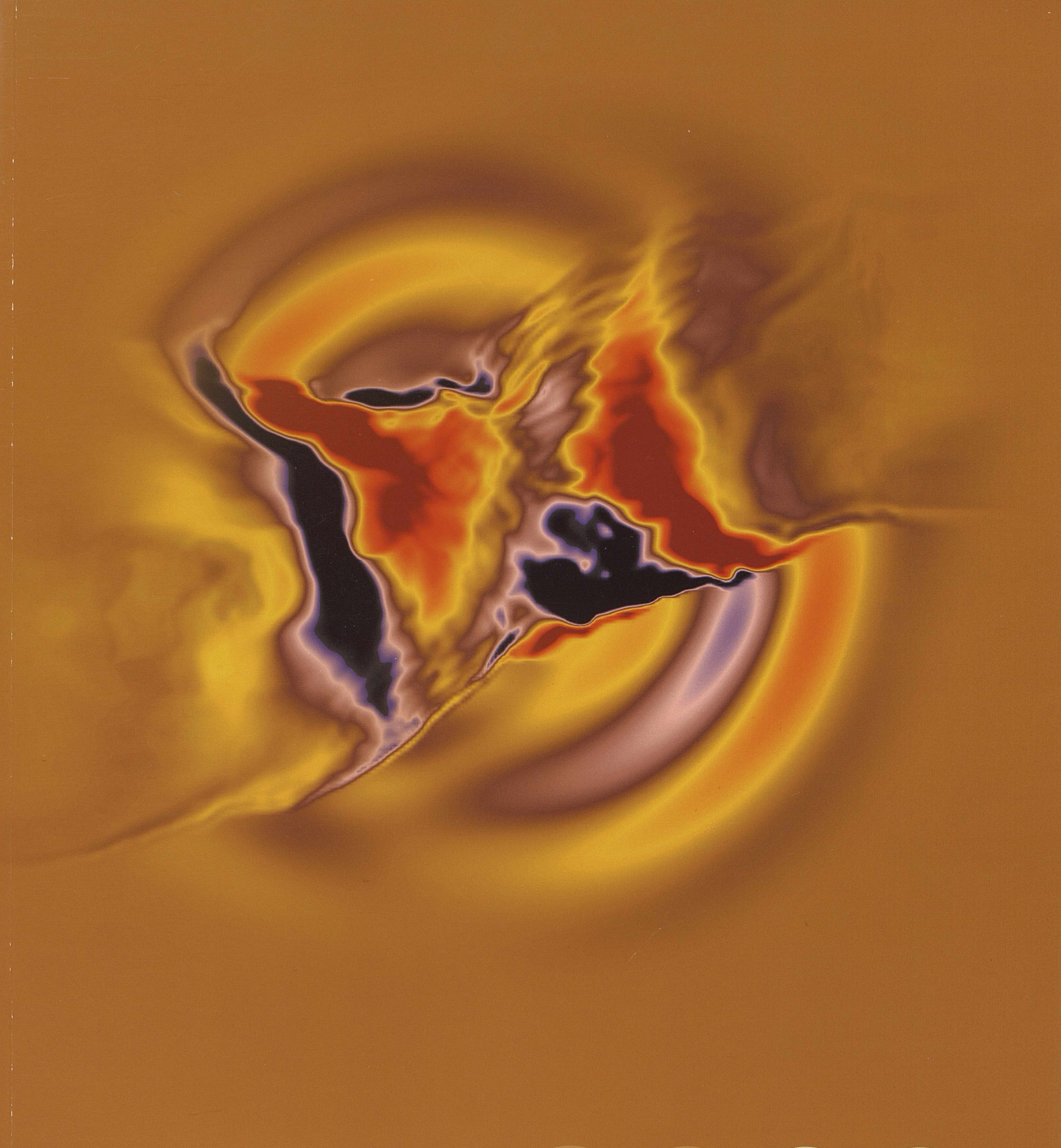
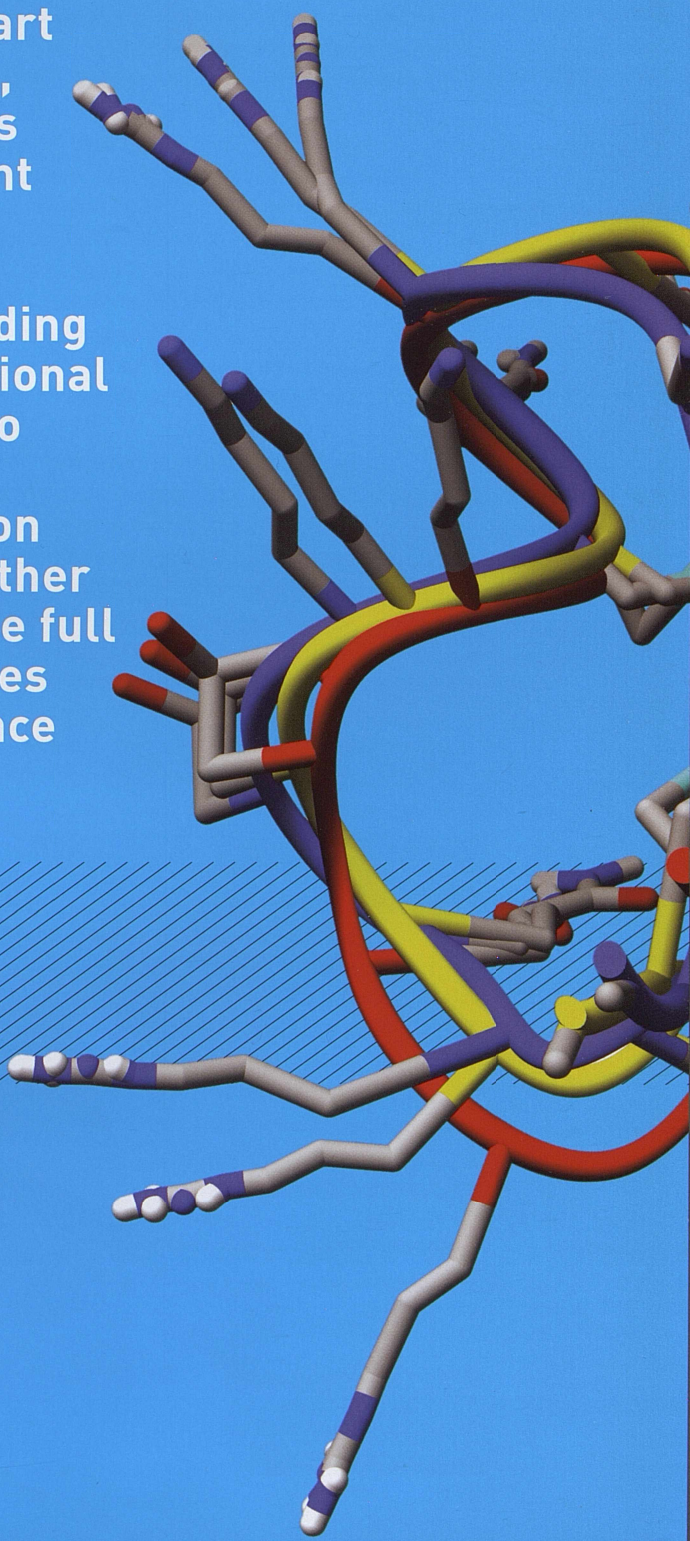


PITTSBURGH SUPERCOMPUTING CENTER 2009

PROJECTS IN SCIENTIFIC COMPUTING

PSC.EDU/09

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

Once again, we're pleased to highlight some of the remarkable scientific work at Pittsburgh Supercomputing Center and also to announce our partnership with David E. Shaw Research in making an innovative new system available to the U.S. biomedical research community (p. 4).

Our biomedical visualization experts, Art Wetzel and Greg Hood, this year fortuitously struck up collaboration with Clay Reid of Harvard, one of the leading visual cortex scientists in the world. When Reid began doing high-throughput transmission electron microscopy of the mouse visual cortex, he needed help assembling and processing massive amounts of imaging data. Wetzel and Hood and PSC's National Resource for Biomedical Supercomputing have proved to be the resource he needed, and it's a pleasure to be able to report on this work (p. 20).

Tom Cheatham at the University of Utah has grown up, scientifically speaking, with the NSF supercomputing centers program and, more recently, the TeraGrid (p. 5). He's pushed the limits of what's possible with molecular dynamics simulations for many years, and his current work with the RNA of the hepatitis C virus has potential for groundbreaking strides in new drug therapies for viral disease (p. 24).

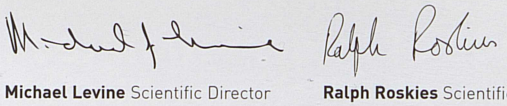
Zulema Garraffo of the University of Miami has used PSC resources for several years to refine HYCOM, a state-of-the-art ocean model. She and her colleagues have shown the importance of high resolution in ocean modeling as it bears on some of the challenging scenarios about how climate-induced changes in the Gulf Stream could affect Northern Europe (p. 28).

Two Carnegie Mellon projects, led respectively by Alessandro Acquisti and Colin Morningstar, exemplify how high-performance computing opens doors to new understanding in very different fields — Internet privacy (p. 32) and quantum chromodynamics (p. 40).

With effective collaboration between experiment and computation, John Yates, Ken Jordan of the University of Pittsburgh and colleagues Peter Maksymovych and Dan Sorescu of the National Energy Technology Laboratory made new findings in the chemistry of molecular chain reactions (p. 36).

PSC continues to be a vital resource for research and education in Pennsylvania (p. 6) and, through the Super Computing Science Consortium (p. 8), we help to promote economic development in southwest Pennsylvania and West Virginia and important work on development of clean-fuel technologies. This year's publication features a new section (pp. 9-11) that highlights our important work within the community to help educate the next generation of scientists and science-literate citizens.

As always, this publication represents the work of PSC's staff, second-to-none as an assemblage of talent and experience in high-performance computing. We are grateful for support from the National Science Foundation, the U.S. Department of Energy, the National Institutes of Health, the Commonwealth of Pennsylvania and many others.



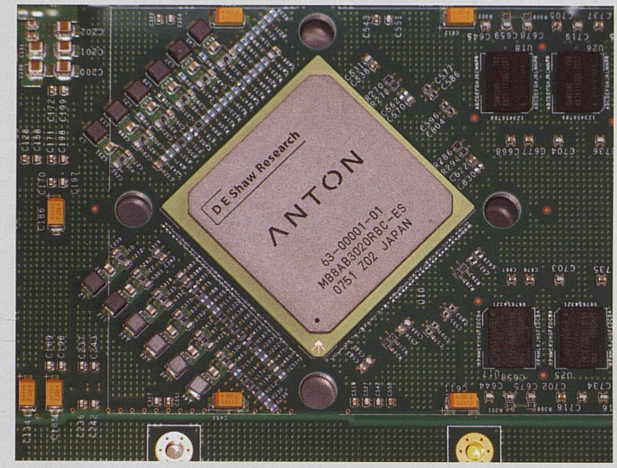
Michael Levine Scientific Director **Ralph Roskies** Scientific Director

MICHAEL LEVINE →
and RALPH ROSKIES
PSC co-scientific directors

TABLE OF CONTENTS

A NEW SPECIALIZED SYSTEM FOR BIOMOLECULAR RESEARCH

A $2.7M GRANT FROM THE NIH WILL ALLOW THE NATIONAL RESOURCE FOR BIOMEDICAL SUPERCOMPUTING AT PSC TO PROVIDE U.S. BIOMEDICAL RESEARCHERS WITH ACCESS TO A NOVEL SPECIAL-PURPOSE SUPERCOMPUTER FOR BIOMOLECULAR SIMULATION, MADE AVAILABLE WITHOUT COST BY D. E. SHAW RESEARCH



The National Institute of General Medical Sciences (NIGMS), part of the National Institutes of Health, has awarded a two-year, $2.7 million grant to the National Resource for Biomedical Supercomputing (NRBSC) at PSC (see p. 14) to host a specialized supercomputer for biomolecular simulation designed by D. E. Shaw Research (DESRES). The machine, called Anton, will be made available without cost by DESRES for non-commercial research by universities and other not-for-profit institutions.

Anton was designed to dramatically increase the speed of "molecular dynamics" (MD) simulations compared with the previous state-of-the-art, allowing biomedical researchers to understand the motions and interactions of proteins and other biologically important molecules over much longer time periods than have previously been accessible to computational study. The machine and the novel algorithms it employs were designed by a team of researchers led by David E. Shaw, chief scientist of DESRES.

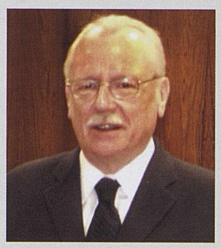
"This is an incredibly exciting project in many ways," said Joel Stiles, the director of NRBSC. "With this very generous gift from D. E. Shaw Research and the funding provided by NIH, we'll be deploying a tool of unprecedented power for the benefit of biomedical researchers nationally. We hope and expect that this project will help to significantly advance our understanding of biomolecular structure and function, and to spur ongoing scientific and technological development in MD research and in other areas of computational biology." In addition, Stiles, a faculty member in Biological Sciences and the Lane Center for Computational Biology at CMU, will collaborate with Christopher Langmead in the CMU School of Computer Science to develop specialized software for analysis of data as it streams off the machine.

The award is one of 14 made by NIGMS using funding from the American Recovery and Reinvestment Act of 2009 for projects the NIH views as "Grand Opportunities" for major scientific progress. "The Grand Opportunities grants fund projects that promise to have a significant impact on a field of biomedical science," said NIGMS Director Jeremy M. Berg. "By closing specific knowledge gaps, creating new technologies, or building community-wide resources, these awards will dramatically propel progress in key scientific fields."

Several Anton machines are currently operational within the DESRES research lab, but the one to be installed at NRBSC will be the first outside of DESRES. Time allocations on NRBSC's Anton machine will be made through a process of peer review. Although the NIH has supported MD-related research by individual scientists for many years, it has never before provided funds to make a supercomputing system available as a national resource.

While experimental methods such as X-ray crystallography can determine rigid molecular structures at near atomic resolution, MD simulations track atomic positions over the course of time. Changes in the shape of a biomolecule are often intimately related to its function, suggesting that atomic-level molecular dynamics simulations may ultimately play an important role in the design of therapeutic drugs. Because atomic-level MD simulations of proteins require an enormous number of calculations to simulate even a very short period of biological time, however, many of the most important biological phenomena have historically fallen outside the reach of even the most powerful general-purpose scientific supercomputers. Anton, however, has now run simulations extending for more than a millisecond of biological time — about 100 times longer than the longest previously published MD simulation.

Jim Kasdorf, PSC director of special projects



CREATING NATIONAL 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

PSC AND TERAGRID

PSC is actively involved in TeraGrid leadership. Scientific director Ralph Roskies oversees PSC contributions to the integrating efforts of the national Grid Infrastructure Group that guides TeraGrid. Co-scientific director Michael Levine is PSC's representative to the TeraGrid Forum — TeraGrid's principal decision-making group.

Other PSC staff with TeraGrid leadership roles include Sergiu Sanielevici, Area Director for User Support and Jim Marsteller, head of PSC's security, who chairs the TeraGrid Security Working Group. Laura McGinnis plays a lead role in TeraGrid education, outreach and training (EOT) activities. PSC director of systems and operations, J. Ray Scott, leads the TeraGrid effort in Data Movement, and PSC director of strategic applications, Nick Nystrom, leads the TeraGrid Extreme Scalability Working Group, which fosters planning to meet the challenges of deploying extreme-scale resources into the TeraGrid.

PSC staff members serve on all of TeraGrid's working groups.

TeraGrid Resource Providers

- 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

TeraGrid™

PSC staff whose work contributes to the TeraGrid

include (l to r, seated) David Moses, PSC executive director; Laura McGinnis, PSC manager of education, outreach and training; Sergiu Sanielevici, PSC director of scientific applications and user support; J.Ray Scott, PSC director of systems & operations. (l to r) Matt Mathis, Chris Rapier, Shandra Williams, Brian Gill, Anjana Kar, Jenda Domaracki, Michael Lambert, Kathy Benninger, Phillip Blood, Chad Vizino, Rich Raymond, James Marsteller, Rob Light, Pallavi Ishwad, Tom Maiden, Ken Hackworth, Nathan Stone, Bryan Webb, Mahin Mahmoodi, Greg Foss, Ed Hanna, Elizabeth Albert, Brian Johanson, Robin Flaus-Scibeck, Kevin Sullivan, Jason Sommerfield, Dustin Sorge, Derek Simmel, Adam Fest, Ed Berger. (top rear: l to r) Ed Wozniak, John Urbanic, Andrew Adams, Robert Budden, Shawn Brown.

SUPERCOMPUTING IN PENNSYLVANIA

WITH COMMONWEALTH OF PENNSYLVANIA SUPPORT, PSC PROVIDES EDUCATION, CONSULTING, ADVANCED NETWORK ACCESS AND COMPUTATIONAL RESOURCES TO SCIENTISTS AND ENGINEERS, TEACHERS AND STUDENTS ACROSS THE STATE

Cheryl Begandy, PSC director of outreach, coordinates PSC's programs of corporate research, education & community outreach in Pennsylvania.

PSC Staff talk with Daniel Lentz (center), principal of Intermediate Programs, Fox Chapel High School, and other guests during Discover 09 Open House. Inset: PSC scientist Marcela Madrid (left) talks with a University of Pittsburgh student about her HIV research.

DISCOVER 09: PSC OPEN HOUSE

With a turnout of 131 people, including students and teachers from local universities and representatives from governmental organizations, PSC's first annual "Discover" Open House introduced high-performance computing to a regional audience. PSC animated visual displays highlighted modeling in many fields, such as storm forecasting, earthquake soil vibration and realistic cellular modeling. "With many contacts made and information shared," said Begandy, "we showed how the revolution of computational science represents an opportunity in many areas, including small business."

K-12 SCIENCE EDUCATION

With dynamic K-12 educational and outreach programs (see pp. 9-11), PSC helps to prepare technology-ready workers and a science-literate populace. PSC workshops this year trained 120 science teachers in 32 school districts in Western Pennsylvania.

School districts in Pennsylvania are organized into intermediate units (IUs), which provide academic, technical and administrative services for the districts. In November 2008, PSC hosted a group of IU directors from across the state. PSC staff described its K-12 outreach programs, including connectivity through 3ROX (see p. 12) and its new CyberSafety program. In February,

Introducing Warhol: New PSC System Supports Pennsylvania Research

Through an arrangement with Hewlett-Packard (HP), PSC in January made available a 64-core, HP Blade System featuring eight nodes — each housing two Intel Xeon quad-core processors and 16 gigabytes of memory — interconnected by an InfiniBand link. Named Warhol, this new PSC system is provided as a resource to researchers in Pennsylvania.

"At PSC we have collaborated with Hewlett-Packard for many years," said J. Ray Scott, PSC director of systems and operations, "in development of innovative software and hardware to support research. In making this new system available to PSC, they further our mutually productive partnership."

PSC repeated this program for the Mercer County IU. Attendees included over 50 administrators, curriculum coordinators and technology directors from 15 districts.

In October 2008, in collaboration with the Leonard Gelfand Center for Service Learning and Outreach at Carnegie Mellon, PSC sponsored a Symposium on Science, Technology, Engineering & Mathematics (STEM) Strategies and Practices. To an audience of 32 school superintendents, science curriculum coordinators and STEM teachers from the region, PSC staff presented an overview of STEM-related programs at PSC and Carnegie Mellon and on STEM initiatives regionally and nationally.

