

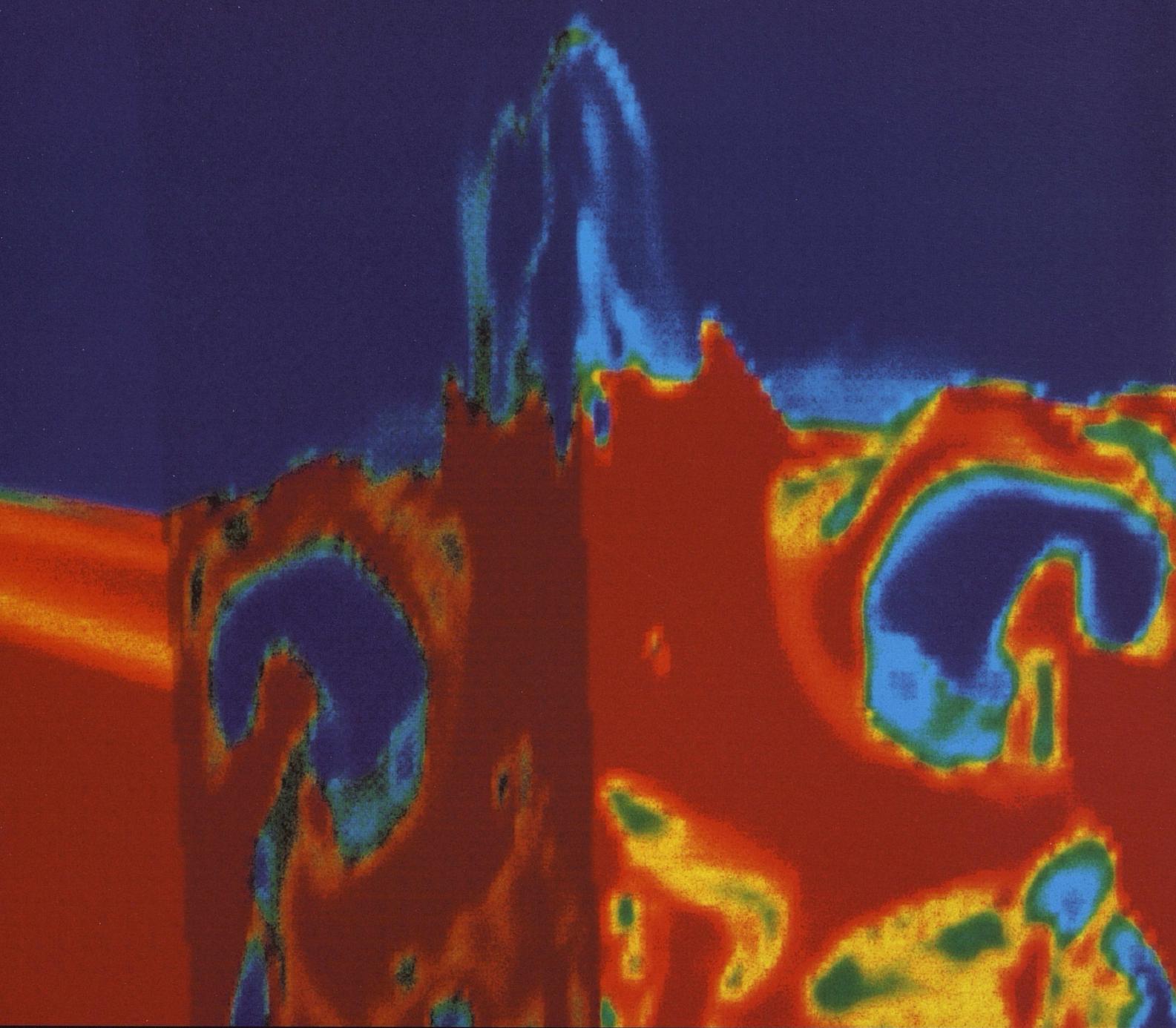
2 0 0 0

Projects in Scientific Computing

PITTSBURGH SUPERCOMPUTING CENTER

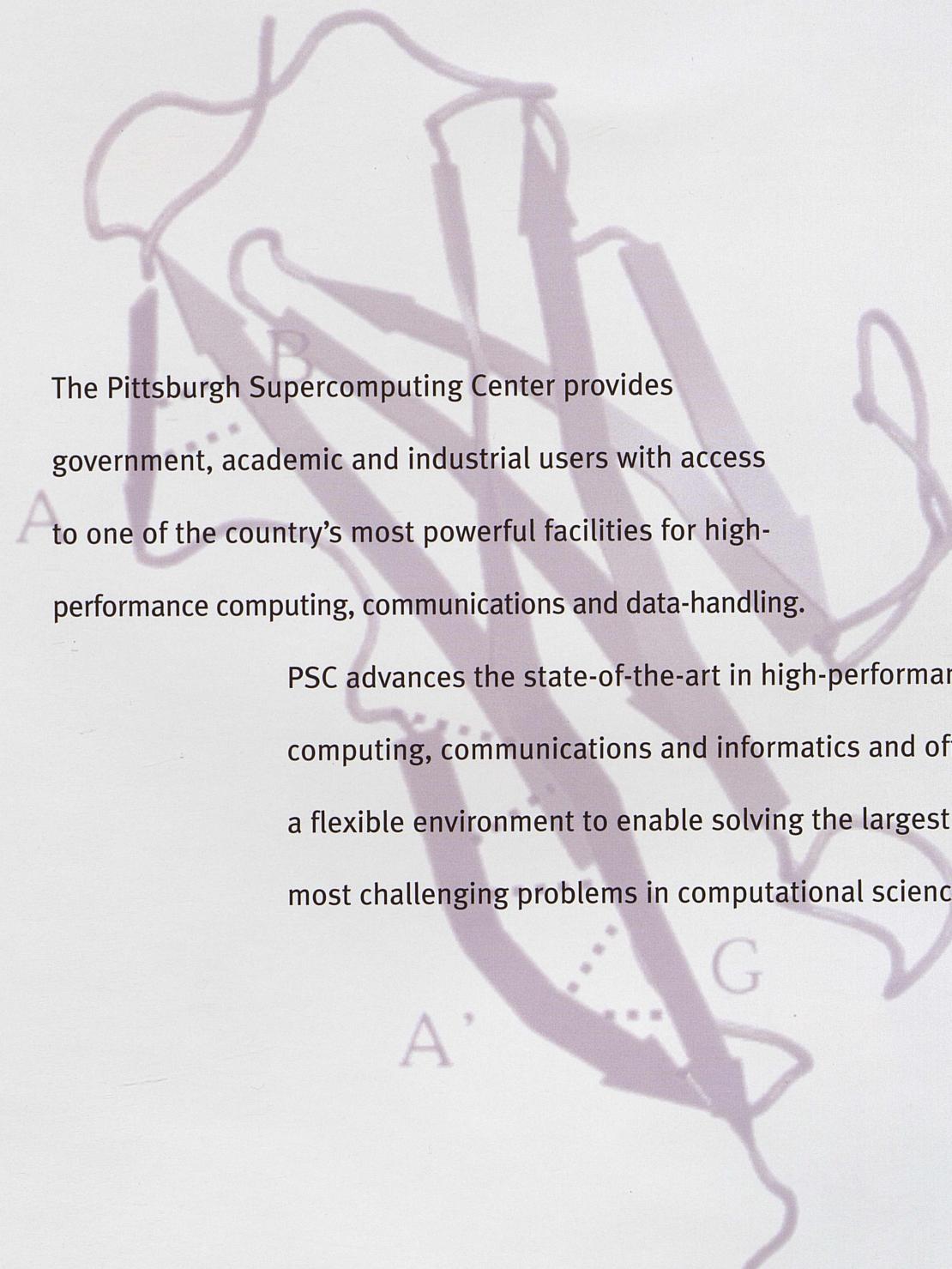


TERASCALE



Contents

<i>Foreword from the Directors</i>	2	<i>Supercomputing in Pennsylvania</i>	12	PATHOLOGY	
<i>Pittsburgh Supercomputing Center, 2000</i>		<i>The Super Computing Science Consortium</i>	14	<i>An Objective View of Cancer</i>	26
<i>Big Iron in the Steel City</i>	4	<i>Projects 2000: Contents</i>	16	AT THE FRONTIER OF PHYSICS AND CHEMISTRY	
<i>Terascale Computing in the PACI Program</i>	6	MANUFACTURING PROCESS DESIGN		<i>Getting the Jump on Superacids</i>	28
<i>by Robert R. Borchers, National Science Foundation</i>		<i>Clean Power</i>	18	EVOLUTION AND STRUCTURE OF THE UNIVERSE	
<i>Research Notes & Highlights</i>	8	STRUCTURE OF PROTEINS & DNA		<i>The Dance of Two Black Holes</i>	30
		<i>The Rude Mechanicals</i>	22		
		<i>Getting a Grip on AIDS</i>	24		



The Pittsburgh Supercomputing Center provides government, academic and industrial users with access to one of the country's most powerful facilities for high-performance computing, communications and data-handling.

