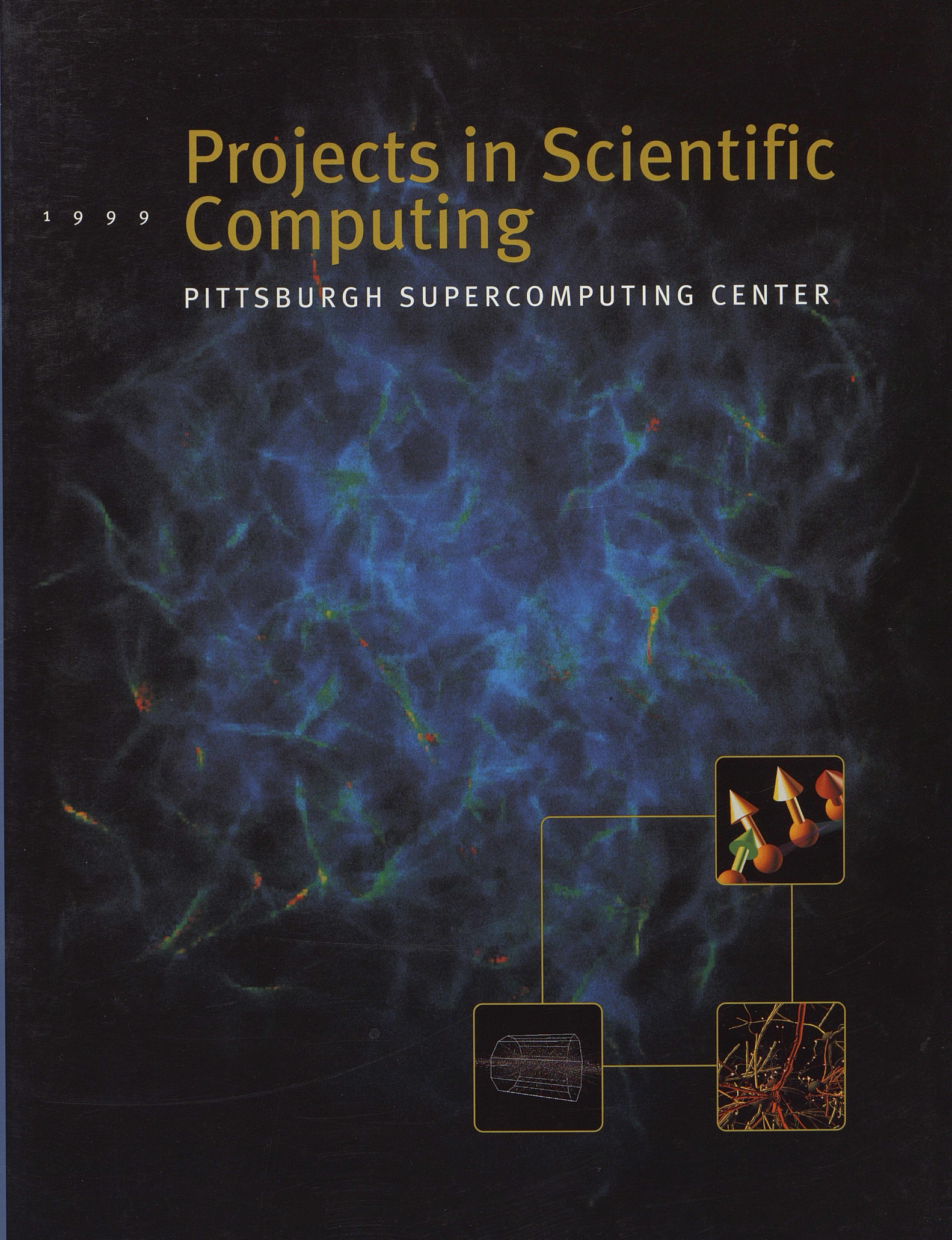
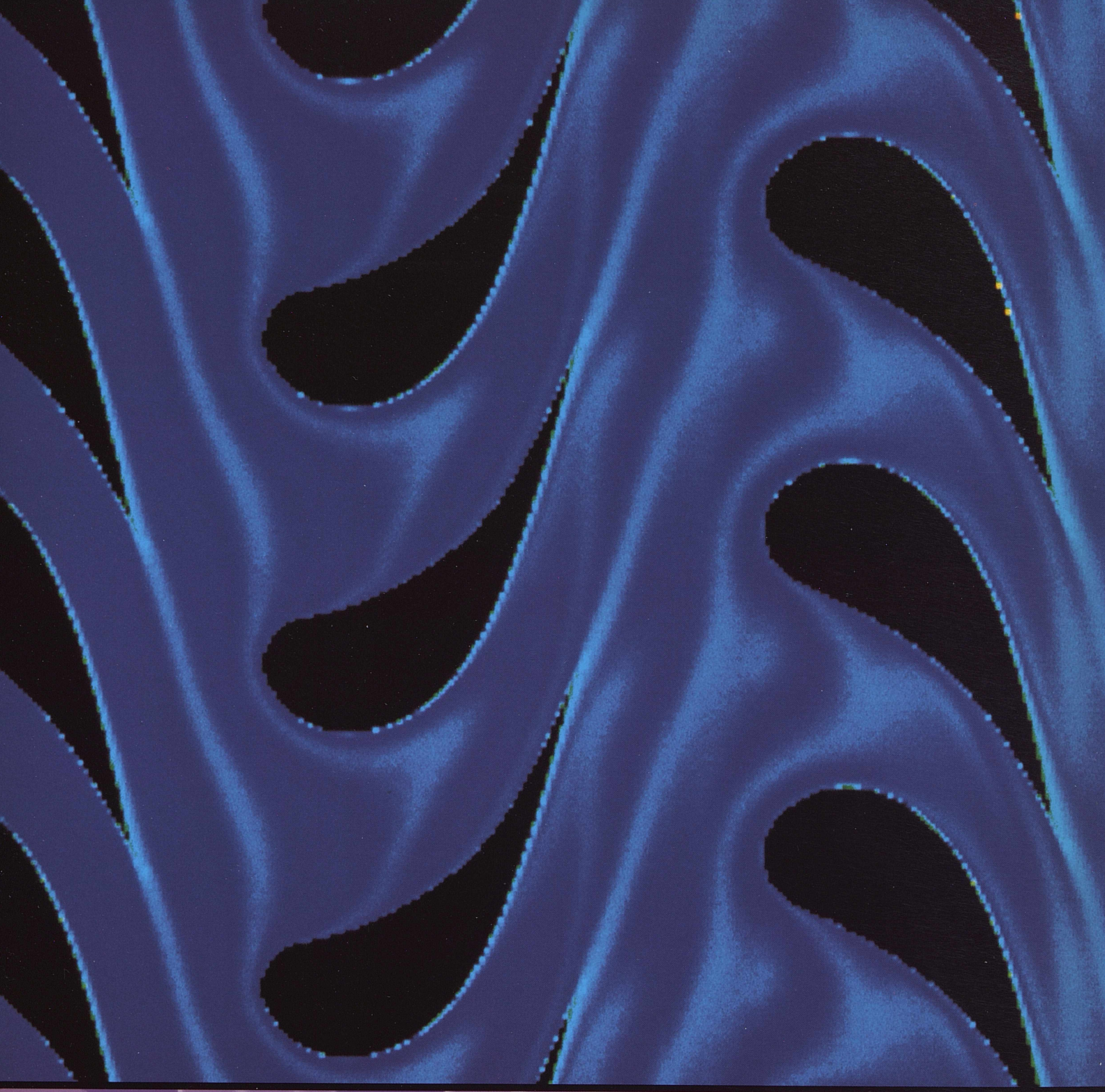
1999

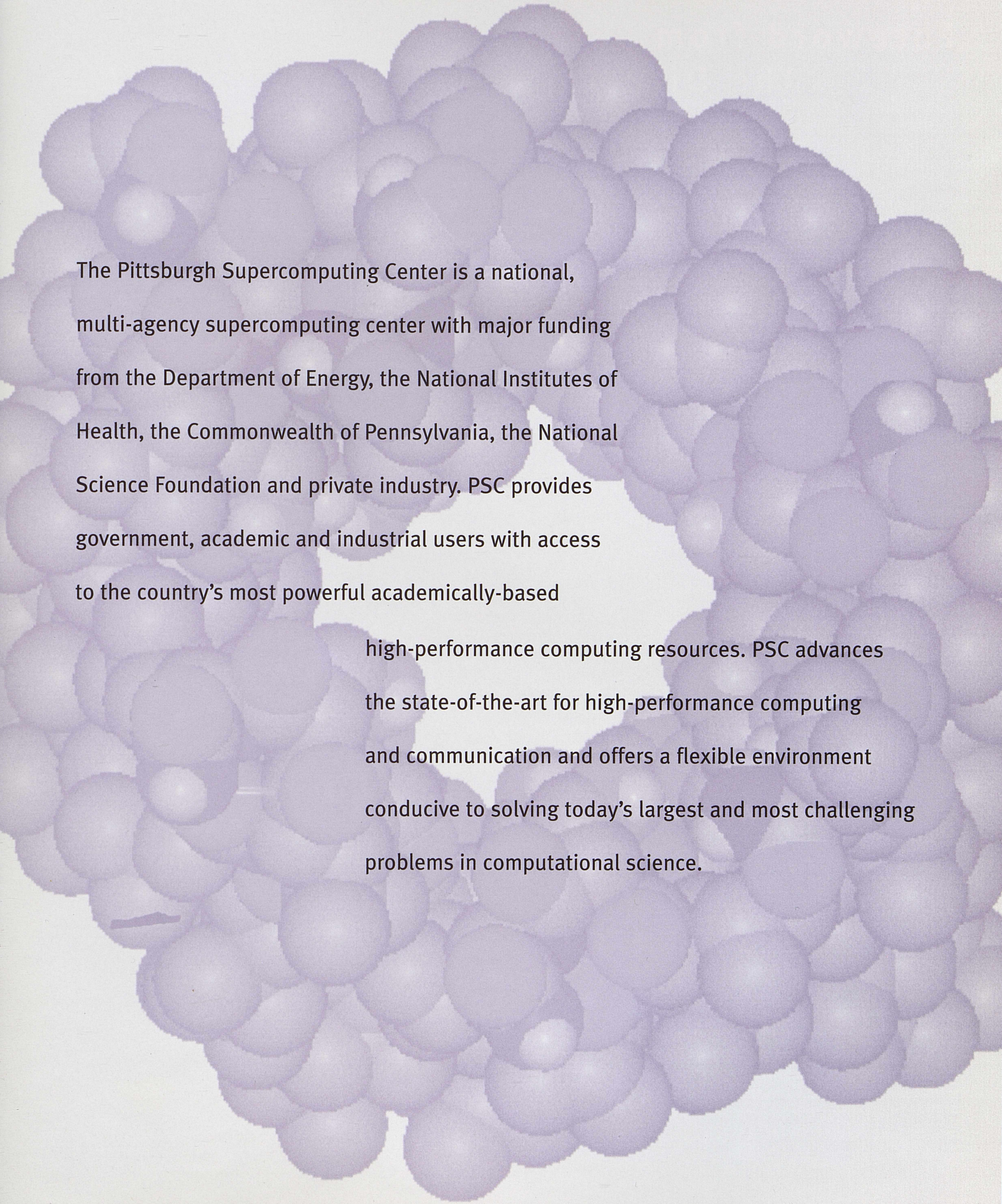
Projects in Scientific Computing

PITTSBURGH SUPERCOMPUTING CENTER





Contents



The Pittsburgh Supercomputing Center is a national, multi-agency supercomputing center with major funding from the Department of Energy, the National Institutes of Health, the Commonwealth of Pennsylvania, the National Science Foundation and private industry. PSC provides government, academic and industrial users with access to the country's most powerful academically-based high-performance computing resources. PSC advances the state-of-the-art for high-performance computing and communication and offers a flexible environment conducive to solving today's largest and most challenging problems in computational science.

Foreword from the Directors

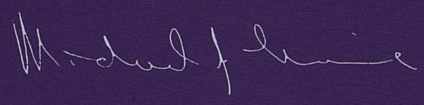
The work of the Pittsburgh Supercomputing Center in this past year has again demonstrated our commitment and ability to support the advancement of science and engineering through the application of *capability* computing, large-scale data processing and high-performance networking, to challenging problems facing the academic community, federal agencies and leading corporations.

The examples presented in this year's *Projects in Scientific Computing* span the scale from both theoretical and experimental work on quarks (pp. 15 & 20-21) through biologically significant, molecular level simulations (pp. 24-27) to the astrophysical-scale work on the solar dynamo (pp. 14-15) along with examples from materials sciences (pp. 18-19 & 22-23) and computational fluid dynamics (pp. 16-17).

Last year's landmark protein folding work of Kollman and Duan received additional recognition this year as a 1999 Computerworld Smithsonian Award finalist (p. 5). Noteworthy visualization (Gray Matters, p. 7) and image analysis of tissue samples (p. 6) work extend PSC's impact in biomedical research and education.

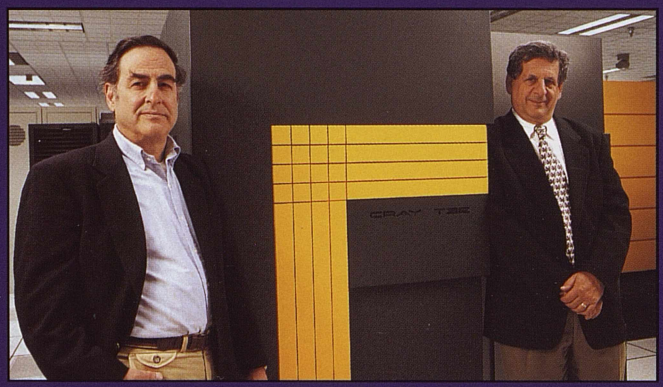
PSC's CRAY T3E-900, Jaromir, serves the computational science community as the top-ranking, academically-based U.S. machine. With powerful internal communications facilitating its use as a whole on single problems, Jaromir has been the necessary support for much of the work described herein. Also targeted to be used, as a whole, on one or a few problems at a time, PSC's Intel-based Dual Boot Cluster provides additional resources for work less demanding of internal communications. The Compaq AlphaServer 8400 supports much of PSC's intensive efforts in bioinformatics. Coupled through excellent local and wide-area high-performance networks to outstanding, large-scale data storage and retrieval facilities as well as to a nationally distributed user community, PSC systems and expertise work to advance the nation's scientific research and development activities.

This work has received support from multiple sources. The Department of Energy has sponsored much of the physics and materials science work in support of both scientific understanding and of increasing our overall capability in numerical simulation. The National Institutes of Health has supported the application of high-performance computing, communications and informatics technology to biomedical research and education. The National Science Foundation sponsors a great deal of research, support and development work for national networking. The Commonwealth of Pennsylvania sponsors activity in research, education and technology development in support of educational institutions and advanced technology activities within the State. Corporations benefit from PSC's facilities and staff expertise in using those facilities to improve product design.



Michael J. Levine, *scientific director*

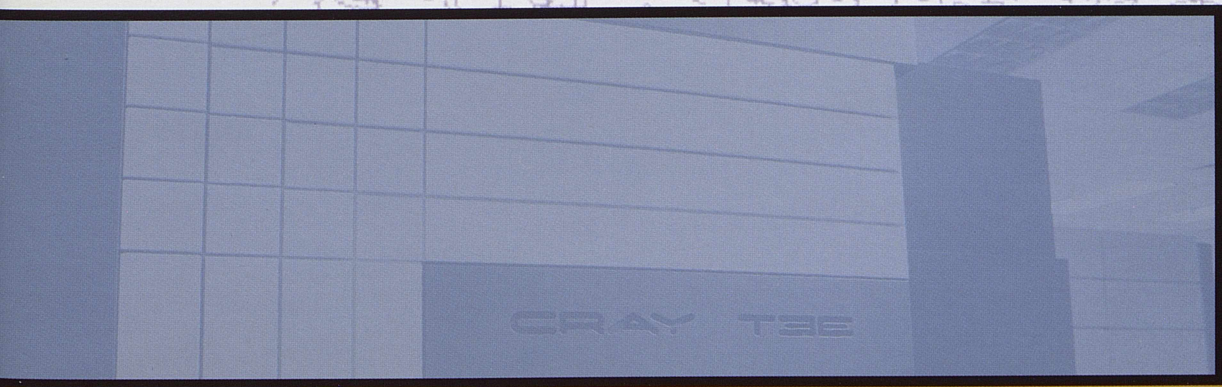
Ralph Z. Roskies, *scientific director*



► ***Michael Levine (left) and Ralph Roskies, scientific directors, Pittsburgh Supercomputing Center***

[1 9 9 9]

Pittsburgh Supercomputing Center



Pittsburgh Supercomputing Center, 1999

At the edge of the new millennium, the Pittsburgh Supercomputing Center is one of the leading centers of high-performance computing and communications in the world. Through a productive collaboration with the U.S. Department of Energy, PSC has during the past year contributed significantly to a high-priority national research program. Much of that work is detailed (pp. 14-19) in this publication.

