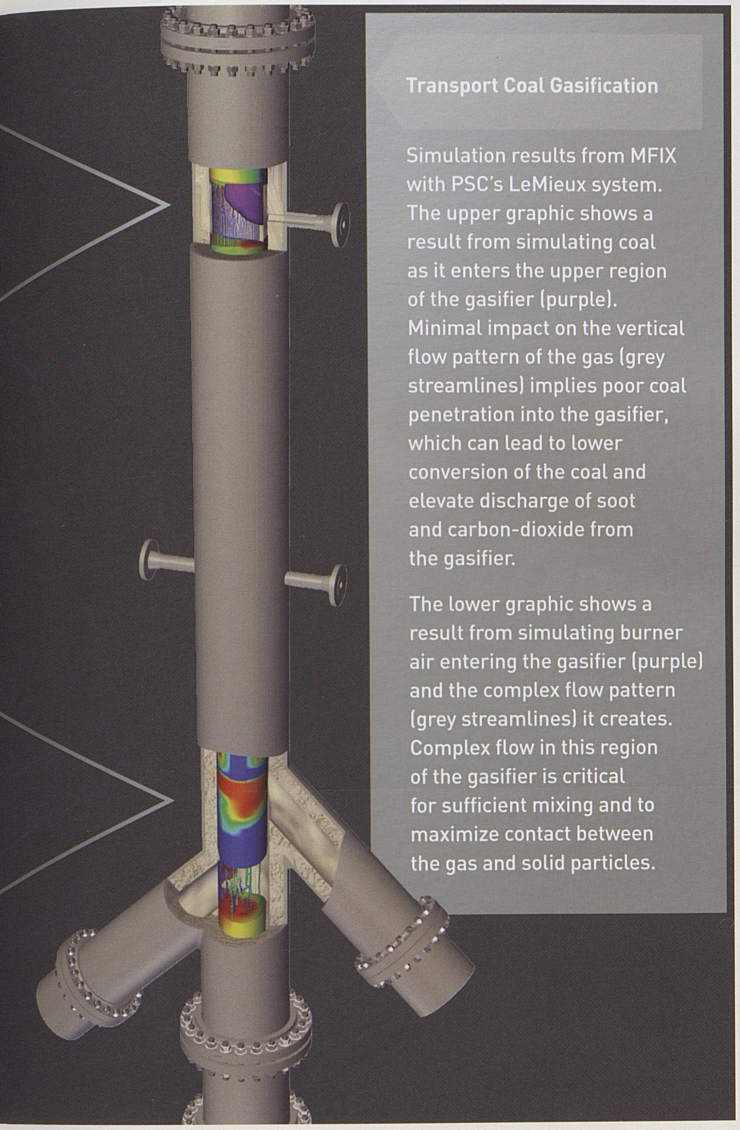
Guenther's group had a powerful tool at their disposal: MFIX (Multiphase Flow with Interphase Exchanges), software developed at NETL to reduce cost in development of clean-coal technologies. It describes the hydrodynamics, heat transfer, and chemical reactions in fluid-solid systems such as coal gasification. MFIX's ability to simulate various coal-combustion processes and its usefulness as a design tool were recognized this year by R&D Magazine, which gave MFIX an R&D 100 award as one of the top 100 technologies of the year.

Using LeMieux over the past five years, Guenther's team validated MFIX's ability to accurately predict conditions at the PSDF. Via this process they developed a patent-pending coal-chemistry module to be used with MFIX. "LeMieux's speed allowed us to quickly make whatever changes we wanted in the virtual world and then tell the plant engineers at PSDF what to expect if they changed this or that variable," says Guenther. "With LeMieux we had the CPU power, and the basic physics-based equations were there in the software."

They used from 16 to 64 of LeMieux's processors. "The support staff at PSC often gave us a high priority," says Guenther, "which was really critical for our work. Sometimes MFIX was running absolutely continuously around the clock at PSC for every bit of an entire work week or longer."

PSC scientific-visualization specialist Kent Eschenberg created a software routine to take data from the LeMieux runs and translate it into a format Guenther could use directly with advanced 3D visualization software, so he could see and analyze the results quickly.

When the PSDF work began, however, the MFIX team lacked solid information on crucial "reaction rate" parameters involved in gasification, and it required iterative, careful work over time — with many simulations on LeMieux — to improve the software's predictive ability, work which led eventually to some surprising findings.



Transport Coal Gasification

Simulation results from MFIX with PSC's LeMieux system. The upper graphic shows a result from simulating coal as it enters the upper region of the gasifier (purple). Minimal impact on the vertical flow pattern of the gas (grey streamlines) implies poor coal penetration into the gasifier, which can lead to lower conversion of the coal and elevate discharge of soot and carbon-dioxide from the gasifier.

The lower graphic shows a result from simulating burner air entering the gasifier (purple) and the complex flow pattern (grey streamlines) it creates. Complex flow in this region of the gasifier is critical for sufficient mixing and to maximize contact between the gas and solid particles.

KEYING ON KINETICS

MFIX has modeled many gas-solids reaction processes, but it had not before been used to model a transport gasifier — so-called because the hot gas and unburnt solids travel upward through the long, pipe-like reactor. In this environment, coal particles react with steam, air, and other chemicals in many reactions that lead ultimately to a mixture of carbon monoxide, hydrogen, and other gas species called "syngas" — short for "synthesis gas." Unreacted coal particles recirculate from the "riser" — a vertical tube at the top — back to the reactor's "mixing zone." Eventually, all the coal particles transform into syngas, and the coal is gasified.

THE XT3 OPENS UNPRECEDENTED ABILITY TO DO LOCALIZED PREDICTIONS FOR THE REACTOR AT VERY HIGH RESOLUTION

A critical variable in this environment where multiple reactions take place simultaneously is "kinetics" — how fast the reactions proceed. "This transport gasifier was a fairly new piece of technology," says Guenther, "and when we began there were no reaction rates in the open literature that covered our conditions. The kinetics for detailed gas-phase combustion is out there, but detailed reaction mechanisms for devolitization, tar cracking, and gasification under transport conditions are not readily available."

By adjusting the kinetics in the MFIX code incrementally through a series of simulations, Guenther and his colleagues began to get results that matched what was happening in the PSDF. It was a back-and-forth process. The engineers in Alabama would make a gasification run, and the NETL researchers would run the MFIX model on LeMieux to simulate the same conditions.

They knew they were getting it right when they pointed out a potential problem to the PSDF engineers. Simulations showed "oxygen breakthrough" — a stream of oxygen that did not react with the recirculated coal, but reacted instead with fresh coal higher up in the riser. This could cause the coal to combust rather than gasify, resulting in higher temperatures and elevated levels of CO_2. "We made some predictions that at first the plant engineers found hard to believe,' says Guenther, "but later accepted."

Another leap occurred when the PSDF was shut down eight months for modifications and NETL ran simulations to see what effect the modifications would have. "When they were back up and running they were getting numbers like we predicted," says Guenther. "That was a milestone for us in showing that MFIX running on LeMieux is a valuable tool for commercial-scale design."

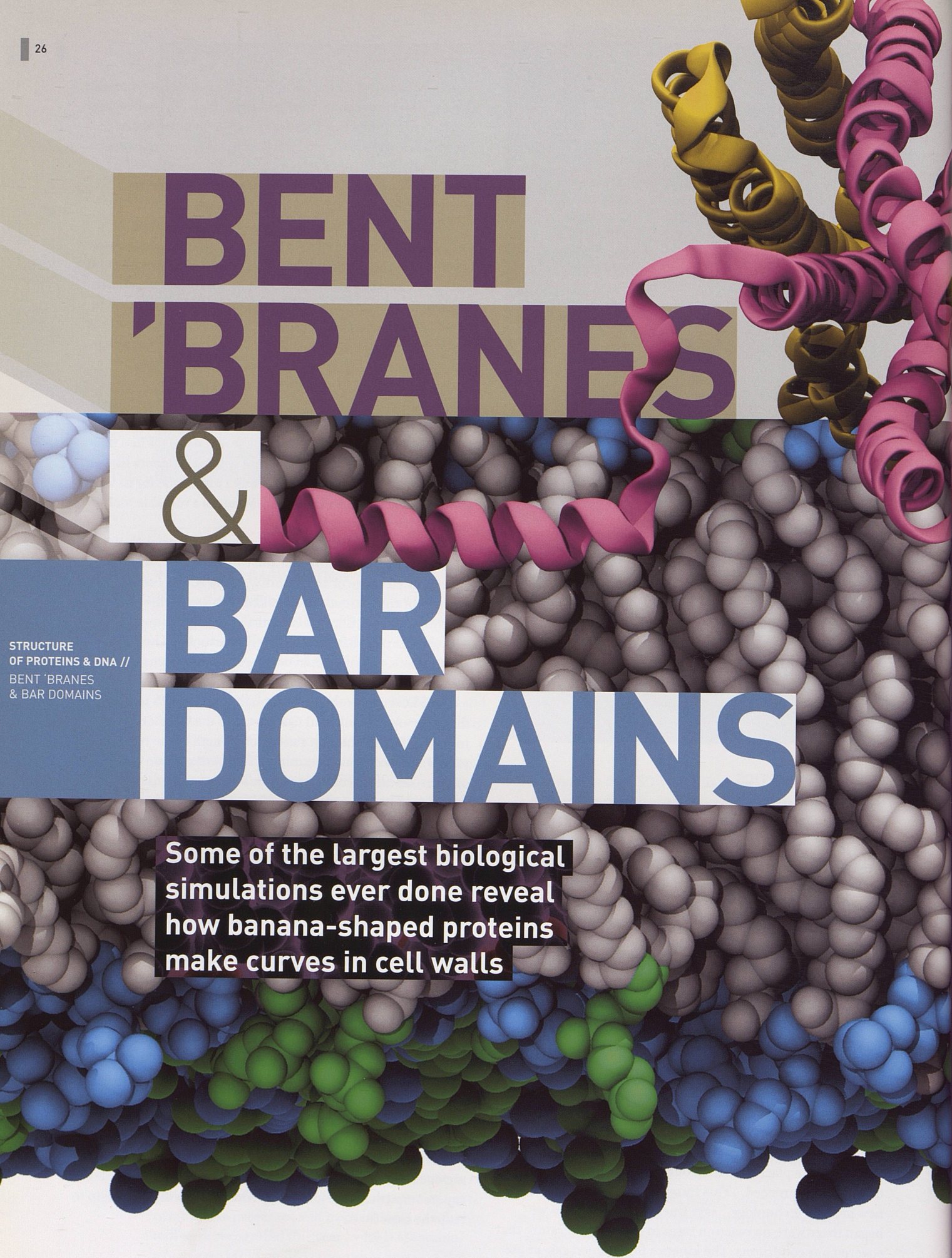
The Orlando project's partners agree. "Previously we had to rely on pilot-plant data and theoretical models that had a limited range of applicability," says Nicola Salazar, director of coal monetization at KBR. "But with this cutting-edge tool, we are able to predict the effects more quickly and with greater confidence. Ultimately, the tool will help us to develop safe, reliable, and energy-efficient designs for our commercial gasifiers."

MFIX is currently running on PSC's Cray XT3 to continue the scale-up that will lead to start-up of the Orlando plant in 2010. "What's the best ratio of the length of the reactor to its diameter?" says Guenther. "Is it better to use 100-micron-diameter coal particles or 300-micron-diameter coal particles? These are the kind of "what if" scenarios we can investigate."

With the XT3, Guenther can run MFIX efficiently on up to 1,024 processors, which opens unprecedented ability to do localized predictions in specific areas of the reactor at very high resolution. "For localized predictions, we are using 12-million computational cells, seven times more resolution than we did with LeMieux. To my knowledge, commercial-scale simulations at this fine a resolution have never been done before, and it means we can answer much more in-depth questions about reactor performance."