PSC advances the state-of-the-art in high-performance computing, communications and informatics and offers a flexible environment to enable solving the largest and most challenging problems in computational science.

Foreword from the Directors

This past year has been one of remarkable achievement for the Pittsburgh Supercomputing Center. The National Science Foundation chose PSC as the site of what will become the world's most powerful unclassified supercomputer, to be built in collaboration with Compaq Corporation (p. 4). The National Institutes of Health renewed, for five more years, funding for PSC's biomedical Research Resource (p. 8). The Department of Energy's newest national laboratory, the National Energy Technology Laboratory (NETL), based in Pittsburgh and Morgantown, West Virginia, expanded its partnership with PSC and regional universities to speed the development of next generation turbines (p. 12). The National Science Foundation also awarded PSC's networking group a major grant to improve the efficacy of computer networks (p. 10). These awards are all testimony to the talents of PSC's staff.

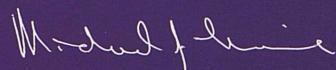
This booklet marks the 15th year in which PSC has documented how it advances science and engineering in the nation through the application of capability computing, large-scale data processing, and high-performance

networking. This past year PSC has also increasingly applied its technological capabilities to regional economic development, through its work with NETL, and its work with Pennsylvania corporations. In the HUBS project (p. 10), PSC has begun a partnership with the University of Pittsburgh Medical School, Johns Hopkins University in Maryland, the University of Medicine and Dentistry in New Jersey, and the University of Pennsylvania in Philadelphia to apply high-performance computing and communication to improving health care. PSC extended its outreach by working with Carnegie Mellon University and the Carnegie Science Center to develop a museum show on the Human Brain which has been seen by thousands of school children and their parents.

The examples presented in this year's booklet demonstrate the variety of PSC's activity: Research on muscle proteins (p. 22), on HIV (p. 24) and on pathology (p. 26) illustrate biomedical research supported by the National Institutes of Health's National Center for Research Resources. Research on advanced turbine simulation (p. 18) and fluidized-bed combustion illustrate the contributions of PSC to NETL's mission, both in

large-scale computing and in advanced communication. Reports in advanced areas of chemistry and physics—superacids (p. 28) and general relativity (p. 30)—illustrate how support from the Commonwealth of Pennsylvania has advanced the research of Pennsylvania scientists, and improved their ability to compete for national funds.

PSC is striding into the new millennium with a new sense of mission and conviction that its efforts are vital to transforming science and engineering. And the best is yet to come.



Michael J. Levine, *scientific director*



Ralph Z. Roskies, *scientific director*



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

[2 0 0 0]

Pittsburgh Supercomputing Center



Big Iron in the Steel City

On August 3, the National Science Foundation awarded \$45 million for PSC to provide "terascale" computing capability for U.S. researchers in all science and engineering disciplines. PSC will collaborate with Compaq Computer Corporation to create a new, extremely powerful system for the use of scientists and engineers nationwide.

Terascale refers to computational power beyond a "teraflop"—a trillion calculations per second. While several terascale systems have been developed for classified research at national laboratories, the PSC system will be the most powerful to date designed as an open resource for scientists attacking a wide range of problems. In this respect, it fills a gap in U.S. research capability—highlighted in a 1999 report to President

Clinton—and will facilitate progress in many areas: novel computer-science applications such as tele-immersion and areas of significant social impact, such as the structure and dynamics of proteins useful in drug design, storm-scale weather forecasting, and modeling of earthquakes and global climate change.