PSC's partnership with the National Institutes of Health has trained thousands of biomedical scientists and produced important research. PSC's networking group exerts a leadership role nationally in the development of advanced network technology. In 1999 PSC takes pride that its resources, human and technological, make it possible for U.S. researchers to attack the important, difficult problems of engineering and science.

Networking the Future

In March 1999, the National Center for Network Engineering, PSC's networking resource group, with a decade of experience in network technology, took on a new leadership role. Along with its support for the very high-performance Backbone Network Service (vBNS), NCNE now also provides engineering services for Abilene, an advanced network for educational and research applications. Both these networks can move data more than 10 times faster than the current Internet.

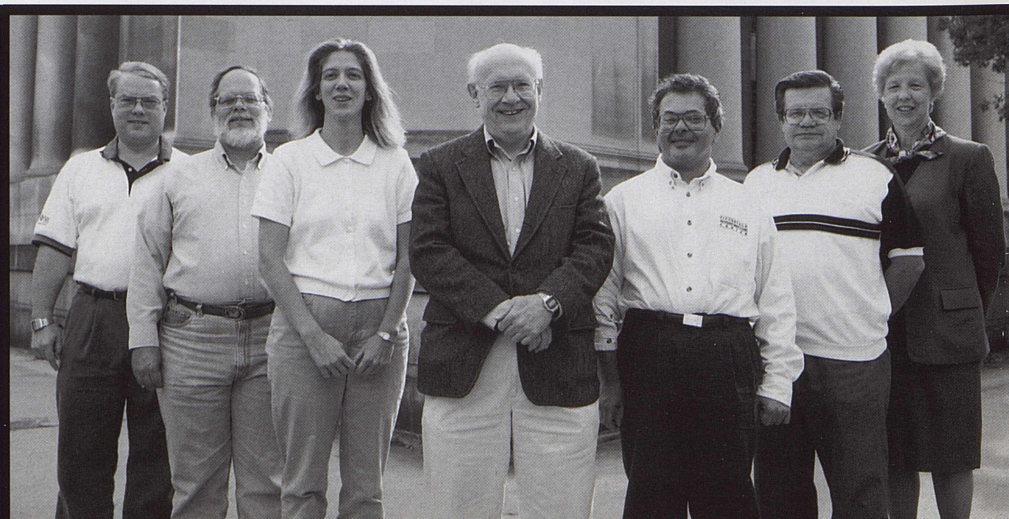
With Abilene, NCNE extends its role as the engineering support component of the National Laboratory for Advanced Network Research. Through NLANR, PSC staff serve as a technical resource for over 100 institutions connected to vBNS. "It's an opportunity and a challenge," says NCNE director Gwendolyn Huntoon, "to participate in forging the network of the future."

NCNE also operates the Pittsburgh gigaPoP, a high-performance network hub providing connections to Abilene and vBNS for western Pennsylvania and surrounding regions. Universities connected through the Pittsburgh gigaPoP include Carnegie Mellon University, the University of Pittsburgh, Pennsylvania State University and West Virginia University.

In November 1998, over 150 network engineers from around the country gathered at PSC to learn about emerging network technologies, one of five such meetings organized by PSC's networking staff since January 1998 at sites around the country. For a June 1999 workshop, PSC staff implemented a high-quality video stream across vBNS, Abilene and the Internet to broadcast a presentation live from Sweden.

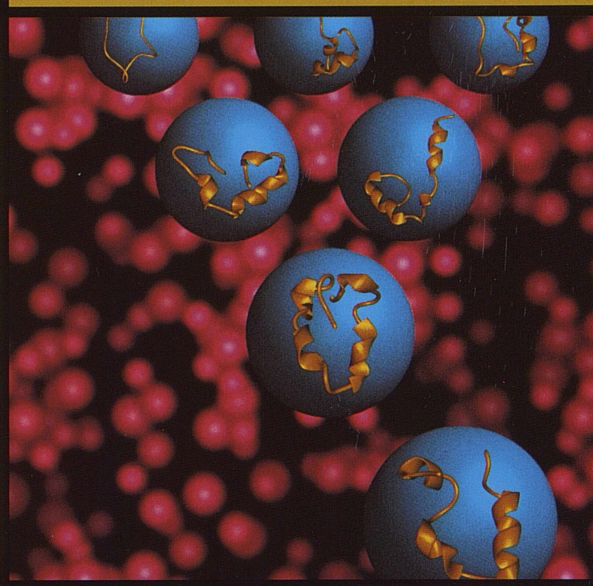
More information on NCNE: http://www.ncne.nlanr.net

[the most powerful academically-based supercomputer in the United States]



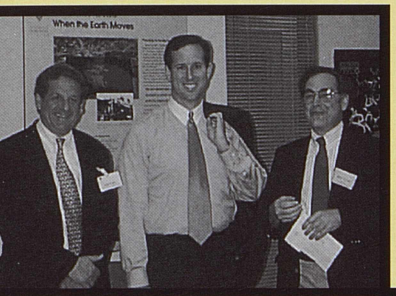
The PSC operational management team (l to r): J. Ray Scott, assistant director, systems and operations; David Deerfield, assistant director, biomedical initiative; Gwendolyn Huntoon, assistant director, networking; Bob Stock, associate director; Sergiu Sanielevici, assistant director, scientific applications and user support; Rich Raymond, manager, user support; Elvira Prologo, manager, administrative staff. Absent: Janet Brown, manager, networking.

◀ *This Pittsburgh black-and-gold CRAY T3E, the most powerful academically-based supercomputer in the United States, is known to users as Jaromir, in honor of the Pittsburgh Penguins high-scoring forward. With 512 separate processors that can work simultaneously on the same problem, Jaromir at its peak can do 460 billion computations per second, 80 times more processing power than every person on Earth holding a calculator doing one calculation per second.*

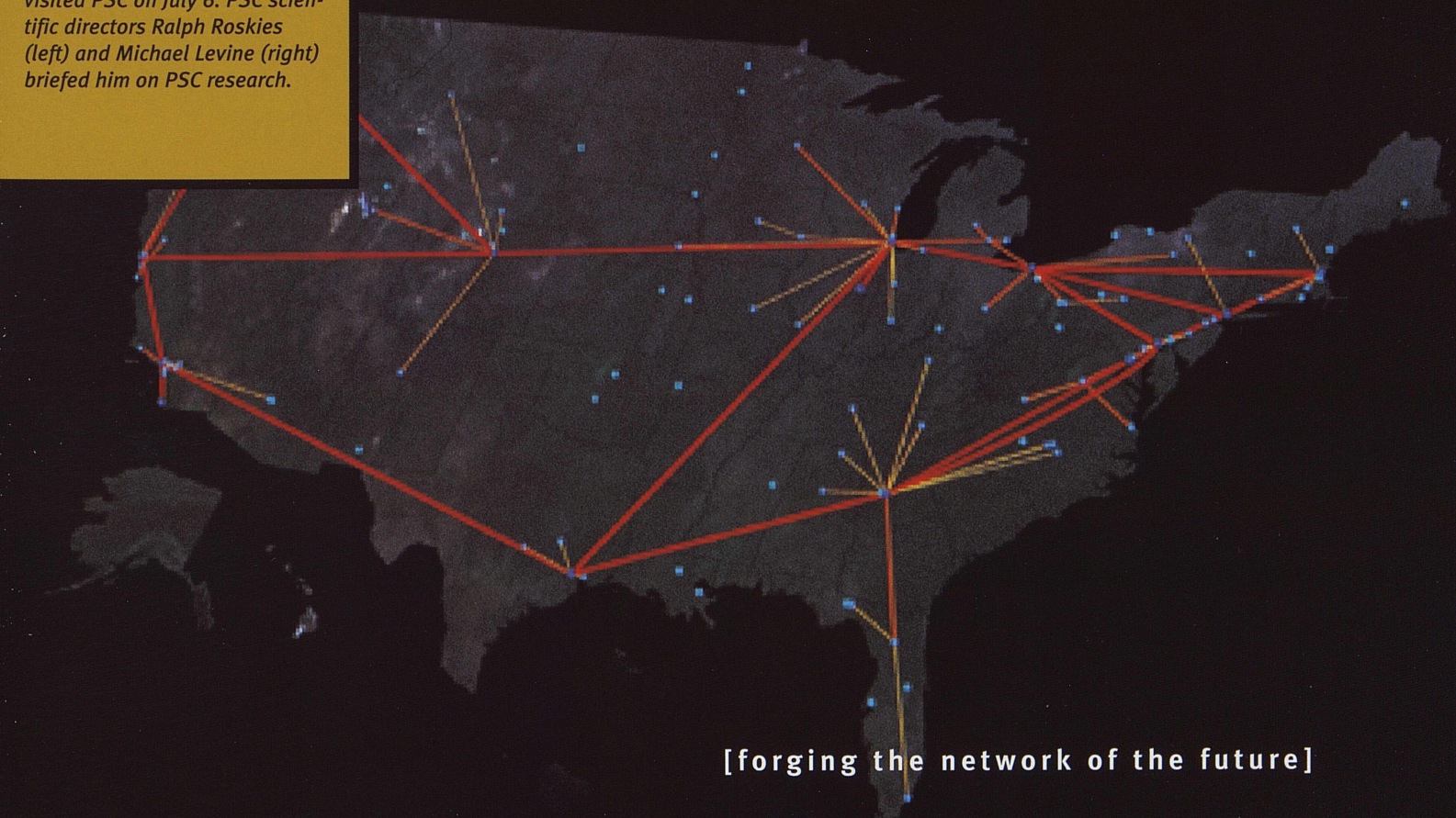


▶ *For simulations of the "protein-folding problem," PSC and the University of California, San Francisco were finalists for the 1999 Computerworld Smithsonian Award in Science. Using software they developed at PSC, Peter Kollman and Yong Duan of UCSF, simulated the folding of a small protein in water for a full microsecond, 100 times longer than prior similar simulations.*

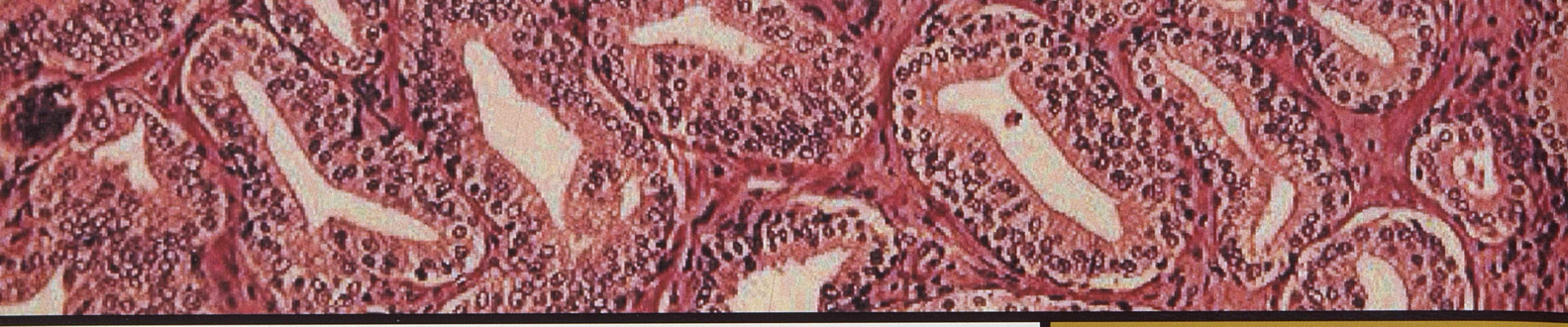
More information:
http://www.psc.edu/science/kollman98.html



▲ *U.S. Senator Rick Santorum visited PSC on July 6. PSC scientific directors Ralph Roskies (left) and Michael Levine (right) briefed him on PSC research.*



[forging the network of the future]



Biomedical Supercomputing

Established in 1987 with support from NIH's National Center for Research Resources, PSC's biomedical initiative was one of the first high-performance computing programs nationally to focus on biomedical research. Its extensive training activities reach hundreds of researchers each year. PSC scientists have developed techniques for harnessing high-performance computing in DNA and protein sequence analysis, and—with support from the National Human Genome Research Institute—PSC workshops on these techniques help researchers nationwide cope with the explosion of genome-sequencing data.

In collaboration with other scientists as well as in PSC core research, PSC biomedical scientists have produced important results in a range of problems, as represented in this booklet (pp. 24-27). In 1998, a collaboration between PSC and scientists at the University of California, San Francisco led to the longest-ever simulation of how a newborn protein folds into its mature shape. Reported in two leading U.S. scientific journals, the results offer new insight into the "protein-folding problem," one of the premier challenges of molecular biology. This research was a finalist for the 1999 Computerworld Smithsonian Award in Science.

In 1998, the National Center for Research Resources awarded a four-year, multi-million dollar grant for PSC to assess technologies that can improve the ability of biomedical researchers at separate locations to collaborate. Through this program, PSC scientists are testing a range of collaborative tools—including desktop conferencing and shared software revision control—for use in two ongoing PSC collaborations. As part of this project, the Center for Biomedical Informatics at the University of Pittsburgh Medical Center is assessing the technologies and social factors that affect collaborative success.

▲ This microscope image of a cancerous prostate cell represents research in which PSC scientists are collaborating with pathologists at the University of Pittsburgh Medical Center and researchers at the National Cancer Institute to develop automated techniques that improve the speed and accuracy of tissue diagnosis. With these methods, statistical analysis of relative cell locations can provide an objective and more reliable evaluation of malignancy. PSC is also implementing collaborative tools to enhance the ability of these researchers to work together across geographical distance.

For more information on:
Biomedical supercomputing at PSC—
http://www.psc.edu/biomed/biomed.html

Biomedical collaboratories—
http://collaboratory.psc.edu/

[advanced tools to cope with the explosion of genome data]

▲ A workshop in progress at the PSC Computer Training Center.

Pittsburgh Supercomputing Center Workshops (1998-1999)

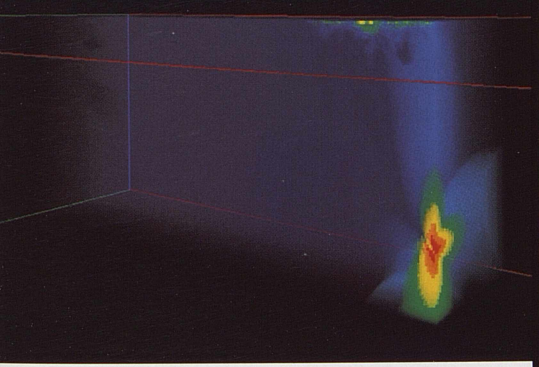
Supercomputing Techniques: Parallel Processing
Parallel Programming Techniques
Recommended Programming Practices
Nucleic Acid and Protein Sequence Analysis
Biomedical Image Analysis and Visualization
Single Particle Reconstruction from Electron Microscope Images
Molecular Mechanics and Dynamics of Biomolecules
Developing and Using PC Clusters as a Biomedical Research Tool

Gray Matters: A Museum Show

Take the journey inward with light impulses from the retina as they travel the optic nerve to the visual cortex. With the brain's neuron structure projected onto the three-dimensional, virtual-reality space of a planetarium dome, this is the experience of "Gray Matters," an interactive, multi-media museum show to entertain and educate children and adults about the brain.

Developed through a collaboration among PSC and Carnegie Mellon's Studio for Creative Inquiry and Center for the Neural Basis of Cognition, "Gray Matters" includes an interactive segment during which the audience in effect becomes a brain. As each person controls a separate neuron, the audience learns to orchestrate neural activity in parallel as the brain does.