The Orlando plant will be the world's first commercial-scale plant that incorporates a transport gasifier with integrated gasification combined cycle technology. Syngas from the transport gasifier will pass through a gas turbine to generate electricity, then heat water for a steam turbine that generates more electricity. Along with more electricity per ton of coal, there are essentially no emissions of sulfur dioxide and particulates, appreciably lower oxides of nitrogen, and 25-percent less carbon dioxide. With help from supercomputing, Florida energy consumers will soon begin to reap the benefit. (TP)

MORE INFORMATION:
www.psc.edu/science/2007/coal.html



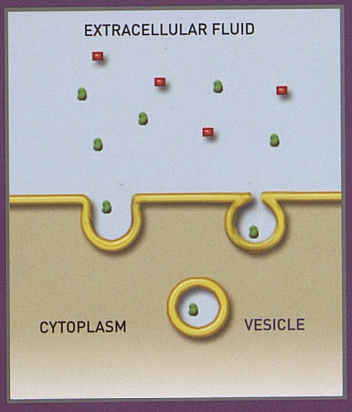
BENT 'BRANES & BAR DOMAINS

Some of the largest biological simulations ever done reveal how banana-shaped proteins make curves in cell walls

Gregory Voth

Phil Blood

Think of "The Blob" in that bad old movie — a giant amoeba-like alien that globs onto and absorbs humans. Endocytosis, the life-sustaining process by which cells absorb large proteins and other molecules, is something like that, and it's an amazing ability. A squishy, water-resistant wall that isolates and protects the cell's innards, the cell membrane can also pucker into a mouth-like opening that engulfs external molecules and closes off into a spherical chamber called a "vesicle" that's absorbed into the cell.



Similar abilities of the cell wall to form a vesicle are involved in transporting molecules out of the cell, such as neurotransmitters bursting out of nerve cells. Other processes, such as those that generate tubules in muscle cells, also involve curving of the cell wall. All these processes have many steps, but they all start with "remodeling" — a change in membrane curvature.

"Membrane remodeling is one of the most important phenomena in cellular biology," says University of Utah chemist Gregory Voth. "It's fundamental to vesicle budding, a key feature of many processes, one of the most well known being neurotransmitter release in the synapses."

Experiments show that these complex events are triggered by an ensemble of proteins, and they show also that a banana-shaped protein segment called the BAR domain is a key player in inducing the cellular membrane to curve. But how? The experiments show that the BAR domain is involved, but they don't show the structural and dynamic bottom-line, what individual atoms are doing as curvature proceeds, which is where Voth and Phil Blood — then a University of Utah Ph.D. candidate, now a PSC scientist — came into the picture. "What we've been able to do via molecular dynamics simulations," says Blood, "is to see the atom-by-atom details of how these proteins generate membrane curvature."

Using PSC's Cray XT3 as their workhorse, along with TeraGrid systems at SDSC, NCSA and the University of Chicago/Argonne, Voth and Blood did molecular dynamics (MD) simulations that track the atom-level interactions between a BAR domain and a model of cellular membrane. Their simulations — involving 738,000 atoms, using more than two-million XT3 processor hours — are among the largest biological computations ever performed.

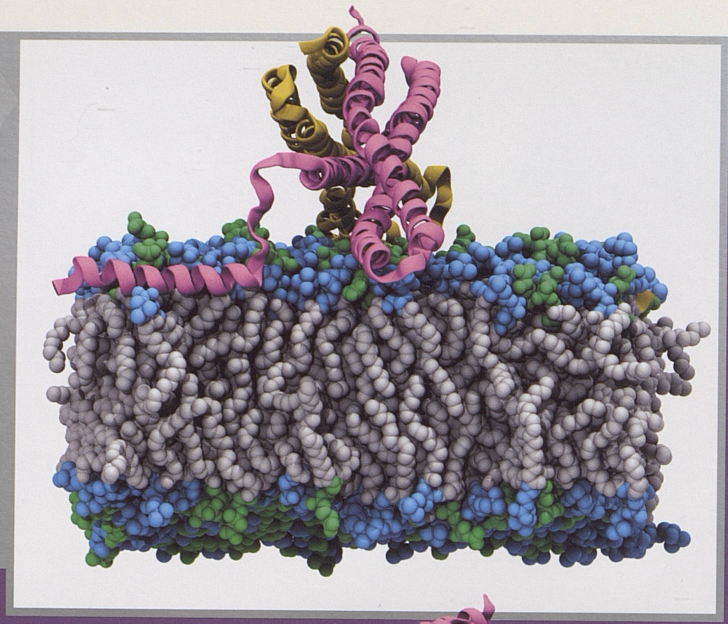
Their results, reported in *Proceedings of the National Academy of Sciences* (October 10, 2006), confirm and go beyond experimental results, showing that BAR domains act as a scaffold, forcing the membrane to adopt their curvature, and that the orientation of the BAR domain as it attaches to the membrane determines the degree of curvature.

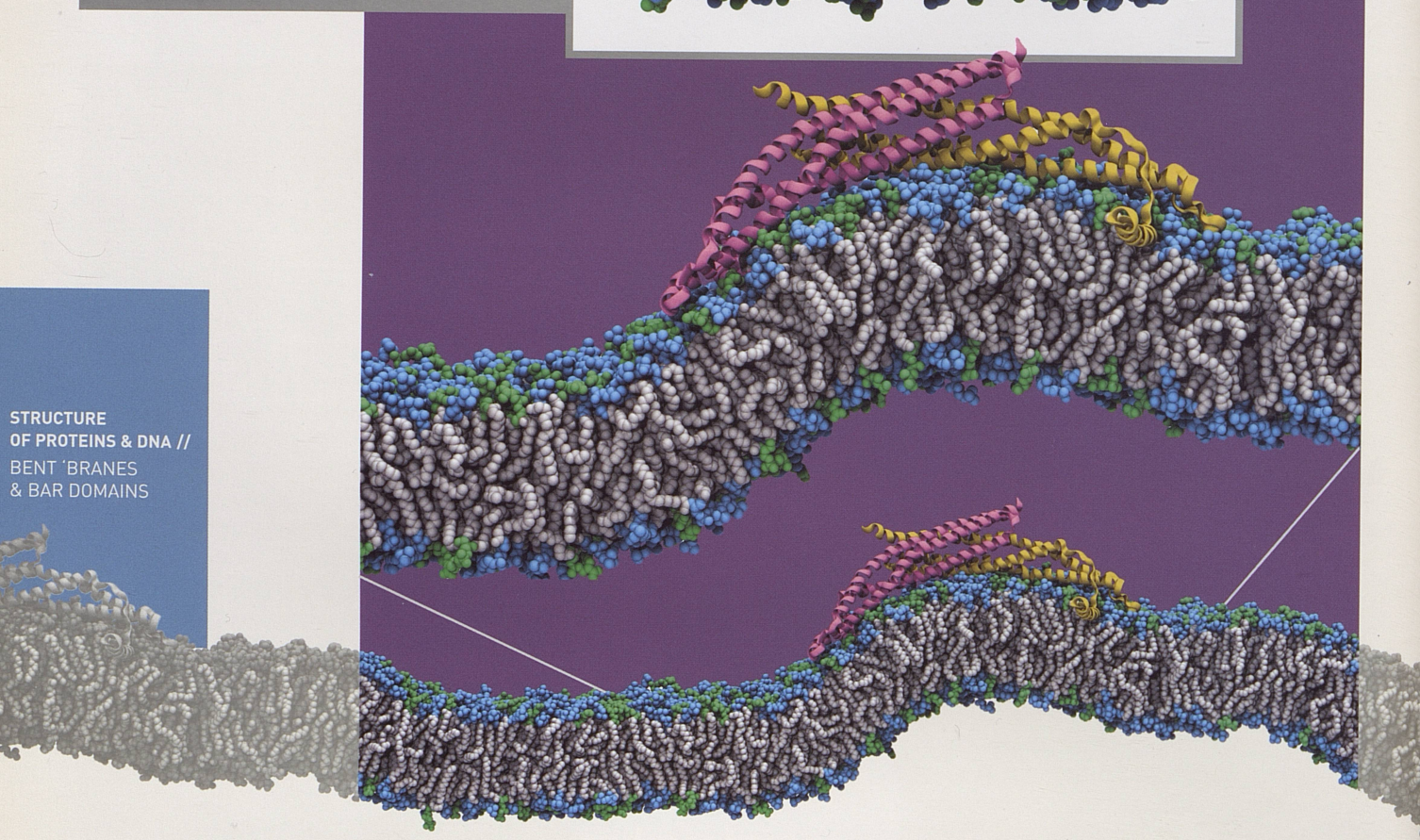
Extending this success from 2006, Voth and Blood this year refined their simulations to look more deeply and see more exactly which parts of the BAR domain structure drive curvature. They found — contrary to what had been surmised from experiments — that helical structures at each end of the BAR domain do not by themselves initiate curvature.

"With this kind of work," says Voth, "we're beginning to see how the modern capabilities of molecular dynamics simulation touch very fundamentally important biological processes. There's been this investment both in hardware and software. This combination is paying off. With this kind of problem, we're getting at the interplay between proteins and membranes at a level beyond binding studies and ion-channel transport, and we're able to see whole-scale changes in the membrane morphology driven by these interactions."

Bending the Membrane

This front view (below) from a simulation on PSC's Cray XT3 represents a BAR domain protein segment (pink & yellow helices) bending a segment of cellular membrane with negatively-charged lipid head groups (green) after 30 nanoseconds of simulation. Loops on each end of the BAR domain bind tightly to the descending sides of the curved membrane. The N-terminal helix (yellow) is embedded in the bilayer. The side view (right) shows a cutaway of the embedded N-terminal helix and also shows the loop structure that binds to the membrane surface.





TOOLS FOR THE JOB

In 2004 a team of British scientists laid the groundwork for Voth and Blood's simulations by working out the three-dimensional structure of the BAR domain, work that appeared on the cover of Science. "They solved the crystal structure for this protein," says Blood, "and it was quite interesting to see that it looks like a banana, a crescent shape that immediately says to you, 'I'm a curvature forming protein.'"

Beginning with its shape, the BAR domain suggested a number of possibilities for how it could generate curvature. The concave side of the crescent carries a positive electronic charge, which could bind to the negatively-charged surface of a cellular membrane and mold it to a matching curvature.

BAR domains also have a helix structure at each end, and experiments suggest that these "N-terminal helices" insert into the membrane, anchoring the protein at its base on both ends, which could help to induce puckering of the membrane. "There are many nuances," says Blood, "but nothing had been shown explicitly."

As Blood began to build a computational model, using the DataStar at SDSC and IA64 at NCSA along with a system at Argonne, he knew he'd taken on a demanding task. "The TeraGrid has been essential to making this possible." With MD software called NAMD (Not just Any Molecular Dynamics program), developed by Klaus Schulten's group at the University of Illinois, Urbana-Champaign, along with the XT3, Blood was working with the most powerful tools available for MD simulations at the large scale his problem demanded. "NAMD has been amazing," he says, "because it scales well even with the long-range electrostatics."

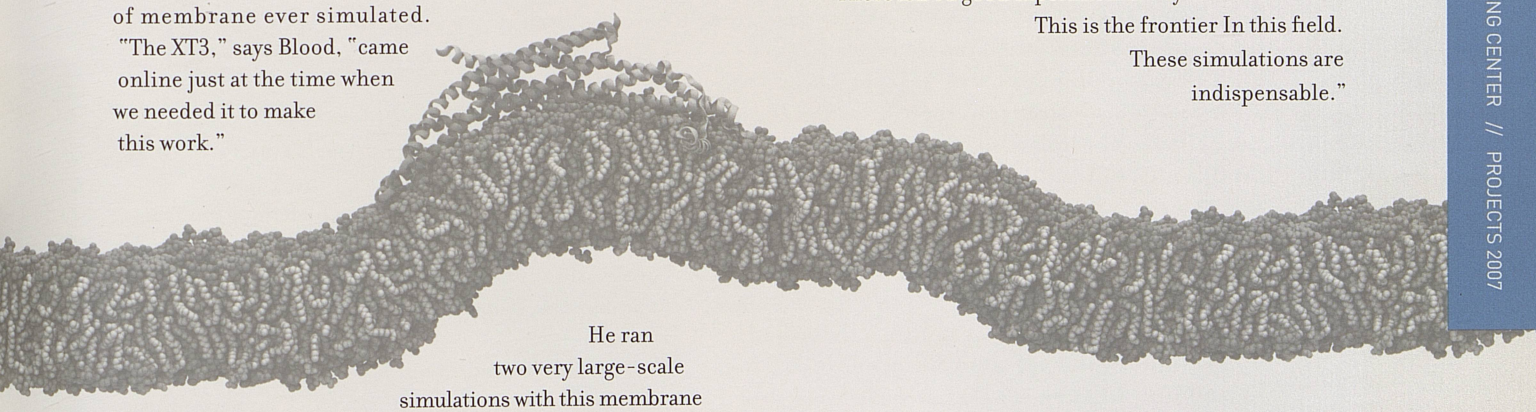
The problem required accurate treatment of the electronic charges between atoms, and the MD software tracks these forces as they vary between every atom in the simulation at each time step as the simulation progresses.