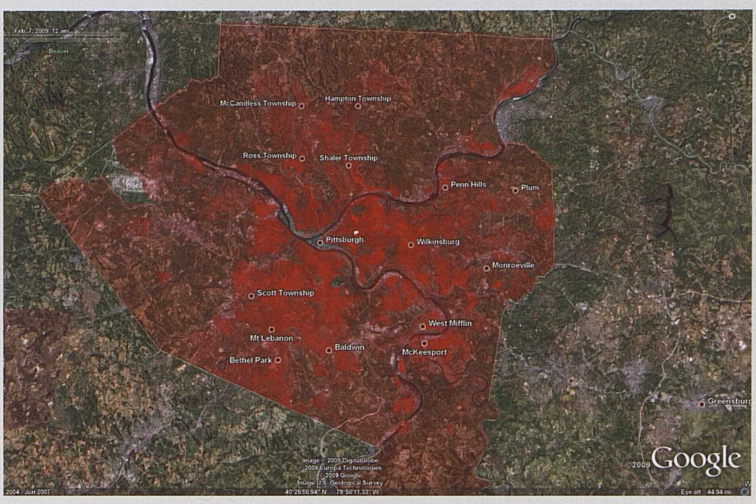
COMMUNITY OUTREACH

PSC staff take part in numerous outreach programs, locally and nationally. PSC exhibited at the SciTech Initiative in March at the Carnegie Science Center. With a theme of "What is supercomputing and how is it used?," PSC's booth — visited by more than 400 attendees — included hands-on materials to stimulate student thinking. As part of this event, PSC was a bronze sponsor of the Pittsburgh Regional Science and Technology Fair in April, presenting awards to students whose projects best exemplified computational science. One of these students, with PSC encouragement, took his project to the Intel International Science & Engineering Fair and won third place in Mathematical Sciences.

PRIVATE-SECTOR & UNIVERSITY RESEARCH

Research by university scientists in Pennsylvania supported by PSC is exemplified by several projects in this booklet:

- Wiring Diagram of the Brain: Greg Hood & Art Wetzel, Pittsburgh Supercomputing Center in collaboration with Clay Reid, Harvard University (p. 20).
- Identity Theft Vulnerability: Alessandro Acquisti & Ralph Gross, Carnegie Mellon University (p. 24).
- Reactivity of Molecular Chains: Ken Jordan & colleagues, University of Pittsburgh (p. 36).
- New Findings in Quantum Chromodynamics: Colin Morningstar & colleagues, Carnegie Mellon University (p. 40).
- PSC this year entered a partner agreement with the Pennsylvania NanoMaterials Commercialization Center: www.pananocenter.org/



Modeling the Spread of H1N1

On Pittsburgh's WTAE-TV this October, PSC scientist Shawn Brown demonstrated PSC modeling of the spread of the H1N1 virus in the Pittsburgh metropolitan area. Brown's work is part of the National Institutes of Health's Models of Infectious Disease Agent Study (MIDAS) project, which supports research to simulate disease spread, evaluate intervention strategies and help inform health policymakers. He collaborates with the Pittsburgh MIDAS Center of Excellence, led by Donald Burke, M.D., of the University of Pittsburgh Graduate School of Public Health.

Research & Training at Pennsylvania Companies, Colleges & Universities, 2008-2009

From July 1, 2008 through June 30, 2009, PSC provided nearly 12.5 million total processor hours to 670 individual Pennsylvania researchers from 36 institutions. PSC workshops in high-performance computing reached 175 Pennsylvania grad and undergrad students. If purchased from a commercial provider, this computing time would be valued, conservatively, at nearly $12 million. The following Pennsylvania universities and colleges used PSC resources during this period:

Allegheny-Singer Research Institute
Bloomsburg University of Pennsylvania
Bryn Mawr College
Bucknell University
Cabrini College
Carnegie Mellon University
Cedar Crest College
Cheyney University of Pennsylvania
Drexel University
Duquesne University
Franklin and Marshall College
Haverford College
Indiana University of Pennsylvania (all campuses)
Lehigh University
Lock Haven University
Pennsylvania State University (all campuses)
Shippensburg University of Pennsylvania
Temple University
Thomas Jefferson University
University of Pennsylvania
University of Pittsburgh (all campuses)
Ursinus College
Villanova University
Waynesburg College

THE SUPER COMPUTING SCIENCE CONSORTIUM

PENNSYLVANIA-WEST VIRGINIA PARTNERS IN DEVELOPMENT OF CLEAN POWER TECHNOLOGIES

(SC)² co-chairs Lynn Layman, PSC (left) & Bob Romanowsky, NETL.

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 resource that supports and encourages companies to collaborate with local universities in southwest Pennsylvania and West Virginia and to have access to PSC.

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.

In August, PSC scientists provided a series of presentations to NETL scientific staff on PSC resources and capabilities in visualization, data analysis, parallel performance optimization, emerging software and trends in high-performance computing.

(SC)² Partners

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

MORE INFORMATION
www.sc-2.psc.edu

(SC)² Super Computing Science Consortium

PSC & (SC)²: RESEARCH FOR CLEAN ENERGY

Since the 1999 founding of (SC)², 51 (SC)² researchers have used PSC systems for a range of clean-energy related projects, using more than 5.7-million hours of computing time, over 330,000 hours within the past year.

This work includes:

High-Fidelity Simulation of Turbulent Combustion
www.psc.edu/science/2008/sc2/

Clean Liquid Fuel from "Syngas"
www.psc.edu/science/2006/sc2/

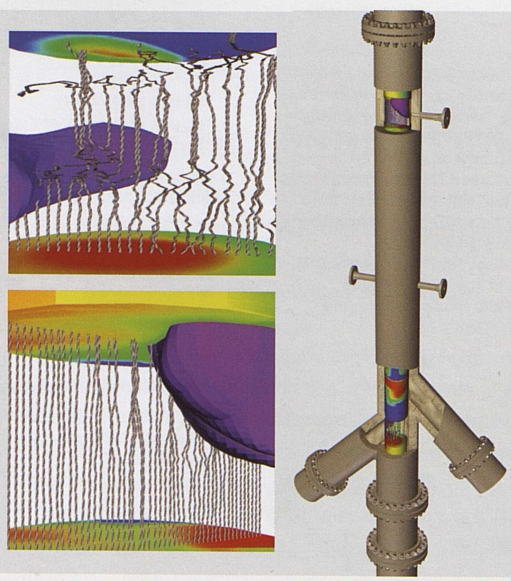
Fuel-Quality Hydrogen from Fossil Fuels
www.psc.edu/science/2005/sc2

Gas from Black Liquor
www.psc.edu/science/2004/sc2/

Fluidized-Bed Combustion of Silane
www.psc.edu/publicinfo/netl/

Lean-Fuel Mixes in Next-Generation Power-Generating Turbines
www.psc.edu/science/Richards/clean_power.html

A New Design for a Power-Generating Turbine
www.psc.edu/science/cizmas2002.html



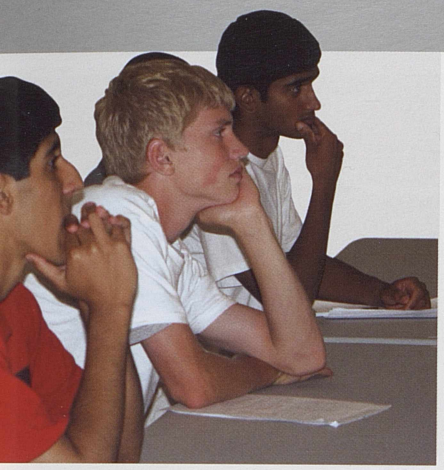
Modeling an Operational Clean-Coal Power Plant

This graphic represents results from 2007 NETL research that used PSC resources to model coal gasification for a power plant in Florida, anticipated to be the world's cleanest coal-fired plant when it comes online in 2010. See: www.psc.edu/science/2007/coal

The top image shows 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 coal to gas and elevated discharge of soot and carbon-dioxide. The lower graphic represents burner air entering the gasifier (purple) and the complex flow pattern (grey streamlines) it creates. Complex flow in this region is critical for sufficient mixing and to maximize contact between the gas and solid particles.

ENERGIZING SCIENCE LEARNING

WITH A QUARTET OF PROGRAMS IN SCIENCE EDUCATION, PSC GIVES THE PITTSBURGH REGION A JUMPSTART TOWARD A CYBER-SAVVY WORKFORCE



"It's one thing to read about it, and another thing entirely to see it happening," says Marian Opest, a biology teacher at Penn Hills High School, a suburb of Pittsburgh. What stirs her enthusiasm is CMIST (Computational Modules in Science Teaching), one of four programs in secondary science education that PSC has introduced to teachers and to science classrooms in the Pittsburgh region and beyond over the past four years.

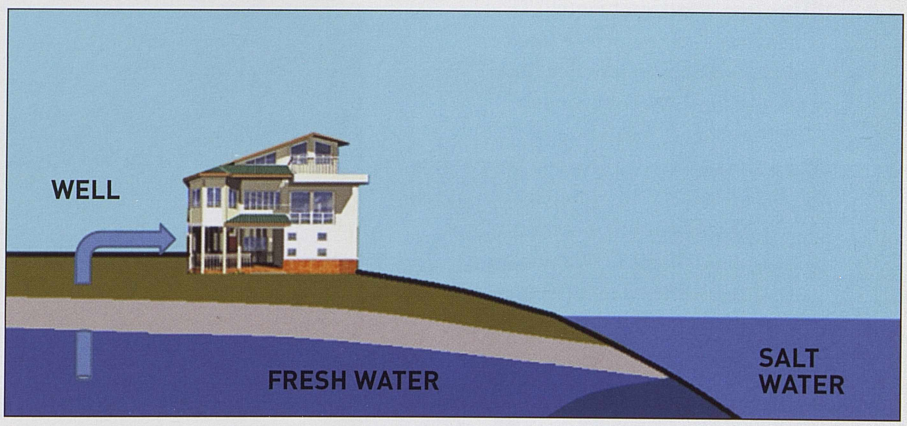
"It's the difference between reading a textbook and visually experiencing a topic such as diffusion or osmosis," says Pallavi Ishwad, education outreach specialist for the National Resource for Biomedical Supercomputing (NRBSC), PSC's biomedical program. A former high-school biology teacher herself, Ishwad has seen first-hand the difference that computational tools, such as vivid 3D animations produced by supercomputer simulations, can make in science learning.

"Introducing 'cool' technology into the classroom engages students," says PSC's director of outreach and education, Cheryl Begandy, "and increases their willingness to stay with subjects they may otherwise find too complicated or just uninteresting." For Begandy and Ishwad along with other PSC staff the goal is to help in re-imagining high-school science instruction so that it better attracts future scientists, engineers and educators. Ultimately the goal is to create the cyber-savvy workforce demanded by the 21st-century marketplace.

CAST

PSC's first venture into science education, Computation and Science for Teachers (CAST) has over the past three years introduced teachers to easy-to-use modeling and simulation tools. Based on ideas pioneered in the Maryland Virtual High School Project, the CAST approach to "computational thinking" can be incorporated in classroom use across the math science curriculum.

Over 40 teachers from southwestern Pennsylvania have participated in CAST's weeklong summer workshop followed by quarterly sessions throughout the school year. "We actually built the model of the water cycle," says CAST participant Jim Lear, physics teacher at Pittsburgh's Oakland Catholic High School. "It was really neat to see the variables change and the effect on the amount of water as vapor, liquid, and vapor precipitating back to liquid in our model."



SITUATION

Before you buy a home, you should always consider the location, drainage, and water resources. Consider the situation above. This is a beautiful house, near the beach in Louisiana. The property includes the well and an in-ground septic system. The closest property is a mile away and is an old gas station that closed in 1989. What are some of the concerns that you should investigate before you purchase?

CMIST

Introduced in 2007, CMIST provides multi-disciplinary teaching modules — including lecture slides, animations and lesson plans — as ready-to-use DVDs. Produced with high-quality, biologically realistic 3-D animations, the modules are geared to lead students toward an integrated understanding of biology, chemistry, physics, math and computation.

The initial CMIST module, "Molecular Transport in Cells" — produced with software called MCell and DReAMM, co-authored by NRBSC director Joel Stiles — presents important principles of osmosis and diffusion. It has prompted enthusiastic feedback and created interest nationally. "I passed the DVD to some teachers," wrote Manorama Talavier, a school district curriculum administrator from Virginia. "They love it and want more."

To date over 500 people locally and nationally have attended CMIST presentations. A second module, "Big Numbers in Small Spaces: Simulating Atoms, Molecules, and Brownian Motion," premiered in May, with a talk by Stiles at the National Science Bowl Finals in Washington, DC. Ishwad introduced it to Pittsburgh-area teachers at a one-day July workshop.

AVAILABLE HERE (FREE)
www.nrbsc.org/cmist.

From "Big Numbers in Small Spaces," this view of the CMIST virtual laboratory from inside a glass beaker shows the splash from a drop of water fallen into the beaker, and then the CMIST movie zooms in to show the quantity of water molecules contained within a single droplet from the splash.

PSC's Pallavi Ishwad with participants in a BEST workshop.

BEST

Begun in 2007 by Ishwad, Better Educators of Science for Tomorrow (BEST) introduces high school teachers to a bioinformatics curriculum adapted from an NRBSC program called MARC (Minority Access to Research Careers) for undergrad and graduate science students. Drafted by an interdisciplinary group of teachers representing physics, chemistry, biology, mathematics and technology who participated in MARC, the BEST curriculum facilitates a cross-disciplinary approach and brings awareness to high school students about bioinformatics. BEST offers ready-to-use lesson plans for single-subject trained teachers to teach a multidisciplinary subject like bioinformatics in high schools.

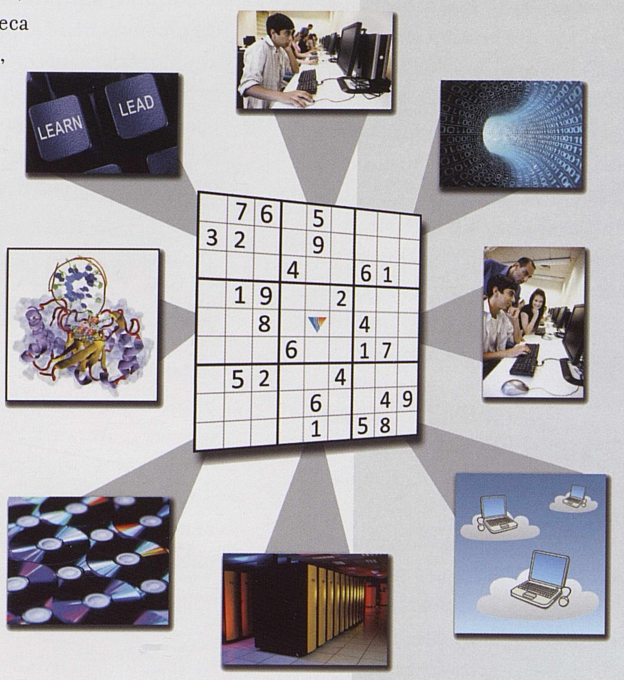
Ishwad piloted BEST in three schools this year. "I am excited to use what I learned here," said Dean Walker, a physics teacher at Seneca Valley High School outside Pittsburgh, "to both inspire students who love biology but hate physics, as well as provide students with a better feel for contemporary positions in the science job market."

CompEx

Making its debut in 2009, Computation Exploration (CompEx) is a high-performance computing teaching module for high-school computer science. Developed by PSC outreach staff and scientists Robin Flaus, Phil Blood, Bryon Gill, Tom Maiden, Nathan Stone and John Urbanic, with input from Carnegie Mellon student Srihari Seshadri, CompEx presents basics of message passing (MPI) in parallel processing and guides students through development of their own parallel program to solve a Sudoku puzzle.

PSC piloted CompEx in February at Upper St. Clair High School south of Pittsburgh. Nine students completed the course, writing programs and running them on Amazon's Elastic Compute Cloud. They were also introduced to data analysis with widely used software called Hadoop. Said one student, "I enjoyed working with high level computers that we would not have had access to without a fat wallet."

CompEx students with computer science teacher Todd Ollendyke.



Funding for these PSC education and outreach programs has come from Pittsburgh-area foundations — The Buhl Foundation, Grable Foundation and Heinz Endowments — and from the NIH along with additional support from the Commonwealth of Pennsylvania and the NSF TeraGrid program.

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
www.psc.edu/networking/

FROM PITTSBURGH TO ABU DHABI: FIRST INTERNATIONAL TELEPRESENCE SESSION VIA NLR

PSC this year played a lead role in demonstrating "TelePresence," a video teleconferencing capability, developed by Cisco Systems. In April, PSC coordinated the first international TelePresence session over research and education networks via National LambdaRail (NLR), a major initiative of U.S. universities and the private sector to provide networking infrastructure for research and education. NLR and ANKABUT, a science initiative of the United Arab Emirates, arranged a TelePresence link between PSC and the campus of Khalifa University of Science, Technology and Research in Abu Dhabi.

In May, NLR demonstrated multi-point TelePresence, with a session linking the Renaissance Computing Institute (RENCI) of Chapel Hill, North Carolina and Pennsylvania State University's College of Information Sciences and Technology with PSC. "The demonstration of multi-point TelePresence reinforces that TelePresence is the leading-edge in live video tele-conferencing," said Wendy Huntoon, PSC director of networking. "For PSC, this extends our ongoing productive collaboration with Cisco and close partnership with NLR.

PSC network engineer Steve Cunningham (left in photo) was "in the room" in Abu Dhabi via TelePresence.

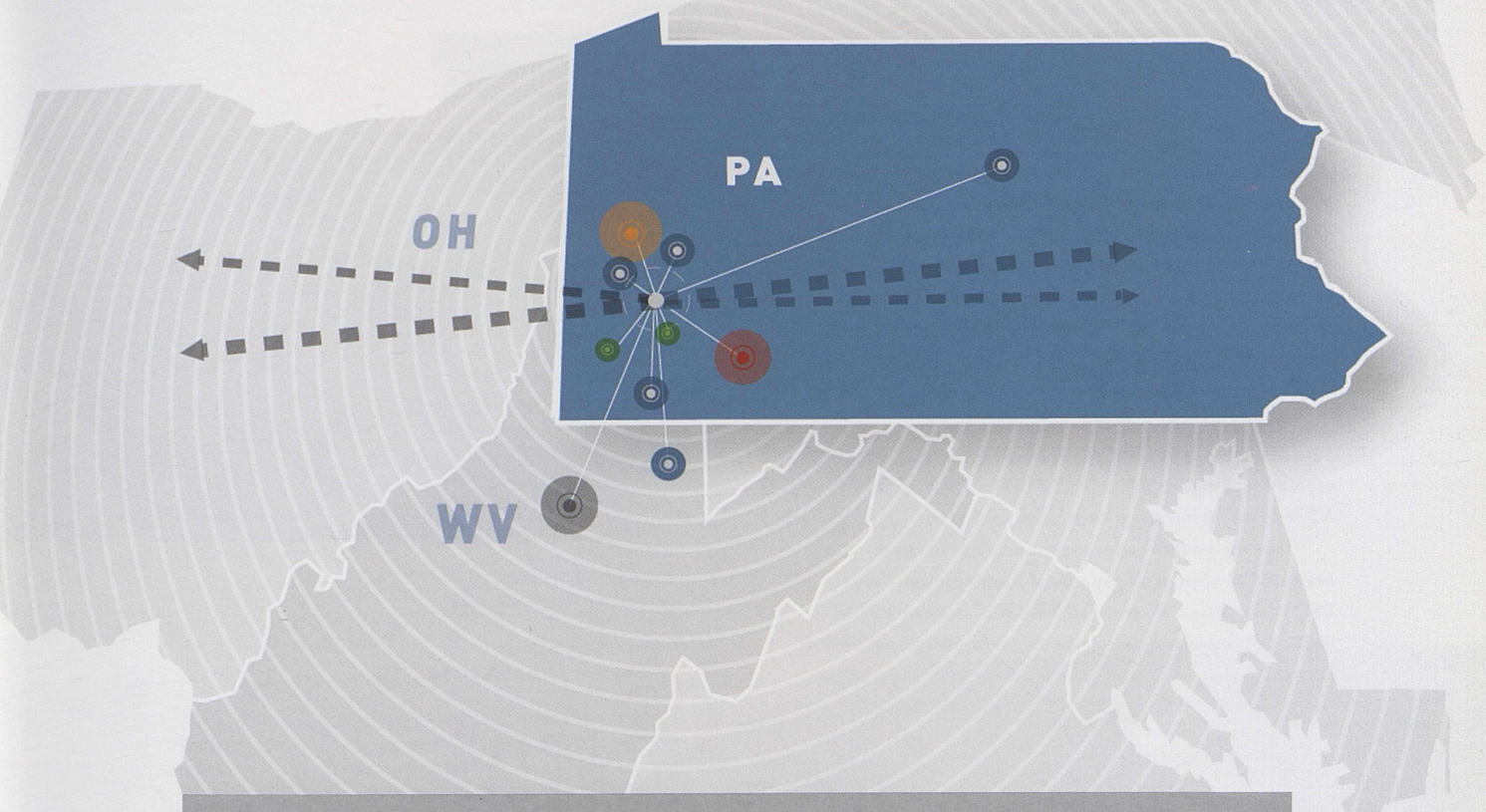
3ROX PLANS FOR STATEWIDE BROADBAND

Through 3ROX, the Three Rivers Optical Exchange, PSC connects universities and public schools in Pennsylvania and West Virginia to high-performance networks, such as Internet2, which links leading U.S. universities, corporations, government research agencies, and not-for-profit networking organizations.