▶ In the images shown here—from an animation developed by PSC scientist Greg Hood along with John Burkardt and visualization specialist Greg Foss—spark-like flashes represent impulses transmitting along the complex, branched structure of neurons in the visual cortex. After its first exhibition—at the Carnegie Science Center in Pittsburgh—"Gray Matters" will travel to other museums in the United States. A condensed version will be shown at Expo 2000 in Hannover, Germany.



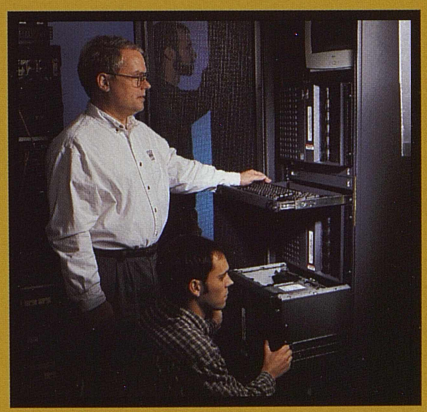
When the Earth Moves

▲ In collaboration with an earthquake-modeling group led by Jacobo Bielak of Carnegie Mellon, PSC visualization specialist Greg Foss has produced animations from CRAY T3E computations. This frame, from simulation of the Jan. 17, 1994 Northridge, Calif. quake, shows ground motion 19.5 seconds after the quake began, when intense motion at the epicenter has reached the surface. Color corresponds to the intensity of ground motion. In September 1999, the National Science Foundation awarded a new, three-year $2 million grant to Bielak's team to continue this work.

Unveiling the PSC Dual Boot Cluster

▶ In March 1999, the PSC Cluster became operational as an alternative resource for users. This cluster of 10 four-processor Intel SMP (symmetric multi-processor) machines provides an aggregate system of 40 Intel PII 400 MHz processors. Uniquely configured as a "dual boot" system, the PSC Cluster runs either the Linux or Microsoft Windows NT operating system and switches easily between them. Users submit jobs and the batch scheduling system reboots the operating system as necessary. "The goal of the cluster," says J. Ray Scott, PSC assistant director of systems and operations, "is to serve as a production high-performance computing resource, not merely backup or load-balancing for other PSC systems."

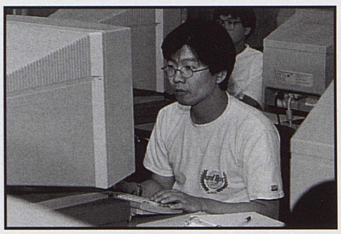
The PSC Cluster stems from a spring 1997 Intel Technology for Education 2000 grant to Carnegie Mellon, and its implementation draws on more than 10 years of PSC experience with cluster technology. After more than a year of testing prototype hardware and evaluating hardware and software issues, PSC staff in January 1999 settled on system design and acquired the 10 Intel SMP machines. Four separate workstations—cluster support servers—link to the cluster through high-speed switches. To date, researchers have used the PSC Cluster for a range of computations, including quantum studies of magnetism, computational fluid dynamics, and molecular dynamics.



Supercomputing in Pennsylvania

▲ *Beverly Clayton, PSC executive director, coordinates PSC's program to provide advanced training and high-performance computing resources for researchers in Pennsylvania.*

With support from the Commonwealth of Pennsylvania, PSC provides education, consulting, advanced network access and computational resources to scientists and engineers in Pennsylvania. This program serves not only academic researchers but also corporations who can exploit the capability of high-performance computing to enhance their competitiveness.



Education

PSC workshops provide industry researchers as well as university faculty and students with training in the most advanced techniques of computational science. They include extensive hands-on sessions, either in PSC's Computer Training Center or at corporate and academic sites around the state.

During 1998-99, PSC workshops in biomedical computing at Penn State's main campus and at Hershey Medical College trained more than 100 Pennsylvania students and scientists in computational methods for biomedical research. A series of four workshops on parallel computing—two at PSC and one each at Shippensburg University and Lehigh University—drew more than 70 participants from 12 Pennsylvania colleges and universities.

To request a PSC workshop, e-mail workshop@psc.edu

Pennsylvania Research: Understanding Ion Channels

Using PSC's CRAY T3E, researchers led by distinguished Pennsylvania scientist Michael Klein, director of the Laboratory for Research on the Structure of Matter at the University of Pennsylvania, are addressing a range of problems in solids, fluids and biological systems. In two of these projects, scientists are simulating "ion channels," one of the most important constituents of living cells.

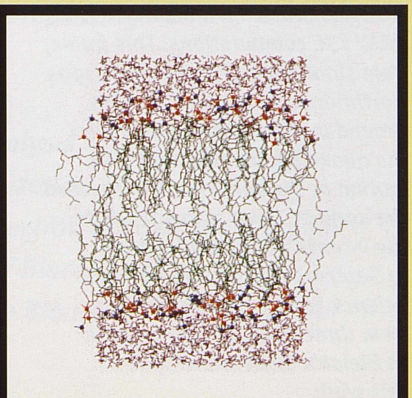
As nature's ubiquitous switch for controlling electrical current, ion channels play a key role in conducting nerve stimuli, regulating heartbeat and toxic response, and are a site for drug therapy. They perform their biological role by regulating the flow of ions—such as calcium and phosphorous—back and forth across cell walls.

To gain detailed understanding of the structure and current-carrying properties of ion channels, researchers in Klein's group at Penn use the T3E for detailed simulations. Ion channels are very large, complex molecular assemblies. Even

◀ *The LS3 ion channel simulated by Qingfeng Zhong of Klein's group. The pore structure of the channel is shown, with atoms—oxygen (red), carbon (gray), nitrogen (blue), represented by space-filling spheres. Zhong's molecular dynamics simulation of the LS3 channel is one of the longest in time ever carried out, extending more than seven nanoseconds, with results that correlate much better with the experimental current-conducting properties of the channel.*

▶ *A frame from simulation of a "lipid bilayer," the native cellular environment for an ion channel called M2. The top and bottom layers of the bilayer are water, with an inner structure of lipids. Atoms of the "headgroups," where the water meets the lipid, are nitrogen (blue), phosphorous (orange), oxygen (red), carbon (green).*

Preston Moore of Klein's research team is carrying out simulations of the M2 ion channel inside a lipid bilayer, work which can provide data to complement laboratory studies. "Experimentalists find it extremely difficult to look atom-by-atom like we can," says Moore, "to see which amino-acid residues are in contact with the lipid headgroups, which are solvated in the waters, and similar details."



◄ *Alcoa, the world's largest aluminum company, used PSC resources to design parts for the Plymouth Prowler. An engineering tour-de-force, the Prowler is the first U.S. car engineered from the ground up to exploit aluminum technology. At 2,800 pounds, it weighs 50% less than it would with traditional steel design, making it quick, fuel efficient and impervious to rust.*

Alcoa has pioneered sophisticated computational modeling that can predict how aluminum will perform in products. In recent work, Alcoa engineers are using the CRAY T3E for detailed understanding of how atmospheric elements, contaminants and lubricants involved in processing can change the properties of the aluminum surface, knowledge that directly affects the cost, quality and performance of products. With the T3E, complex simulations that used to take days or weeks are done in hours.

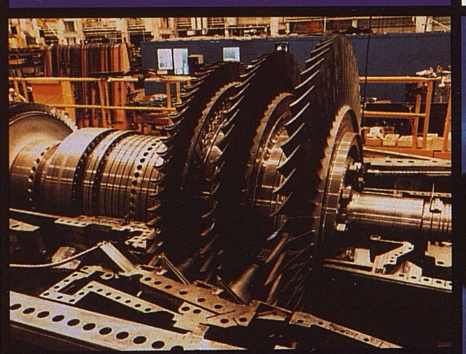
More information: http://www.psc.edu/science/chu.html

simplified versions used in experiments involve tens of thousands of atoms. These simulations can only be carried out with high-performance systems like the T3E.

"Our group has used supercomputers for more than 15 years," says Klein, "and we've pioneered novel methods in both quantum and classical simulations. Our work is well suited to PSC's CRAY T3E, and access to PSC facilities has enabled us to make significant progress."

Economic Development
PSC provides high-performance computing and networking to boost competitiveness of Pennsylvania business and industry. Along with major corporations such as Alcoa and Westinghouse, PSC provides resources to smaller Pennsylvania businesses including Concurrent Technologies Corp., Form Soft, Knowledge Systems, Mine Safety Appliances and Tissue Informatics Inc.

[boosting competitiveness of Pennsylvania business and industry]



► *To gain an edge in the fiercely competitive multi-billion dollar global market for electrical power, scientists at Westinghouse Science and Technology Center collaborated with PSC scientists to take advantage of PSC's CRAY T3E. For these turbines, even slight improvements in efficiency translate into significant reductions in the cost of generating power.*

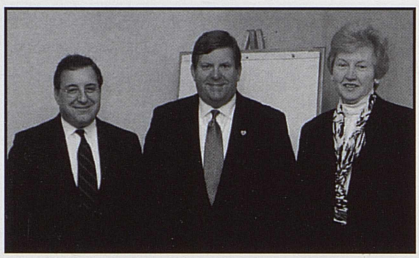
By converting the previously "sequential" software into a parallel form that can exploit the T3E, the simulations have become more realistic—more accurate as a design tool. At the same time, work that would have taken three months can now run in under 12 hours. The company cited this breakthrough as a factor leading it to retain 200 high-tech jobs in Pittsburgh. With this capability, Westinghouse is positioning itself as a world-class supplier of engineering analysis to the utility industry.

More information: http://www.psc. edu/science/cizmas.html

The Super Computing Science Consortium

On Aug. 31, 1999, officials from Pennsylvania and West Virginia signed a five-year agreement establishing a regional partnership, the Super Computing Science Consortium, or $(SC)^2$. The partners are the Federal Energy Technology Center (FETC), a U.S. Department of Energy (DOE) research laboratory with campuses in Pittsburgh and Morgantown, PSC, Carnegie Mellon University, West Virginia University and the West Virginia Governor's Office of Technology.

$(SC)^2$ paves the way for research collaboration between FETC and PSC and for a high-performance network linking West Virginia with the high-performance network hub at PSC. It also opens cooperative channels among the partners with the objective of enhancing research in the West Virginia-Southwest Pennsylvania region. Through the $(SC)^2$ framework, the partners will provide intellectual leadership to apply high-performance computing and communication to problems in medicine, pharmaceuticals, energy and the environment.



▲ *Officials gathered Aug. 31 at FETC in Morgantown to sign the agreement establishing $(SC)^2$. Paul Christiano (left), provost, Carnegie Mellon University; David Hardesty, president, West Virginia University; and Rita Bajura, director, FETC.*

Toward Low-Cost Clean Power

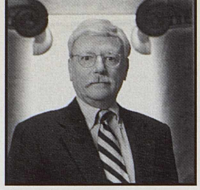
A grant from DOE to support research between PSC and FETC laid the groundwork for $(SC)^2$. During the summer of 1999, PSC scientists collaborated on several projects related to producing electrical power with greater efficiency and lowered environmental impact. Along with the three projects described here, another project (with Stanford University) is described later in this booklet (p.17).

Unsteady Flow in a Gas Turbine

PSC scientist Ravi Subramanya and Paul Cizmas of Texas A&M University used software they developed to simulate gas turbines. These huge jet-engine like machines do the heavy-duty work of converting fossil fuel into megawatts of electricity. Slight improvements in turbine efficiency can greatly reduce the cost of electrical energy.

The software—PaRSI3D (Parallel Rotor Stator Interaction 3D)—exploits the parallel-processing capability of machines like the CRAY T3E by using multiple processors to simultaneously calculate the flow at each turbine blade. This approach reduces simulation elapsed time compared to "serial" turbine-simulation, which sequentially calculates the flow at each blade. Along with greatly reducing turnaround time, PaRSI3D also produces more accurate results.

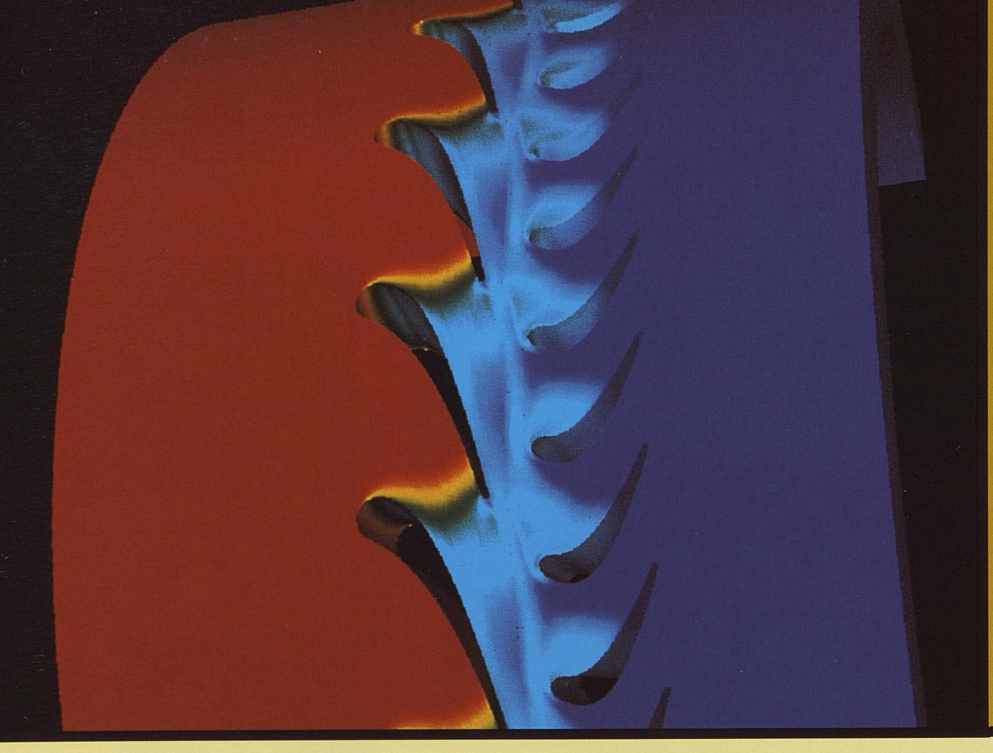
Cizmas and Subramanya first developed their parallel approach to turbine simulation in two dimensions. With PaRSI3D, they extended it to the much more challenging problem of three dimensions. Enhancements to the software provide useful information for turbine design, such as radial flow features. These simulations help to optimize the shape and relative positioning of turbine blades, design factors that can improve turbine efficiency.



▲ *Jim Kasdorf, PSC director of special projects. "FETC is a national laboratory in our own backyard, and the $(SC)^2$ partnership provides a regional focus for PSC—to use our knowledge and resources to work with others as a catalyst for high-technology development."*

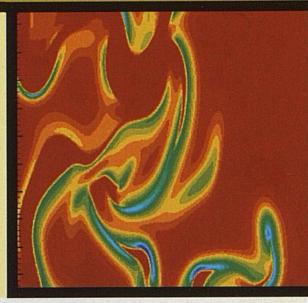
"The Pittsburgh Supercomputing Center is one of the most important technology assets in this region," says Matthew Tunnell, director of policy and technology for the Pennsylvania Department of Community and Economic Development, who represented the Commonwealth of Pennsylvania at the inauguration of $(SC)^2$. "It has been vital to the Southwest Pennsylvania technology renaissance. Our ability to expand access to this valuable asset will lead to innovation that creates wealth for this region."

[PSC will provide computational muscle for the Federal Energy Technology Center and a network hub for West Virginia University.]



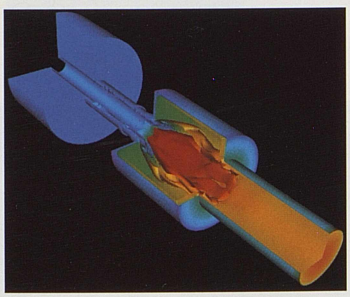
◀ *A frame from turbine simulations with PaRSI3D. The turbine has one stage (two rows) with a ratio of one stationary blade to two rotating blades. Color indicates flow density and shows how it decreases (from red to blue) when combustion gas speeds up through the stationary blades before striking the rotating blades.*

▼ *This frame from simulation of reactive turbulent combustion with two flame fronts shows heat-release contours in the flow field.*



Lean Combustion for Gas Turbines

As part of DOE's Advanced Turbine Systems Program, FETC researchers have used the CRAY T3E to simulate the FETC Dynamic Gas Turbine Combustor. With this experimental facility, FETC studies "lean, pre-mix combustion"—burning relatively low quantities of fuel relative to air, which reduces emissions of nitrogen-oxide pollutants.