Scaling, a crucial feature for a large problem such as this, means that as you recruit a large number of processors for the computation — Blood used 512 XT3 processors — the per-processor performance remains high, doesn't take a prohibitive hit due to the inter-processor communication involved in tracking the atom-to-atom dynamics.

Among MD programs, NAMD is exceptional in its ability to run efficiently with hundreds or thousands of processors, and this is especially true on the 4,000-processor XT3, which has outstanding inter-processor bandwidth. "With the XT3, we haven't found a hard upper limit on scaling," says Blood. "Our simulations have run well with 512 processors, and we've used up to 1,024 processors on this system with little dropoff."

AT THE FRONTIER

As Blood built his computational model and explored the size of the problem, he began with a patch of membrane — 342,000 atoms, 25 nanometers in length — that was in itself a demanding computation. Trial studies showed, however, that this wasn't enough length for curvature to become apparent. Not long after the XT3 became available as a production system in October 2004, Blood more than doubled the size of his membrane model to 738,000 atoms — a 50-nanometer length of membrane, the longest patch of membrane ever simulated. "The XT3," says Blood, "came online just at the time when we needed it to make this work."



He ran two very large-scale simulations with this membrane and the BAR domain, tracking 27 nanoseconds (billionths of a second) of the dynamics, at time steps of two femtoseconds (millionths of a nanosecond). The two simulations resulted in two different orientations of the BAR domain, with the membrane curving to match the protein in each case. A control simulation with the membrane alone, no protein — on the DataStar system at SDSC — produced no similar curvature.

THIS IS THE FRONTIER IN THIS FIELD. THESE SIMULATIONS ARE INDISPENSABLE.

These results showed convincingly that the BAR domain generates membrane curvature and that the degree of curvature depends on the BAR domain's orientation. For Blood, they also raised a question. "How much of that curvature is due to the N-terminal helices and how much is due to the charged concave surface?"

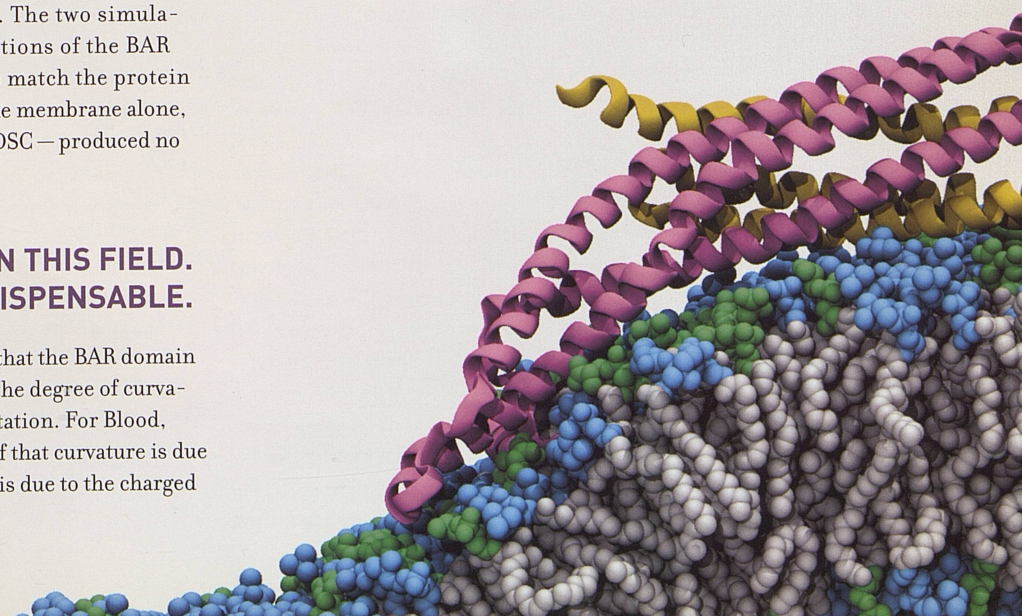
Some experiments suggested that the helices were the main driving force, with the concave surface providing little more than a scaffold. Another series of simulations, mainly on the XT3, showed that — at least for an individual BAR domain — this is not the case. These simulations showed, furthermore, that a particular structural element of the BAR domain, positively charged loops near the end, are crucial, binding to and stabilizing the BAR domain on the membrane surface, and without the binding of these loops curvature doesn't occur. "From these studies," says Blood, "it looks like the N-terminal helices are not primarily responsible for driving curvature."

In a follow-on study (*Biophysical Journal*, May 2007) Voth, Blood and their colleague Gary Ayton used the MD results as input to a larger-scale model of membrane shape change. "We use the molecular dynamics to feed a model at the meso-scale," says Voth, "and show how the interactions generated at the atomistic level propagate and scale up to a larger scale of remodeling."

Based on their success with NAMD and the XT3, Voth and Blood are looking ahead to even bigger, more realistic simulations. "The next generation of these simulations," says Voth, "will push up to 10-million atoms, which we'll need to look at multiple interacting BAR domain proteins and to develop accurate but simpler coarse-grained models for these systems. In nature, it's the collective behavior of many of these proteins interacting with the membrane that creates curvature. There isn't a good experimental way to look at this. This is the frontier In this field. These simulations are indispensable."

MORE INFORMATION:

www.psc.edu/science/2007/bardomain.html



STRANGE ACTION AT THE ACTIVE SITE

Quantum simulation of an enzyme reaction gives initially baffling results that point toward a surprising new insight

For the sake of discussion, let's say that enzyme biochemistry is like football. You can represent an enzyme's reaction mechanism with diagrams that resemble the Xs and Os of a football play. For both football and enzymes, there's a series of steps that have to happen for things to work. A ball carrier has to get the handoff just in time to hit the hole, for instance, and the hole has to open when he gets there.

Within a living cell, the atoms of an enzyme are like superfast, supersmall players. Enzyme reactions often occur in timeframes approaching a nanosecond (a billionth of a second) and distances of an angstrom (a hundred-millionth of a centimeter) or less. And if something isn't working as it should, the result can be disease or death.

Researchers can't see enzyme reactions within a living cell, but with help from powerful supercomputing systems they can simulate them and see what happens in minute detail. They can make animations from the simulation and look at the reactions as if watching a movie. Using these sophisticated and powerful tools, PSC scientist Troy Wymore and University of Pittsburgh biochemist John Hempel have simulated reactions in an important enzyme family called ALDH (aldehyde dehydrogenase).

Genetic malfunctions in human ALDH lead to a variety of debilitating disorders. ALDHs also affect the cancer-fighting activity of one of the most-used chemotherapy drugs. In 2002, with better understanding of ALDH reactions as the goal, Wymore, Hempel and their PSC colleagues David Deerfield and Hugh Nicholas focused on a particular ALDH in humans, ALDH3. Their detailed quantum-mechanical simulations identified a novel mechanism for a step in the enzyme reaction that had been generally disregarded but which their work showed to be crucial.

The mechanism they found — a proton transfer from the "backbone" of the enzyme's structure — may be a key to understanding ALDH-related disease. "We think these diseases occur because something interrupts this proton transfer," says Wymore. "What we think we've done is give an *atomic* rationale for ALDH-related metabolic diseases. It opens the door to possible intervention strategies."

To further explore their hypothesis, in 2006 they used PSC's newest system, the Cray XT3, for more simulations. This time they focused on another ALDH, ALDH2, and the simulations showed an unexpected intermediate reaction step, one never seen experimentally and completely unexpected — to such an extent that the researchers had to pause and think about the validity of the ALDH2 model they used. It seemed like a back-to-the-drawing-board moment.

Not long after, however, Hempel learned of a laboratory study that showed the same never-before-observed reaction step. This not only confirmed their surprising simulation result, it also suggested they had flipped a page in ALDH chemistry to uncover a more complex and more complete description of ALDH's reaction mechanism.

"This shows," says Hempel, "that simulations of enzyme chemistry have to be taken seriously. They reveal details you can't otherwise see, and as a result we find that now we have to look closely at things nobody thought needed to be considered.

Shawn Brown, Troy Wymore, John Hempel

A DEEPER LOOK

At least 18 different versions of ALDH are known in plants and animals, and they all do essentially the same thing. They take a toxic molecule called an aldehyde, produced during metabolism, and change it (oxidize it) into a form (carboxylic acid) that can pass safely out of the cell and into the bloodstream. In humans, malfunctioning ALDH is involved in two known diseases, one of which, an inherited disorder called Sjögren-Larsson syndrome, leads to skin scaling and mental retardation.

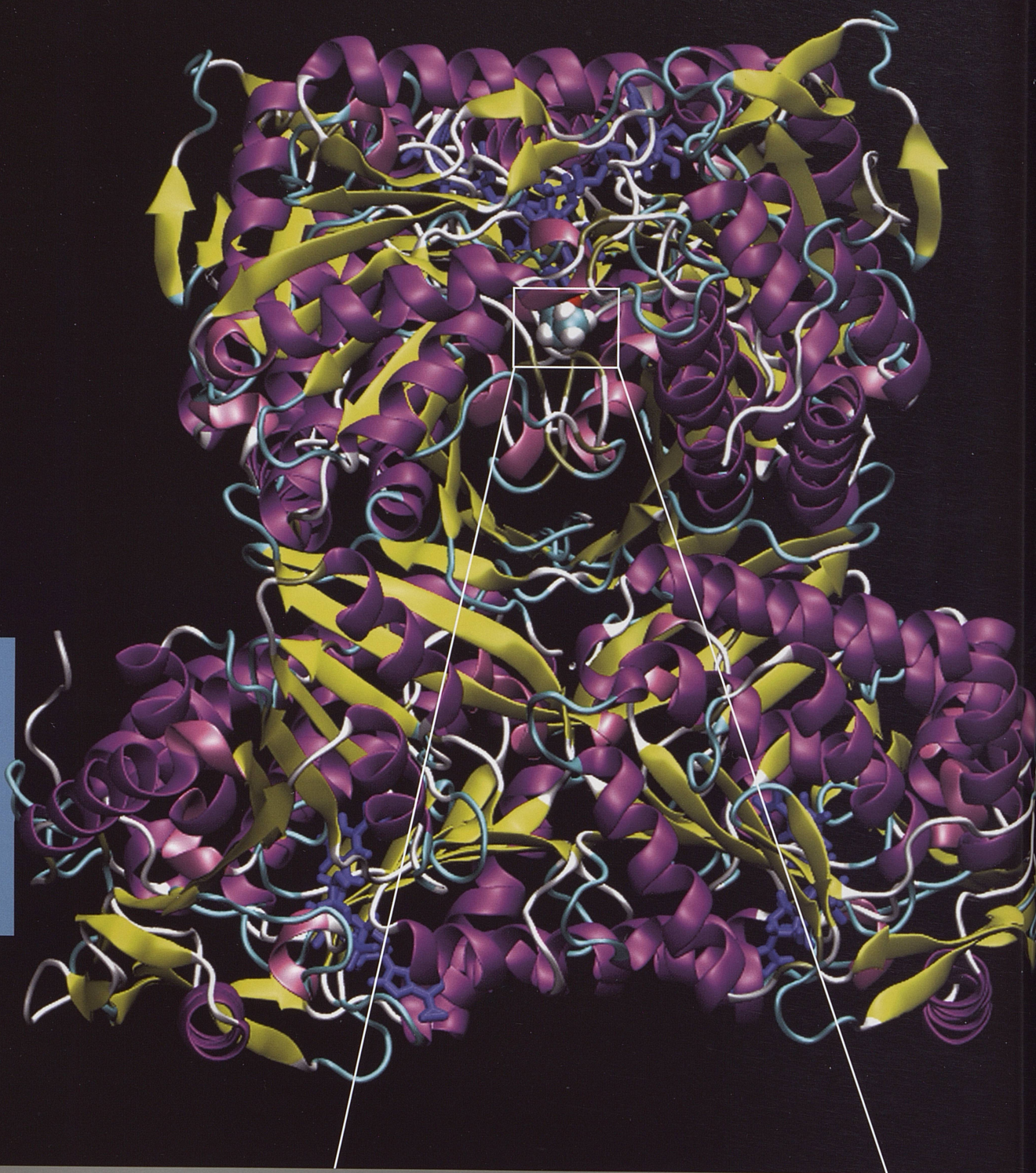
ALDHs are also involved in cancer therapy. A widely used chemotherapy drug breaks down in the body to an aldehyde, and the normal action of ALDHs interferes with the drug's ability to destroy cancer cells, requiring higher doses with hard-to-tolerate side effects. In this context, researchers would like to be able to create drugs that reduce ALDH's ability to react with aldehydes, at least in the local area of a tumor under treatment.