This year 3ROX joined with a coalition of Pennsylvania colleges and universities, heathcare and economic development organizations to form the Pennsylvania Research and Education Network (PennREN). PennREN applied for $100 million in federal stimulus money through the American Recovery and Revitalization Act and $29 million in private funds to build and maintain a broadband network for expanded educational opportunity and healthcare services across the Commonwealth of Pennsylvania. The proposed network would reach every region of the state, providing access to more than two-million households and 200,000 businesses. "The system would rival any in the United States," says Huntoon, "and would provide the capability to connect regional networks across the Commonwealth."

NETWORK RESEARCH: MEASUREMENT LAB

In January, Google launched its new network measurement tool, M-Lab, which includes the PSC-NCAR developed Network Path Diagnostics tool (NPAD) as one of four key network measurement instruments. A web-based deployment of servers and tools for study of broadband networks, M-Lab includes PSC senior network engineer Matt Mathis on its steering committee.



3ROX MEMBERS

Universities

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

NLR Member Institutions

3ROX (PSC), Carnegie Mellon University, Case Western Reserve University, OneCommunity, University of Pittsburgh, Pennsylvania State University, Indiana University

K-12 Institutions

Allegheny Intermediate Unit (AIU3), Arin Intermediate Unit (IU28), Beaver Valley Intermediate Unit (IU27), Intermediate Unit One, Northwest Tri-County Intermediate Unit (IU5), Riverview Intermediate Unit (IU6), City of Pittsburgh School District (IU2), Woodland Hills School District

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, ESnet — 1 Gbps, National LambdaRail PacketNet — 10 Gbps, TeraGrid Extensible Backplane Network — 30 Gbps

Other Network Connections

Southern Crossroads (SOX) — 1Gbps, TransitRail — 1Gbps, OARnet — 1 Gbps, FrameNet — 10 Gbps

National Commodity Internet Networks

Global Crossing — 1 Gbps; Sprint — 1 Gbps

Pittsburgh Local Exchange Networks

Comcast

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

NOTES & HIGHLIGHTS

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. This September NCRR awarded just under $800,000 as a supplemental grant for 2009-10, part of which supports storage and analysis of massive brain-imaging data from NRBSC's collaboration with Harvard (p. 20).

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

In September, the National Institute of General Medical Sciences (NIGMS), part of NIH, awarded $2.7 million over two years to the NRBSC to support a partnership with D. E. Shaw Research to make an innovative new computing system available to U.S. biomedical scientists (see p. 4).

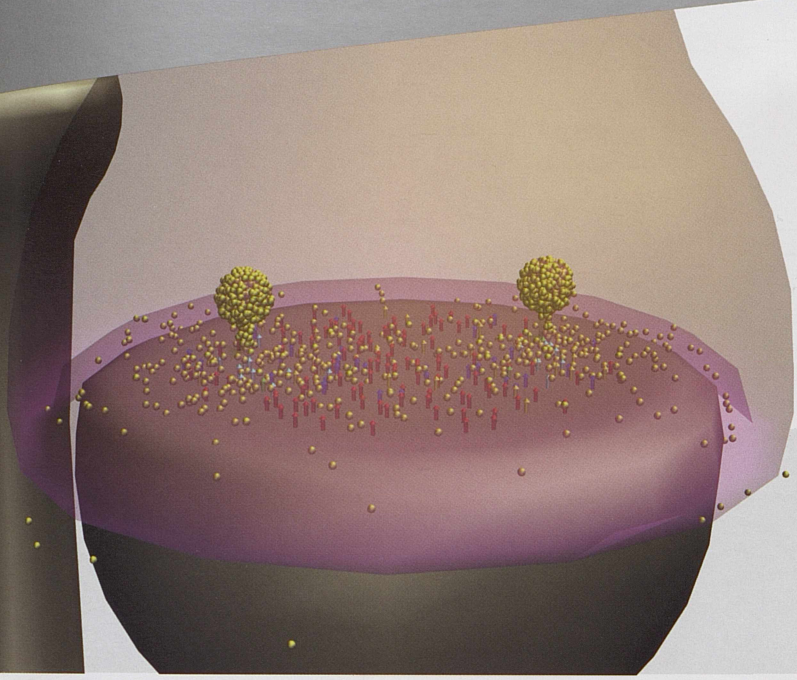
The NRBSC and PSC have developed educational programs, CMIST and BEST (see pp. 9-11), for high school and undergraduate biology, chemistry, physics, computer science and math that have provided training to students and educators in the Pittsburgh region and nationally.

MORE INFORMATION
www.nrbsc.org

NRBSC Biomedical Collaborations

Albert Einstein College of Medicine
Carnegie Mellon University
Cornell University
Duke University
Harvard 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

The NRBSC team: (seated, l to r) Boris Kaminsky, Pallavi Ishwad, Jenda Domaracki, (standing) Art Wetzel, James Keener, Jack Chang, Markus Dittrich, Aji Janis, Troy Wymore, Jun Ma, Christal Banks, Alex Ropelewski, Joel Stiles, Hugh Nicholas, Adam Kraut, Greg Hood, Gary Blumenthal. Not pictured: Jacob Czech

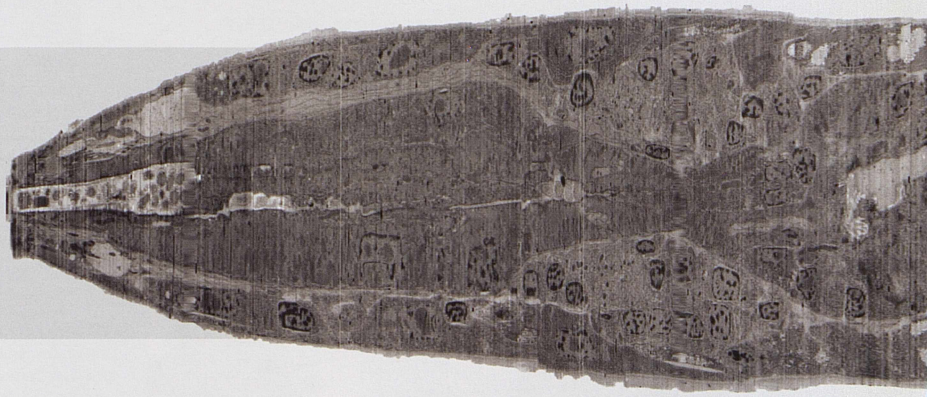


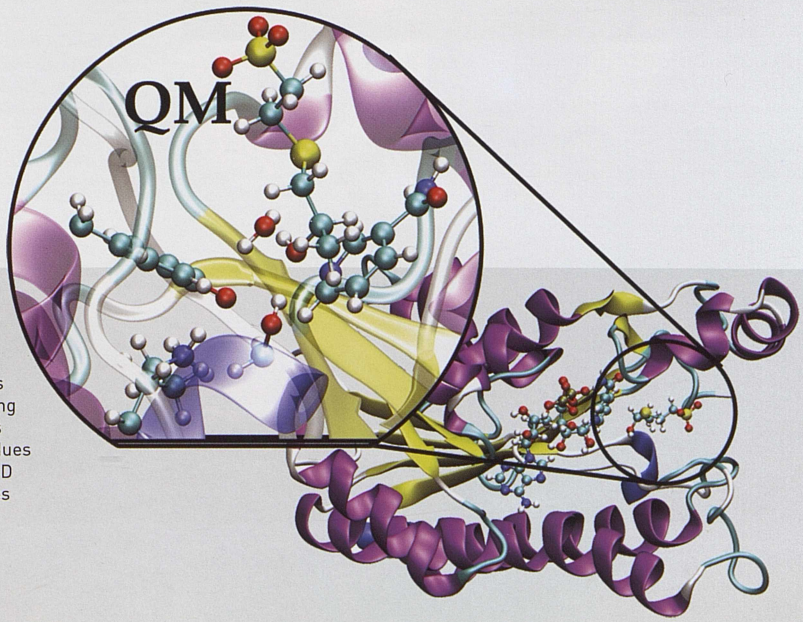
RESEARCH

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. MCell, DReAMM and PSC_DX software is developed at the NRBSC and used to model and visualize events such as this image, which represents neurotransmitter release in one dendritic spine.

VOLUMETRIC VISUALIZATION using the NRBSC's PSC_VB software enables 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 transverse section (from a dataset captured by Richard Fetter in Cori Bargmann's laboratory) of C. elegans, a roundworm much studied as a model organism, was aligned with programs developed by Greg Hood at NRBSC.





NRBSC STRUCTURAL BIOLOGY focuses on developing software for quantitatively accurate enzyme reaction simulations and integrating the results with sequence-based bioinformatics studies. This PSC-developed software is enabling more accurate simulations of enzyme reactions and insight into the function of amino acid residues outside the active site. This image shows the 3-D structure of R-HPCDH, an enzyme that catalyzes a coupled proton/hydride transfer.

COMPUTATIONAL SERVICE & TRAINING

Since NRBSC's inception, PSC and NRBSC together have provided access to computing resources for more than 1,600 biomedical research projects involving more than 4,200 researchers at 274 research institutions in 46 states and two territories. Among these are several projects featured in this booklet (pp. 20 & 24).

NRBSC training activities reach hundreds of scientists each year. More than 3,600 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
  compbio.cmu.edu
- the University of Pittsburgh Department of Computational Biology
  www.ccbb.pitt.edu

Pittsburgh Supercomputing Center Workshops (2008-2009)

Computational Methods for Spatially-Realistic

Microphysiological Simulations

Summer Institute in Bioinformatics
(for minority-serving institutions)

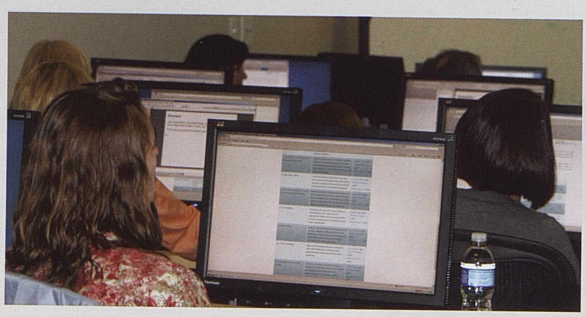
Methods and Applications of Hybrid QC/MM Simulations to Biomolecular Systems

Bioinformatics Internship Program

Multi-Core Programming and Performance Tuning

Parallel Programming and Cluster Computing

TeraGrid New User Training



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 (l to r), who oversee the day-to-day operations of the center: David Moses, executive director; John Kochmar, manager of high-performance computing facilities; J. Ray Scott, director of systems & operations; Elvira Prologo, manager of administration; David Kapcin, director of financial affairs; Wendy Huntoon, director of networking; Nick Nystrom, director of strategic applications; Bob Stock, PSC associate director; Sergiu Sanielevici, director of scientific applications & user support;

Not pictured: Janet Brown, manager of networking; Cheryl Begandy, director of outreach; Richard Raymond, manager of user support; Joel Stiles, director of NRBSC.

PROJECTS IN SCIENTIFIC COMPUTING

2009

PROJECTS 2009 CONTENTS

IN PROGRESS

WIRING

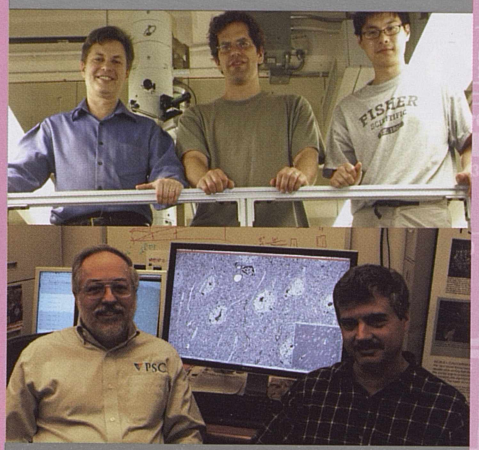
A HARVARD-PSC COLLABORATION IS CAPTURING AND PROCESSING MASSIVE AMOUNTS OF HIGH-RESOLUTION IMAGE DATA FROM WHICH TO TRACE A WIRING DIAGRAM OF THE VISUAL CORTEX

A convoluted mass of tissue comprising roughly as many nerve cells, called neurons, as there are stars in the Milky Way, about 100 billion, the human brain may be the most complex structure in the universe. How much do we know about this intricately interconnected biochemical-electrical processing network that makes cognition possible — everything we call thought, creativity, emotion, memory, vision and more?

The answer, despite amazing advances in brain science over the past half century: not much, really. Probably the most studied brain activity is vision, with the most studied region being the primary visual cortex — an area in the back of the brain that processes visual stimuli. Brilliant experiments over the past 50 years have shown that neurons are organized in the visual cortex according to function — the ability to recognize particular kinds of visual information. Nevertheless, we have almost no understanding of how the neurons interconnect.

"What determines what a cortex does?" asks Clay Reid, professor of neurobiology at the Harvard Medical School and Center for Brain Science. "What makes us human? I think it's almost unarguable that what determines the differences — either between different cortices that do different things in the human brain, or the difference, for instance, between the human brain and the mouse brain — is the connections."

For Reid, the next big step is to identify the brain's wiring diagram at the level of the individual neuron connections. During the past year, Reid, Ph.D. student Davi Bock, and postdoctoral fellow Wei-Chung Lee have used innovative, high-throughput transmission electron microscopy (TEM) to capture high-resolution images of sections from a mouse visual cortex. These sections — from a volume of brain containing 100 identified neurons and portions of many more — correspond to live experiments showing which neurons perform particular visual operations.



↓ **CLAY REID** (left), **DAVI BOCK** and **WEI-CHUNG LEE**
Harvard University in their electron microscopy lab at Harvard

↑ **ART WETZEL** (left) and **GREG HOOD**
National Resource for Biomedical Supercomputing, Pittsburgh Supercomputing Center

Collaborating closely with Reid's group, scientists Art Wetzel and Greg Hood from PSC's National Resource for Biomedical Supercomputing (NRBSC, see p. 14) received and processed about a terabyte of TEM data per day between April and September of this year. The data comprise thousands of high-resolution sections, each about 40 nanometers thick (a few hundred atoms). With funding from the NRBSC and Harvard, they managed the prodigious data transfer and developed software to process and assemble the images for viewing and analysis.

Having assembled an unprecedented dataset, this Harvard-NRBSC-PSC collaboration is now positioned to begin reverse engineering the mouse visual cortex — to get the neuron-to-neuron wiring diagram. "We're at a very exciting point," says Reid. "We've demonstrated the ability to generate a new class of data, and we're beginning to do some science. We're looking toward synaptic-level anatomical reconstruction of cortical circuits of known function — the wiring diagram."

DIAGRAM

OF THE BRAIN

SELECTIVITY OF NEURAL FUNCTION

Reid's work on the visual cortex follows from renowned work at Harvard by David Hubel and Torsten Wiesel in the late 1950s and early 60s — for which they won the 1981 Nobel Prize in Physiology or Medicine. By inserting electrodes into the brain of an anesthetized cat, Hubel and Wiesel did breakthrough experiments showing that the primary visual cortex is organized in relation to features seen in the visual field of each eye.

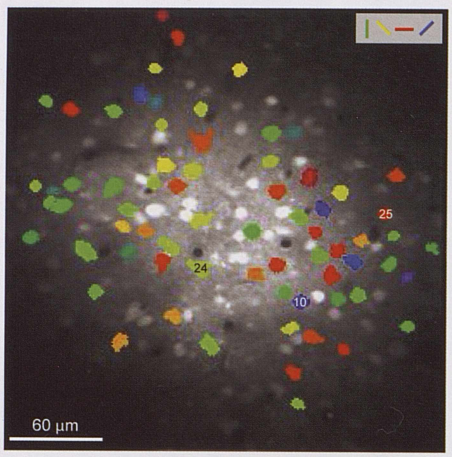
Neurons that respond to vertical features — trees or telephone poles, for instance — are grouped together in patches called orientation columns. Other orientation columns respond to horizontal lines, or forty-five degree lines. Hubel and Wiesel showed that from simple parts of the overall visual stimulus the cortical system builds up the complex image we see in our mind's eye.

This work showed, says Reid, that something in the cortical network creates this ability of neurons to function selectively, and to understand how this happens will require knowing in detail how the neurons connect. "We know the wiring diagram in very broad strokes," he continues, "but not at the level of single neurons, and that's what we're interested in."

About five years ago, Reid began a series of experiments with a technique called two-photon calcium imaging — a powerful method that makes it possible to see the activity of distinct neurons in a living brain. Rather than inserting an electrode, as Hubel and Wiesel did, Reid uses fluorescent dye that's sensitive to calcium levels, an indicator of a firing neuron. As an anesthetized small mammal sees the vertical, horizontal and oblique bars used by Hubel and Wiesel, the researchers focus the microscope (through a small hole in the skull) on the visual cortex, at single-cell resolution to a depth of 400 micrometers (400 millionths of a meter). Via the fluorescent indicators, they watch as individual neurons fire in response to the visual stimuli.

Reid's findings from two-photon imaging — reported in *Nature* (2005) — show distinct differences between rats and cats in how neurons are functionally grouped. For Reid, this work also brought into focus the prospect of wiring diagrams. The high resolution possible with the two-photon microscope showed, he says, that "you can actually build a 3D model of where neurons are in the living brain."

"These results," he wrote in *Nature*, "indicate that cortical maps can be built with single-cell precision."



NEURON FUNCTION

From two-photon calcium imaging of a mouse visual cortex, this image shows individual neurons color-coded according to orientation of visual stimuli to which they respond: vertical (green), horizontal (red), left oblique (yellow) or right oblique (blue).