▲ *This frame from simulations by FETC researchers (rendered by PSC scientist Nick Nystrom) depicts temperature during methane combustion in the FETC Dynamic Gas Turbine Combustor.*

Lean-fuel mixes often lead to combustion instability, rapid pressure oscillations that can create problems in the turbine. Using FLUENT (a commercial fluid-dynamics package) on the T3E, FETC researchers replicated a series of experiments using different combustor geometries at varying flow rates and fuel-air mixtures. Prior research indicated that the magnitude of the instability depends on the delay between when fuel is injected and when the mixture reaches the flame. The simulations confirm this and suggest that the delay varies with flame shape and size and that these flame characteristics themselves vary with combustor geometry and operating conditions.

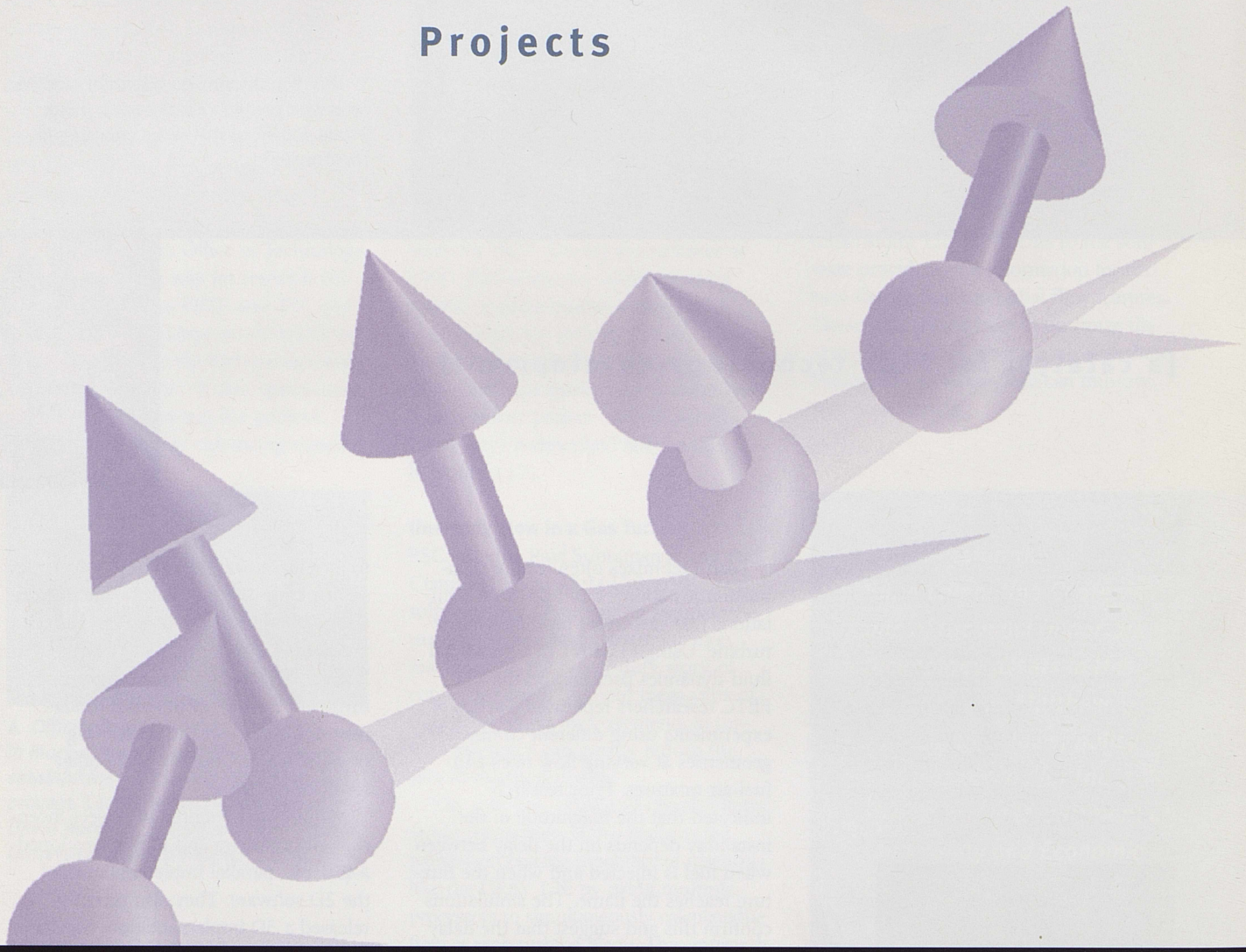
Improved Software for Turbulent Combustion

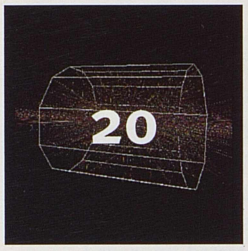
In another project of interest to FETC, PSC scientists Ravi Subramanya and Raghu Reddy collaborated with scientists at the Combustion Research Facility at Sandia-Livermore National Laboratory to develop a parallel implementation of software for direct numerical simulation (DNS) of combustion flow. The basic problem is what happens when gas or liquid fuel burns under conditions of turbulent flow. DNS provides fundamental insight into the underlying physics, which helps to verify other forms of modeling.

Within three months of project inception, Subramanya and Reddy developed a scalable, parallel implementation of the 2D software. They also recently released a 3D implementation. Sandia researchers can now use parallel systems like the CRAY T3E to tackle complex problems that previously would have exceeded computational capability. Tests show excellent scaling on the T3E—increasing speedup with added processors, so that problems taking a year with the serial software now take a day on the 512-processor T3E.

[1 9 9 9]

Projects





The Edge of Reality

In February 1998, PSC began a collaboration with the U.S. Department of Energy (DOE) that has in a short span produced impressive results across a range of projects. Some of this work, the STAR project (p. 20) and a partnership with Oak Ridge National Laboratory to simulate complex states of magnetism (p. 22), is detailed elsewhere in this booklet.

A large part of the DOE-PSC effort is university-based research under the Accelerated Strategic Computing Initiative (ASCI), a high priority national program using high-performance computing to replace nuclear testing as a way to assure viability of the nuclear stockpile. Through ASCI's Academic Strategic Alliances Program, research groups at the University of Chicago, Stanford, Caltech and the University of Illinois have worked with PSC staff and its CRAY T3E to attack complex problems in astrophysical thermonuclear flashes, turbulence in turbine engines, shock waves in gas, and design of solid-propellant rockets.

Another ASCI-related research group at Cornell and the Massachusetts Institute of Technology has worked with PSC to predict how metals perform at extremes of temperature and pressure. For all of these projects, the objective is to develop highly realistic simulation technologies.

"The DOE programs we support press the limits of large-scale simulation," says Jim Kasdorf, PSC director of special projects, who coordinates the DOE program. "PSC has been a leader in this area for more than a decade, and we're enabling the highest capability academically-based scientific computing in this country."

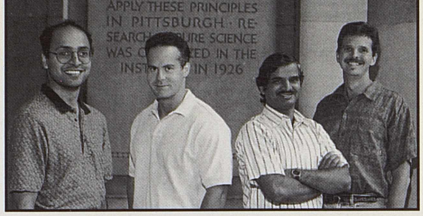
Improved Actions for Staggered Quarks

Quantum chromodynamics, or QCD, is the theory of the strongest force in nature, the force that holds the nucleus of an atom together. QCD tells us that protons and neutrons are made of bundles of particles called quarks held together by other particles called gluons—the strongest imaginable glue. According to QCD, quarks and gluons are the most fundamental particles in nature, the essential building blocks of all matter.

With support from DOE's Office of Science, a nationwide group of physicists, the MILC (MIMD Lattice Computation) collaboration, have used PSC's CRAY T3E to develop computational methods for QCD. The theory has withstood many tests, says MILC physicist Robert Sugar of the University of California, Santa Barbara, but it's been difficult to extract predictions that can be measured against experiment. Because of the complexities of the theory, QCD is one of the most demanding computational projects in science.

QCD theory is defined in the space-time continuum, but to do numerical simulations requires treating space and time as finite, on a four-dimensional grid called a lattice. The lattice, however, introduces inaccuracies—called finite-lattice spacing effects. To obtain

▲ *Robert Sugar, University of California, Santa Barbara, at a PSC seminar.*



▲ *This group of PSC scientists—(l to r) Ravi Subramanya, John Urbanic, Raghu Reddy and Nick Nystrom—collaborated with ASAP research teams to adapt software to the CRAY T3E environment, troubleshoot problems and produce visualizations of the results.*

reliable results, therefore, the lattice spacing must be decreased (and the number of grid points correspondingly increased), which quickly eats up computing capability.

To escape this conundrum, the MILC collaboration and other lattice-gauge theory groups have worked to develop more sophisticated schemes—called "improved actions"—for putting the theory on the lattice. Through an extensive series of computations, MILC researchers have tested improved actions for a particular formulation of QCD known as Kogut-Susskind or staggered quarks. "We've used PSC's T3E and other resources," says Sugar, "to develop a family of improved actions for the Kogut-Susskind formulation of lattice quarks." Tests show a significant reduction in finite-lattice spacing effects. As a result of this success, MILC is embarking on a major program of production runs.

Magnetism in the Solar Dynamo: University of Chicago

What happens on the surface of the Sun as the fusion cooker at its core converts matter to energy at the rate of four-million tons per second? An array of complex phenomena occurs in stars like the Sun—nuclear burning, turbulence and convection at extremes of temperature and pressure, magnetic-field generation, star expansion, star death and many more. How can we understand this physics, which simply isn't subject to experiments in the normal sense? You can't build a laboratory model of a star.

The answer, say scientists at the Center for Astrophysical Thermonuclear Flashes at the University of Chicago, is large-scale numerical simulation. Through the ASCI program, this group of researchers has harnessed PSC's CRAY T3E to simulate various stellar phenomena,

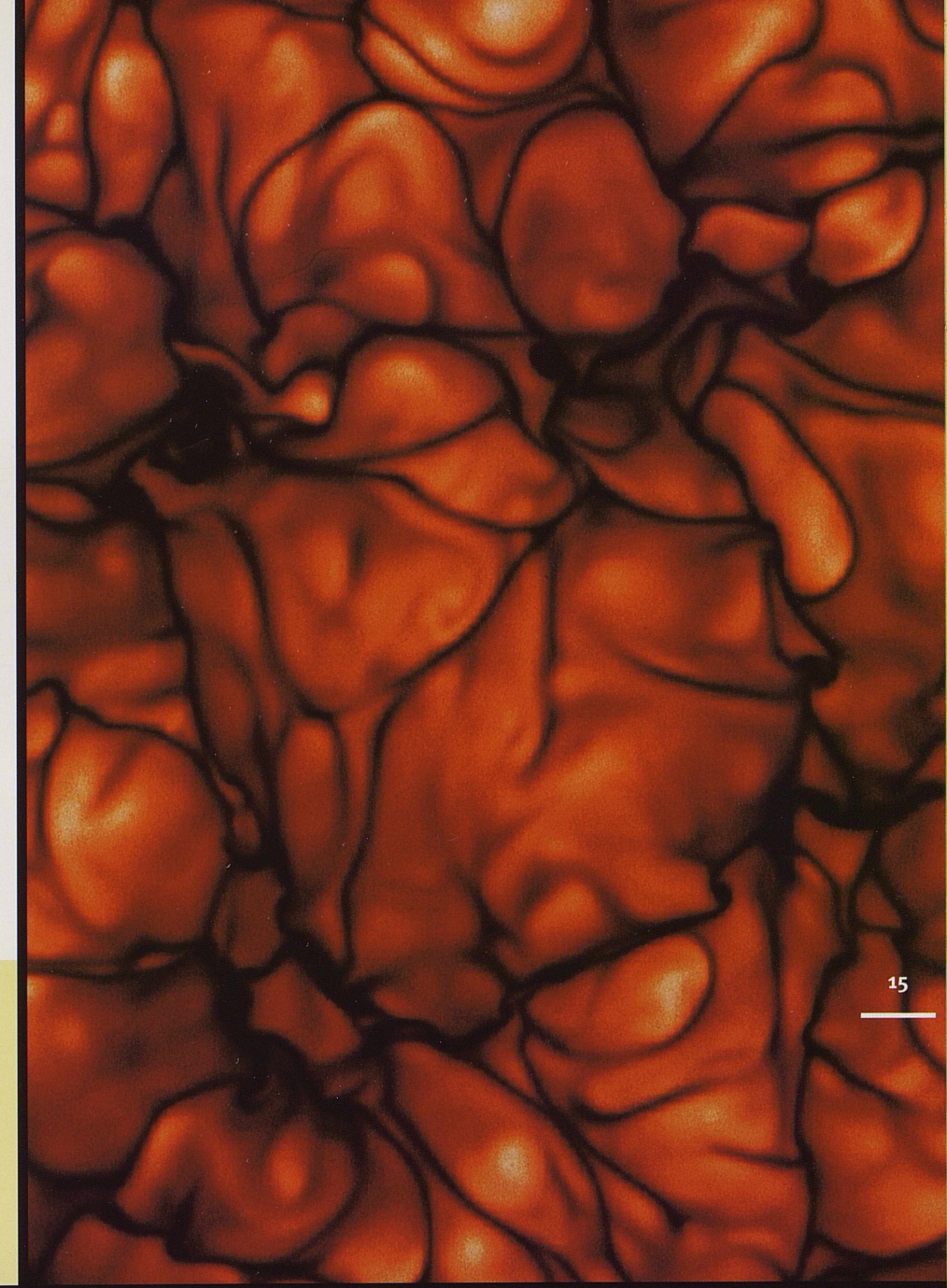
[a new theory of magnetic fields on the Sun]

including x-ray bursts from neutron stars and, recently, a large-scale study of how stars like the Sun generate magnetic fields.

In a series of computations, physicist Fausto Cattaneo simulated the "dynamo action" of the Sun. His question—what is the structure of magnetic fields generated by overturning turbulent layers of ionized gas (convection) at the Sun's surface? Large-scale solar magnetic fields result from rotation patterns in so-called active regions of the Sun, but recent theory has suggested that smaller-scale magnetic fields may also be generated by convection at the Sun's surface.

Essentially, Cattaneo created a box of solar fluid at high temperature inside the computer, and—because of the capability of the CRAY T3E—his box of fluid simulated the physics of the solar dynamo with much greater realism than previously possible. "This calculation," says Cattaneo, "was 1,000 to 10,000 times more ambitious than previous studies." The results—from a series of 512-processor runs—confirm recent theory and indicate that turbulent convection can be an efficient generator of small-scale magnetic fields.

"The T3E is a wonderful machine," says Cattaneo, "and PSC has it set up very well—with the least amount of headache to get it to work. And when it works it's like a dream. The mass-storage system is easy to use. You can migrate data place-to-place without getting stuck. Of all the places I've computed, PSC has been outstanding."



▲ *Temperature fluctuation in solar convection. This panel, a horizontal plane near the upper boundary of the volume, shows hot (red) and cold (black) fluid regions. The granular structure is typical of convection.*

◀ *Distribution of magnetic energy from solar convection. This frame from a video animation (produced by PSC visualization artist Greg Foss) shows positive (orange) and negative (blue) magnetic field intensities.*

Rocket Science:
University of Illinois Urbana-Champaign

Solid-propellant rockets are the heavy lifters of the aerospace industry. They provide the immense thrust needed to launch large payloads into Earth orbit or outer space. Design of these rockets is a sophisticated problem drawing on a range of subdisciplines: ignition and combustion of composite energetic materials; solid mechanics of the propellant, its case, insulation and nozzle; fluid dynamics of the interior flow and exhaust plume; quantum chemistry of energetic materials; aging and damage of components; and analysis of potential failure modes.

The goal of the Center for Simulation of Advanced Rockets at the University of Illinois Urbana-Champaign is detailed, whole-system simulation of solid-propellant rockets at normal and abnormal operating conditions. Comprehensive simulation will provide a much safer, less expensive approach to the technological issues than traditional rocket design methods of trial and error. CSAR is using the solid rocket motor of the NASA space shuttle as its simulation vehicle, and scientists there have employed PSC's CRAY T3E to test and develop a range of simulation tools.