In both cases, what's needed is deeper understanding of how ALDH works. With the promising clue of their work on ALDH3, a complicated proton-relay mechanism in the enzyme's "active site" — where the reactions occur, Wymore decided to look at ALDH2 for similar results. Again, as with ALDH3, he employed a powerful approach called quantum mechanics/molecular mechanics (QM/MM) — a hybrid approach that tracks the movement of electrons and protons with quantum theory in the active site and uses a less computationally demanding method to keep track of the atoms in neighboring parts of the enzyme.

Wymore used a precise structure of ALDH2 obtained recently with x-ray crystallography. He worked closely with PSC scientist Shawn Brown, who modified software called DYNAMO to run QM/MM efficiently (with linear scaling) on the XT3. To take advantage of the XT3's parallelism, Brown divided the simulation — a total of 36,738 atoms — among 900 XT3 processors.

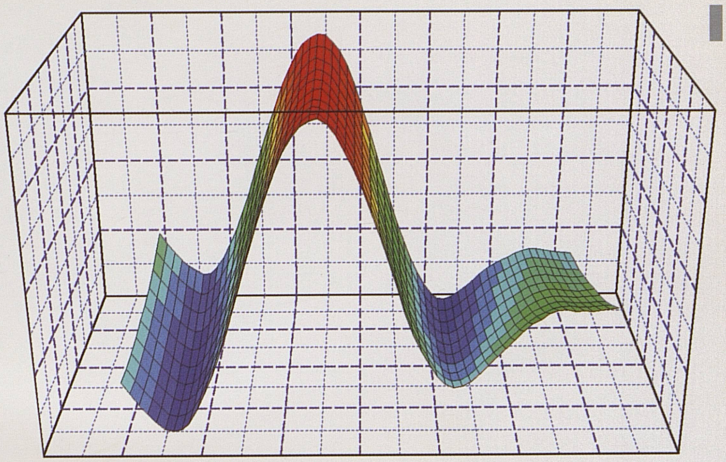
The new simulations included a "substrate" molecule — an aldehyde — and also included a co-enzyme, a molecule that sits in the active site and must be present for the reaction to occur. The co-enzyme, NAD (nicotinamide adenine dinucleotide), was expected to react with a hydrogen ion (hydride) to form NADH.

STRUCTURE OF PROTEINS & DNA //
STRANGE ACTION AT THE ACTIVE SITE



Zooming-In on the Active Site

This closeup of the active site of human ALDH2, shows the region of Wymore's QM/MM simulation in which the enzyme reaction was simulated quantum-mechanically, to capture the breaking and making of bonds as the reaction proceeds from reactants to products. The stick-figure closeup shows the NAD co-enzyme (upper right) and the aldehyde substrate (center), held in the active site by a carbon-sulfur bond and with hydrogen bonds (dotted lines) before this stage of the reaction begins. Colors represent atoms: N (dark blue), S (yellow), H (white), C (light blue), O (red).



Free Energy

This graphic is a "free-energy profile" that represents by vertical height the amount of energy required for a reaction to occur. (Colors correspond to energy, increasing blue-to-green-yellow-orange-red.) This profile shows results of Wymore's simulations of the proton transfer from ALDH's backbone (right side of plot) to an oxygen atom on the aldehyde substrate (left side of the plot), an intermediate step in the overall ALDH reaction mechanism. The front-to-back dimension corresponds to distance between a carbon atom in the aldehyde substrate and the active-site sulfur (in a cysteine amino-acid residue) to which it binds. The profile shows that the C-S bond becomes stronger (increased slope in the low-energy valley) as the proton moves from the nitrogen to the oxygen atom.

A NEW PICTURE EMERGES

The calculation ran for 9,450 XT3 processor hours. The surprising result showed a sulfur atom from ALDH2 reacting with NAD to form NAD-S, a reaction that, as far as Wymore and Hempel knew, had never been seen experimentally. Their first thought was that the ALDH2 structure from x-ray crystallography may have been flawed for their purposes, slightly shifted from its living-cell configuration in which the reaction would proceed as expected.

It was at this point, however, that Hempel saw a manuscript about a related ALDH by Sergey Krupenko's group at the University of South Carolina Medical School. These scientists reported a laboratory study of an ALDH reaction that also produced NAD-S, as the XT3 simulations predicted.

Although this laboratory work showed that NAD-S forms only in small amounts, it confirmed what Hempel and Wymore had seen in simulations. And it forced new thinking about ALDH's enzyme mechanism. In what cases did NAD-S form instead of NADH?

IT COMES DOWN TO WHETHER OR NOT A PROTON MOVES HALF AN ANGSTROM.

Using PSC's Jonas system, Wymore mounted further simulations with the aim of answering this question.
The researchers turned to Jonas because of its large memory. "We ran on Jonas," says Wymore says, "because we need 40-60 gigabytes of memory." They used 64 Jonas processors, and with data from many of these runs, four to five years of processor time, a new picture of the ALDH mechanism emerged.

"It looks like a more ordered mechanism," says Wymore, meaning that it requires a greater number of sequenced steps. "Until our work on this, the accepted thinking was that the initial reaction always happens with NAD in position to receive the hydride." Based on their new simulations, their working hypothesis is that a staged process must occur, starting with the proton transfer they found previously, for the reaction to form NADH.

NAD is located slightly away from the active site most of the time and when it sweeps in at the correct instant — after the proton transfer from the backbone stabilizes the aldehyde (which then binds with the sulfur atom) — it is converted to NADH. When this sequence of events plays out, the average ALDH enzyme does its job and converts a molecule of aldehyde into carboxylic acid.

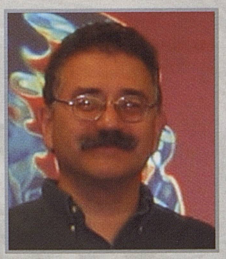
Occasional missteps in the mechanism, however, block the proton transfer, and this accounts for formation of small amounts of NAD-S, as both the simulation and experiment show. "Basically," says Wymore, "it comes down to whether or not a proton moves half an angstrom." In rare cases, mutations in ALDHs block the proton transfer and render the enzyme useless. People with Sjögren-Larsson syndrome have structural changes (mutations) in their ALDH that disrupt this transfer entirely.

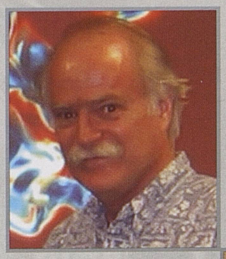
Along with the therapeutic gains that can result from deepened knowledge of the enzyme mechanism, the work by Wymore, Hempel and Brown contributes to enzyme modeling and simulation science. A goal of this research, supported by NIH's National Center for Research Resources, is to develop tools that other researchers can use to advance from simpler simulations to QM/MM studies that uncover the transient, hard-to-observe details of enzyme reactions. "It's really hard to set up these systems," says Wymore. "We're trying to develop tools to make these simulations more accurate and easier for other researchers to use. We're a long way from thorough understanding of how enzymes function." (TP)

MORE INFORMATION:
www.psc.edu/science/2007/aldh.html

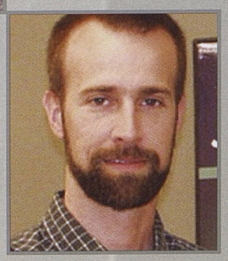
BURSTS *of* STELLAR TURBULENCE

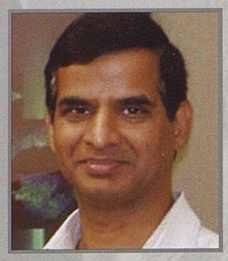
With PSC's Cray XT3, Paul Woodward and David Porter figured out how to do the hardest thing, run small to medium-sized problems really fast





David Porter and Paul Woodward





Nathan Stone and Raghurama Reddy, PSC
developed PDIO, software that routes data from the XT3 in real-time to remote users.

What if as a computational scientist you had the ability to ask "what if" questions and get answers not in days or weeks but minutes? And what if you could see the results from your questions in real time and, depending on what you see, shift your inquiry on the fly and get answers almost as quickly as you're able to process what you see. It would be like having a powerful computer wired into your inquiring mind, and your ability to ask new questions as they occur to you, based on rapid results, would transform your ability to do science.

Using PSC's Cray XT3 and with help from PSC-developed software, Paul Woodward and David Porter have been doing this, running interactive simulations of turbulence and "steering" them to explore the features that most interest them. "The XT3 has the fastest processor interconnect in a machine with thousands of CPUs," says Woodward, "and this feature is all important to enable interactive steering of flow simulations."

For many years, Woodward and Porter, astrophysicists at the University of Minnesota's Laboratory for Computational Science and Engineering (LCSE), have studied turbulent astrophysical flows. "The consistent thread in our research," says Woodward, "has been using large-scale supercomputer simulations to understand and model turbulent convection in stars."

Their ambitious long-range goal is to accurately simulate in detail the turbulent dynamics of an entire giant star, stars similar to the sun. As a step toward this, their recent work has focused on small-scale turbulence, studies that can identify parameters from which to build an accurate model of turbulence on large scales. "This need to represent the large-scale effects of small-scale turbulence," says Woodward, "arises in many areas of science and engineering, from flow in pipes and combustion engines to meteorology and ultimately to the sort of 'stellar weather' we have been simulating for many years."

Their focus on small-scale turbulence led the researchers to a significant breakthrough in massively parallel processing. With the Cray XT3, they've solved the problem of "strong scaling" — which, essentially, is getting a large number of processors to work together efficiently on a moderate-sized problem. "Getting a small problem done fast is the hardest thing," says Woodward.

In January 2007, Woodward and Porter used 4,096 XT3 processors (plus eight input/output nodes), the whole system, to simulate turbulent shear between two fluids. They used a computational grid of 576^3 cells, fine enough to resolve the small-scale turbulence they want to understand — a run that would take weeks or months on an average cluster. With performance of 2.32 gigaflops (billions of calculations per second) on each XT3 processor, 9.5 teraflops overall, the run of 6,000 time steps completed in 7.7 minutes.

The drop in per-processor efficiency in using the whole machine was less than 5-percent — a credit to the XT3's interprocessor communication. "That's pretty damn amazing," says Woodward. "It's a testimonial to the effectiveness of your Cray XT3 interconnect."

Although impressive, running a problem of this size in eight minutes, Woodward acknowledges, is overkill. The real significance lies in the potential it demonstrates. "You can take that and say 'Suppose we do a bigger problem not by assigning more work to each processor but by having more processors.' That leads to petascale computing that runs really well on reasonably sized grids. The principle challenge of petascale computing is strong scaling."

Shear Between Two Gases

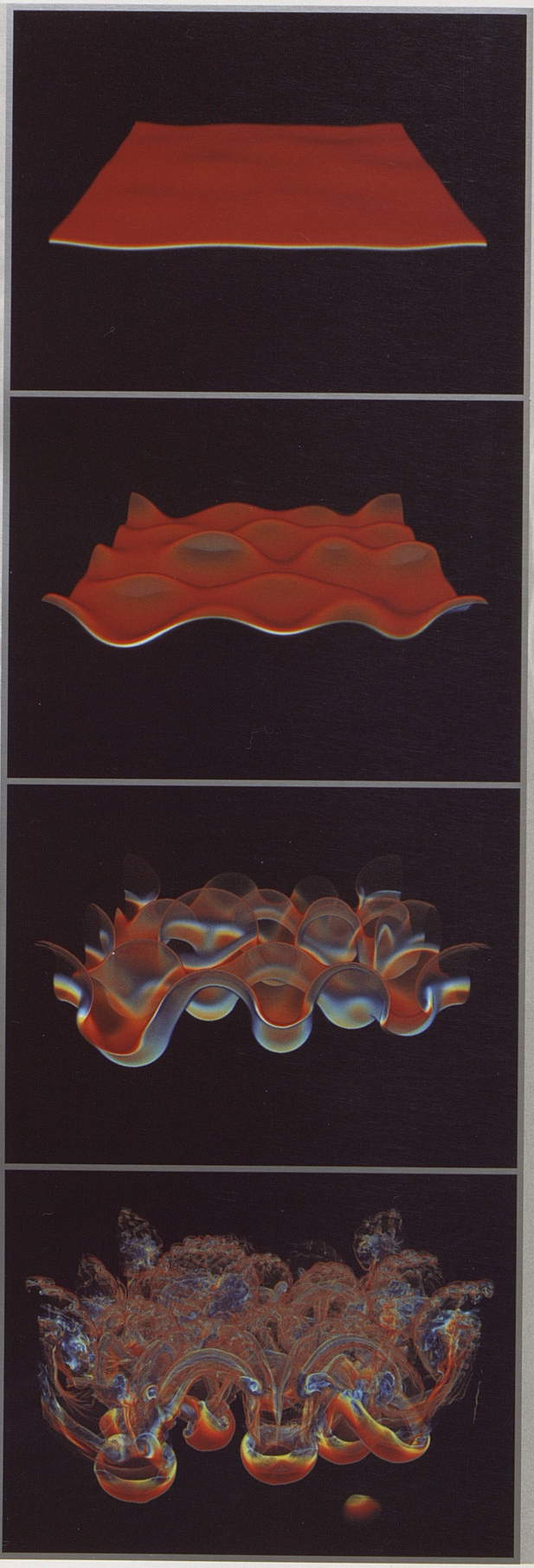
From simulation of turbulent mixing between initially isolated layers of two gases of different density traveling across each other at Mach 1/2, these graphics show the "shear layer" that forms between them. The lower fluid is 2.5 times more dense. Only mixtures are visible, with regions of pure denser or lighter fluid transparent. Color goes through yellow to red where the denser fluid predominates and aqua to blue where the lighter fluid predominates.