A TERABYTE A DAY

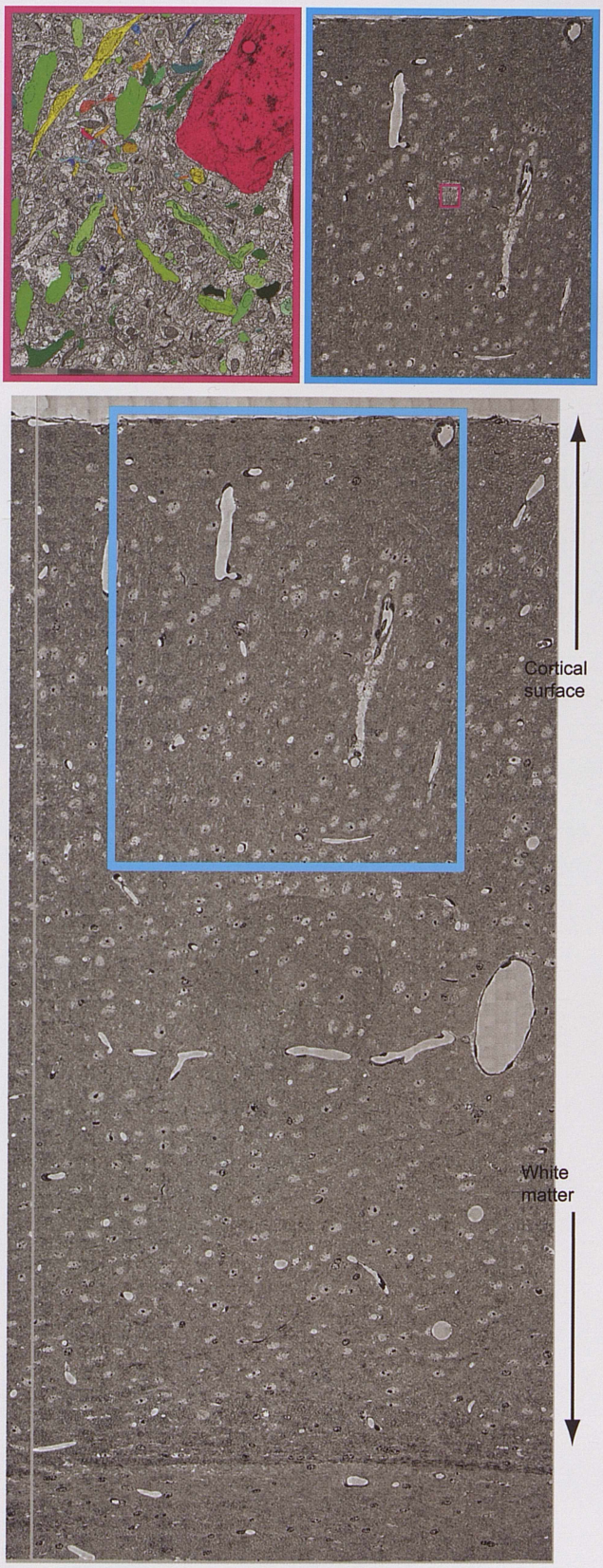
To take the next step, Reid, Bock, Lee and their collaborators at Harvard built a customized TEM with a four-camera array that can image cortical circuits over an area encompassing hundreds of micrometers. "We need very large images," says Reid, "and many of them." The result — a system that can capture a terabyte per day of image data from serial sections of the visual cortex.

Over the six-month run of data transmission, Wetzel and Hood, with assistance from PSC systems and networking staff, transferred TEM datasets daily from a small Linux computer at Harvard to PSC. PSC network engineers worked with staff at Harvard to tune Internet protocols and used PSC's HPN-SSH patch to improve performance in the OpenSSH program used for secure data transfer. In total the PSC-NRBSC team gathered more than 110 terabytes of raw and semi-processed TEM images. "This is near the limits," says Wetzel, "of what can be sent using commodity best effort network service — susceptible to variations in bandwidth due to other traffic."

At PSC this data goes to an archiver, while portions used for the next step — image alignment — get copied to an NRBSC cluster and specialized deskside machines. Because the four-camera parallel TEM imaging field captures thousands of images (up to nearly 14,000 frames) for each serial section, the frames must be stitched into a single-section mosaic before they can be stacked into a reconstructed 3D cortical volume.

Wetzel has focused on the stitching, the first step of which is to detect and correct for variations due to motion as a section repositions in the camera field and also for deformations that can happen as a section is in vacuum under the high-energy electron beam. "The most pervasive distortion that occurs," says Wetzel, "is warping — like localized stretchings of a rubber sheet. These can't be completely identified in individual planes but require comparison with nearby planes to partly separate changes due to distortion and those due to actual features of the tissue."

To get an idea of the amount of cortical information captured in each section, Reid makes an analogy to slicing a wedge of cheese. The cortical tissue is sectioned with an "ultramicrotome" — a high precision diamond knife. If each slice were a millimeter thick like a thin slice of cheese (instead of 40 nanometers), and the lateral dimensions increased by the same proportion, each slice would cover an area bigger than an NBA basketball court.



ZOOMING IN ON THE VISUAL CORTEX

The vertical extent of a section (bottom) encompasses depth from the cortical surface (grey matter) to pia (white matter). At low resolution, you see only the blood vessels (lighter features). Zooming in (top right) by a factor of roughly five shows the blood vessels more clearly, and begins to show neuron nuclei (dark spots). Zooming in by another factor of 10, you begin to see "what we care about," says Reid, "the wires — axons and dendrites, color coded yellow and green respectively," with magenta representing the neuron cell body.

Wetzel uses various software methods — fast Fourier transform correlations and other search methods — to find features in overlapping tiles for alignment into a single mosaic. His approach first produces low-resolution estimates — as much as 80X reduced — so that tile matching can bridge regions that appear "large" and mostly featureless at full resolution (capillaries, for example), and then can progress hierarchically to encompass the full dataset.

IF EACH SECTION WERE A THIN SLICE OF CHEESE, IT WOULD BE BIGGER THAN AN NBA BASKETBALL COURT.

Hood has focused on finding the best ways to stack the thin section images into 3D volumes for viewing and analysis — the proof of the pudding for wiring diagrams. He applies a pairwise non-linear method to register each reconstructed section with its neighboring section. He then uses a multi-resolution algorithm to "relax" the dewarping across the pairwise registrations, producing a global alignment for the entire stack. A test alignment of five 2D sections (143 gigabytes of raw data) has shown the viability of this approach. All the alignment processing is parallelized and scales well to large stacks of large-area sections.

"We have this huge, unprecedented dataset with the resolution of electron microscopy," says Reid, "that encompasses 12 out of 100 cells for which we know the function. We're beginning to make 3D models of this data, to analyze in detail, and to start drawing the wiring diagram."

MORE INFORMATION

www.psc.edu/science/2009/brain/

HOW TO KNOCK OUT HEPATITIS C

HAVING DETERMINED STRUCTURE OF THE VIRAL RNA'S ENTRY SITE BOUND WITH A DRUG COMPOUND, UNIVERSITY OF UTAH SCIENTISTS MAY HAVE THE BLUEPRINT FOR KNOCKING OUT HEPATITIS C

↓ THOMAS CHEATHAM
University of Utah

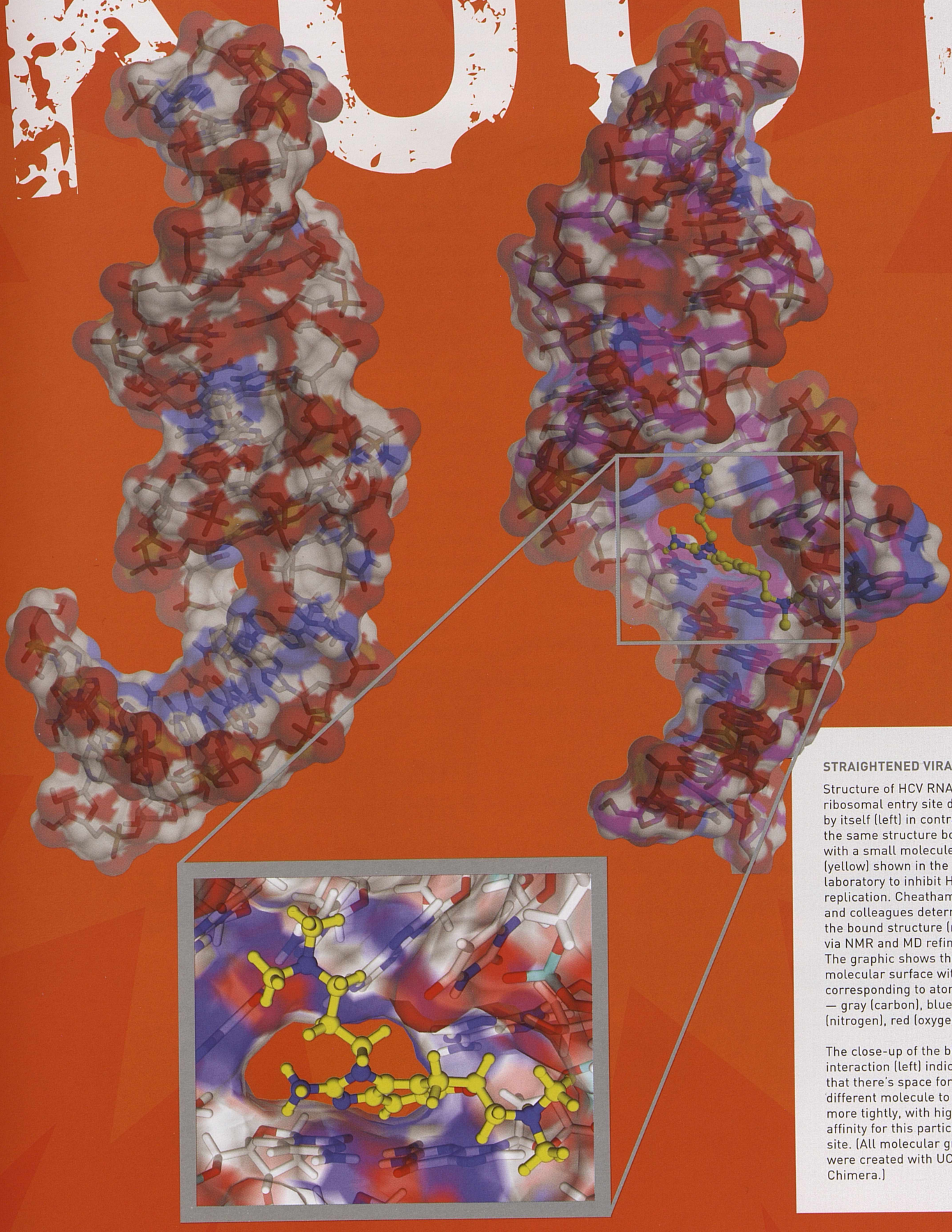
The hepatitis C virus infects an estimated 170 million people worldwide, with two to three million new cases each year. Persistent HCV infection causes liver cancer, the main cause of liver transplants, and in the United States alone leads to about 10,000 deaths annually.

Standard treatment is a cocktail of anti-viral medications that for many patients have a range of side effects, often including oppressive flu-like symptoms. In most cases, treatment must continue for a year or more to be effective, and even then succeeds little better than half the time, with recurrence common. "There's a tremendous need," says University of Utah computational biochemist Thomas Cheatham, "for more effective therapeutics."

As the intense effort on HIV has underscored, finding drugs that deliver a knockout punch to a virus — as opposed to managing symptoms — is a Mt. Everest research problem. A skilled team led by Cheatham's colleague Darrell Davis at the University of Utah, one of the leading pharmacy schools in the country, has mounted an expedition to climb this mountain.

Their ambitious plan is a direct thrust at HCV's genetic material, its RNA, by which it copies itself inside cells. While there's a long way to go, the good news is the Utah team has established a base camp on the slope of the mountain beyond where anyone else has been able to go. Using a combination of advanced experimental methods and powerful computational resources at PSC and other TeraGrid sites, Davis, Cheatham and their colleagues have for the first time determined the atom-by-atom structural details of a critical part of HCV's RNA bonded with a small molecule that has demonstrated ability to inhibit HCV replication.

The small molecule is from a group of similar molecules that Davis has tested in his laboratory for RNA binding activity. "This structure along with our HCV compound library," says Davis, "places our research team in a unique position to develop a new class of RNA targeted therapeutics."



STRAIGHTENED VIRAL RNA

Structure of HCV RNA's ribosomal entry site domain by itself (left) in contrast to the same structure bound with a small molecule (yellow) shown in the laboratory to inhibit HCV replication. Cheatham, Davis and colleagues determined the bound structure (right) via NMR and MD refinement. The graphic shows the molecular surface with color corresponding to atoms — gray (carbon), blue (nitrogen), red (oxygen).

The close-up of the binding interaction (left) indicates that there's space for a different molecule to bind more tightly, with higher affinity for this particular site. (All molecular graphics were created with UCSF Chimera.)

PINNING DOWN RNA

Aside from being a big step toward new drugs for HCV treatment, this structure is a notable advance, say the researchers, for structure-based, computer-aided drug design (CADD) in general. Having played a role in developing Viagra, HIV-inhibitors and other therapeutic drugs, structure-based CADD has within the past decade gained a significant role in pharmaceutical research — reducing lengthy and costly trial-and-error laboratory work. Nearly all of the CADD effort, however, has focused on proteins, virtually none on nucleic acids such as RNA.

In many pathogens, RNA molecules present an inviting target for therapeutic drugs, but they have not been conducive to CADD. "By its inherent nature," says Cheatham, "RNA is much harder to precisely define structurally than a protein. It's flexible, very sensitive to surroundings, and binding of small molecules often radically changes the structure. As a result there's not been a lot of effort in CADD with RNA or other nucleic acids."

To overcome these challenges, the Utah team applied advanced techniques of NMR (nuclear magnetic resonance), a widely used method of obtaining structural information from crystallized molecules — basically it's like doing MRI on a molecule and using the information gained to deduce its three-dimensional structure. Along with distances between hydrogen atoms, usually available from NMR, Davis obtained "residual dipolar couplings" (RDC) — NMR information that defines directional alignment of one part of the molecule with another.

The next step is to refine the structure — a process for which Davis and Cheatham relied on molecular dynamics (MD), a computational method that models the forces and distances between atoms. With MD computations, the scientists could test potential structures — among many that meet the NMR constraints — to find which have the lowest energy expenditure, the native state of a molecule, and align best with the NMR information, including RDC.

One of the major problems with RNA structure refinement has been the applicable MD "force fields" — expressions that describe the atom-to-atom interactions, and a goal of Cheatham's HCV work is to improve these force fields. "RNA can adopt many different conformations," says Cheatham. "A loop can fall apart because it's moving. We know there's systematic problems, and we can test the force fields so that we know we've got the correct one."

STRAIGHTENING THE BEND

The Utah scientists focused on a particular stretch of HCV's RNA, its internal ribosomal entry site (IRES), a loop-like structure near one end that is, in effect, a marker for where the RNA-copying process begins. Prior research showed that there's a helical bend — a more than 90° turn — in this loop of RNA when it's not bound with another molecule, resulting in an overall L-shape for the IRES region.

Working with the NMR data, Davis first used an MD program called Xplor to refine the HCV IRES and small molecule bound together by themselves — without surrounding water molecules and ions as in the natural cellular environment. Cheatham then used TeraGrid systems to run the more exacting AMBER (Assisted Model Building with Energy Refinement) MD package, which he has helped to develop. With an extensive series of runs, he corrected problems with the RNA force fields and, taking as input the results from Davis's work, further refined the structure of the IRES-drug complex, this time including surrounding water and ions, to arrive at a more accurate set of final structures.

The refined structure shows that binding of the small molecule straightens the bend of the IRES loop, a dramatic structural shift that explains the biological effect of inhibiting replication. With this success as a starting place and demonstration of method, the Utah team is working to narrow the candidates of similar small-molecule drug compounds that have ability to stop the virus, which can then be synthesized and laboratory tested in virus cultures to find the most potent possible molecules to develop as drugs.

"This gives us a starting place," says Cheatham, "to further optimize this class of compounds, to increase the affinity with the IRES binding site and the specificity of its binding to only the RNA target we're aiming at, ultimately leading to a new class of highly potent HCV therapeutics."

> **"THIS STRUCTURE PRESENTS A NEW CHEMICAL PARADIGM FOR SMALL MOLECULE INHIBITORS OF STRUCTURED RNAS."**

A number of other viral RNAs — including yellow fever, dengue and swine fever — are known to have an IRES region like that of HCV, suggesting a structure-based CADD program to develop RNA drugs for a range of viral diseases. The Utah scientists have mapped a path that suggests this Mt. Everest will be climbed. "This structure," says Davis, "presents a new chemical paradigm for small molecule inhibitors of structured RNAs."

MORE INFORMATION

www.psc.edu/science/2009/hcv/

PAINKILLER SNAILS

Along with their work on hepatitis C, Thomas Cheatham and his colleagues at Utah are investigating potent snail venoms that show promise of treating chronic pain without addiction. The same cone-shaped shells that populate beaches are home to cone snails, carnivorous hunters that can paralyze small fish almost instantly with a neurotoxic venom.

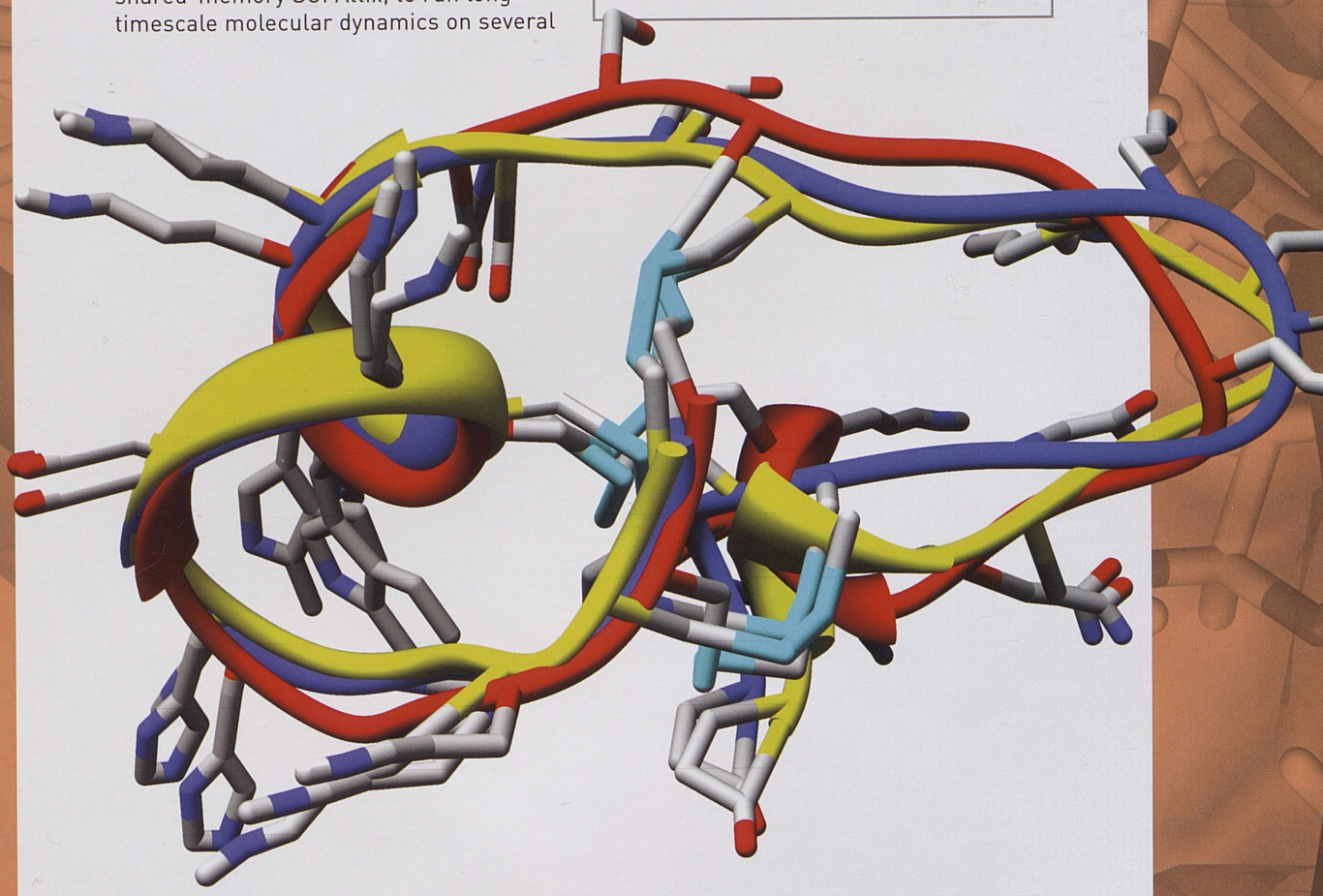
These venoms, known as conotoxins, have shown significant potential to treat "neuropathic pain" — chronic long-term pain, usually involving nerve damage. Unlike morphine, conotoxins don't affect opioid receptors and therefore offer the tantalizing prospect of suppressing pain without addictive side effects.