In large-scale calculations, Dinshaw Balsara and colleague C. W. Shu of Brown University tested a class of numerical schemes—algorithms—that can simulate turbulent flow which includes very large and very small eddies at the same time—complex structure like that occurring in the core of solid-propellant rockets. This kind of simulation is very demanding of computational capability, and high-quality algorithms use less computational firepower to accurately represent a wide spectrum of turbulence. The calculations by Balsara and Shu implemented effective strategies for "parallelizing" the software, achieving overall exceptional performance (32.6 gigaflops on 512 processors) with efficient scaling to very large systems, while allowing extension to increasingly high orders of accuracy.

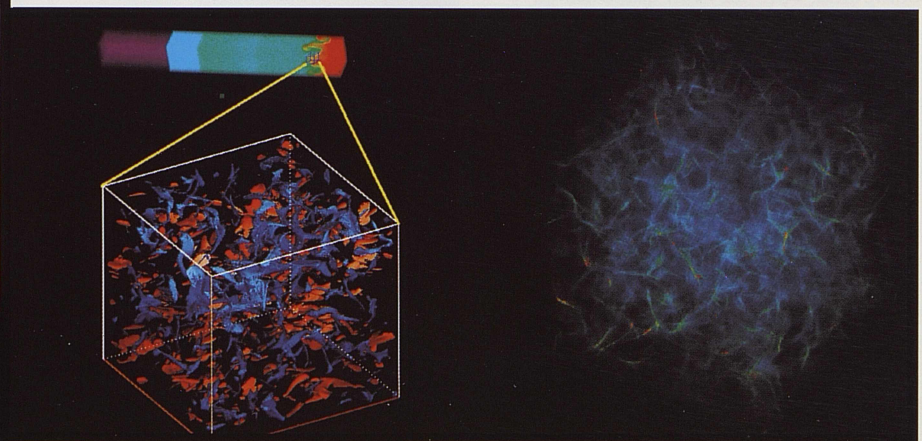
"Being able to use PSC's T3E was extremely valuable," says CSAR managing director William Dick, "in our efforts to develop code that's easily portable to any parallel machine—an important goal of our ASCI research. The network connection to Pittsburgh is excellent, and support from PSC staff has been outstanding across all aspects of our work, including code optimization, parallelization, debugging—you name it. Our success in this work is also a PSC success."

▼ *October 29, 1998: The NASA Space Shuttle blasts off with Senator John Glenn on board.*

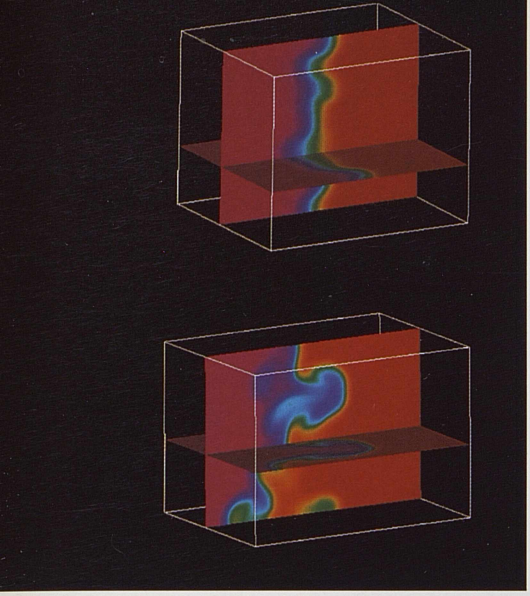
▶ *Turbulent mixing from reshock. These two snapshots represent simulation of a Mach 10 shock wave moving left to right through light gas (violet) in a square channel ramming into gas three times denser (red). Density contours at the interface become more complicated (below), improving mixing between the fluids, after the transmitted shock bounces off the side wall and "reshocks" the interface.*

Shock Waves in Gas:
California Institute of Technology

A shock wave traveling in air at supersonic speed rams into a contained volume of high-density air. The turbulent mixing that occurs in shock interactions at the interface between gases of different density is an important problem in technologies such as scramjets—propulsion at hypersonic speeds—and power generation by inertial confinement fusion. It also occurs in natural phenomena such as the violent star explosions called supernovas.



▲ *Zooming in on shocklets. At fine scale, simulations show vorticity (red) and divergence (blue)—a measure of how compressed a fluid is. In the companion graphic, showing divergence of the flow field, the stringy structures represent "shocklets"—very compressed regions, or high negative divergence (red).*



[outstanding code optimization and parallelization]

This problem is the focus of research at Caltech's Center for Simulation of Dynamic Response of Materials. Using PSC's CRAY T3E, Caltech scientists Daniel Meiron and Ravi Samtaney have carried out a series of calculations aimed at developing realistic simulation capability to complement experiments in Caltech's T5 hypervelocity shock tunnel.

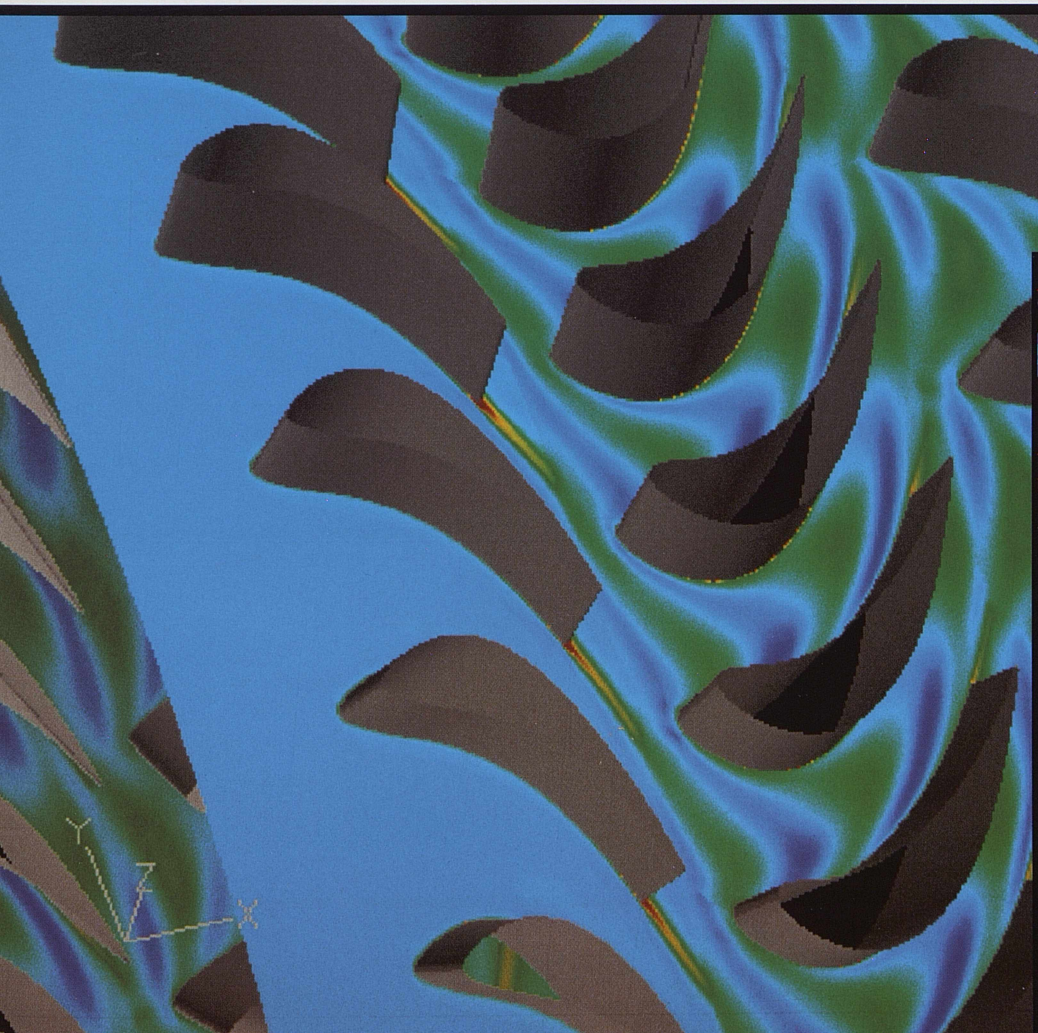
In 1998, with a series of 512-processor calculations, Samtaney and Meiron explored the turbulent mixing that occurs after the shock crosses the interface and bounces off a surface to "reshock" the mixing zone. These very large calculations were among the first to explore this phenomena in three dimensions.

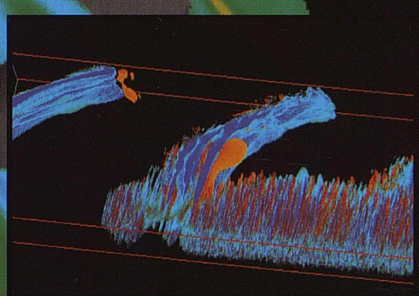
In recent work, Samtaney zoomed in on the details of turbulence at the interface. These direct numerical simulations at fine scale will help verify the accuracy of larger-scale modeling. At much higher resolution than prior similar studies, these simulations provide a more accurate picture of reality and, as a result, show the development of "eddy shocklets"—small regions of highly compressed fluid, essentially a fine-scale version of shock waves.

"The T3E is fantastic," says Samtaney. "In sheer speed, this machine is blazing compared to others I've used. We've been able to push this kind of simulation, called isotropic turbulence, to higher levels of turbulence where you see these regions of high compression, and it's important that our models be able to handle that."

Turbines and Turbulence: Stanford University

Gas turbine engines are the reliable workhorse of modern civilization. They produce a large portion of our electrical energy and they keep global business on schedule by lifting giant airliners into the air. The basic idea is as old as water wheels and windmills—rotating blades translate raw energy into circular motion. In the modern version, the raw energy is fossil fuel, and the design involves complex fluid dynamics. Stationary

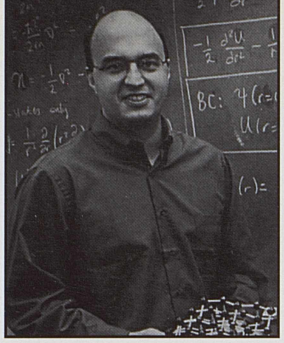




▲ *Transition from smooth to turbulent flow along the surface of a turbine blade. The rear wall of the 3D volume represents the blade surface. Turbulence can be seen increasing from left to right. These simulations seek to isolate the process that triggers turbulence.*

◀ *Unsteady flow of the Aachen experimental compressor as simulated by Jixian Yao of Stanford on the CRAY T3E.*

[seeing the transition to turbulence with unprecedented detail]



▲ *Tomás Arias, Cornell University. "We're very pleased with the quality of the computing environment and user support at PSC, which is truly unparalleled. It would be very difficult to maintain this level of production anywhere else. It's great to be cranking out results like this."*

blades (stators) channel high-speed flows into rotating blades (rotors). Variations in the shape or number of blades along with other variables govern how well the turbine works.

The Center for Integrated Turbulence Simulations at Stanford University directs its research toward realistic simulations for the design of the compressor, combustor and turbine components of aircraft gas turbine engines. Ultimately, the goal is to integrate these component simulation technologies and to accurately simulate the complex situations governed by interactions among components. An array of phenomena present themselves as challenges—rotating stall in the compressor, blade vibrations, combustor instabilities due to coupling between heat release and acoustic modes, and heat transfer from the combustion gas to the first blade rows of the turbine.

In one large-scale study at PSC, CITS researcher Paul Durbin and colleagues attacked the problem of "wake-induced transition" from smooth to turbulent flow along the surface of a rotor blade. As flow goes from one stage of blades to the next, the wake from upstream blades strikes downstream rotors, disrupting smooth flow along the rotor surface and creating flows that are inherently complex and unstable. In a series of direct numerical simulations, solved using 52 million grid points, the researchers looked at this problem with an unprecedented level of detail.

In another series of CITS studies at PSC, Jixian Yao carried out calculations on the unsteady flow in a multi-blade-row compressor. These simulations aim at validating a particular "flow solver," and his computations simulate the geometry of an experimental compressor, the Aachen Compressor, composed of three blade rows.

"PSC has been a tremendous resource for our program," says CITS director Bill Reynolds. "For the kind of highly structured code we write, the CRAY T3E is an excellent machine that scales well as you go up in processor count. The consulting people were extremely helpful. PSC is a great asset to the national research effort and has enabled us to make substantial progress."

Other turbine research at PSC is described in this booklet (pp. 10-11).

A New Picture of How Metals Deform

"Our calculations at PSC are turning the last 30 to 40 years of thinking about the origins of plasticity in a large class of transition metals on its head," says physicist Tomás Arias of Cornell University. He and MIT graduate student Sohrab Ismail-Beigi are using PSC's CRAY T3E to study fundamental properties of molybdenum and tantalum, both of which are "transition metals," a subgroup of elements related in atomic structure that hold promise for high-strength, high-temperature technological applications.

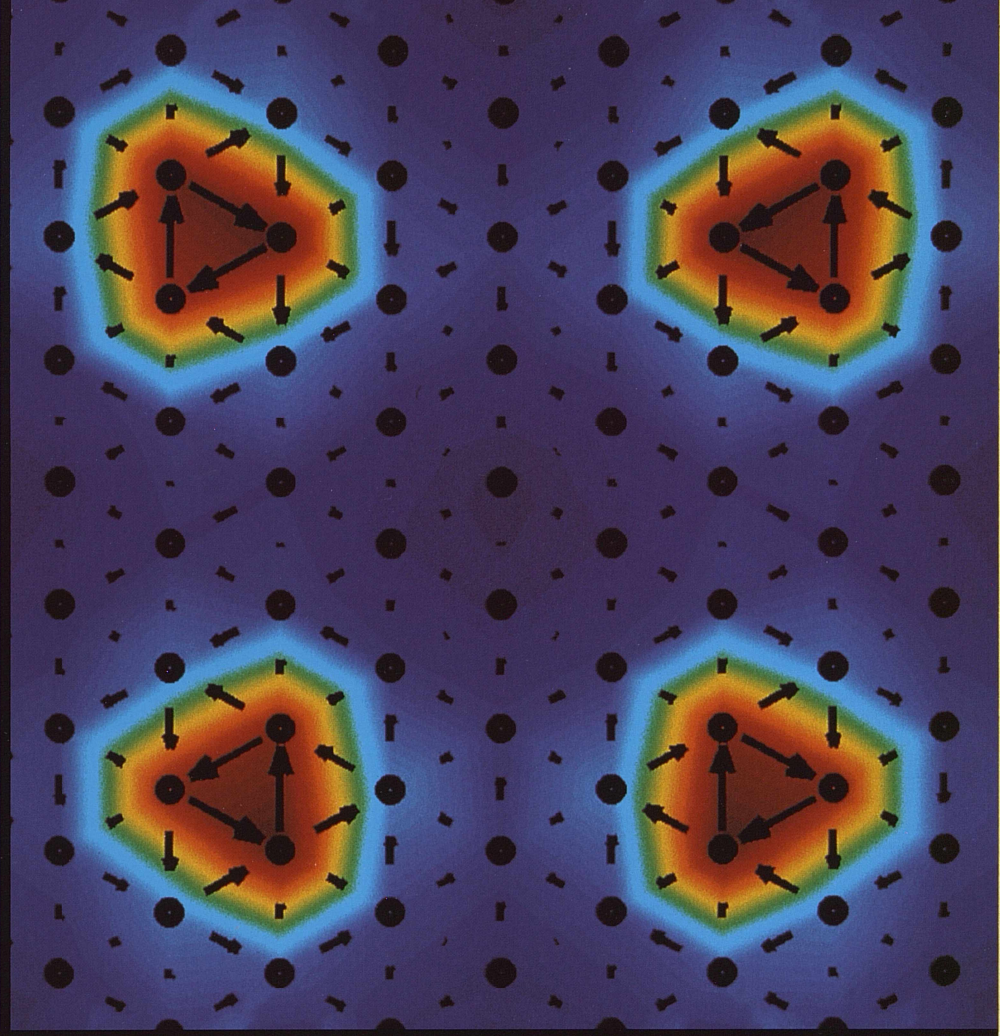
Their PSC work represents the first calculations applying density-functional theory—a quantum-theory based approach to representing the interactions between atoms and electrons—to "screw dislocations" in these metals. Dislocations are microscopic structures that govern how the metal deforms, and accepted thinking has been that a dislocation in these materials must have a particular atomic structure—known as a symmetry breaking core.

The calculations at PSC appear to overturn this view. They show that for the screw dislocation in molybdenum and tantalum symmetry isn't broken. They also improve on the accepted theoretical view by indicating better agreement with experiments on the amount of stress required to deform the metals. "Our results," says Arias, "indicate a screw dislocation core without broken symmetry and with energy scales in much better accord with the experimental data."