CONVERGING ON TURBULENCE IN GIANT STARS

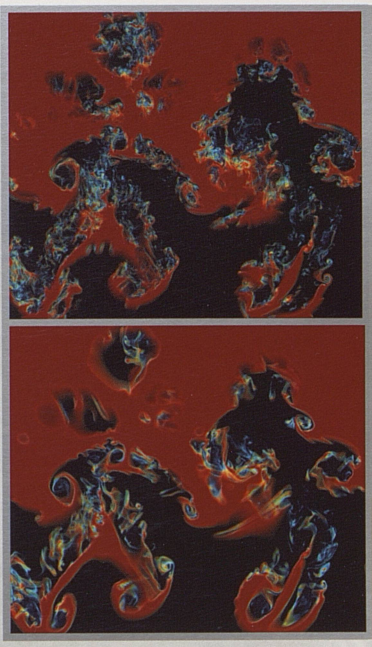
As an ultimate goal, Woodward and Porter want to simulate an entire giant star, which means modeling the swirling, pulsating bands of gas that surround a star's thermonuclear core. These volatile turbulent phenomena underlie and govern the birth and death of stars, which provide not only the light energy that sustains life, but also the elements that are its constituents. The primary means scientists have to learn about these processes is computational simulation.

As a necessary step toward simulating an entire giant star, their recent work with PSC's XT3 has focused on "subgrid-scale" turbulence. To radically simplify, subgrid-scale means turbulence that happens in very small spatial dimensions compared to the large domain of the overall problem, but which nevertheless affects the large scale and must be accounted for in a large-scale model.

With subgrid-scale turbulence, their aim is to run simulations at fine enough resolution (large quantity of grid cells) that the results correspond to what happens in nature. Their test is "convergence" — getting results that no longer significantly change (as measured statistically) as the grid becomes progressively more fine.



One kind of turbulent mixing, known as Rayleigh-Taylor instability, occurs due to gravity when a heavy gas is on top of a lighter one. This sequence shows four time intervals from an XT3 simulation (on a 768^2 x 1536 grid), with regions of predominantly heavier fluid (red), lighter fluid (blue), more even mixture (yellow), and 50-50 (white). By the time of the fourth image, the mixing, though confused, shows descending "mushroom caps" that produce organized very fine-scale mixing.



This thin slice (above left) through the mixing domain at a late time interval from an R-T instability simulation on a fine grid (768^2 x 1536 cells) shows striking agreement in larger features with the same slice (below left) from the same simulation on a coarser grid (384^2 x 768). The qualitative differences are minor, which suggests convergence — that all reasonable simulation codes would arrive at similar conclusions. "If so," says Woodward, "we may be confident in the simulated results and use them to validate subgrid-scale models of turbulent multifluid mixing."

several ways that they demonstrated first at conferences, iGrid2005 in San Diego and again at SC05 in Seattle.

To do simulations in real time as exploratory runs — to ask "what if?" questions on the fly — was a new capability in turbulence research. To make it possible, Woodward and Porter worked closely with PSC scientists Nathan Stone and Raghurama Reddy, who developed specialized software — called Portals Direct I/O (PDIO) — to route simulation data from the XT3 in real time to remote users.

As numbers crunch in Pittsburgh, PDIO assembles the resulting data streams and routes them for Woodward and Porter to volume-render and display as real-time images. "The ability to have instant response from supercomputer simulations is very useful," says Woodward. "You can, for instance, change the Mach number, and almost immediately see what that does. This is a whole pipeline of utility programs that we tied together in an automated fashion. The support from PSC is outstanding."

TO ASK "WHAT IF?" QUESTIONS ON THE FLY IS A NEW CAPABILITY IN TURBULENCE RESEARCH.

PETASCALE & STRONG SCALING

To appreciate the leap to strong scaling, it helps to know about its opposite, weak scaling. Over the next few years, supercomputing infrastructure will evolve to "petascale" — the ability to do a quadrillion (10^{15}) calculations per second. This amount of computing — which will enable major advances in many areas of science — will come from systems comprising tens or hundreds of thousands of processors linked with each other so that scientists can use them in teams to address very large problems. In computational lingo, this is "scaling" — as you "scale up" a computational problem, you apply more processors.

Weak scaling is the easy approach, at least relatively speaking. It means, basically, that as you scale up you make the problem larger, but each processor does the same amount of computing. If you double the size of a 3D grid in each dimension, for instance, you need eight times more processors.

Strong scaling, on the other hand, as Woodward says, is not easy. It means that for a given problem as you scale up, say, from 100 to 800 processors you apply this greater number of processors to the same grid, so that each processor is now doing 1/8 as much work. You would like the job to run eight times faster, and to do that requires restructuring how the program divides the work among processors and increases the communication between them.

THE XT3 & RESTRUCTURED CODE

Woodward and Porter appreciated the XT3 — fast processors with a fast interconnect — from its inception as a TeraGrid resource in 2005, and worked with it from early on. Using their turbulence code, PPM (Piecewise Parabolic Method), they developed the ability to do interactive, steerable runs remotely with real-time visualization — a breakthrough in

With real-time interactive capability in place, over the past year Woodward and Porter extended their effort to rapidly solve small-scale problems and made a dramatic leap when they found a way to restructure their PPM code. The happy discovery came as result of working with the Cell processor, a multi-core processor — multiple processors on one "chip" — with much potential for supercomputing applications.

The approach they arrived at, says Porter, is "maximum cache reuse" — not easy to explain, but it has to do with "vector length," strings of identical mathematical operations. From early days of supercomputing, code performance was optimal with vector-lengths as long as possible. With the Cell processor, however, Woodward and Porter found — to their surprise — that shortening the vector-length to conserve cache also speeded up performance.

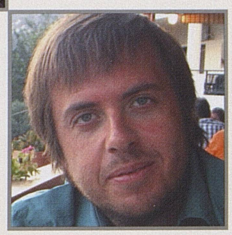
"Over the past year," says Woodward, "we have completely redesigned our PPM simulation codes, and improved performance — about one gigaflop per processor — by dramatically reducing the data granule on which the processors operate." The performance improvement depends on the XT3's fast interconnect, because maximum cache reuse makes increased demands of inter-processor communication. The combination of restructured code and the XT3, as Woodward and Porter demonstrated with 9.5 teraflops sustained on 4,000 processors, is a big step in the direction of petascale.

MORE INFORMATION:
www.psc.edu/science/2007/turbulence.html

MYSTERY OF THE MISSING GALAXIES

Simulations show how the darkest galaxies in the universe get that way and appear to resolve a serious problem with the prevailing theory of how the universe evolves

Stelios Kazantzidis



Lucio Mayer

Who knew? The Milky Way, our own galaxy, is a neighborhood bully. For 10 billion years, it's been picking on small galaxies that pass near, huffing and puffing and blowing away their interstellar gas.

Left with only a few stars and nearly emptied of the gas they need to form new ones, these small galaxies become ghostly shadow galaxies, composed almost entirely of "dark matter" — the mysterious invisible matter that has a huge effect on the structure of the universe. They are the "darkest" galaxies known to exist, so faint that — although it's likely they are speckled throughout the universe — only a few have been detected, all satellite galaxies of the Milky Way and nearby Andromeda.

Until recently, scientists lacked an explanation for how these dark galaxies come into being that could account for both their exceptional darkness and their nearness to much larger host galaxies. With innovative thinking and powerful supercomputing tools to test their ideas, a team of physicists led by Lucio Mayer of the University of Zürich and Stelios Kazantzidis of Stanford University appear to have solved this riddle.

Mayer, Kazantzidis and their colleagues — Chiara Mastropietro (University of Munich) and James Wadsley (McMaster University, Canada) — used supercomputers in Zürich and relied heavily on LeMieux, PSC's terascale system (now decommissioned), for simulations over a two to three-year period. Their findings, reported this February in *Nature*, show that a triad of physical processes — ultraviolet radiation, gravity and the pressure exerted by galactic gas — transform a normal gas-dominated, small galaxy into a dark-matter shadow of its former self. Their solution to the puzzle of very small, dark galaxies — called "dwarf spheroidals" because of their size and rounded shape — has large implications for what we know about the universe. The leading theoretical model for evolution of the cosmos, the cold dark-matter model (CDM), predicts that dwarf galaxies are the building blocks of large galaxies like the Milky Way and that these dwarves should be ubiquitous. Scientists, however, have so far detected very few of them, and this mismatch between theory and observation — often called the "missing satellites problem" — has been a major weakness of the CDM model.

Mayer and Kazantzidis' simulations suggest strongly that the missing satellites are dwarf spheroidals not yet detected. "There has been this problem," says Mayer, "that when you look at the number of luminous galaxies surrounding the Milky Way you see fewer than CDM theory predicts. Many people have used this to say that probably the CDM model is wrong. With this new work, we're saying, 'Look, these small faint galaxies exist, we simply haven't found them yet.'"

THE KEY INGREDIENT: RAM PRESSURE

Before Mayer and Kazantzidis took aim at the problem, the prevailing theory for the origin of dwarf spheroidals was that diffuse radiation — the "cosmic ultraviolet background" — from galaxies like the Milky Way evaporated the gas in small nearby galaxies. Detailed calculations, however, made it clear that UV evaporation by itself can't cause the gas loss from dwarf galaxies like those near the Milky Way. "The biggest galaxy that can undergo this kind of evaporation," says Mayer, "is about 10 times smaller than the dwarf galaxies we see."

Earlier work of Mayer's on dwarf galaxies gave him a clue that other forces might be involved. He and his colleagues wanted to understand why dwarf galaxies close to the Milky Way were spheroidal while those at greater distance are disk-shaped. The answer, they found, is gravity. Because the distant dwarves feel less gravitational pull from the Milky Way they maintain their disk structure, while closer dwarves that begin as disks morph into spheroids.

Could other interactions between large and much smaller galaxies, Mayer began to wonder, also affect other dwarf properties, such as gas and dark matter content? "I realized that a major ingredient was missing from the physics," says Mayer, "and that was ram pressure, the 'sweep' produced by gas around the big galaxy that blows away gas in the small galaxy." Similar to drag that an airplane experiences as it flies in Earth's atmosphere, ram pressure is force from a large galaxy's envelope of hot gas that pushes against a smaller galaxy moving through it. "This kind of mechanism has been known for awhile," says Mayer, "as an effect in galaxy clusters, where you have thousands of galaxies. But it had been overlooked for our own galaxy and its satellites."