"This snail produces thousands of different types of these toxins," says Cheatham, "and they can lead to total paralysis. If you break down the separate groups, each affects a different part of your nervous transmission. We might be able to design these to treat a particular pain or disease state in selective ways, with minimal side effects."

In 2008, Cheatham and grad student Pawel Gruszczynski from the University of Gdansk in Poland used Pople, PSC's shared-memory SGI Altix, to run long-timescale molecular dynamics on several versions of conotoxins. As they reported in *ChemMedChem* (March 2009), their analogue conotoxin model retained the ability to produce analgesia, and suggests one model for new painkiller therapeutics.

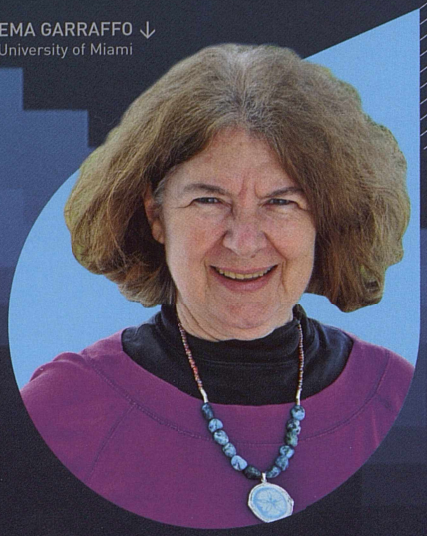
In another study, they modeled the binding site of two different "pain channels" and demonstrated conotoxin selectivity. "One conotoxin binds well to one channel and doesn't bind well to the other and vice-versa," says Cheatham. "Additionally, we think there are multiple binding modes — which not only helps explain the experimental data but also helps to explain the diversity of activity of these compounds." Their results from both studies, analyzed in collaboration with University of Utah experimentalists Greg Bulaj and Toto Olivera, who lead the field of conotoxin research, are steps toward mapping the molecular features responsible for the drug's biological effect.

A conotoxin average structure from 200 nanoseconds of MD simulations by Cheatham and colleagues. The side chains show the heavy atoms —carbon (gray), nitrogen (blue), oxygen (red). The backbones, a different color (red, blue and yellow) for each structure, show motion, and the side chains show differential localization.



DEEP WATER BLUES

Benjamin Franklin published the first map of the Gulf Stream, drawn for him by a Nantucket sea captain in 1770 and later printed in France.



ZULEMA GARRAFFO ↓
University of Miami

OCEAN SCIENTISTS FIND BETTER RESOLUTION IS NEEDED TO MODEL GULF STREAM COOLING SCENARIOS

Who first mapped the Gulf Stream? Clue: As colonial postmaster, he wondered why packet boats from England took two weeks longer to get to America than merchant ships piloted by Nantucket sea captains. With help from one of these captains, who as a whaler knew about the North Atlantic's "river of warm water," America's first great scientist, Ben Franklin, published a chart of the strong west-to-east current that British captains didn't know they were stemming in their established route to New York.

Nearly 250 years later, a supercomputing system named for Franklin, PSC's BigBen, has helped to show why it's still important — maybe more than ever — to accurately map currents of the North Atlantic. What's critical nowadays are deep-down pathways of the "Nordic overflows" — dense, cold waters that cascade as if over the lip of a bathtub as they flow over underwater ridges out of the Nordic Seas into the North Atlantic, where they feed deep circulation patterns that convey cold water southward while tropical water flows northward via the Gulf Stream.

The worry is this: This ocean conveyor belt system, known to scientists as the Atlantic Meridional Overturning Circulation (AMOC), may be threatened by global warming. Much climate modeling suggests that increased cold, fresh water from glacial melting in Arctic regions and a warmer atmosphere will slow down the AMOC during the next few decades. Some scenarios include the possibility of an AMOC shutdown as an extreme case.

There are many scenarios, with predictions that are controversial and far from clear. What's needed is reliable climate modeling, and that in turn requires accurate representation of the ocean circulation, including the deep-down Nordic overflow pathways. The problem — as University of Miami oceanographer Tamay Özgökmen and collaborators Zulema Garraffo, Yeon Chang and Hartmut Peters reported in *Ocean Modelling* (March 2009) — is that existing models of these flows aren't up to the task.

Using BigBen, a resource of the NSF TeraGrid, and a state-of-the-art ocean model called HYCOM, Garraffo — of the University of Miami's Centre for Computational Science — and collaborators simulated the Nordic overflows at three different grid resolutions. Only at the highest resolution — more than 10 times finer than typical climate models — did their ocean model begin to represent flow volume and location and other properties in reasonable agreement with observed data. "This work quantifies the importance of model resolution to solve for the deep-flow pathways of the Atlantic," says Garraffo. "There are many aspects of the circulation for which conclusions cannot be drawn from coarse-resolution models."

For reliable predictions of how Nordic overflows can affect the AMOC, the researchers conclude, it is important to reproduce the ocean processes that occur at small scales. "Climate models," says Özgökmen, "are typically based on very coarse resolution. We show that lower resolutions fail quantitatively or even qualitatively to capture the structure of the overflows. The details of topography are essentially endless, and unless you capture this, you won't know what happens with the deep flows."

"THIS WORK QUANTIFIES THE IMPORTANCE OF MODEL RESOLUTION TO SOLVE FOR THE DEEP-FLOW PATHWAYS OF THE ATLANTIC."

BIGBEN & HYCOM

Since even before BigBen became a TeraGrid production resource in 2005, Garraffo — working with colleagues George Halliwell and Eric Chassignet and with PSC consultant John Urbanic and other PSC staff — has used this system to model the Atlantic Ocean. Her aim has been to validate and improve the performance of HYCOM (HYbrid Coordinate Ocean Model) — a distinctive model with a pedigree going back to the mid-1990s, when its forerunner, MICOM (Miami Isopycnic Coordinate Ocean Model), running on a Cray T3D at PSC, became the first model to correctly capture the "separation" of the Gulf Stream at Cape Hatteras, where it veers from a shoreline-hugging course northeast into the open sea.

The key to this breakthrough, as with the recent HYCOM modeling of the Nordic overflows, was sufficient computational capability to increase the grid resolution beyond what had before been possible. With the T3D, researchers could run the model on a massively parallel system and at a horizontal resolution of 1/12° latitude and longitude (equivalent to about six kilometers) for a decade of simulated ocean time. At lower resolutions, the coarseness of the model prevented it from accurately representing even such prominent features as the direction of flow past Cape Hatteras.

The problem is that improvements in resolution, such as from 1/3° (about 30 kilometers) to 1/12°, greatly increase the amount of computing required. Only with the availability of BigBen did it become feasible to use HYCOM to simulate the entire Atlantic Ocean for 40 years of ocean time even at moderate resolution. Garraffo's series of runs at 1/3° showed good agreement with historical observations and, further, test runs at 1/12° — using 1,936 BigBen processors — showed excellent agreement for sea-surface temperature, sea-surface height, and current transports.

The distinctive feature of HYCOM, inherited from MICOM and giving it a significant advantage in accuracy over other models, is that it is "isopycnic" — which means constant density. With this mathematically sophisticated approach, developed by University of Miami ocean scientist Rainer Bleck, the ocean is divided into vertical layers (HYCOM is typically used with 32 layers) so as to prevent spurious heat diffusion from the surface to depth as the model progresses in time, with each layer preserving its own water mass.

HYCOM builds on MICOM by extending its usefulness to shallow coastal areas, where it allows finer vertical resolution to capture the turbulence of near-shoreline effects. The advantages of HYCOM led to its being chosen as the next-generation ocean model by the U.S. Naval Oceanographic Office and by the National Oceanic and Atmospheric National Centers for Environmental Prediction.

NORDIC OVERFLOWS

In 2008, Garraffo began collaborating with Özgökmen, Chang and Peters to address the Nordic overflows. Is it possible, they asked, to realistically model the deep pathways of these overflows and, if so, what horizontal resolution is required?

Garraffo and Chang used computers at the University of Miami to run HYCOM for the North Atlantic, including part of the Norwegian Sea, first at 1° resolution (about 100 kilometers) then at 1/3°. For the high resolution run at 1/12°, Garraffo turned to BigBen at PSC.

The model seafloor was crucial. Prior to actually initiating the simulation of flows, the model used interpolation routines to represent seafloor topography from "bathymetric" maps of the underwater channels and ridges. The researchers found that for low resolutions the standard interpolation routines seriously distorted these features. "At low resolution," says Garraffo, "the model can't see all the channels because they are under resolved."

One of the main overflow passages is a narrow, deep channel north of the Faroe Islands, the Faroe Bank Channel (FBC). While the 1/12° model essentially captured the FBC topography, the 1/3° model underrepresented its depth by 200 meters, and at 1° the FBC vanished from the model.

Chang, Peters, Özgökmen, and Garraffo carefully examined all the available deep-flow observations for comparison with the modeling. They found that the three simulations showed clear and strong resolution-dependent differences, and the differences increased — as the models ran for 19 months of ocean time. At 1° the overflow can't find its path through the FBC, and the cold water masses go the wrong direction. Although the overflow pathways at 1/3° are more realistic, the model results still disagree with observations, with most of the overflow water ending up in the wrong ocean basin. "The 1/12° resolution," says Garraffo, "is not excellent, but at least it allows you to see the current and eddies."

"We find that the mean structure of the overflows in Denmark Strait and Faroe Bank Channel are simulated only at the highest resolution," says Özgökmen. "Severe problems with the lower resolution cases extend far beyond the actual overflows to large parts of the deep circulation."

In an experiment to see if they could improve the results for deep circulation resulting from topographic errors, the researchers manually corrected the topography in several ridges and channels for 1° resolution. This simulation reduced errors in some areas, but increased discrepancies in other parts of the overflow region. Manual correction to bathymetry could have some limited usefulness, the researchers conclude, and they suggest that standard topography-generating

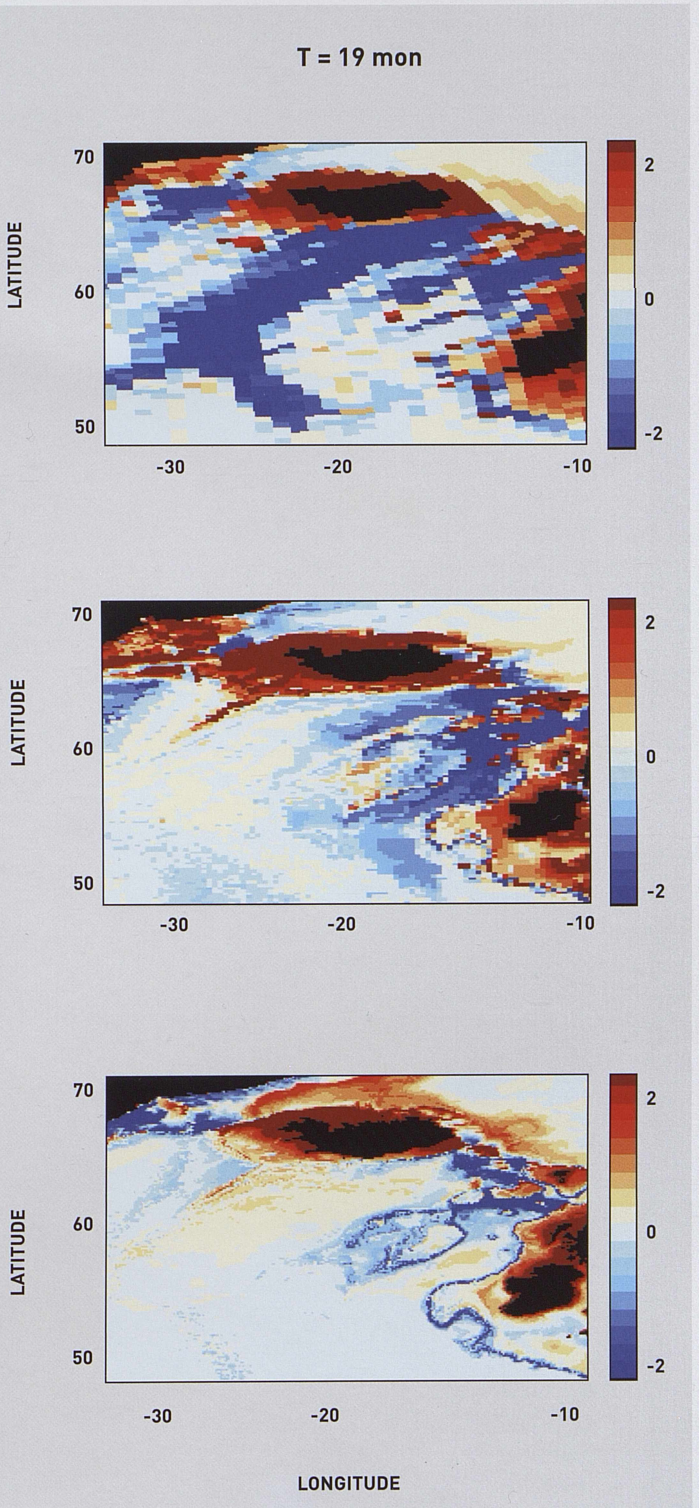
algorithms be used carefully with coarse grids.

Their main conclusion, nevertheless, is that higher resolution, hence more powerful computing, is needed for climate modeling. "These results," the researchers say in their paper, "demonstrate the importance of an accurate representation of the domain geometry, in particular the channels of the complex Iceland-Scotland ridge system, in order to reproduce the pathways of the deep AMOC."

"Climate or ocean models," says Özgökmen, "run at 1° can't resolve these channels and canyons. They don't get the deep water right, and so they cannot describe correctly the deep pathways and transports. To produce fairly realistic overflow and AMOC, the resolution needs to be an order of magnitude larger."

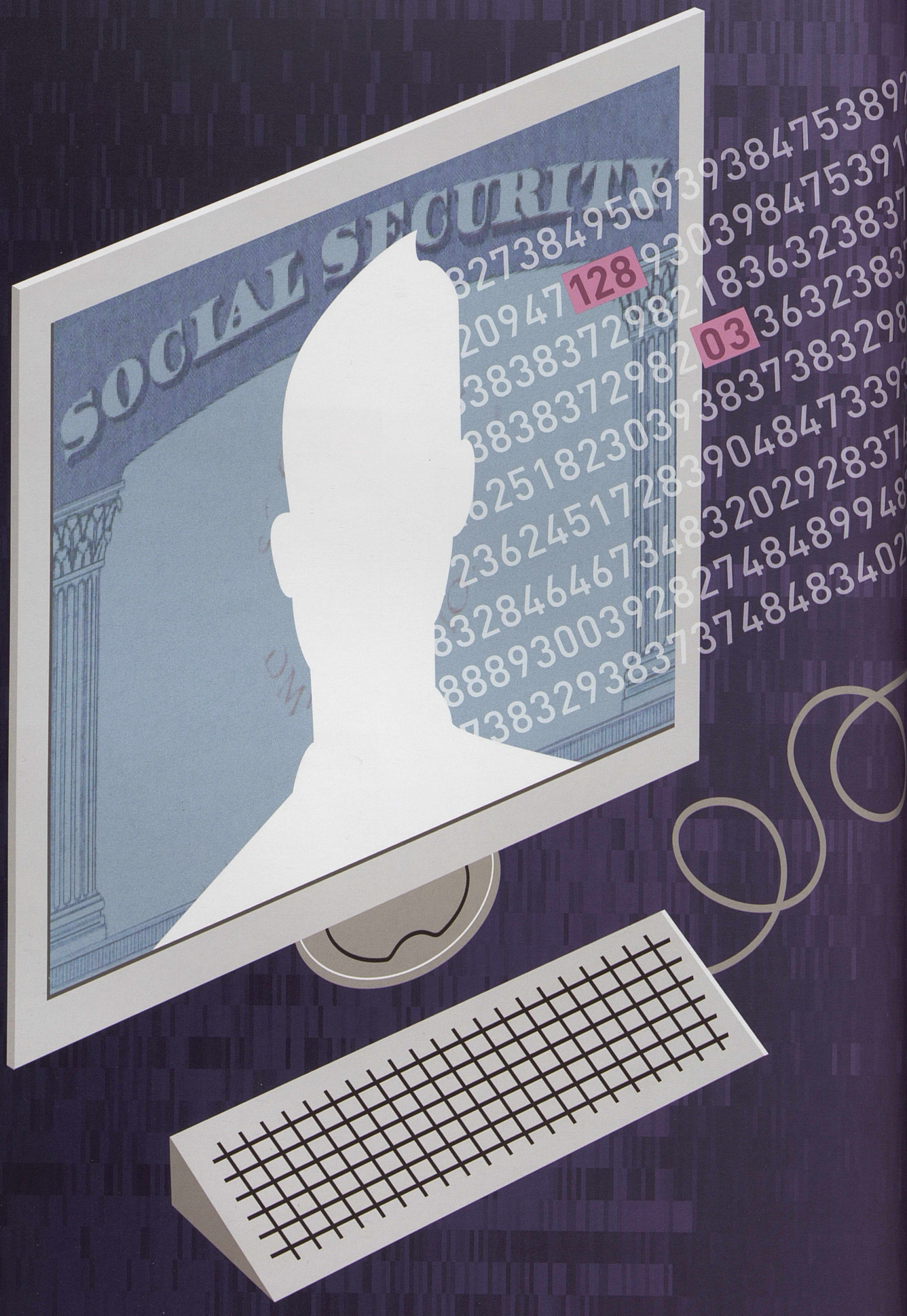
MORE INFORMATION

www.psc.edu/science/2009/deepwater/

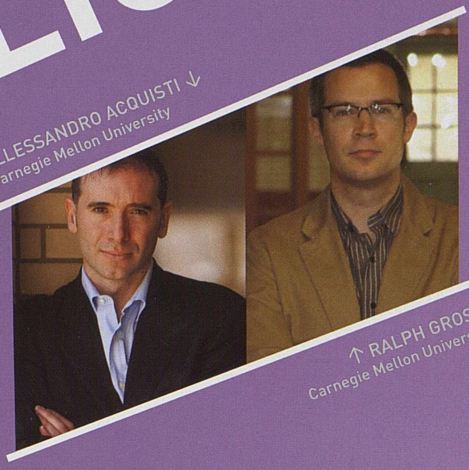


SEA-SURFACE TEMPERATURE CONTOURS

Temperature differences of the ocean surface indicate flow patterns of the cold overflows out of the Nordic Seas into the North Atlantic. These three views show results at three different resolutions (low to high) after 19 months of simulated overflows. At 1°(top) most of the cold water masses, rather than passing through the Faroe Bank Channel (FBC), are pushed south over the wider Iceland-Faroe Ridge. At 1/3° (middle) only a small part of the overflow turns toward the northwest to pass the partially blocked FBC. At 1/12° (lower) there are no unrealistically strong currents, and the overflow, as it should be, is concentrated in the FBC.