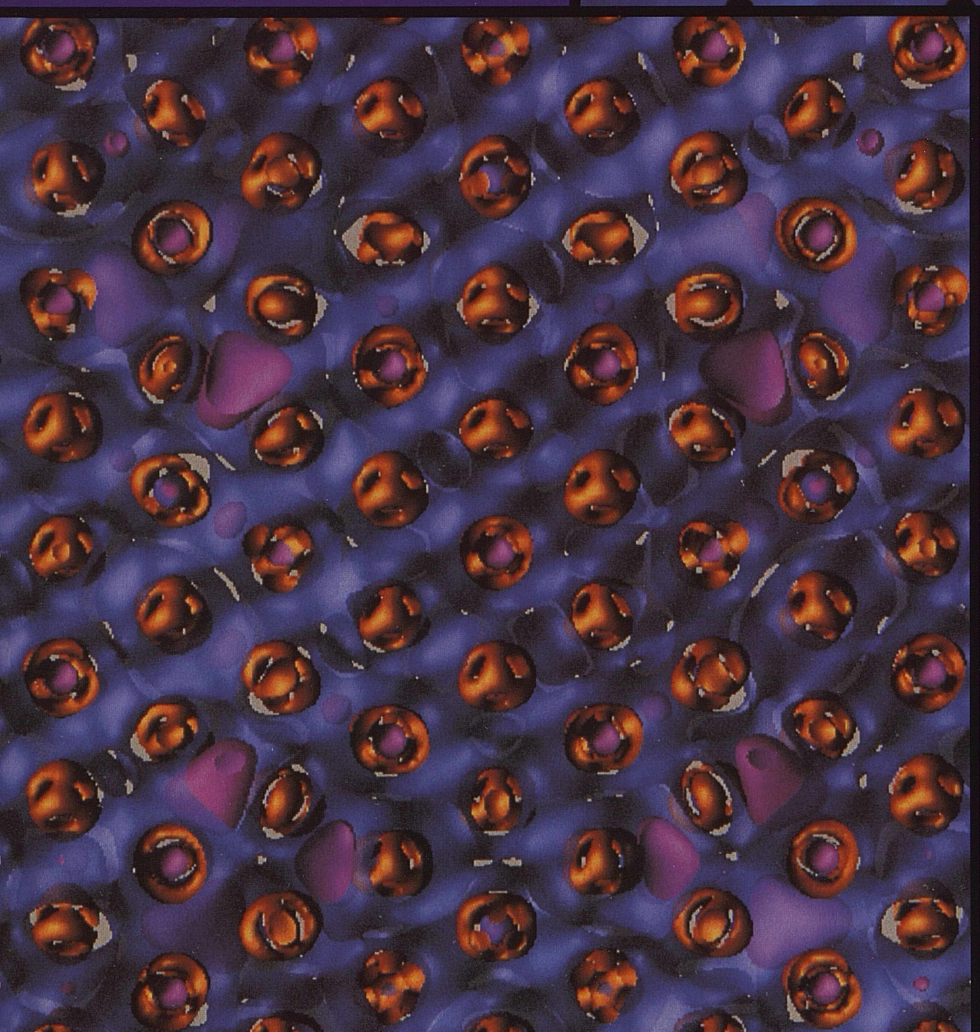
Arias and Ismail-Beigi also computed the energies and electronic structures associated with two different configurations of the screw dislocation in molybdenum and tantalum. For both metals, there's a low energy, relatively stable configuration known as the "easy core" and a higher energy, less stable state known as the "hard core." PSC calculations show the energy difference between the two states to be two to three times less than computed with the accepted models.

"This information is critical," says Arias, "to constructing large-scale theories of dislocation motion. These results are changing our understanding of the basic physics of these systems, and they provide important information for the experimental part of the ASCI program."

[new understanding of the basic physics of metals]



▶ *Elastic energy density in the "easy core" configuration of tantalum. High energy regions (red) surround the four dislocations in this 90-atom system. Vectors indicate the displacement of atoms (strain) from their ideal positions.*



◀ *Electron density of the "easy core" screw dislocation in tantalum as calculated by Arias and Ismail-Beigi. A sea of electrons (blue) of medium density moves through the metal except in voids surrounded by low density regions (magenta). Other electrons at high density (gold) remain attached to the atoms.*

A Taste of Quark Soup

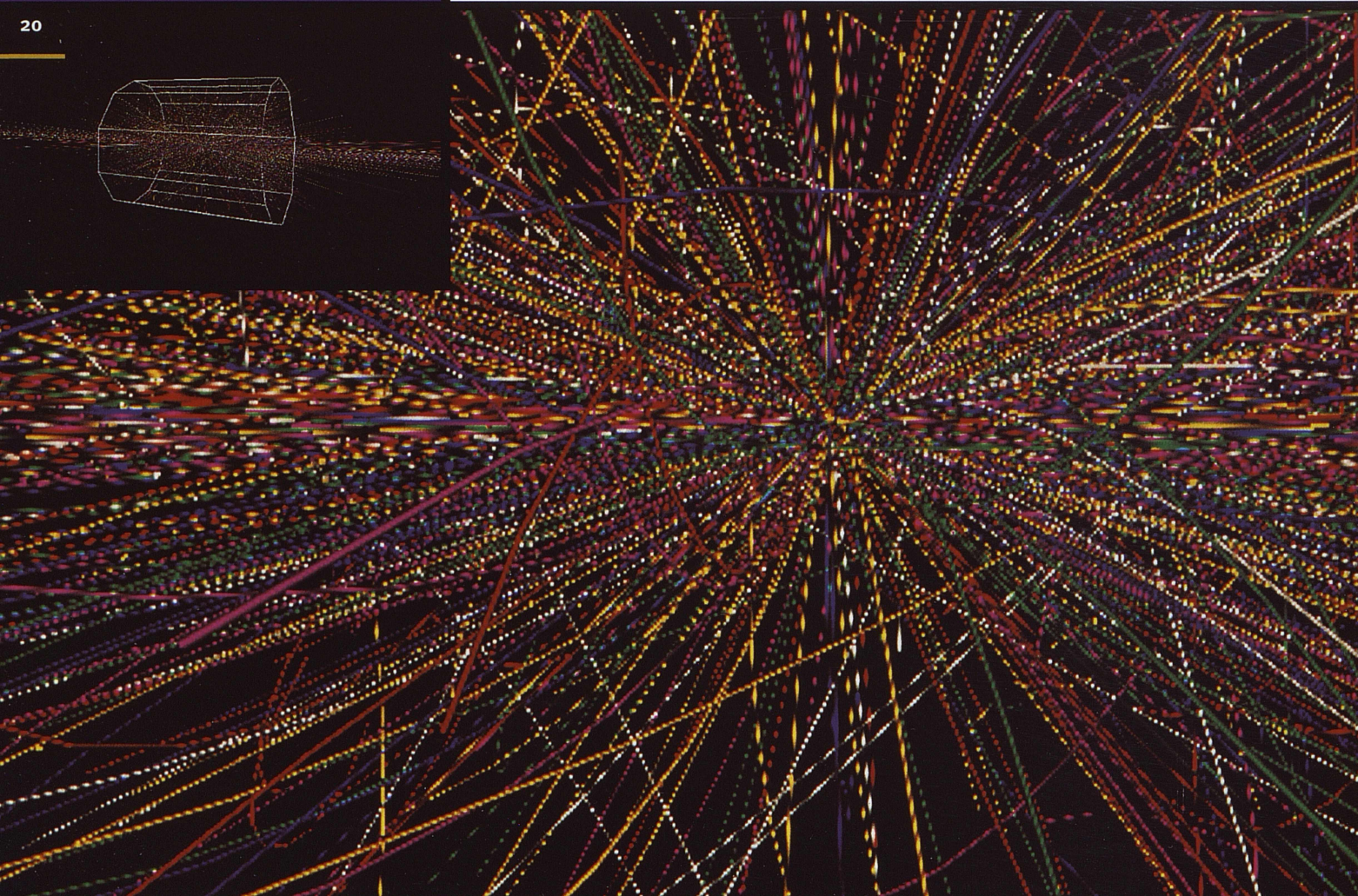
To recreate in a contained space conditions like those during the first microsecond after the birth of the universe—that's what physicists are up to on Long Island as the millennium winds down. The Relativistic Heavy Ion Collider, RHIC, a $600 million project at Brookhaven National Laboratory is now finishing tests in preparation for the first actual physics runs, scheduled for January 2000.

"There's a tingling in the air," says Mort Kaplan, professor of chemistry at Carnegie Mellon. "We're going to see phenomena we've never seen before." Kaplan is a founding member of the STAR (Solenoidal Tracker at RHIC) collaboration, which involves 37 institutions and more than 400 researchers worldwide. One of two large detectors in the new collider, STAR will track the profusion of particles liberated when heavy ions smash into each other at near light speed.

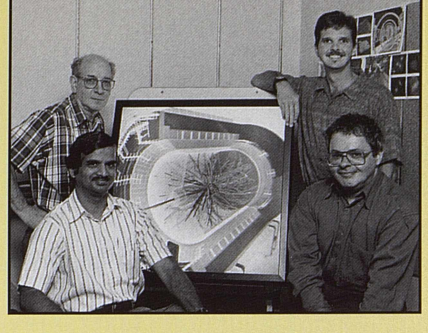
Beginning in summer 1998, Kaplan and his colleague Dan Russ collaborated with PSC scientists to conduct a series of simulated collider events to help researchers and their processing systems get ready for the real thing. These "mock data challenges" relied on PSC's CRAY T3E and also used a T3E at NERSC (National Energy Research Scientific Computing Center) in California.

The MDCs produced nearly 300,000 simulated heavy-ion collisions, captured as 3.9 trillion bytes (terabytes) of data, which gives an idea of the magnitude of the computing challenge associated with STAR. Run-throughs of the data storage and analysis processes led to software changes that sped-up the process and reduced the amount of memory required. "The availability of extensive PSC resources," says Kaplan, "not only the T3E, but also the massive data-storage capability and consulting staff close at hand—was crucial to the success of this very large-scale series of simulations."

▼ *Images from simulation of a gold-gold head-on collision. The STAR detector is shown in outline. The same event is also shown closeup. Colors represent different charged particles, with about 20 represented.*



[PSC resources were crucial]



▲ *Mort Kaplan, Carnegie Mellon, a founding member of the STAR collaboration (standing left) with PSC scientists Nick Nystrom (standing right), Sergiu Sanielevici (seated right) and Raghu Reddy.*

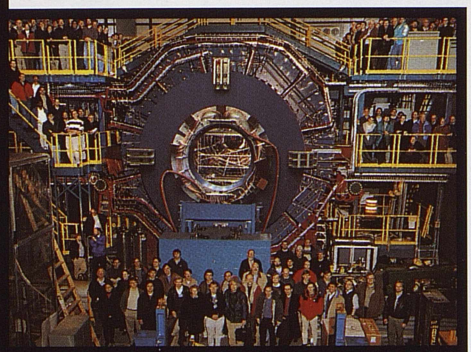
When Heavy Ions Collide

The "HI" in RHIC means heavy ion, and that's the key to this collider for the new millennium. RHIC will be the first facility to smash heavy ions into each other, and prior collision experiments by comparison are relatively tame. Many RHIC experiments will use gold, a "heavy element"—with a massive nucleus of 79 protons and 118 neutrons. Strip away the electrons and you have a heavy ion. RHIC will accelerate gold ions around two nearly circular 2.4-mile tracks in opposite directions to reach 99.99 percent the speed of light.

Compared to other accelerators, RHIC collisions will create much higher nuclear densities. The colliding nuclear particles will flatten like pancakes and pass through each other for an instant before they explode, an ultra-high density state equivalent to turning up the heat to ten trillion degrees Kelvin, 10,000 times hotter than the center of the sun. The resulting blast of particles, to put it nicely, is messy. A head-on gold-gold collision is likely to produce 5,000 to 10,000 individual subatomic particles shooting off in every possible direction.

Somewhere in that instant of chaos may exist a state of matter that mirrors the universe at the instant after the big bang. At these extreme conditions, the most fundamental particles—quarks—are expected to be released from the gluons that bind them into neutrons and protons. That's what the physicists are looking for—a state of matter never before observed called quark-gluon plasma, known less technically as quark soup.

If quark soup is there, it will be up to STAR to detect it. The $60 million STAR detector is a 1,200 ton instrument about the size of a house. At its heart is the Time Projection Chamber, essentially 140,000 wires that operate like a 3D digital camera. Flying particles induce current in the wires, allowing the camera to record them. If quark soup is there, this recorded data will enable scientists to sniff it out.



▲ *The STAR detector at RHIC.*

Calibrating the STAR Trigger

STAR presents a supreme challenge in data acquisition and processing as well as physics. "There will be up to 1,000 collisions per second," says Kaplan. "And the fastest you can write to magnetic tape is one collision per second." Each collision yields about 20 megabytes of data, about the amount that can be recorded in a second.

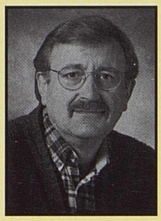
To sift through this overabundance of data and identify events that may yield new knowledge involves applying theory-based physics models of what will happen. "Most of what the detector will see isn't interesting," explains Kaplan. "Fast decisions have to be made online, simple measurements that allow you to decide whether to eliminate an event and go on to the next one."

This process, controlled through a series of software pipelines, is called the STAR trigger. The question is how to decide in real time what data to collect and what to discard, and the mock data challenges were crucial in setting up algorithms to do this. But the real test is yet to come. No one knows with certainty what will happen when RHIC flings gold ions at each other for the first time.

The theoretical models, emphasizes Kaplan, are based on long extrapolations from known results. For this reason, a large part of the analysis will be to carry out new simulations to understand the data. Computational simulations will continue to be crucial to STAR's search for quark soup.

More information:
http://www.psc.edu/science/star.html

Magnetic Moments



▲ *Malcolm Stocks, Oak Ridge National Laboratory.*

▲ *Yang Wang, Pittsburgh Supercomputing Center.*

Where would we be without magnets? Bulk magnets similar to those that transform refrigerators into photo displays play a little-noticed role in modern life, but they're essential in power generation, electrical motors, the sensors in cars, and many home appliances.

Other magnetic materials, thin films and composites, are ubiquitous in consumer electronics. The growth of home entertainment, from cassettes to videos, has been driven by progress in magnetic-recording technology, and the information age we're living in would still be a futurist dream (and programmers would still be shuffling punch cards) were it not for amazing magnetic devices like the read-heads on disk drives.

Despite these advances, we know less than you might think about how magnetism works at the level of atoms and electrons. "Even refrigerator magnets are very complicated," says Malcolm Stocks. "There are big pieces of the underlying physics we don't understand." With research support from the U.S. Department of Energy, Stocks, a physicist at Oak Ridge National Laboratory, leads a team of scientists working to unlock magnetism's puzzles.

As a member of this team since 1993, PSC scientist Yang Wang helped develop software called LSMS—the locally self-consistent multiple scattering method—that simulates the interactions between electrons and atoms in magnetic materials. In 1998, LSMS became the first research software to break the teraflop barrier in supercomputing.

Breaking the Teraflop Barrier

On Nov. 9, 1998, running on a 1,480 processor CRAY T3E system at SGI/Cray, LSMS sustained performance of 1.02 teraflops (trillions of calculations per second), the first time an actual research program ran at teraflop speed. For an earlier 657 gigaflop run with the same software, Stocks' team—which includes researchers from Oak Ridge, the National Energy Research Scientific Computing Center (NERSC), the University of Bristol (UK) and PSC—won the 1998 Gordon Bell Prize for best achievement in high-performance computing.

Wang's contribution played a crucial role. "He took the code to Pittsburgh and made it run on the T3E," says Stocks. Wang came to PSC in 1996 and in this position collaborated with Oak Ridge scientist Bill Shelton to translate LSMS to run on PSC's CRAY T3D, predecessor to the T3E.

LSMS can simulate unit cells (the irreducible unit of atomic structure for solid materials) ranging from hundreds to thousands of atoms, well beyond most existing methods, and close to the range of realistic magnetic materials. The Nov. 9 calculation simulated a 1,458 unit cell of iron. "This is not much smaller than nano-particles being made today," says Stocks.