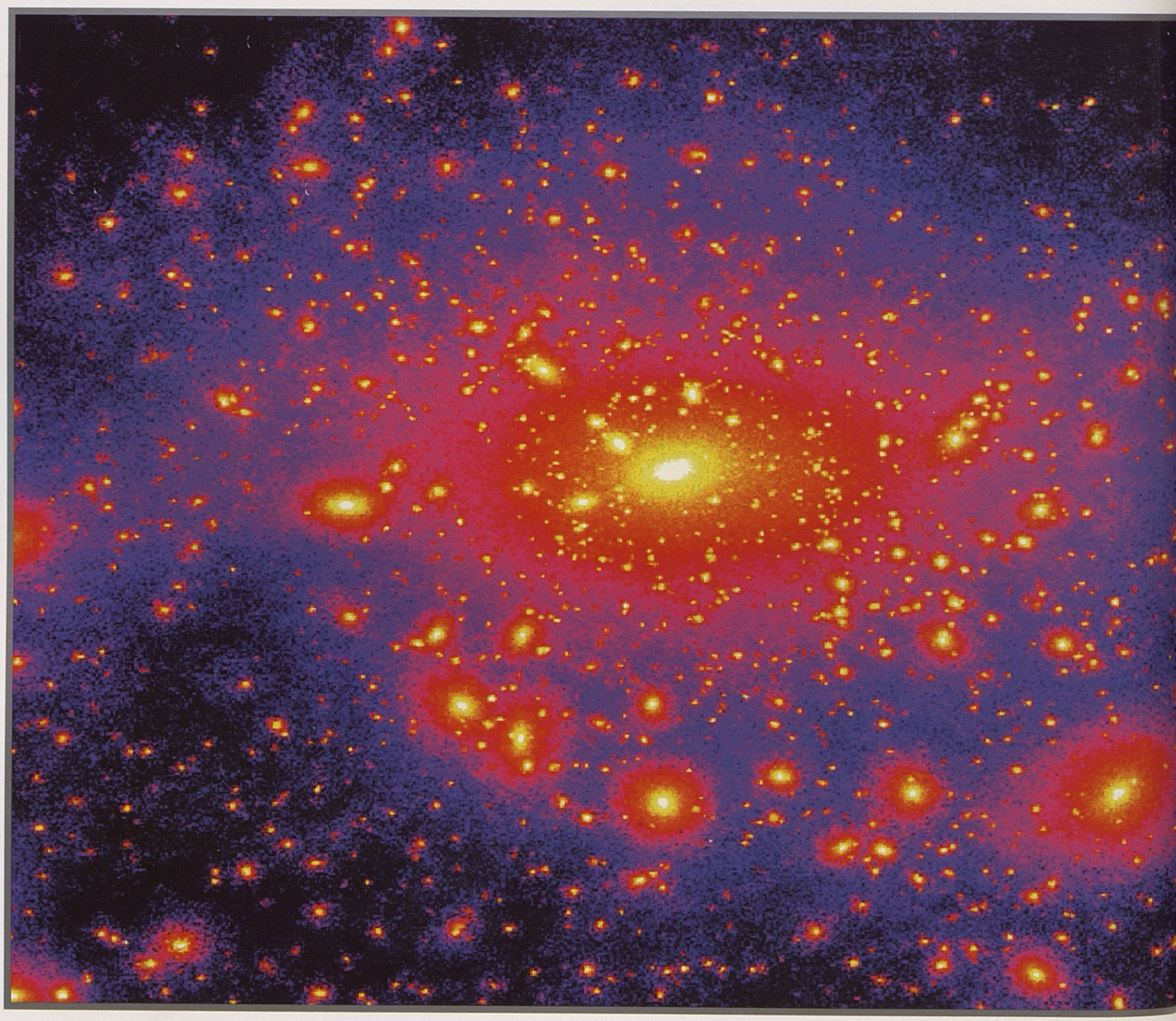
GONE WITH THE WIND

With ram pressure as a big clue to go on, Mayer, Kazantzidis and their colleagues revised a cosmological modeling program called GASOLINE, which they initially developed in the 1990s at the University of Washington with astrophysicist Thomas Quinn. Starting as a gravitational model, GASOLINE has expanded and become more diverse as the demand arose to include more physics. Wadsley added an approach called smoothed-particle hydrodynamics (SPH), which describes fluid properties, like interstellar gas, using millions of particles as tracers. Mayer contributed radiative cooling and heating.

The emphasis was on flexibility, so GASOLINE could run on a range of supercomputers. This led to its being used on projects as diverse as planet formation, asteroid collisions, and now dark matter in spheroidal dwarf galaxies. "Our work relies on fast, reliable supercomputing systems," says Kazantzidis. "GASOLINE is optimized to run efficiently on LeMieux, and this system has been ideal for these simulations."

The researchers added the physics of ram pressure to GASOLINE. They also accounted for tidal forces — like those that pull Earth's oceans toward the moon and create tides. The gravitational pull of the Milky Way similarly distorts nearby dwarf galaxies, stretching their gas and making it less dense, more easily acted on by ram pressure. The researchers also kept the physics of the cosmic ultraviolet background in the model. Although a relatively minor effect, the UV background heats the gas, keeping it loosely bound and thus more easily swept away by tides and ram pressure.

To represent a spheroidal dwarf galaxy and the outer portion of the Milky Way, the simulation included 3.2 million SPH particles and tracked them through 10 billion years of galactic evolution. For the final calculations, GASOLINE ran on 64 LeMieux processors for roughly 30,000 processor hours. Over a period of years, the dwarf-galaxy project used 150,000 to 200,000 LeMieux processor hours.



MERGING GALAXIES FORM PAIRED SUPERMASSIVE BLACK HOLES

In another project simulating galaxy interactions, Mayer, Kazantzidis, Thomas Quinn and colleagues relied on GASOLINE and PSC's LeMieux to arrive at the first simulated formation of paired supermassive black holes, a binary SMBH. This work, reported in *Science Express* (June7, 2007), simulates a Milky Way-size galaxy merging with its twin, forming a new type of structure — a central disk of gas from a hundred to a few thousand light-years wide and from a few hundred to a billion solar masses.

The simulations followed these processes over a range of spatial and temporal scales, finding that when this gas disk forms the SMBHs in most cases form a binary system. Earlier simulation studies of black-hole binaries, says Kazantzidis, have been restricted to late stages of the merger, after the black holes had already formed a loose pair. Understanding of binary black-holes is critical in testing the existence of gravitational waves, as predicted by Einstein's General Theory of Relativity. (MS)

The results were conclusive. As gas-rich progenitors of dwarf spheroidals fell into the Milky Way 10-billion years ago, radiation from the cosmic UV background heated the small galaxy's gas. The process of tidal shocks shakes away some of its stars and stretches and loosens the gas density. At the same time, ram pressure strips away the small galaxy's gas. After 10-billion years a dwarf spheroidal galaxy has replaced the gas-rich progenitor. "The winds are so strong in the universe around galaxies like the Milky Way," says Mayer, "that they completely blow away the gas from the small galaxy."

ASTRONOMERS ARE ALREADY LOOKING FOR MORE DARK GALAXIES LIKE THOSE PREDICTED BY THIS WORK.

All that remains are the few stars that had already formed in the center of the dwarf galaxy and a huge halo of dark matter. The dark matter stays because ram pressure and UV radiation don't affect it. This is one of the mysterious properties of dark matter — it interacts gravitationally with matter we see, but not in other ways, such as magnetically or chemically. The size of the spheroidal dwarf galaxies predicted by this simulation matches those of observed dwarf galaxies, lending further credence to the model.

"Due to the combined action of three environmental effects — ram pressure, tidal shocking and the cosmic ultraviolet background," says Kazantzidis, "a dwarf galaxy transforms into a dwarf spheroidal. No previous work has been able to elucidate the combined effect of these various environmental mechanisms on the structure of dwarf galaxies." Observational astronomers are already looking for the tiny, dark galaxies predicted by this work. If found, they will lend new force to the CDM model that successfully explains many aspects of cosmic evolution.

Because dwarf spheroidals are the most dark-matter dominated galaxies in the universe, they are ideal targets, says Kazantzidis, for experiments that try to identify the elusive dark-matter particle. "Dark matter is one of the grandest challenges of modern-day science," he says. "This new understanding of the darkest galaxies known may lead to fundamental insights into the nature of dark matter." (TP)

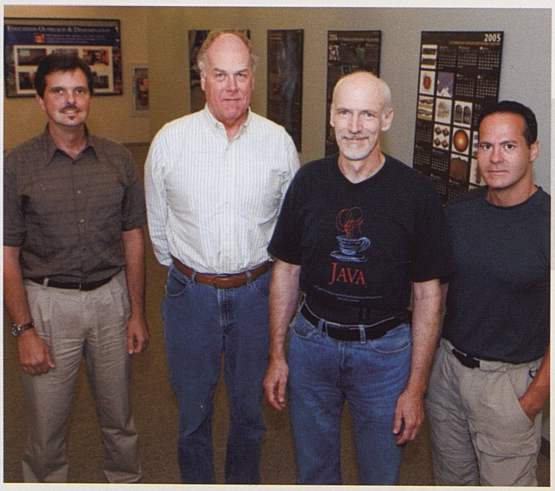
MORE INFORMATION:
www.psc.edu/science/2007/darkgalaxies.html

In this image from Mayer and Kazantzidis' simulation, brightness (blue-to-violet-to-red-to-yellow) corresponds to increasing concentration of dark matter. The bright central region corresponds roughly to the Milky Way's luminous matter of gas and stars. The blue outline represents a boundary of the galaxy's dark-matter halo, with dark-matter satellites shown as bright clumps. The simulation predicts that such dark-matter halos are "lumpy," filled with hundreds of small satellites of dark matter, a finding that presents a potential resolution to the "missing satellites problem" of the CDM model of the universe.

SCIENTIFIC VISUALIZATION

VISUALIZING OCEANS, QUAKES & BLOOD

PSC computational scientist consultants work closely with discipline scientists to provide tools for visualization and analysis



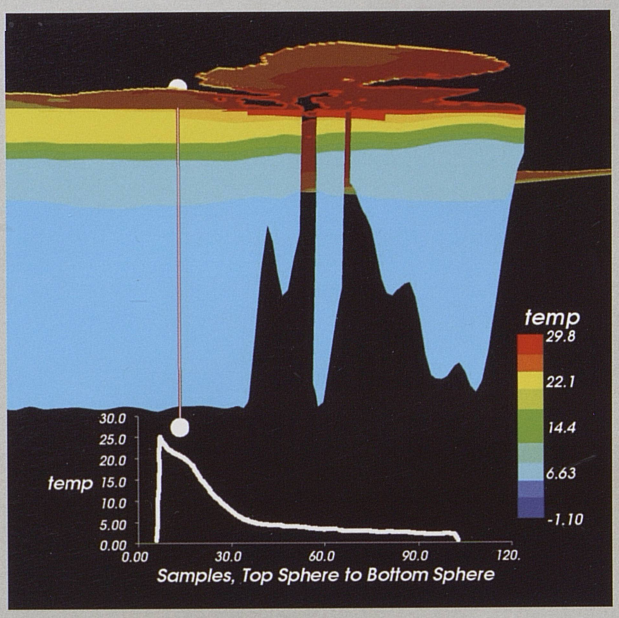
PSC scientists **Nick Nystrom,** director of strategic applications, **Kent Eschenberg, Joel Welling** and **John Urbanic,** who have consulted and provided visualization support for a number of research projects at PSC.

When it comes to a scientist's desire to simulate physical phenomena in realistic detail, there may be no such thing as "too much information." To do simulations with massively parallel computational engines like PSC's Cray XT3, nevertheless, brings enormous challenges in analysis and interpretation of the information produced. Regardless of the target problem — fluid dynamics of blood flow in an artery, galaxy formation in the early universe, evolution of tornados, and many others — such computational experiments often produce terabytes of data.

How much information is that? One terabyte is a trillion bytes. In printed form, that's about 50,000 trees worth of paper. Ten terabytes would represent the entire printed contents of the U.S. Library of Congress.

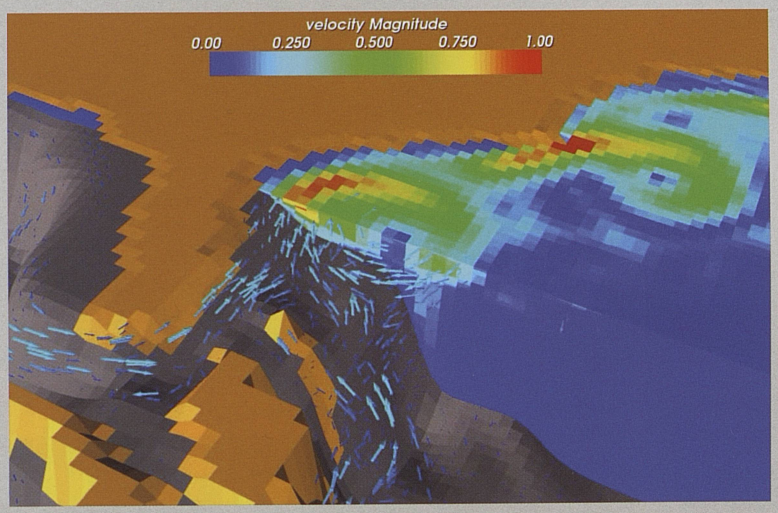
How is it possible to assimilate, analyze, and arrive at conclusions from such mountains of information? A large part of the answer is "scientific visualization" — a range of software tools that can manipulate data from simulations and present it in visual form. Looking at images rather than piles of numbers, scientists rely on the brain's natural ability to interpret visual information. Much as if they are observing a phenomenon in nature, they can see what happens in their simulation and extract insights in seconds that might otherwise take days or weeks.