SOCIAL SECURITY
128
03

WAKEUP CALL FOR PUBLIC DATA



ALLESSANDRO ACQUISTI ↓
Carnegie Mellon University

↑ RALPH GROSS
Carnegie Mellon University

WITH DATA-INTENSIVE COMPUTING ON PSC'S POPLE, CARNEGIE MELLON RESEARCHERS EXPOSE VULNERABILITY IN ONLINE INFORMATION

Remember when everyone paid for groceries at the supermarket with cash or a check? "Identity theft" in those halcyon days might have been a story idea for an episode of the Twilight Zone. Now it's one of the fastest growing crimes in America, costing Americans almost $50 billion in 2007 alone.

The information age has changed commercial life irrevocably in favor of rapid transactions at high volume, and inherent in this transformation is the risk that highly skilled, dishonest people can obtain identity information and use it fraudulently. Further complicating the situation are social networks such as Facebook. As much as they add to our ability to communicate, they also make information public that exacerbates the potential for criminals to create financial chaos.

Alessandro Acquisti is acutely aware of these risks. A professor of information systems and public policy at Carnegie Mellon University, Acquisti focuses his work on "the behavioral economics of privacy" — the trade-offs, incentives and cognitive biases associated with protecting and revealing personal information. "Social networking sites make it so easy to broadcast personal information," says Acquisti, "and we forget that what was hastily published today on a profile intended for our peers may be saved for an unpredictable amount of time, and perhaps used — or abused — by others in a completely different context."

This past July these concerns hit the front page of the *New York Times*. Using Pople, PSC's Silicon Graphics Altix 4700, for an intensive search and analysis of huge amounts of data, Acquisti and post-doctoral researcher Ralph Gross showed that with information gleaned from various online sources it's possible to predict most — and sometimes all — of an individual's Social Security number (SSN). "In a world of wired consumers," says Acquisti, "it is possible to combine information from multiple sources to infer data that is more personal and sensitive than any single piece of original information alone."

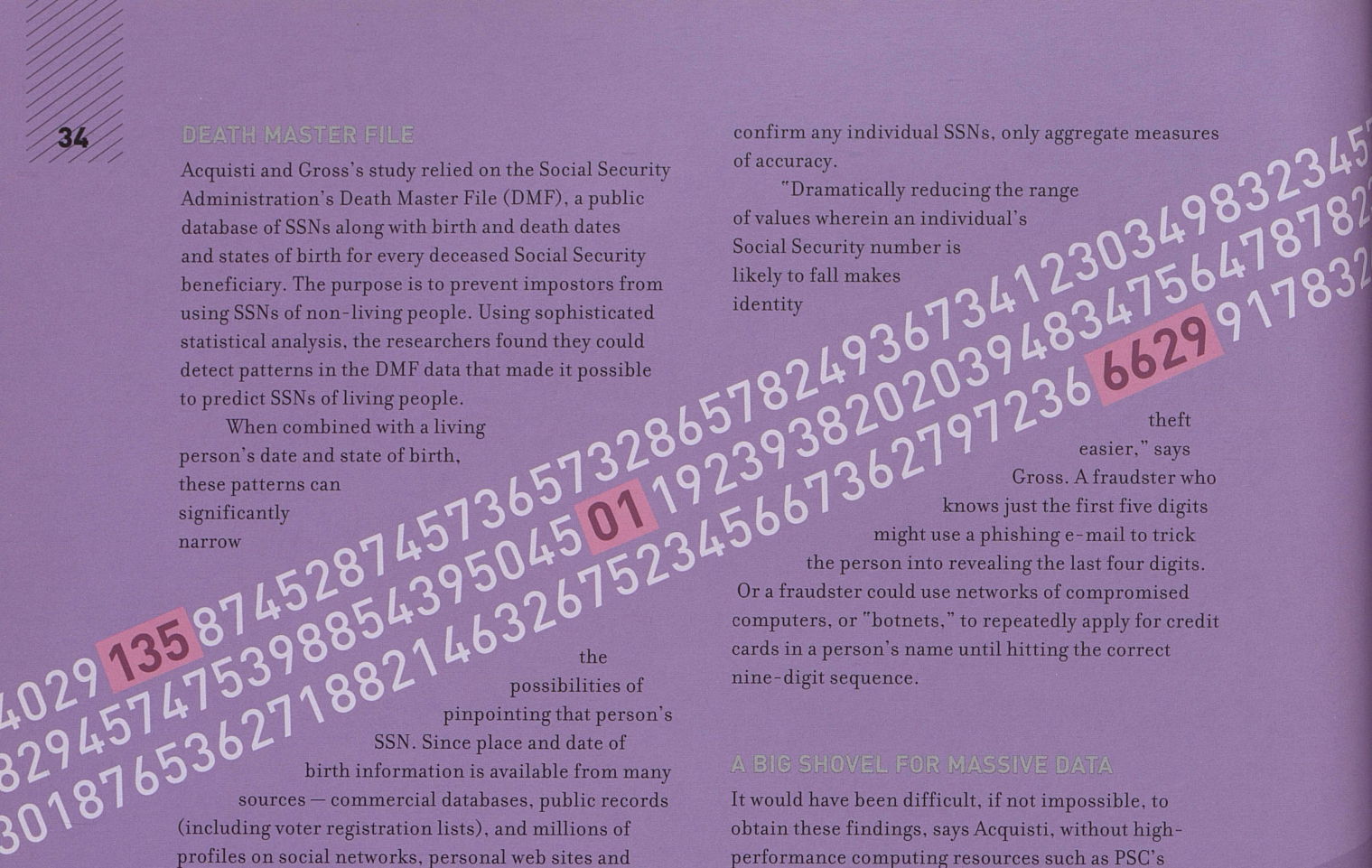
The *Times* and other national media reported their findings — published in *Proceedings of the National Academy of Sciences* (July 7, 2009) — as a major news story. Simply put, Acquisti and Gross show that privacy safeguards from decades back are no longer adequate. Since many businesses use SSNs as passwords or other authentication — unanticipated when Social Security began — the SSN predictability they show poses a risk of identity theft. Their published findings exclude sensitive details of their ability to predict SSNs, and the intent is to alert both individuals and policy-makers.

Privacy experts described Acquisti's findings as a wakeup call. "Social Security numbers are an aging technology," said Peter Swire,of Ohio State University, as reported in the *Times*, "and we have to do serious planning for what will come next."

DEATH MASTER FILE

Acquisti and Gross's study relied on the Social Security Administration's Death Master File (DMF), a public database of SSNs along with birth and death dates and states of birth for every deceased Social Security beneficiary. The purpose is to prevent impostors from using SSNs of non-living people. Using sophisticated statistical analysis, the researchers found they could detect patterns in the DMF data that made it possible to predict SSNs of living people.

When combined with a living person's date and state of birth, these patterns can significantly narrow the possibilities of pinpointing that person's SSN. Since place and date of birth information is available from many sources — commercial databases, public records (including voter registration lists), and millions of profiles on social networks, personal web sites and blogs — the possibility to predict SSNs becomes a reality.



The statistical patterns and birth information can predict SSNs because the Social Security Administration's methods for assigning numbers, based partly on geography, are well known. For most individuals born since 1989, furthermore, SSNs are assigned shortly after birth, making those numbers easier to predict than for earlier birth years.

Acquisti and Gross tested their method, a complex statistical algorithm, on DMF records for people who died between 1973 and 2003. For people born after 1988, they could with a single attempt identify the first five digits for 44 percent of people. For people born between 1973 and 1988, they could identify the first five digits in 7 percent of the cases. With more attempts, up to but fewer than 1,000, they were able to identify all nine SSN digits for 8.5 percent of people born after 1988.

For smaller states and recent years of birth, their accuracy was higher than for large states and earlier years. They needed 10 or fewer attempts, for instance, to predict all nine digits for one out of 20 SSNs issued in Delaware in 1996. "If you can successfully identify all nine digits of an SSN in fewer than 10 or even fewer than 1,000 attempts," said Acquisti and Gross, "that Social Security number is no more secure than a three-digit PIN."

When the researchers tested their method on non-DMF data — using birth dates and hometowns that students had self-reported on popular social networking sites — the results were almost as good despite the typical inaccuracies of social-network data. The researchers used enrollment records to confirm the accuracy of their predictions, though they didn't confirm any individual SSNs, only aggregate measures of accuracy.

"Dramatically reducing the range of values wherein an individual's Social Security number is likely to fall makes identity theft easier," says Gross. A fraudster who knows just the first five digits might use a phishing e-mail to trick the person into revealing the last four digits. Or a fraudster could use networks of compromised computers, or "botnets," to repeatedly apply for credit cards in a person's name until hitting the correct nine-digit sequence.

A BIG SHOVEL FOR MASSIVE DATA

It would have been difficult, if not impossible, to obtain these findings, says Acquisti, without high-performance computing resources such as PSC's Pople. After first working with desktop computers and coming to a bottleneck in their work, the researchers approached PSC. This was mid-2008, at an important phase in the project. "At that stage," said Acquisti, "we had a rough idea of the results, but to go forward we had to try many different variations of the algorithms. It would have been incredibly difficult to do this, or taken much, much longer without access to this system."

Acquisti and Gross and several graduate students who worked with them turned to Pople. Named for Nobel laureate chemist John Pople of Carnegie Mellon), this system features 768 cores (processors) and 1.5 terabytes of shared memory (all of memory accessible from each core). Working with a core dataset of about eight gigabytes, the researchers used 100 processors for up to eight hours for each of a series of seven runs.

PSC staff installed Octave — an open-source version of the programming language MATLAB

> "GIVEN THE INHERENT VULNERABILITY OF SOCIAL SECURITY NUMBERS, IT IS TIME TO STOP USING THEM FOR VERIFYING IDENTITIES AND REDIRECT OUR EFFORTS TOWARD SECURE, PRIVACY-PRESERVING AUTHENTICATION METHODS."

— and wrote a script to submit a large number of parallel Octave jobs simultaneously. This facilitated the Acquisti team's interactive process, which involved doing many runs representing different states and computational strategies, checking and analyzing results and rethinking before running more variations. PSC's consulting, said Acquisti, was "extremely helpful."

"This project," said Sergiu Sanielevici, PSC director of scientific applications and user support, who also leads user support and services for the TeraGrid, "exemplifies how powerful systems like Pople can open doors to data-mining and data-centric research in fields not traditionally associated with HPC, such as the social sciences, and make it possible to get answers that would otherwise be impractical or impossible."

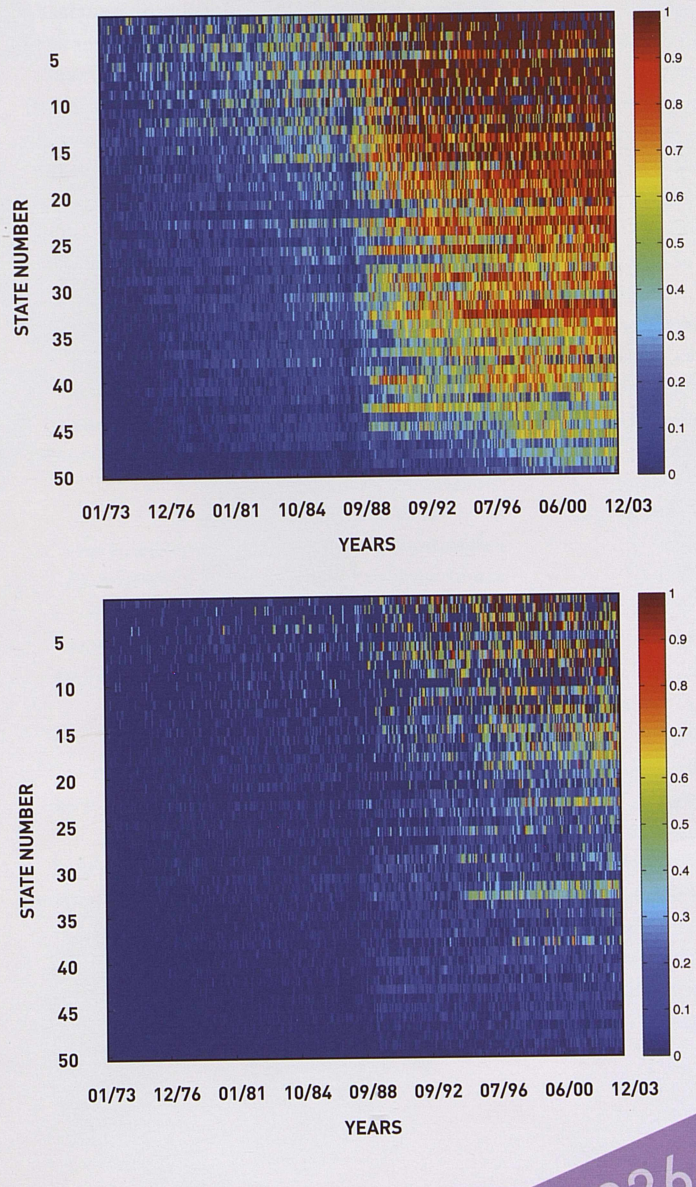
Carnegie Mellon graduate students Jimin Lee, Ihn Aee Choi, Dhruv Deepan Mohindra, and Ioanis Alexander Biternas Wischnienski collaborated in this research and did much of the hands-on computational work.

Future SSNs could be made more secure, say Acquisti and Gross, by switching to a randomized assignment scheme. Protecting people who already have been issued numbers, however, is more difficult. Given the ease with which SSNs can be predicted — particularly the first five digits and particularly for the millions of Americans born since 1988 — legislative and policy initiatives aimed at removing the numbers from public exposure, or redacting the first five digits, added Acquisti, may be well-meaning but misguided.

"Given the inherent vulnerability of Social Security numbers," says Acquisti, "it is time to stop using them for verifying identities and redirect our efforts toward implementing secure, privacy-preserving authentication methods." Methods to consider include two-factor authentication, similar to the PIN number-card combinations used for bank accounts, and digital certificates.

MORE INFORMATION

www.psc.edu/science/2009/privacy/



PREDICTION SUCCESS RATIOS

Prediction accuracies for Death Master File records with January 1973 to December 2003 birthdays across the 50 states. Ratios of accurate prediction for the first five digits (top), and ratios of accurate prediction for the complete SSN with less than 1,000 attempts (bottom). In each quadrant, columns represent months, and rows represent states (sorted by their 1973 births, lowest to highest). The colors in each cell represent ratios (from 0 through 1, dark blue through red) out of monthly SSN counts. This fairly unassuming graphical figure, notes Acquisti, represents results of "more than 700,000 regressions over a very large data set."

CHAIN, CHAIN, CHAIN

WITH EXPERIMENTS BACKED UP BY COMPUTATION, SCIENTISTS FOR THE FIRST TIME OBSERVE THE MOLECULE-BY-MOLECULE TRANSFORMATIONS OF A MOLECULAR CHAIN-REACTION

Free radicals can cause explosions. This is not only a socio-political statement (and theme of a Dostoevsky novel), but also a chemical phenomenon. "Flames and explosions that occur, usually in the gas phase, involve what are known as 'free radical chain reactions,'" explains John Yates, now a chemistry professor at the University of Virginia, "a class of reactions that's been known for maybe 90 years."

While scientists have known of these molecular chain reactions, they had never observed the atom-by-atom details of how the molecules break into "free radicals" — molecules with an unpaired electron, which makes them highly reactive — and form new bonds. Recent work by Yates and colleagues, reported in *Science* (December 2008), has rendered the unseeable seen. In a fruitful collaboration between laboratory experiment and supercomputing, Yates — formerly at the University of Pittsburgh, founding director of its Surface Science Center (SSC), and graduate student Peter Maksymovych (now a Wigner fellow at Oak Ridge National Laboratory) teamed with University of Pittsburgh theoretical chemist Ken Jordan and Dan Sorescu of the Department of Energy's National Energy Technology Laboratory.

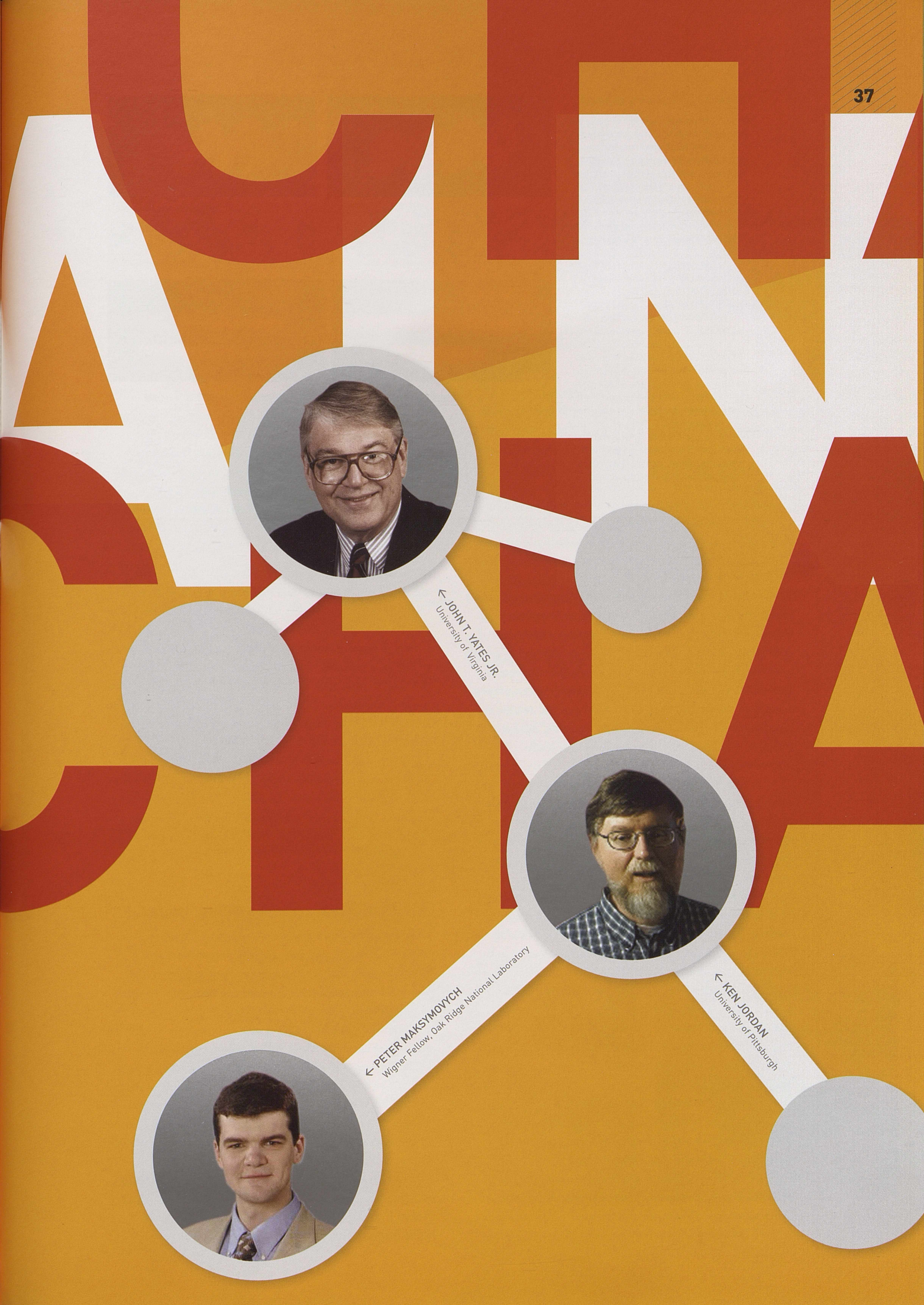
Yates and Maksymovych used the SSC's ultra-high vacuum scanning tunneling microscope (STM) to induce a reaction in a chain of dimethyldisulfide (CH_3SSCH_3) molecules self-assembled on a surface of gold. The sulfur-sulfur bonds broke and reformed into new DMDS molecules via a propagating chain reaction. To follow-up on these intriguing results, Jordan and Sorescu did theory-based computations on PSC's BigBen. The computations provided new insight and support the conclusion of a free-radical chain reaction, the first time that such a reaction has been produced with molecules on a surface.