The Terascale Computing System

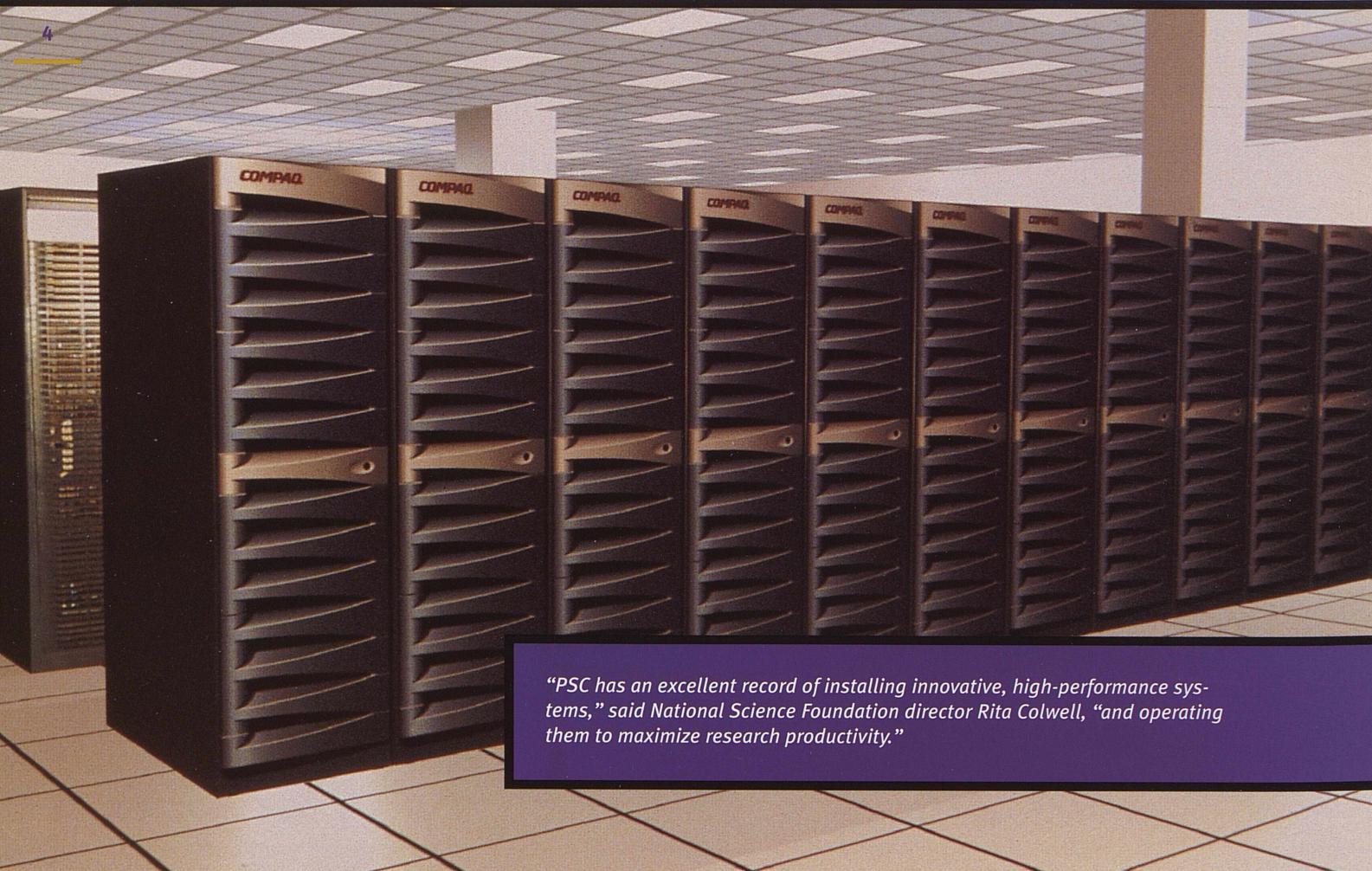
The three-year award, effective Oct. 1, is based on PSC's proposal to provide a system, installed and available for use in 2001, with peak performance of six teraflops. PSC and Compaq will collaborate to develop the system, which although based on existing or soon to be available components is unprecedented in scale. Its design is optimized to the computational requirements

An unprecedented system will be the most powerful in the world available for public research.