Investigating Noncollinear Magnetism

Up till now, to a large degree, computational studies have looked at how magnetism works under limited conditions, so-called ground states, at absolute zero temperature. Under these conditions, the magnetic moments of the atoms—which occur due to unpaired electrons in the outermost orbits—are aligned, all pointing in the same direction or at least parallel to each other.

In real-world conditions, magnets aren't maintained at absolute zero, and the magnetic moments change direction. At high enough temperature, the moments become randomly disordered, eliminating the magnetic field. The objective of LSMS is to track such disordered magnetic states, called noncollinear magnetism. "Now is the time to do real-world problems," says Wang, "where the magnetic moments aren't all up or down, which is what LSMS and parallel systems allow us to do."

LSMS relies on density-functional theory (for which physicist Walter Kohn won the 1998 Nobel Prize in chemistry) to describe the interactions between electrons and atoms in magnetic materials. DFT alone, however, can solve only aligned states of magnetism. To simulate noncollinear states, Stocks' team developed a complex variation of the basic theory, called "constrained DFT," that applies a separate magnetic field to each atom in the unit cell.

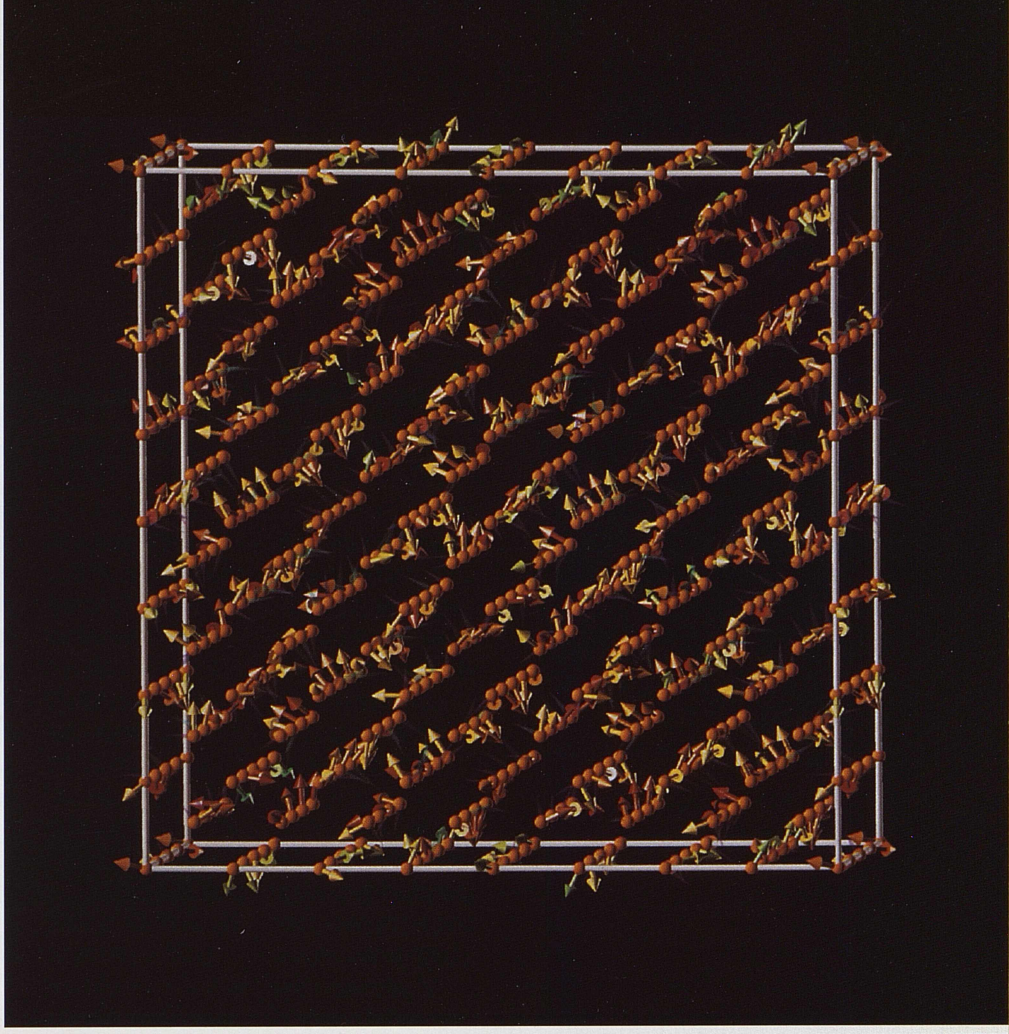
In a recent calculation, the group applied its constrained DFT model to a 512-atom unit-cell of iron. The method keeps track of how a change in the magnetic moment of one atom affects other atoms in the vicinity, and solves the relevant quantum-mechanical equations for each atom of the unit cell. The results,

[The 1998 Gordon Bell Prize for best achievement in high-performance computing]

using a 512-processor CRAY T3E at NERSC, validate LSMS as a powerful new tool for studying the finite temperature properties of metallic magnets.

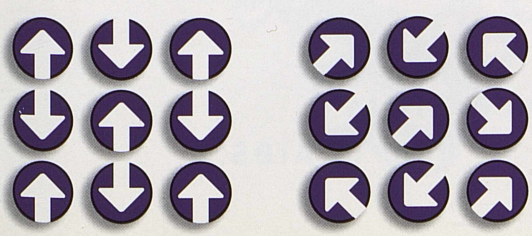
In future calculations, Stocks, Wang and their colleagues plan to address problems in magnetic multi-layers and magnetic nano-particles. Their work is basic research, aimed not at technology per se but at new scientific understanding. At the same time, however, the potential to bring about new technologies is real. Even small improvements in the information density of magnetic-storage devices, for instance, could radically transform computing. "If you could even slightly increase the maximum energy product of a permanent magnet," says Stocks, "it would have enormous effect on technology."

More information:
http://www.psc.edu/science/wang.html



▶ *Results from computing the magnetic moments for a 512-atom unit-cell of iron above its Curie temperature, when the magnetic field disappears. As the closeup shows, the magnetic moment at each atom (arrowhead vectors) has a corresponding constraining field (translucent cones) as a result of the constrained DFT model. Colors indicate vector magnitude (see color bars).*

▼ *Aligned moments create a magnetic field (left). Parallel non-aligned moments (center) are "antiferromagnetic"—there's no magnetic field. Random moments (right) exemplify non-collinear magnetism.*

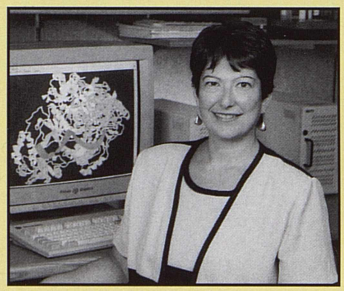


Getting a Thumb-Lock on AIDS

▲ *This ribbon representation of the RT enzyme shows the hand with fingers (blue), palm (red) and thumb (green), with and without DNA (yellow).*

It will soon be 20 years since AIDS sprang into medical awareness. It remains one of the principal threats to human health worldwide. Over 20 million people live with AIDS, and recent data for the United States alone indicate 40,000 new infections annually.

The culprit is human immunodeficiency virus, HIV, an infinitesimal particle of proteins and ribonucleic acid that invades immune-system cells. Limited success in treating AIDS has been achieved with drugs that slow HIV's ability to reproduce. Many of these target reverse transcriptase, an HIV enzyme carried into host cells by the virus. RT does a copy-and-paste of HIV's single-stranded RNA to form double-stranded DNA, which assimilates into the host-cell DNA. By this maneuver, HIV co-opts the host's genetic machinery to reproduce itself.



▲ *Marcela Madrid, Pittsburgh Supercomputing Center.*

"If you can inhibit RT from working, you can cure AIDS," says Pittsburgh Supercomputing Center scientist Marcela Madrid. RT may be HIV's Achilles heel, but potent drug cocktails that attack it haven't been a cure-all. HIV has a protean ability to mutate and outmaneuver drugs, and for this reason especially scientists seek detailed knowledge of RT.

In recent work, Madrid, whose expertise is using supercomputers to simulate the structure and movement of biological molecules, collaborated with a team of HIV researchers. Using the CRAY T3D and T3E, she carried out large-scale "molecular dynamics" simulations of RT, one of the largest enzymes ever simulated this way. The results supplement insights from crystallographic studies with new detail, and they show that computer simulations have a major role to play in helping to develop drugs that shut down HIV replication.

Thumbs Up/Thumbs Down

Since the late 1980s, several teams of structural biologists have focused on RT, using x-ray crystallographic methods to determine the 3D structure of the enzyme. In 1992, a team led by Thomas Steitz at Yale achieved the first success. Another team, led by Eddy Arnold at Rutgers, followed soon after with a structure of RT bound with double-stranded DNA, in effect a snapshot of RT in action.

In the next few years, the Arnold team produced a series of variant RT structures, including drug-resistant mutant versions and others of RT linked with inhibitors, compounds that block the enzyme's ability to carry out its role. Ultimately, this body of work led the way to RT-inhibitor drugs, including AZT, that have helped many HIV-infected people.

In the middle 1990s, a member of Arnold's team, Alfredo Jacobo-Molina, continued his work on RT at the Monterrey Institute of Technology in Mexico. Through a collaborative program between Monterrey and Carnegie Mellon, Jacobo-Molina met Madrid and told her about an interesting feature of RT's structure.

Arnold's group had produced two complementary RT structures—with and without DNA—that revealed a difference. If you visualize the active subdomain of RT as a hand—with palm, fingers and thumb, the thumb extends when the hand grasps DNA, making space for the DNA to fit into the palm. Without DNA, the thumb closes.

This joint-like flexibility has implications for HIV transcription. Only in the closed-thumb position, it's believed, can RT slide along a single strand of RNA to find the starting place for transcribing and building DNA. Some biologists believe that one class of drugs works by lodging in the palm of the active subdomain, like sand in a gear, locking the thumb in an open position, so the enzyme can't function normally.

[New knowledge about the Achilles heel of AIDS.]

A Movie: How the Thumb Closes

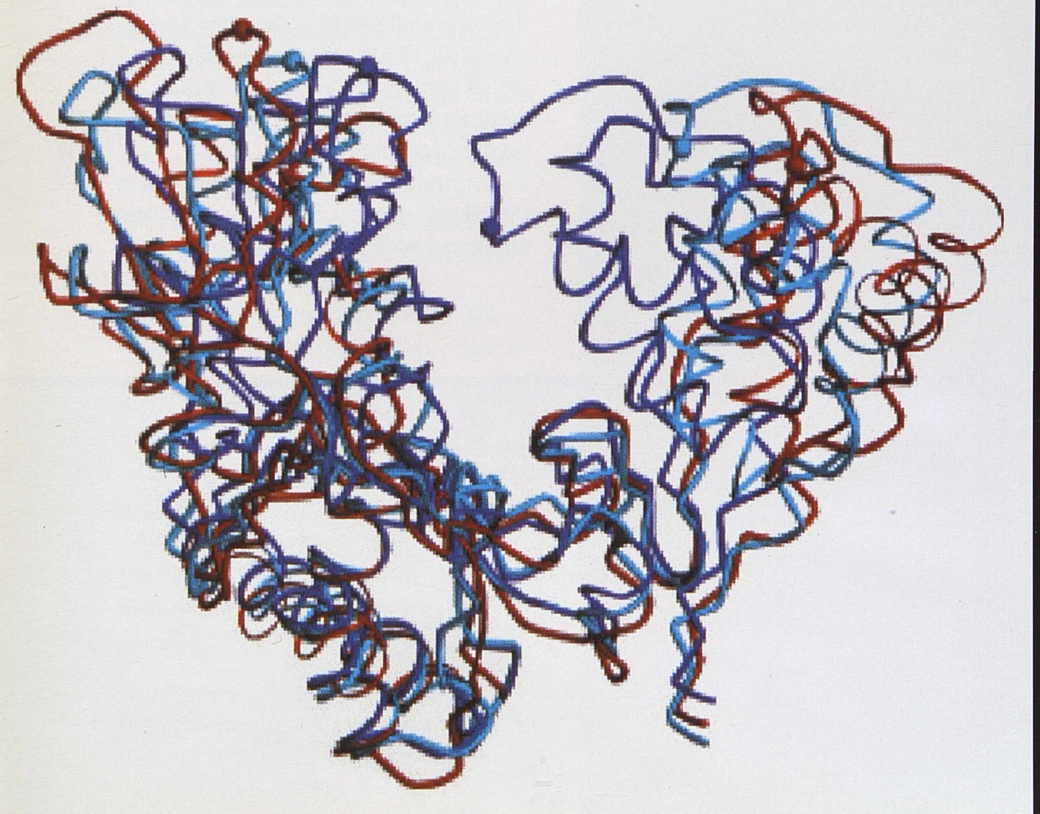
From crystallographic studies, researchers essentially had still photos of a molecular hand in two positions. What they lacked was knowledge of how the thumb moved from one position to another. "It's difficult to predict how hinging motions occur in proteins," says Arnold, "especially with a molecule of this size." Could supercomputers fill in the picture? Working with Arnold and Jianping Ding, Madrid recognized it as a daunting challenge, due to the immensity of the enzyme—more than 1,000 amino acids. But worth a try.

"It was extremely exciting," says Jacobo-Molina, "to see that the simulations backup what we saw experimentally. These calculations describe a huge movement, an entire subdomain rearrangement, for a very large enzyme. Because of the capabilities required, there are few places in the world that can carry out a simulation of this magnitude."

Perhaps most importantly, Madrid's results show that high-performance computing has a vital role to play in AIDS research. Arnold found himself pleasantly surprised that simulations could yield informative results. "My long-term view is that this kind of simulation will become more and more informative as we learn to do it better."

Madrid sees her results as a first step. A necessary next step is to include surrounding water molecules. In further work, she expects to investigate how RT moves with and without attached inhibitor drugs. "This will allow us to test hypotheses about how drugs function, and to help evaluate what approaches may be worth further research."

More information:
http://www.psc.edu/science/madrid.html



◀ *This image from the simulations shows the thumb subdomain in motion toward the fingers, which remain stationary. The structure is traced as it moves starting at one picosecond (red), 20 picoseconds (turquoise) and 30 picoseconds (blue).*

▼ *The closed-thumb structure at the end of the simulation (green) compares well with the unliganded crystal structure (red).*

All in the Family

▲ *Hugh Nicholas, PSC and John Hempel, University of Pittsburgh.*

If there were a mantra for molecular biologists, it might be this: Structure and function are related, two sides of a coin. To know how a protein does its job inside a living cell, look at how it's put together. The twists and turns of a protein's 3D shape define what biomolecules it can interact with and how.

In the 1990s, scientists are mapping and sequencing the genes of many organisms, and the resulting data corresponds to the root of protein structure: the linear sequence of amino acids, like beads on a chain, that precedes and determines 3D shape. As this data streams in, a few scientists, like Hugh Nicholas and John Hempel, are using it to open new understanding of how amino-acid sequence interrelates with 3D structure.

Hempel, a University of Pittsburgh biologist, has for 20 years focused on a family of enzymes called aldehyde dehydrogenase (ALDH). In the early 80s, he worked out the first two ALDH sequences to be solved and, in collaboration with Ron Lindahl at the University of South Dakota, has continued this work. In the late-80s, Hempel began collaborating with Nicholas, a PSC scientist who specializes in analyzing relationships among nucleic-acids (DNA and RNA) or proteins through comparison of their sequence data.

In 1997, Hempel collaborated with a University of Georgia research group led by B.C. Wang to solve the first ALDH 3D structure. This made it possible for Nicholas and Hempel to embark on an ambitious project. Beginning in September 1997, Hempel's student, John Perozich, gathered 145 full-length ALDH sequences, the complete pool of sequenced ALDHs at the time. Using the sequence-analysis facility at PSC, the researchers produced one of the largest multiple-sequence alignments achieved to date.

Nicholas and Hempel then applied techniques they developed to identify recurring sequence elements and analyze them in relation to function. They identified 10 sequence motifs, amino-acid patterns, that recur with a high degree of regularity in the 145 sequences. Their analysis of these motifs offers fresh insight into how sequence influences 3D structure.

[Research that offers promise for reducing side-effects from chemotherapy.]