"This is the first time anyone has ever seen a chain reaction molecule-by-molecule," says Yates. "From a historical point of view, this well established type of chemical reaction, previously known only from indirect measurements in the gas phase, has now finally been seen when you pin the molecules to a surface."

Jordan concurs. "What's new is showing that you can get a chain reaction on a surface." He credits the collaboration between experiment and computation. "In many grand challenge problems — and I think this is true of a lot of the work done at PSC — significant progress occurs because computational and experimental researchers have teamed up."

"This was a great example of a collaboration involving the resources of two major centers — the University of Pittsburgh Surface Science Center and the Pittsburgh Supercomputing Center," says Yates. "It was a vibrant collaboration between theorists and experimentalists."

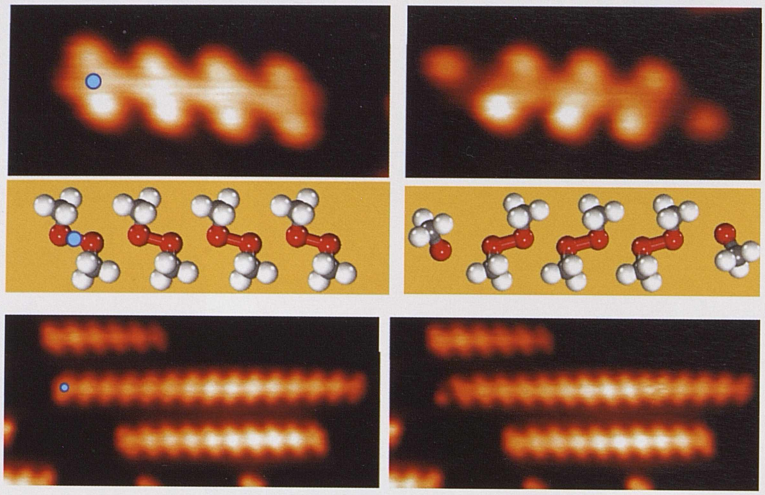
The new findings suggest practical applications related to nanolithography, a light-induced process by which thin, electron-conducting pathways are laid down on semiconductor surfaces in the manufacture of microchips. "The broad field of photochemistry on surfaces," the researchers conclude in *Science*, "will now have to account for chain processes in surface reaction mechanisms."


JOHN T. YATES JR.
University of Virginia
PETER MAKSYMOVYCH
Wigner Fellow, Oak Ridge National Laboratory
KEN JORDAN
University of Pittsburgh

CHAIN REACTION ON A SURFACE

These STM images show the before and after (reacted and unreacted) results of electron-induced dissociation of a single CH_3SSCH_3 molecule and the self-assembled chains on a gold surface. The schematic ball and stick models are shown beside their related STM images. An introduced electron (blue dot) initiates the reaction.

Maksymovych and Yates first performed the STM experiment on a single DMDS molecule, then two molecules, then four (upper right), then a 15-unit chain. This last experiment (lower right) initiated a chain reaction involving 10 of the 15 DMDS molecules in a row, and produced nine new molecules before the gold surface quenched the reaction. "The self-assembled molecular structure," wrote the researchers in *Science*, "thus redirects the energy flow toward a chemical chain reaction involving multiple steps, rather than rapid dissipation into the metal bulk."



SOLID GOLD CHEMISTRY

The secret of this step forward in the chemistry of chain reactions was to confine the reaction to a surface — in this case a surface of solid gold. Actually a little ball of gold presenting single crystal facets on the end of a gold wire, explains Yates. As the part of a solid exposed to the world, a surface is a dynamic place for interactions — such as oxidation, corrosion and catalytic reactions, among others — and a discipline known as "surface science" has grown up around studies of such phenomena.

Before the DMDS results, scientists doubted that a propagating chain reaction could be sustained on a surface. "The traditional view," says Jordan, "is that the excess energy is going to dissipate too rapidly into the atoms in the surface and that this would kill the reaction."

Maksymovych used the SSC's state-of-the-art STM to introduce DMDS molecules onto the gold surface. Yates and Maksymovych chose to work with DMDS because similar molecules were known to self-assemble into chains under the right conditions.

You can think of a DMDS molecule as dumbbell-shaped, with methyl groups (CH_3) as the heavy ends joined by the S-S bond. Through the self-assembly that occurred as the DMDS molecules were deposited, the dumbbells line up side-by-side on the gold surface, leaning to the left. Maksymovych then used the tip of the STM probe to inject electrons into the first DMDS molecule. A short time later, the DMDS molecules were reformed, still in a chain but now leaning to the right, opposite of their starting orientations.

To a non-scientist it may have looked like the molecules did a group shift, left to right, but Yates and Maksymovych saw something else. Their analysis of the STM images revealed that the reformed DMDS molecules were not only tilted but also offset to the right, strong evidence that chemical bonds were broken in the process. From this and other evidence, the scientists deduced this scenario: an electron splits the first S-S bond, creating two CH_3S radicals. One bonds to the gold surface while the other attacks and splits the DMDS molecule next to it in the chain, forming two more radicals. Two of the three then bond to form a new DMDS, while the third radical repeats this step on the next DMDS down the line.

Like a patterned dance figure in two lines of dancers, the molecules broke apart and reformed with different partners down the chain. "It's like unzipping a zipper," says Yates, "a repetitive process down the chain. It's really beautiful — never been seen before, that's for sure."

THE CLINCHER: LOWER ENERGY BARRIERS

Why does this chain reaction happen when no one thought it would, and how can the experimentalists know for sure that it happened as they believe it did? To look more deeply, Jordan and Sorescu went to work to compute the energy changes during the course of the reaction. When graphed, the energy trajectory of a reaction can look like a roller coaster, with peaks representing energy barriers — in this case, the energy needed to break the bond between the two sulfur atoms of DMDS. "The experimental work does not give you the energy barriers directly," says Jordan, "and this is one of the ways that theory — computational work — is very useful."

Their computations relied on density-functional theory (DFT), for which Walter Kohn shared the 1998 Nobel Prize in chemistry (with Carnegie Mellon's John Pople). One of the most used methods in

quantum chemistry, DFT makes it possible to calculate energy barriers and other electronic properties for reactions such as this one, though the computational demands are daunting. "Doing the density functional calculations," says Jordan "is not something you run on your notebook computer. You need a powerful supercomputer."

Using software called VASP (Vienna Ab-initio Simulation Program) and 64 processors of PSC's BigBen system, Jordan and Sorescu calculated the electron density and energy of the DMDS molecules on the gold surface. Their simulation started with a 188 atoms representing the gold surface, together with a chain of self-assembled DMDS molecules to replicate the conditions of the experiment.

The computations showed that the energy required to break the S-S bond in a DMDS chain on a gold surface is much lower than that required to break the S-S bond of an isolated DMDS molecule. In the chain, the S-S bonds of the neighboring DMDS molecule allowed the reaction to proceed with an energy barrier five times lower than that for an isolated molecule. "A change of five in a barrier," says Jordan, "is a big effect."

The low energy barriers validate the conclusion that the reaction moved down the chain's length rather than reacting with the gold surface — and confirm Yates's zipper analogy. "Here we have this self-assembled structure and the bodies are all nestled close together in a certain configuration," he explains, "and now the chain reaction is encouraged by the fact that everything is in the right place to make it unzip."

As the scientists observe in their paper, this first convincing evidence of propagating surface chain reactions could affect microchip manufacturing.

"AFTER ALL, BIOLOGICAL LIFE IS THE ULTIMATE EXAMPLE OF FUNCTIONALITY THAT ARISES OUT OF SELF-ASSEMBLED MOLECULES."

"It could very well be," says Yates, "that when you get down to very thin photoresists, no thicker than a single layer, then suddenly these organized reactions that you hadn't expected will take place."

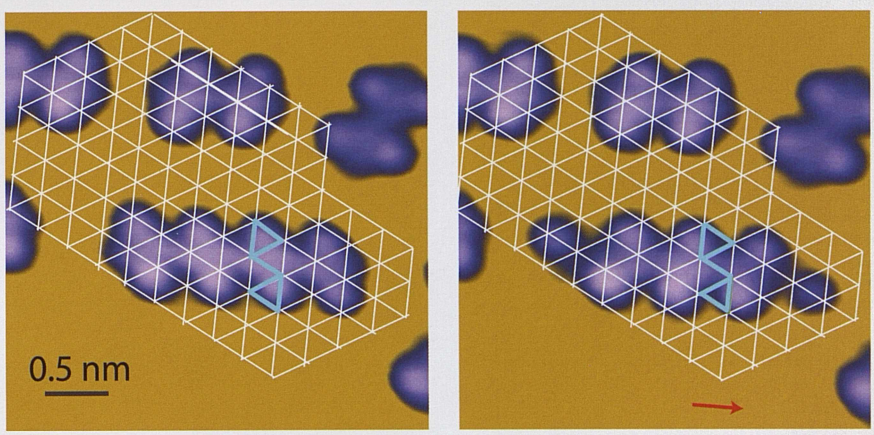
There's much yet to be learned about such reactions, says Maksymovych. "After all," he observes, "biological life is the ultimate example of functionality that arises out of self-assembled molecules. This reaction is also a seminal example of how simple self-organization can give rise to non-trivial, chemical functionality that doesn't exist in the individual building blocks." (TP)

MORE INFORMATION

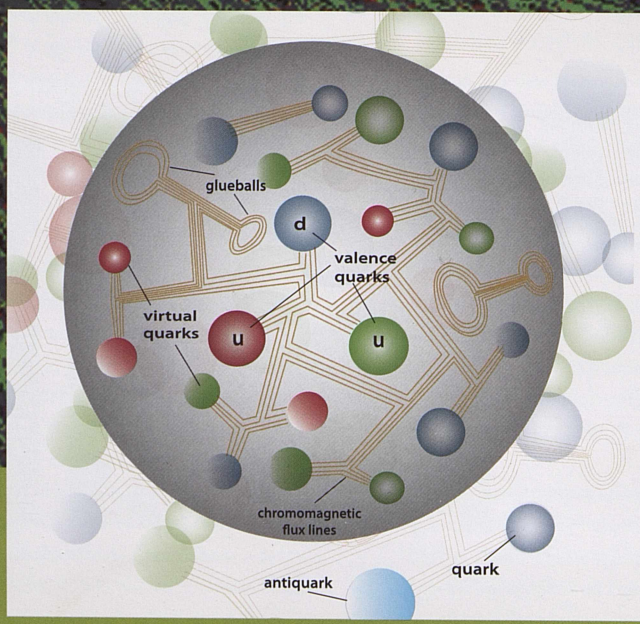
www.psc.edu/science/2009/nanoassembly/

SHIFT TO THE RIGHT

This surface lattice of the CH_3SSCH_3 tetramer (four molecules) before and after the reaction shows that the product molecules (right) have shifted .25 nm (nanometre) toward the right.



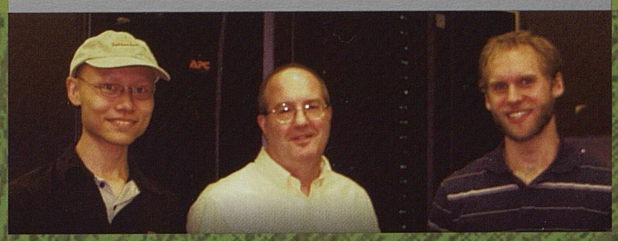
EXCITED STATES OF THE STRONG FORCE



PROTONS

Protons would fly apart except for the strong nuclear force, which is carried by gluons (force lines between quarks). When gluons clump together they form a "glueball." Modern conception of the proton includes more than the three "valence" quarks — a down (d) and two up (u) quarks — which account for only about 2 percent of the proton's mass. The rest comes from a "sea" of virtual quarks and glueballs. Even outside the bounds of the proton, virtual particles spring into and out of existence within the teeming vacuum described by QCD. (Rendering courtesy of Alex Dzierba, Curtis Meyer and Eric Swanson)

↓ COLIN MORNINGSTAR (center)
graduate students CHIK HIM WONG (left) and JOHN BULAVA
Carnegie Mellon University



A TEAM OF PHYSICISTS HAS FOR THE FIRST TIME CALCULATED THE MASS SPECTRA OF "EXCITED STATE" HADRONS

"Nothing exists except atoms and empty space; everything else is opinion," wrote Democritus around 400 B.C. His notion of a single fundamental unit of matter, an atom, lasted more than two thousand years, until during the past century we learned that what we call the atom is itself a world of smaller things — a nucleus of protons and neutrons surrounded by clouds of electrons.

By the 1960s, it became apparent that protons and neutrons are themselves made from smaller things yet. In 1963 Murray Gell-Mann dubbed them quarks, and by the 1970s a powerful theory called quantum chromodynamics (QCD) arose to describe how the strongest force known — the "strong interaction" — acts via things called gluons to bundle three quarks into protons and neutrons.

From collider experiments over many years, scientists learned that protons and neutrons fragment and recombine into a vast array of related particles. As a family, they're called *hadrons* — from Greek, *hadros*, for stout, thick — with the common denominator that they're made from quarks and gluons interacting via the strong force. Unlike a proton — the stable "ground state" hadron — most hadrons exist only at much higher energies for infinitesimally brief flickers of time. To identify these higher-lying states — known as "excited states" — and find them in collider experiments is a major challenge of physics, involving thousands of scientists internationally.

"What are the hadrons? What's the spectrum of QCD?" says Colin Morningstar of Carnegie Mellon University. "What hadrons come from the quark and gluon fields? What particles can we observe in the lab, and what are their properties?" With these fundamental questions as impetus, Morningstar and his colleagues in a team called the Hadron Spectrum Collaboration (HSC) last year achieved a breakthrough. Using resources at several TeraGrid sites — TACC, NICS and PSC's BigBen — and at DOE, they applied a sophisticated, computationally demanding approach called lattice QCD. Doing what no one else had previously been able to do, they successfully calculated the masses of an excited-state spectrum of hadrons.

Lattice QCD computations are essential as a means to reconcile QCD theory with experiments and extract insights into the physics of hadrons, and the HSC team's findings — reported in *Physical Review D* (January 2009) — will guide accelerator experiments, especially at the U.S. DOE Thomas Jefferson National Accelerator Facility in Virginia. "For the experimentalists," says Morningstar, "the effects of these excited states are hidden inside scattering cross-sections, and it's like looking for a needle in a haystack. With our findings, they can do a much better job of figuring out what's there."

QUARKS AND GLUONS ON A LATTICE

The HSC team's success built on decades of computational work in QCD and relied on several innovations of their own. Developed by Nobel-prize winner Kenneth Wilson in the 1970s, lattice QCD makes it possible to solve the complex equations of QCD, using the most advanced supercomputers, by imposing a gridlike structure on the space and time that quarks inhabit.

Within this lattice, quarks exist at the cross points and forces act along the lines between cross points. For their excited-state computation, Morningstar and colleagues created a lattice of 24 points in each of the three spatial dimensions and 64 points in time. A proton has a diameter of approximately one femtometer (one-quadrillionth of a meter, 10^{-15} m), and it's been estimated that a quark is at least 1,000 times smaller than a proton, if it can be measured at all. To model a system this small, the HSC lattice had spacing of 0.11 femtometers in space and 0.033 femtometers in the time dimension.

Using from 2,000 to 4,000 processors on several different systems, including BigBen, they relied on an approach — called Hybrid Monte Carlo (HMC) — implemented through a software package (called Chroma) developed by Morningstar's HSC colleagues and other lattice QCD theorists. The Monte Carlo method, as the name implies, uses random numbers as inputs and from them estimates complicated integrals over the gluon field configurations. "HMC is a sophisticated way of proposing new gluon fields," says Morningstar, "with the appropriate probability distribution, taking into account the effects of the propagating quarks."

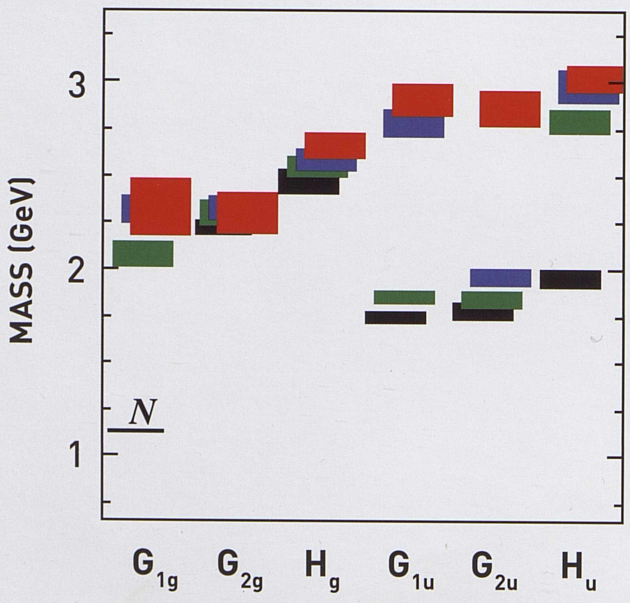
HMC is the first step in the computation, producing an ensemble of possible configurations of the quark-gluon system. For the next step, the researchers use hadron "operators" — mathematical formulations of the quark and gluon fields. They apply these operators to the configurations generated in the first step. Finally, they fit the results to an exponential equation, which yields a hadron's energy, hence its mass — due to the equivalence of mass and energy.

An essential benefit of HMC is that the researchers didn't have to construct a single, correct operator, but could use a set of operators to arrive at the proton's excited states. "We don't know exactly what the proton is," says Morningstar, "and we can't make a proton operator, but we can make an operator that has all the same symmetry properties as the proton. So it will create the proton, but it will also create other particles that have the same symmetry properties. They are the higher lying ones, the excited states."

SMEARED FIELDS, INTACT SYMMETRY

The HSC team's computational advance that made their physics breakthrough possible was their ability to systematically design operators to focus on particular properties. They used a technique called "smearing" to reduce unwanted "noise" — inherent in the statistically based computations. "A lot of work goes into operator construction," says Morningstar. "There's a myriad of excited states, but if you design your operator in the right way, you can kill off couplings to most of the unwanted states."

Smearing involves averaging the field over a larger volume of space, but in such a way that symmetries are left intact. If the field had rotational symmetry before the smearing — meaning it looks the same when you rotate it, say, 90° — it has to retain this symmetry after it is smeared. "It has been known for some time — but we really exploited it here," says Morningstar, "that



THE SPECTRUM OF PROTON AND NEUTRON EXCITED STATES

This mass plot (vertical axis) with different symmetry channels (horizontal axis) represents the spectrum of excited states for the proton or neutron (lower left horizontal bar). Each colored box represents a different nucleon — states of the proton or neutron, with the height of the box indicating the amount of statistical uncertainty associated with it. (Colors make it possible to distinguish different states and have no physical meaning.) Only the proton/neutron and the left-most green box had been identified in QCD calculations before this work. The G_{2g} channel shows spin 5/2 particles, never before calculated in lattice QCD. "It's the pattern that represents the physics," says Morningstar, "and this pattern hadn't been seen before in lattice QCD calculations."

if you make your operators out of smeared quark and gluon fields, you get a signal that plateaus to the state you want much faster."