PSC scientist consultants provide visualization support for many projects that use PSC resources. They often collaborate closely with researchers to devise the best means to represent data so that it will facilitate interpretation and analysis. Within the past year, PSC consultants have provided this kind of support for a number of major projects including ocean modeling, earthquake soil vibration, and turbulent blood flow in human arteries.



The PSC HYCOM Reader

A view of the Gulf of Mexico from underneath, with 3D perspective showing the ocean surface, where water is warmest, and a vertical cross-sectional slice showing temperature (increasing from blue to red) as it varies with depth (the vertical dimension is stretched to aid observation). Areas of no color (black) represent land mass, including the island of Cuba. Sampling along a line between top and bottom spheres (white) generates the plot of temperature versus distance.



Deep Currents

This section of the Atlantic Ocean off the U.S. east coast shows velocity magnitude by color (increasing from blue to red) and velocity direction with arrow-shaped glyphs, that allow visualization of the 3D features of the ocean current.

OCEANS AND CLIMATE

A team of researchers from the University of Miami Rosensteil School of Marine and Atmospheric Science and the Naval Research Laboratory are using PSC's Cray XT3 to simulate ocean climate variability. Their focus is the Atlantic Ocean. Using large-scale models that run for months of XT3 processing time, they simulate the ocean over decades, with the goal of developing realistic models that can couple with atmospheric models and be used to forecast climate change.

Building on a 13-year PSC history of collaboration in ocean modeling, PSC scientist consultants John Urbanic and Kent Eschenberg worked with principal investigator Zulema Garraffo and her colleagues George Halliwell and Alan Wallcraft to develop a 3D visualization capability for their Hybrid Coordinate Ocean Model (HYCOM). Although ocean currents are inherently 3D, HYCOM has until this year relied entirely on 2D contour plots.

Drawing from the Visualization ToolKit (VTK), an open source system for 3D graphics, and ParaView (Parallel Visualization Application), an open-source application that supports parallel systems and large datasets and that provides a user interface for VTK, Eschenberg created a "reader" for HYCOM's hybrid coordinate grid format. Garraffo and her colleagues are integrating PSC's HYCOM reader into an automated workflow for remote visualization of simulation data.

Zulema Garraffo
University of Miami
Rosensteil School of Marine and Atmospheric Science

VISUALIZING OCEANS, QUAKES & BLOOD (continued)

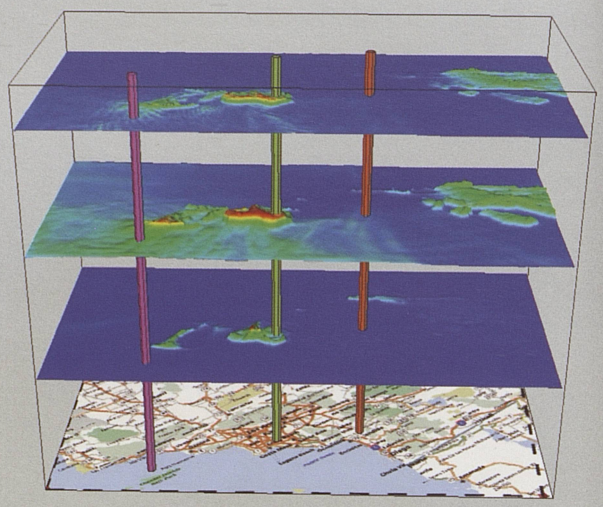
SIMULATED QUAKES

Jacobo Bielak and his colleague David O'Hallaron of Carnegie Mellon University lead the Quake Project, a large collaborative research team that uses sophisticated computational methods to create realistic 3D models of earthquakes. In collaboration with the Southern California Earthquake Center (SCEC), their work aims to provide information that will result in seismic provisions in building codes that will ensure the safest possible structures at reasonable cost.

In November 2006, the Quake team won the HPC Analytics Challenge Award at SC06 in Tampa for "Hercules" — software that coordinates all the stages of large-scale earthquake simulation, from problem definition to final visualization. With this unified framework, all tasks — building a software mesh that subdivides the quake region, partitioning the job among hundreds or thousands of processors, the simulation itself, and visualizing results — are performed in place on the XT3. Relying on software called PDIO developed by PSC staff (see p. 35), Hercules can visualize results in real time as a simulation is running.

In 2007, PSC worked with the Quake group and SCEC to develop tools that compare results between Hercules and TeraShake, earthquake simulation software developed at SCEC using a technique that differs from Hercules. To show how and why results from the two programs vary is important for validation of the models. It requires sophisticated statistical tools, which PSC scientist Joel Welling has developed, along with a comparison viewing capability provided by PSC's Kent Eschenberg, and a new Quake reader for ParaView optimized for their files.

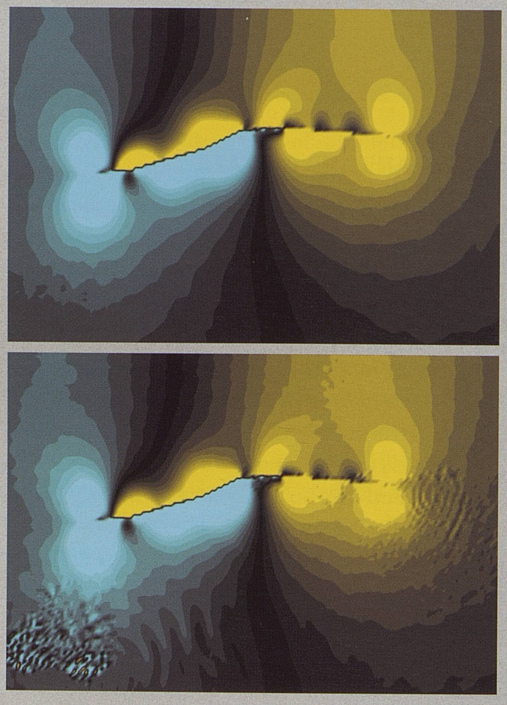
Jacobo Bielak (left) and **David O'Hallaron**



Statistical Comparison

The top two planes of this display, developed by Welling, show a "phase mismatch" and an "envelope mismatch" between two simulations, while the third plane represents a difference between the first two planes.

"Joel Welling's statistical comparison toolset," says Bielak, "provides an outstanding means for assessing the quality of our seismic synthetic datasets against actual earthquake records and for comparing results from different simulation techniques. It greatly facilitates our verification and validation efforts."



Soil Displacement

This graphic shows the displacement from soil vibration in a Hercules run (above) compared to the same region from a TeraShake run (below).

Aortic Arch

Flow through the aortic arch and branching into the brachiocephalic, carotid and other downstream arteries, showing flow velocity (increasing from blue to red).

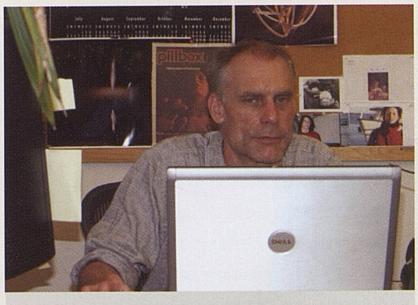
THE ARTERIAL TREE

To model blood flow interactions in different regions of the human cardiovascular system is the goal of research led by George Karniadakis and Leopold Grinberg of Brown University. Their applied mathematics group at Brown, the CRUNCH group, develops algorithms, visualization methods and parallel software for simulations in fluid mechanics. They aim to establish a biomechanics gateway on the TeraGrid with the arterial tree as a simulation framework for research in hemodynamics, disease and drug delivery.

To solve the very complex 3D flow problems of the arterial tree — complicated by many branches and outlets — Grinberg uses NekTar, a program developed at Brown, applied to realistic 3D geometry of the arterial system reconstructed from MRI. Grinberg demonstrated remote visualization with NekTar at SC06 in Tampa. Using PDIO, developed at PSC, to stream data from PSC's XT3 to a remote desktop computer led to a 100-fold speedup in display, compared to the standard data-transfer technique. Such interactive visualizations are a convenient way to present results of blood-flow simulations to doctors.

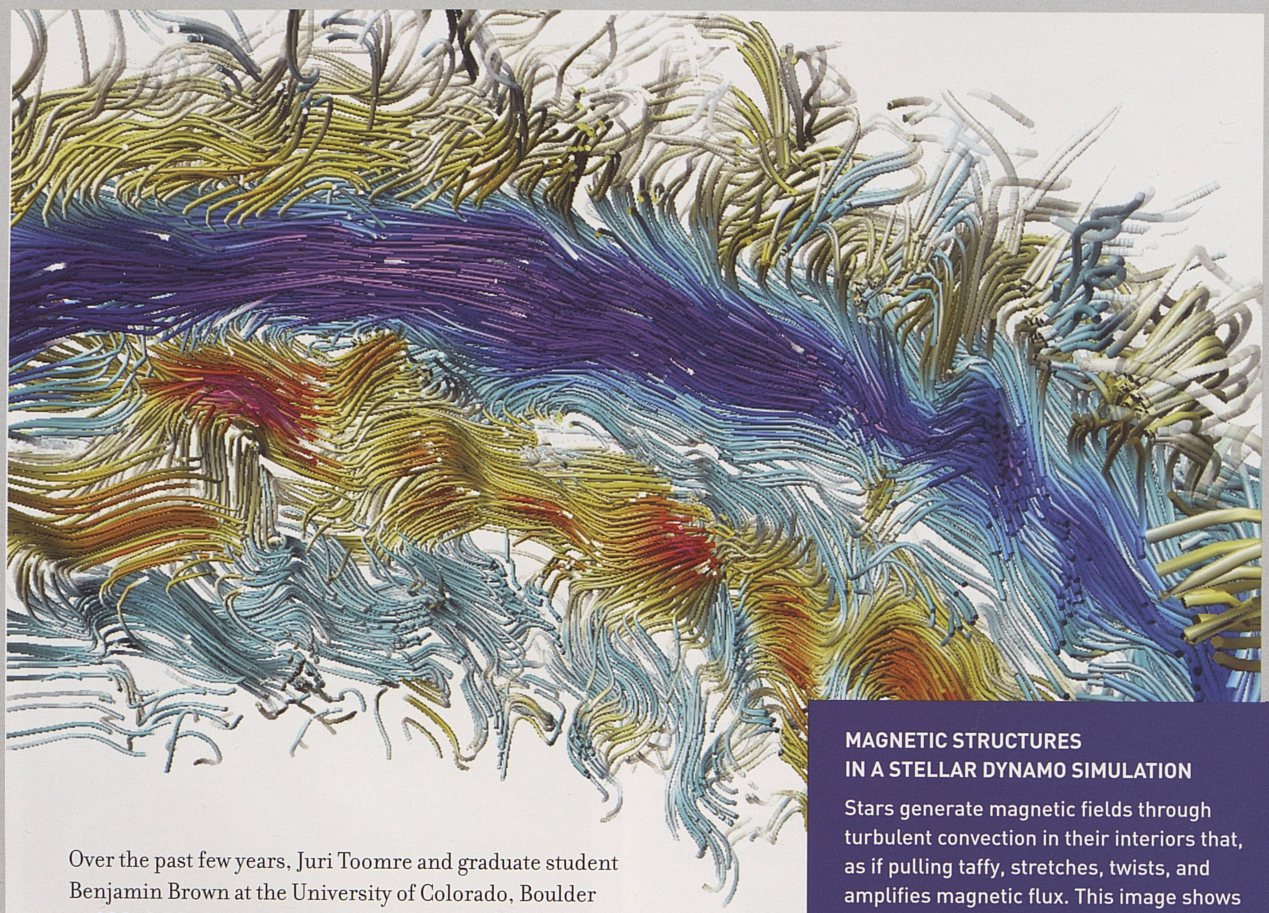
PSC consultants collaborated with Grinberg to help in visualizing fast solution of NekTar simulations, which can involve as many as a billion degrees of freedom and produce up to 1.5 terabytes of data per run. Recent PSC optimizations reduce the overall data volume by more than 80-percent while preserving resolution of the visualization. "With these improvements," says Grinberg, "what would take two hours takes only seconds."