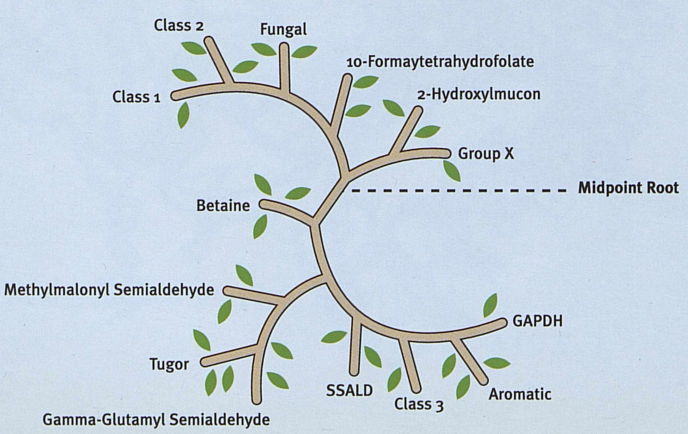
Sequence and Evolution

ALDHs have been found in nearly every form of living thing. Their primary role in humans and other mammals is protecting the body from toxic compounds called aldehydes. Research has found a number of closely related but different ALDH species, over 10 now identified in humans, a number of which are the subject of public-health research.

Hempel and Nicholas classified the 145 sequences, each more than 700 amino-acids long, into sub-families based on evolutionary adaptation. In one kind of adaptation, a gene duplicates within an organism and diverges slightly in structure to take on a modified function, such as to react with a different form of aldehyde. Nicholas and Hempel tracked this by sequence analysis, grouping the 145 ALDH sequences into 13 distinct sub-families.

The researchers also identified conserved residues—amino-acids that stay the same across sequences. They found four 100 percent conserved residues—the same amino-acid at the same position in all 145 sequences—each of which participates in binding with other molecules involved in ALDH's catalytic function.

They extended this kind of analysis to the 13 sub-families, identifying residues conserved within a particular sub-family



▶ *This phylogenetic tree plots evolutionary relations among ALDH sub-families. Each branch represents a point where a gene duplicates and evolves to a new function. Distance between branches corresponds to evolutionary distance as measured by how much the sequences differ.*

and which discriminate that group from other ALDHs. Since a number of chemotherapies work through conversion in the body to an aldehyde that attacks cancer cells, this research offers promise for developing drugs that could reduce chemotherapy side-effects. These therapies lose potency over time because the relevant ALDH increases in concentration, deactivating the aldehyde more quickly. "If you could give the patient an inhibitor for the ALDH that inactivates the chemotherapy," explains Nicholas, "you could get by with a lower dose. This work is a first step in that direction."

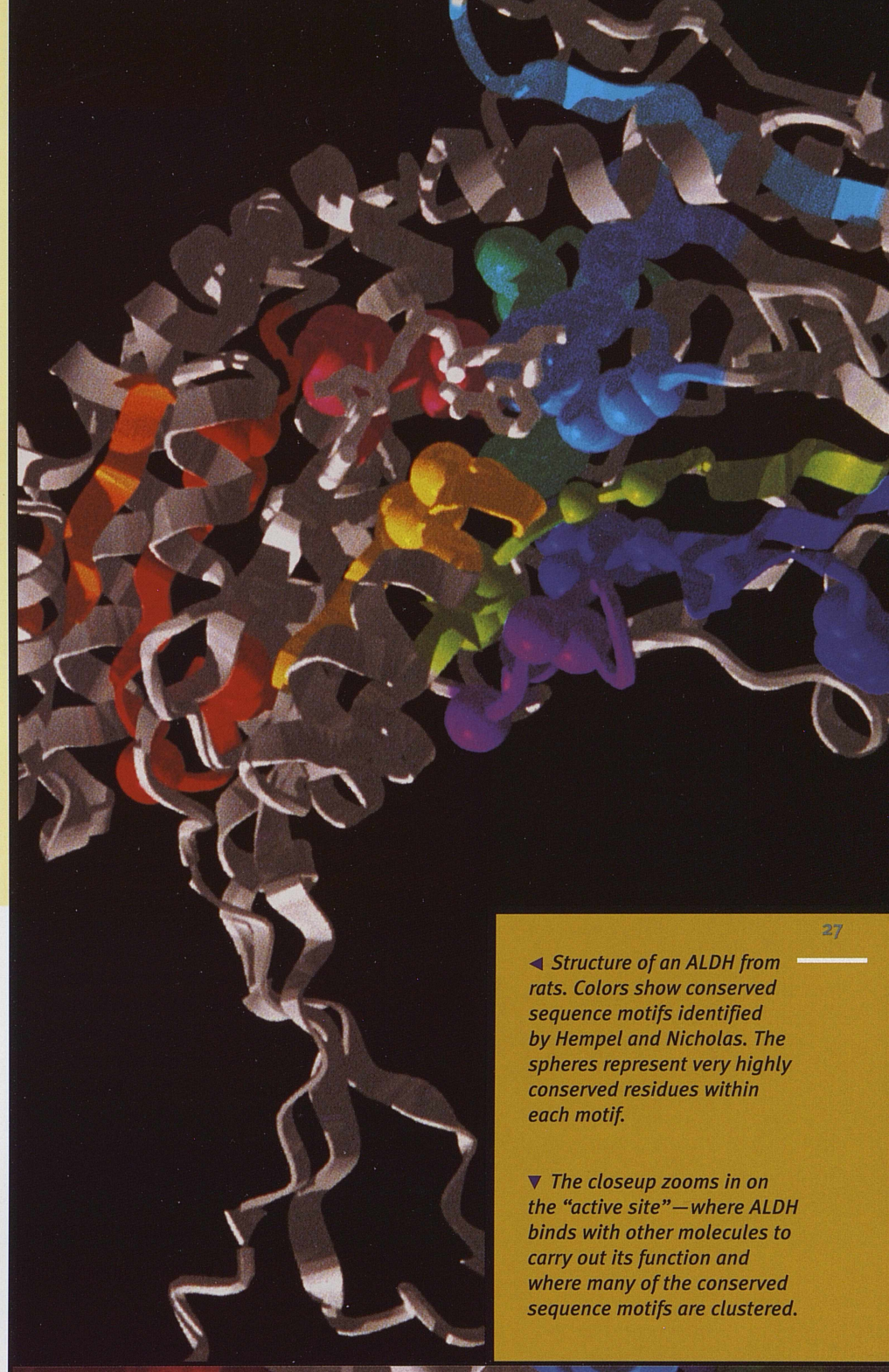
Interplay of Sequence and 3D Structure

A major finding from the analysis is identification of 10 sequence motifs that are themselves highly conserved. "These are stretches of sequence—from five amino-acids up to 14 or 15," says Nicholas. "They're spread uniformly along the entire ALDH sequence, but when you look at the 3D structure they fold back together and come into contact with each other."

This interplay between sequence and 3D structure sheds new light on traditional thinking about protein structure, which tends to focus on 3D patterns (known as secondary structure) such as helices and sheets. "Our results tells us," says Nicholas, "that it's the residues at the ends of these structures that are conserved. Perhaps you can think of a protein not as a bunch of helices, but as defined by the ends of the helices, where it turns. These may be two equally informative ways of thinking, and the genomics data suggests we should look at the turns more than we have."

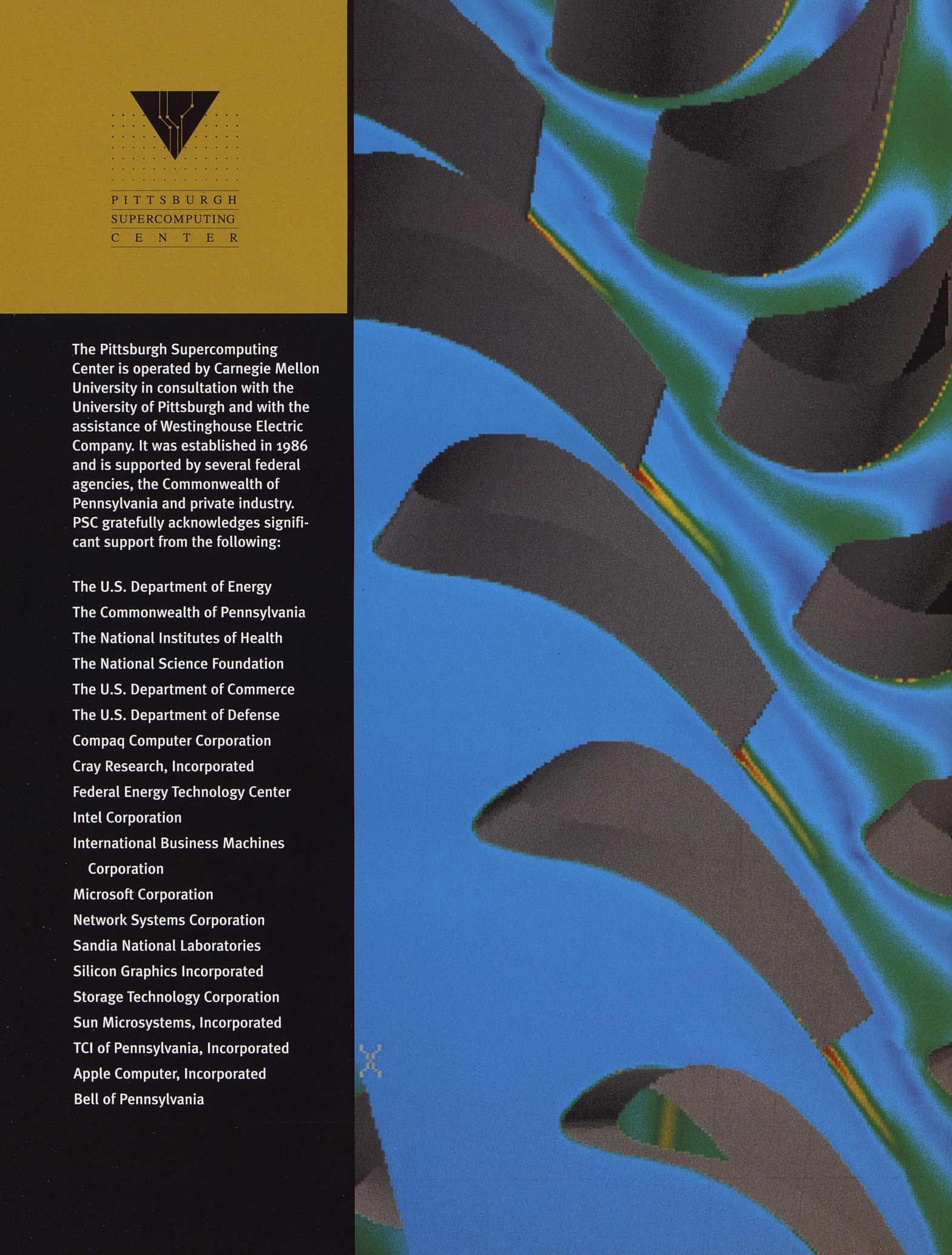
As a goal for this kind of work, the researchers would like to build a sophisticated description based on conserved residues that identifies sequence elements that can predict 3D structure, not only in ALDH but in other protein families. "This provides a model for other systems," says Hempel. "We're developing tools needed to rigorously analyze and determine highly conserved regions and to correlate them with structure."

More information:
http://www.psc.edu/science/hempel.html



◀ *Structure of an ALDH from rats. Colors show conserved sequence motifs identified by Hempel and Nicholas. The spheres represent very highly conserved residues within each motif.*

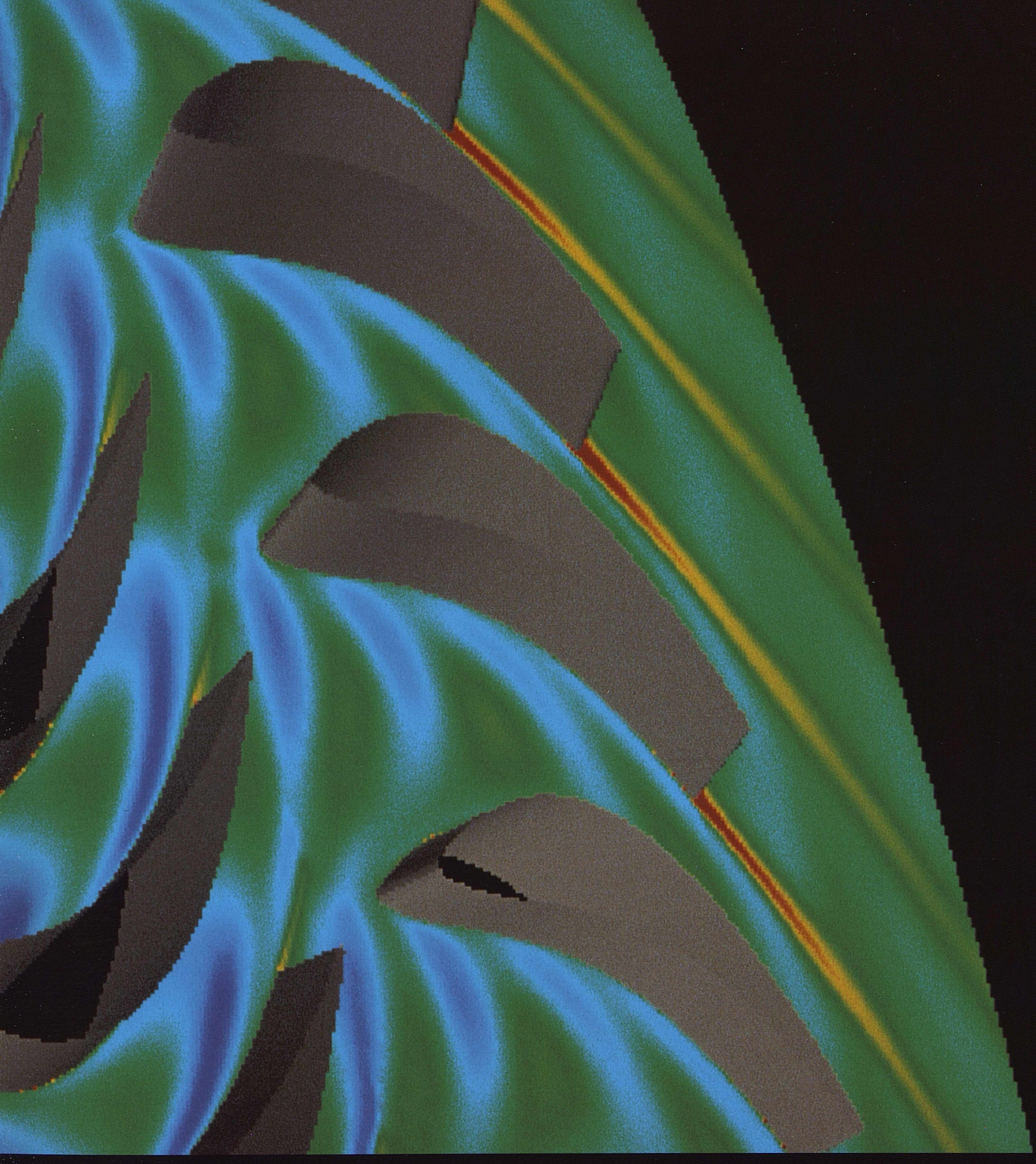
▼ *The closeup zooms in on the "active site"—where ALDH binds with other molecules to carry out its function and where many of the conserved sequence motifs are clustered.*



PITTSBURGH
SUPERCOMPUTING
CENTER

The Pittsburgh Supercomputing Center is operated by Carnegie Mellon University in consultation with the University of Pittsburgh and with the assistance of 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 U.S. Department of Energy
The Commonwealth of Pennsylvania
The National Institutes of Health
The National Science Foundation
The U.S. Department of Commerce
The U.S. Department of Defense
Compaq Computer Corporation
Cray Research, Incorporated
Federal Energy Technology Center
Intel Corporation
International Business Machines Corporation
Microsoft Corporation
Network Systems Corporation
Sandia National Laboratories
Silicon Graphics Incorporated
Storage Technology Corporation
Sun Microsystems, Incorporated
TCI of Pennsylvania, Incorporated
Apple Computer, Incorporated
Bell of Pennsylvania



Credits

Editor/Writer: *Michael Schneider, PSC*

Graphics research, photography direction, copy editing: *Sean Fulton, PSC*

Photography: *Graphic Services at Mellon Institute, Jim Schafer*

Design: *John Brady Design Consultants, Inc. www.bradydesign.com*

Printing: *J.B. Kreider Company Inc.*

Printed on recycled paper.

Pittsburgh Supercomputing Center
Mellon Institute Building
4400 Fifth Avenue
Pittsburgh, PA 15213

Nonprofit Org.
U.S. Postage
PAID
Pittsburgh, PA
Permit No. 251