Besides speed, smearing has the benefit of eliminating signals from most of the myriad interfering, high-energy states. If the quark or gluon field used in the operator is a small, point-like field, it can "couple" to these interfering states. When the field is smeared-out instead of point-like, the interfering state disappears. "The goal of operator construction," says Morningstar, "is to get it to excite the state you want and nothing else; smearing is a crucial part of that."

The HSC group had another trick up its sleeve. They used higher-than-realistic quark masses and smaller-than-normal volumes to simplify calculations that were, nevertheless, enormously demanding. "This was a first demonstration of the computational technology to show you could do this," says Morningstar. "A lot of people in our community didn't think it was possible. We did the simpler calculation first, and now the next step is to go to larger volumes, and to get down to physical quark masses."

Their results correlate well with observed findings, and they expect further computations to improve them. Since its formulation about 30 years ago, lattice QCD has pushed the limits of what's possible with high-performance computing. The HSC group, which uses all the computing it can get,

"A LOT OF PEOPLE IN OUR COMMUNITY DIDN'T THINK IT WAS POSSIBLE."

is already hatching plans for how to tackle lattice QCD with exascale systems. "We want to be able to get out what the physical predictions of this theory are," says Morningstar, "and we just keep applying pressure to it. It's a tough nut to crack." (TP)

MORE INFORMATION

www.psc.edu/science/2009/qcd/

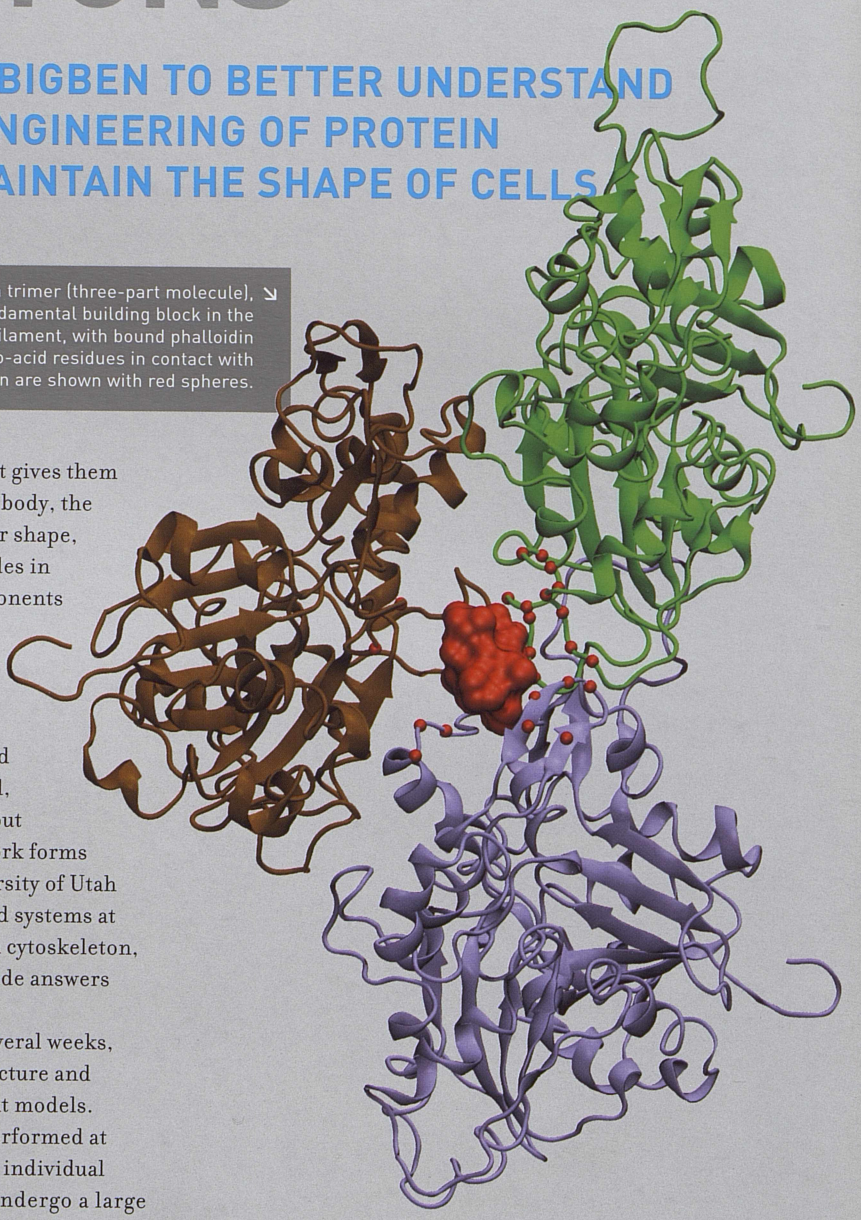
COLLABORATORS:

John M. Bulava
Robert G. Edwards
Eric Engelson
Justin Foley
Bálint Joó,
Adam Lichtl
Huey-Wen Lin
Nilmani Mathur
Colin Morningstar
David G. Richards
Stephen J. Wallace

IN PROGRESS

BENDING IN CELLULAR SKELETONS

RESEARCHERS USE BIGBEN TO BETTER UNDERSTAND THE STRUCTURAL ENGINEERING OF PROTEIN FILAMENTS THAT MAINTAIN THE SHAPE OF CELLS



An actin trimer (three-part molecule), a fundamental building block in the cytoskeletal filament, with bound phalloidin (red). Amino-acid residues in contact with phalloidin are shown with red spheres.

How do cells maintain their shape? What gives them structure? Similar to the skeleton of the body, the "cytoskeleton" of cells maintains cellular shape, protects the cell, and plays important roles in cellular function. One of the main components of the cytoskeleton in eukaryotic cells — cells with membranes, such as human cells — are filaments formed from chains of a protein called actin.

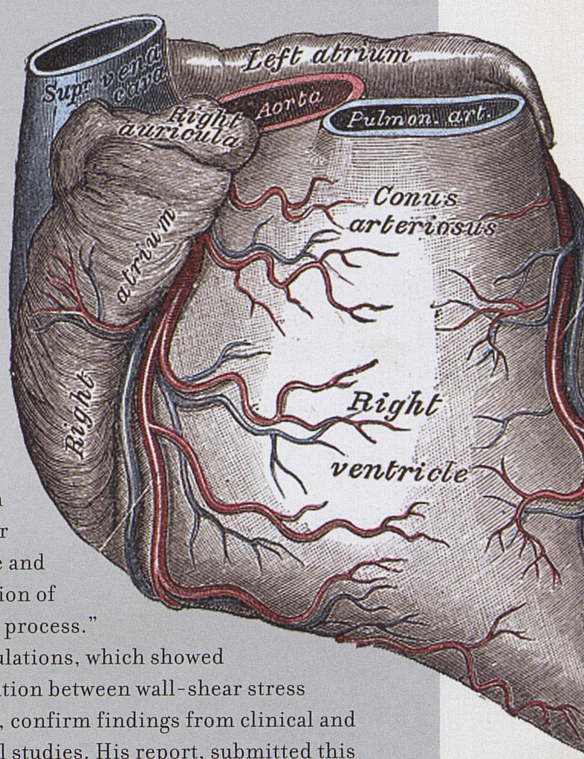
Actin chains form a highly branched network, and though much is understood, many fundamental questions remain about actin filaments and about how this network forms and functions. Researchers led by University of Utah chemist Greg Voth used PSC's BigBen and systems at other TeraGrid sites to simulate the actin cytoskeleton, and their findings are beginning to provide answers to some of these questions.

Using 512 BigBen processors for several weeks, Voth and colleagues investigated the structure and dynamics of two competing actin filament models. "Our simulations of the actin filament performed at PSC have shown us, for example, that the individual actin monomers in the actin filament undergo a large conformational change on relatively short timescale," said Voth collaborator James Pfaendtner, a biomolecular engineer at the University of Washington.

For this work, PSC scientist Phil Blood optimized performance of the latest version of the molecular dynamics program NAMD. Results have also helped the team to shed new light on how the formation of actin chains is involved in ATP hydrolysis — a key issue in cytoskeletal dynamics. The researchers also carried out the first simulations of the actin filament bound with phalloidin — a toxic compound widely used to assist researchers in creating images of the cell, and their results have begun to clarify the atom-by-atom mechanism by which phalloidin acts on the actin filament.

PATHWAYS TO HEALTHY ARTERIES

WITH SUPPORT OF THE TERAGRID PATHWAYS PROGRAM, SHEWAFERAW SHIBESHI EXPLORES HOW TO IMPROVE DIAGNOSIS OF DEADLY ARTERIAL DISEASE

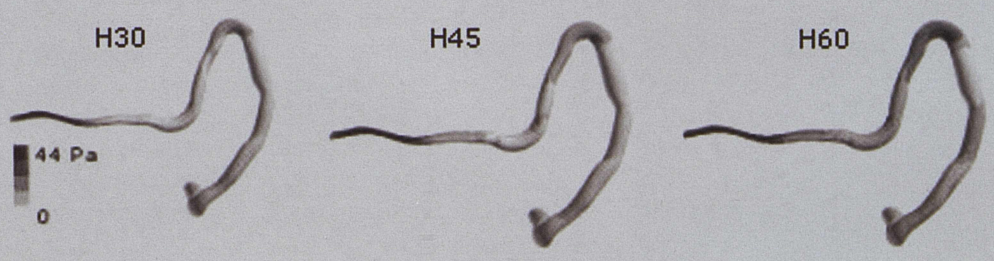


Atherosclerosis, better known as hardening of the arteries, is the leading cause of death in the developed world. In the United States alone, more than half of yearly mortality can be traced to arterial disease, with more than 500,000 deaths each year due to heart attacks from arterial blockage.

In response to this major health problem, physicist Shewaferaw Shibeshi has mounted a series of computational fluid dynamics (CFD) studies of blood flow as it relates to arterial disease. He began this work in 2002, as a Ph.D. student at Howard University, and continued it as a professor at Alice Lloyd College in Kentucky. Last year, through the TeraGrid Pathways program, which supports the use of TeraGrid resources by faculty and students at minority-serving institutions, Shibeshi worked with PSC consultants to use Pople, PSC's shared-memory system, and CFD software called Fluent. With these resources, he simulated blood flow in the right coronary artery to study correlations between the force exerted by flowing blood and "hematocrit" — the proportion of blood volume occupied by red blood cells.

"Blood-flow simulations," says Shibeshi, "commonly consider the blood's viscosity and density, but other blood components, such as hematocrit, can potentially offer a more accurate and useful explanation of arterial disease process." Shibeshi's simulations, which showed a strong correlation between wall-shear stress and hematocrit, confirm findings from clinical and epidemiological studies. His report, submitted this year to *Rheology*, points out that elevated hematocrit promotes plaque buildup, and that hematocrit level, which can be assessed from blood tests, may be useful for improved diagnosis and treatment of this deadly condition.

Increasing wall-shear stress distribution on the wall of the right coronary artery with increase in hematocrit level (30, 45 & 60). Speaking of PSC consultants Rick Costa and Anirban Jana, who worked with him on this project, Shibeshi says, "Whenever I had a question, they answered it in a day or less." ↓

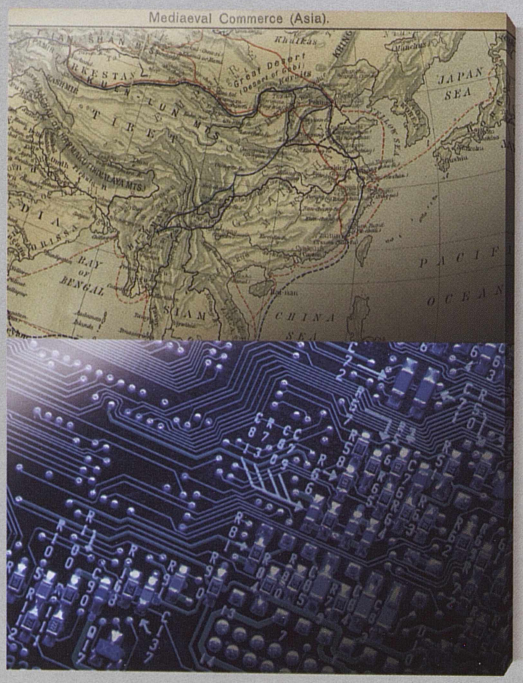


DIGITAL HUMANITIES

THROUGH MINI-RESIDENCIES AT PSC, SCHOLARS ARE OPENING NEW VISTAS IN HUMANITIES STUDY

The powerful tools of information technology have begun to open new horizons in many fields that haven't traditionally been involved with computation. In the humanities, the intersection of digital methods with what has historically been a text-based field has spurred the National Endowment for the Humanities (NEH) to support training for humanities scholars exploring digital possibilities.

Through collaboration with NCSA in Illinois and SDSC in San Diego, PSC has provided mini-residencies this year in which scholars from three NEH digital humanities projects experiment with technology and chart long-range goals. Laura McGinnis, PSC manager of education, outreach and training, coordinates this effort. PSC scientists Phil Blood, Thomas Maiden, Nathan Stone and John Urbanic have helped to define objectives and support needs, which include suitable storage media and software development for a "portal" interface and powerful search methods to access information.



Global Middle Ages:

A collaboration of the University of Minnesota College of Liberal Arts, School of Music, and Department of History and the Johns Hopkins University Department of the History of Art, the Scholarly Community for Globalization of the Middle Ages (SCGMA) is working to develop a new interdisciplinary community of scholars for study of the Middle Ages. SCGMA aims to create an online infrastructure that will support multiple formats and languages in textual, visual and aural resources. As an example, from a database that would include trade records between two civilizations, they would like to develop a map display that shows how trade routes between these civilizations changed over time.

Humanistic Algorithms:

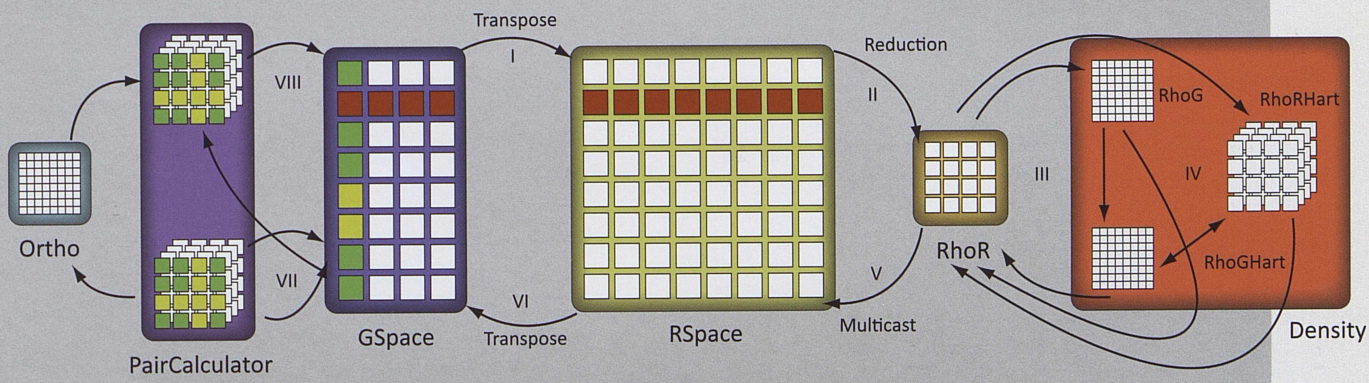
In this project, the University of Southern California's Institute for Multimedia Literary is collaborating with researchers at the University of Illinois, Urbana-Champaign to create a digital archive of multi-media portfolios of faculty and students. The group will use data analytics to extract information from unstructured texts (such as raw textual data like websites) to produce information that can then be used for study to create "meta-analytical" scholarly multimedia.

HistorySpace:

A collaboration of international researchers, led by John Bonnett of Brock University and including participants from the United States, Canada, and the United Kingdom, this project aims to incorporate "virtual worlds" as tools to support research and expression in the humanities. Focusing on history and related disciplines, the HistorySpace effort brings historians and computer scientists together to design workflows and tools in *mWorlds*, an open source virtual world platform.

OPTIMAL PERFORMANCE

BY PAYING ATTENTION TO THE DISTANCE TRAVELED BY MESSAGES FOR INTER-PROCESSOR COMMUNICATIONS, RESEARCHERS SHOW MAJOR GAINS IN COMPUTING EFFICIENCY



Decomposition of the physical system into *chare* arrays (only important ones shown for simplicity) in OpenAtom.

Peace, happiness and supercomputing — to perhaps strain for an analogy, there are some things of which you can never have too much. With supercomputing, despite amazing improvements in capability over more than two decades, demand for the resources always exceeds availability. Because of this, computational scientists are constantly pressed for ways — newer, better algorithms — to gain the best possible performance from their research applications. A speedup of 10 percent in software that may require weeks or months of computing to arrive at useful results is enormous.

During the past two years, PSC staff have worked closely with University of Illinois, Urbana-Champaign (UIUC) graduate student Abhinav Bhatele to fine tune and optimize performance on massively parallel systems for OpenAtom, a quantum molecular dynamics application used to study properties of materials and nano-scale molecular structures. OpenAtom implements a highly useful mathematical formulation of quantum mechanics called the Car-Parrinello method, for Italian physicists Roberto Car and Michele Parrinello who introduced it in 1985.

Bhatele developed and tested a way to assign parts of the overall computing job (called *chares* in the programming model he used, Charm++) to specific BigBen processors. His approach exploits BigBen's network "torus" topology, also used on other systems, to reduce the number of "hops" that a message between processors has to travel. With OpenAtom, this method yielded speedups of up to 20 percent on 1,024 BigBen processors and up to 50 percent on 2,048 processors. His paper reporting these results won the Distinguished Paper Award at the Euro-Par 2009 conference and also won the outstanding master's thesis award at UIUC.

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

Educause

Google

Hewlett-Packard Company

Silicon Graphics, Inc.

The Buhl Foundation

Bill and Melinda Gates Foundation

The Grable Foundation

The Heinz Endowments

EDITOR/WRITER: Michael Schneider, PSC

CONTRIBUTING WRITER: Tim Palucka

GRAPHICS RESEARCH, PHOTOGRAPHY DIRECTION & PROJECT COORDINATION: Shandra Williams, PSC

TRANSCRIBING: Ryan Omecene, PSC

PHOTOGRAPHY: Tim Kaulen, Photography & Graphic Services at Mellon Institute; Jim Burke, University of Pittsburgh Photography Services; Jennifer Wilson, photo of Clay Reid and colleagues.

GRAPHICS: Thanks to the researchers and to Shawn Brown, Greg Foss, Joel Stiles, Art Wetzel and Troy Wymore of PSC for graphics.

COVER GRAPHIC: Rendered by Greg Foss, PSC visualization specialist, from simulation of an earthquake scenario, magnitude 2.6, in the region of Thessaloniki, Greece, by Jacobo Bielak of Carnegie Mellon University and collaborators R. Taborda and H. Karaoglu using PSC's BigBen to run Hercules, the earthquake simulator developed by the Quake Group at CMU.

see *www.psc.edu/science/2008/quake.html*

DESIGN: Wall-to-Wall Studios, Inc.

PRINTING: Broudy Printing

PRINTED WITH SOY INK™

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