"WITH THESE IMPROVEMENTS, WHAT WOULD TAKE TWO HOURS TAKES ONLY SECONDS."



Greg Foss, PSC scientific visualization specialist, who provides visualizations for NekTar simulations.

SOLAR DYNAMO



MAGNETIC STRUCTURES IN A STELLAR DYNAMO SIMULATION

Stars generate magnetic fields through turbulent convection in their interiors that, as if pulling taffy, stretches, twists, and amplifies magnetic flux. This image shows magnetic lines of force in the convection zone of a rapidly rotating star, with colors (red positive, blue negative, white neutral) indicating direction. (Created by Ben Brown using VAPOR.)

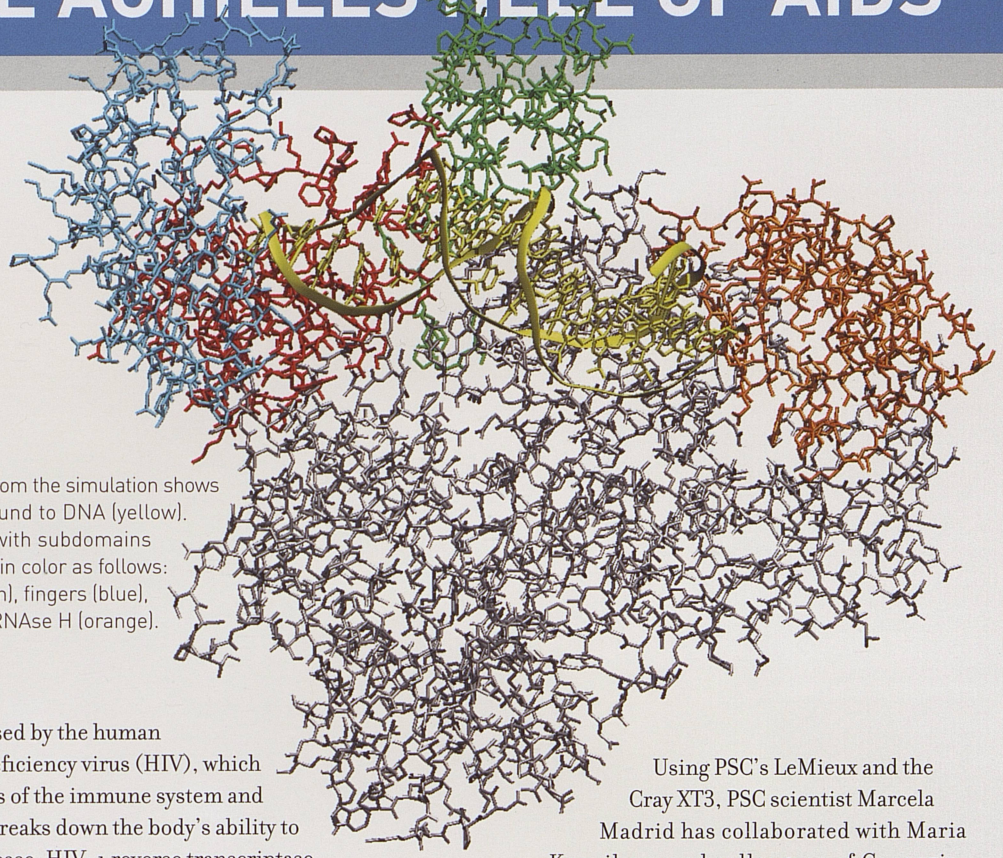
USING THE CRAY XT3 RESEARCHERS EXPLORE GIANT-CELL CONVECTION PATTERNS IN THE SUN

Over the past few years, Juri Toomre and graduate student Benjamin Brown at the University of Colorado, Boulder and Mark Miesch at the National Center for Atmospheric Research (NCAR) have used PSC's LeMieux and BigBen for very large-scale simulations of convection in the deep interior of stars. The convection patterns they model, known as giant cells, influence solar magnetic storms that can affect satellites and electrical systems on Earth.

In the solar interior, convection occurs as hot plasma rises and cooler, more dense plasma sinks. In this convection zone, from about 70-percent of the solar radius outward to the surface, scientists suspect that churning masses of plasma — giant cells — induce a global circulation pattern, moving plasma from the solar equator toward the poles near the surface, and then back to the equator at greater depth. This circulation generates magnetic fields that cause the 11-year sunspot cycle. Better understanding of this cycle and reliable prediction of associated solar storms would help to protect valuable assets in space and on Earth.

Using as many as 2,048 BigBen processors, the researchers modeled the sun's magnetism and internal rotation with a program they developed called ASH (Anelastic Spherical Harmonic). Terabytes of data produced by these simulations reside on disk storage at PSC, and the researchers analyze the data remotely via the TeraGrid network using software developed at NCAR called VAPOR. Their findings, forthcoming in the *Astrophysical Journal*, indicate that at low solar latitudes, plasma sinks along north-south corridors. At higher latitudes, rising and falling areas of plasma meet and create solar cyclones that last for days. "Understanding how stars build magnetic fields," says Miesch, "is a fundamental and enduring mystery that high-performance computing is helping to uncover."

THE ACHILLES HEEL OF AIDS



This image from the simulation shows HIV-1 RT bound to DNA (yellow). RT is silver with subdomains represented in color as follows: thumb (green), fingers (blue), palm (red), RNAse H (orange).

AIDS is caused by the human immuno-deficiency virus (HIV), which invades cells of the immune system and eventually breaks down the body's ability to fend off disease. HIV-1 reverse transcriptase (RT), a multi-functional protein that is part of HIV, essentially copies-and-pastes HIV's DNA, which is then incorporated into immune system cells of the infected person.

Interfering with RT's function would shut down HIV's reproductive capability, which explains why RT is the target of several FDA-approved anti-AIDS drugs.

DETAILED STUDIES OF HIV-1 RT COULD PROVIDE CLUES TO KNOCKING OUT AIDS

A precise, clear understanding of how RT works could make it possible to design more effective anti-AIDS drugs that could give a knockout punch to this worldwide scourge.

Using PSC's LeMieux and the Cray XT3, PSC scientist Marcela Madrid has collaborated with Maria Kurnikova and colleagues of Carnegie Mellon University to carry out molecular dynamics simulations of RT in the absence of DNA, as well as RT interacting with DNA. After long simulation times (beyond five nanoseconds), they observed that regions of RT that resemble the fingers and thumb of a hand tend to move apart — important for accommodating a large molecule like DNA. The DNA undergoes a "twist and slide" motion, which may facilitate its positioning at RT's active site. Interfering with this motion could disrupt RT's function.

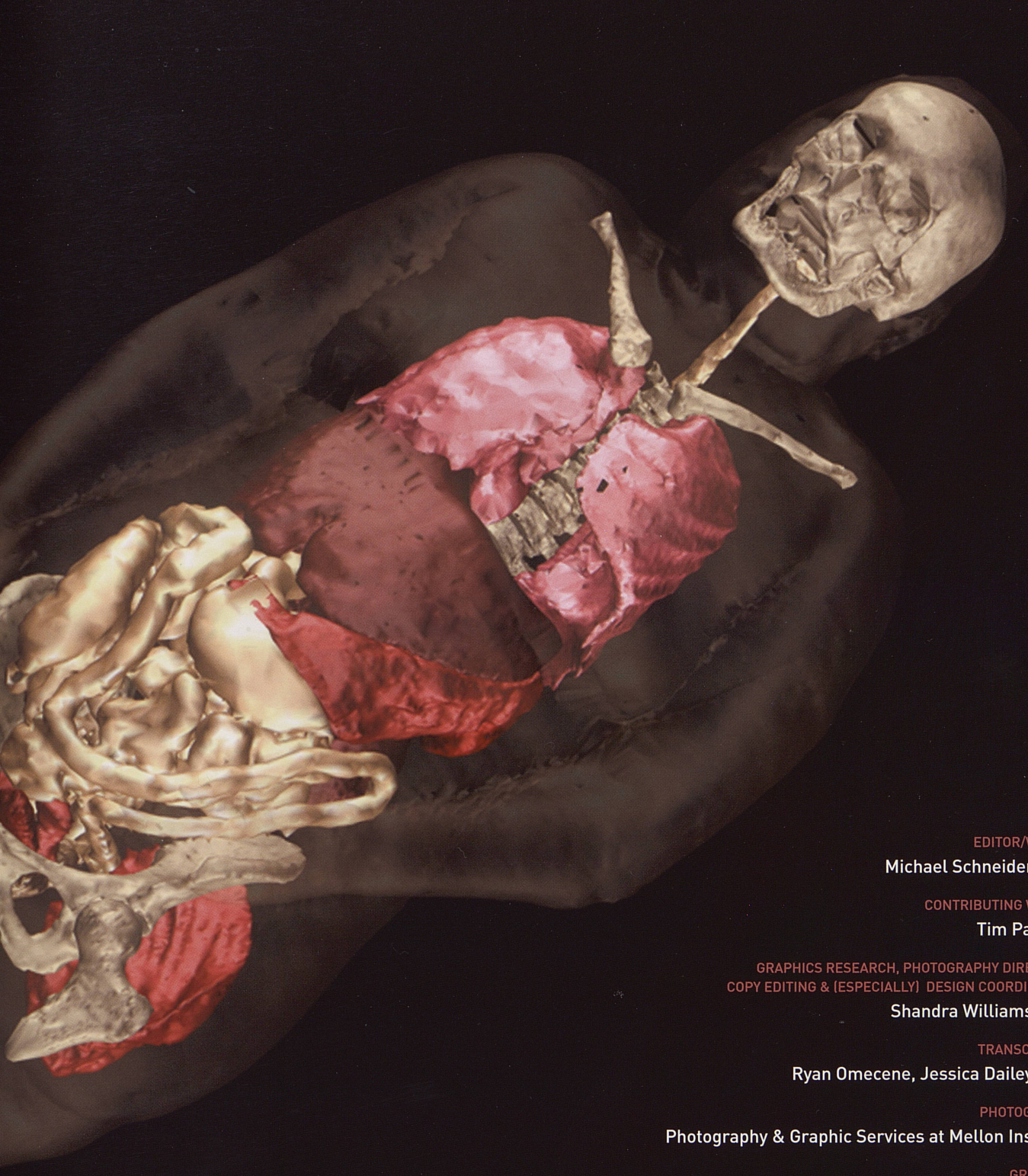
"This work will be important in drug design," says Madrid, "because it shows details of the motion that have not been observed before by any other computational technique."

PITTSBURGH
SUPERCOMPUTING
CENTER

The Pittsburgh Supercomputing Center is a joint effort of Carnegie Mellon University and the University of Pittsburgh together with Westinghouse Electric Company. It was established in 1986 and is supported by several federal agencies, the Commonwealth of Pennsylvania and private industry.

PSC GRATEFULLY ACKNOWLEDGES
SIGNIFICANT SUPPORT FROM THE FOLLOWING

The Commonwealth of Pennsylvania
The National Science Foundation
The National Institutes of Health
The National Energy Technology Laboratory
The National Archives and Records Administration
The U. S. Department of Defense
The U. S. Department of Energy
Cisco Systems, Inc.
Cray Inc.
DataDirect Networks
Hewlett-Packard Company
The Buhl Foundation
The Grable Foundation
The Heinz Endowments



EDITOR/WRITER

Michael Schneider, PSC

CONTRIBUTING WRITER

Tim Palucka

GRAPHICS RESEARCH, PHOTOGRAPHY DIRECTION, COPY EDITING & (ESPECIALLY) DESIGN COORDINATION

Shandra Williams, PSC

TRANSCRIBING

Ryan Omecene, Jessica Dailey, PSC

PHOTOGRAPHY

Photography & Graphic Services at Mellon Institute

GRAPHICS

Thanks to the researchers and to Phil Blood, Kent Eschenberg, Greg Foss, Joel Stiles, Joel Welling, Art Wetzel, and Troy Wymore of PSC for terrific graphics.

COVER GRAPHIC

Detail from simulations of Rayleigh-Taylor turbulence by Paul Woodward and David Porter using PSC's Cray XT3

DESIGN

Wall-to-Wall Studios, Inc.

PRINTING

Broudy Printing

Printed on Sappi McCoy Paper, a premium sheet with 10 percent post-consumer waste fiber, with vegetable-based inks.

PITTSBURGH SUPERCOMPUTING CENTER
300 S. CRAIG STREET
PITTSBURGH, PENNSYLVANIA 15213