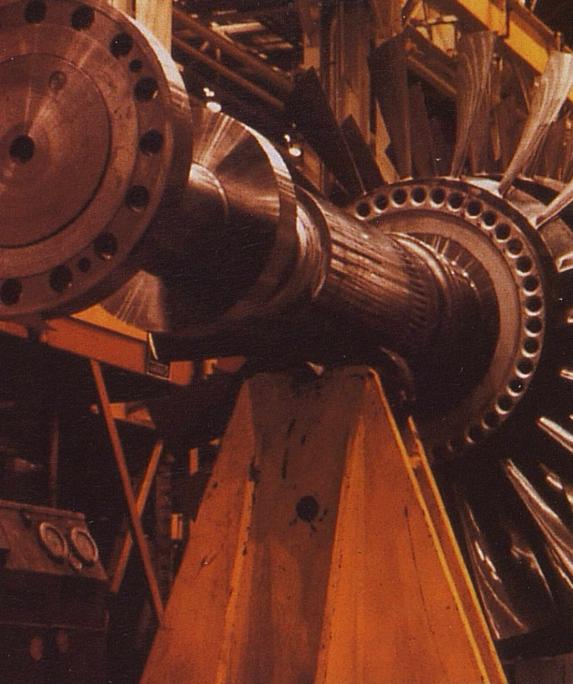
posed by a wide range of research applications, and it pushes beyond simple evolution of existing technology.

The brain of the six teraflop system will be a network of Compaq AlphaServers, 682 of them, each of which itself contains four Compaq Alpha microprocessors. Existing terascale systems rely on other processors, but extensive testing by PSC and others indicates that the Alpha processor offers superior performance over a range of applications.

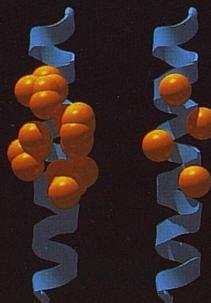
Along with processing power, the terascale system will feature 2.7 terabytes of memory, high-bandwidth, low-latency Quadrics Supercomputer World interconnects among AlphaServers, and remarkable capabilities for large-scale data handling, including the ability to write memory to disk in three minutes,



"PSC has an excellent record of installing innovative, high-performance systems," said National Science Foundation director Rita Colwell, "and operating them to maximize research productivity."



682 compute nodes
2,728 Alpha processors
Peak performance: 6 TFlops
Memory: 2.7 TBytes
Disk: over 50 TBytes
MPI latency: ~5 microseconds
File-server bandwidth: ~18 GBytes/sec.



"I am pleased that the National Science Foundation is expanding its investment in supercomputing systems capable of making trillions of calculations per second," said President Clinton. "This investment will accelerate the pace of discovery in science and engineering—allowing us to better predict tornadoes, speed up the discovery of life-saving drugs and design more fuel-efficient engines."

and to write a terabyte per hour to tape. The system will also employ a tightly coupled visualization system. With these architectural features, the PSC terascale system will promote innovative applications in many areas.

The terascale system will draw on a history of collaboration between PSC and Compaq, and PSC and the computer-science and computational-science community. It represents an extension of PSC's success at installing new systems—resolving the myriad of unanticipated hardware and software glitches that come up—and turning them over rapidly to the scientific community as productive research tools. In fall 2000, PSC will install an initial system with a peak performance of 0.4 teraflops. The six teraflop system, which will use faster Compaq Alpha microprocessors not yet available, will evolve from this.

Technology in the Pittsburgh Region

The terascale system, including software and networking, highlights a significant strength of PSC: its tri-partite affiliation with Westinghouse, Carnegie Mellon, the University of Pittsburgh and the pooled computing-related expertise of faculty and staff at both universities.

"This award, which comes as the culmination of a national competition, recognizes PSC's leadership in high-performance computing and communications," said Jared L. Cohon, president of Carnegie Mellon. "And it provides another key building block for our region's technology future, enhancing our international stature in the development and application of advanced computing technology."

"A gap exists between the computing resources available to the classified world and the open scientific community," said Mark Nordenberg, chancellor of the University of Pittsburgh. "It is ideal that PSC, a world leader in acquiring

and deploying early the most powerful computers for science and engineering, can contribute to filling this gap. This award also demonstrates the unique scientific strengths that exist in Pittsburgh when its major research universities partner with each other and with leaders in industry."

"Today's terascale award is one more in a long list of PSC's major achievements," said Charlie Pryor, president and CEO of Westinghouse Electric Company. "Westinghouse is proud of PSC's contribution to the nation's scientific community and is pleased to have been associated with PSC since its inception."

University of Pittsburgh

**[pushing beyond simple
evolution of existing technology]**

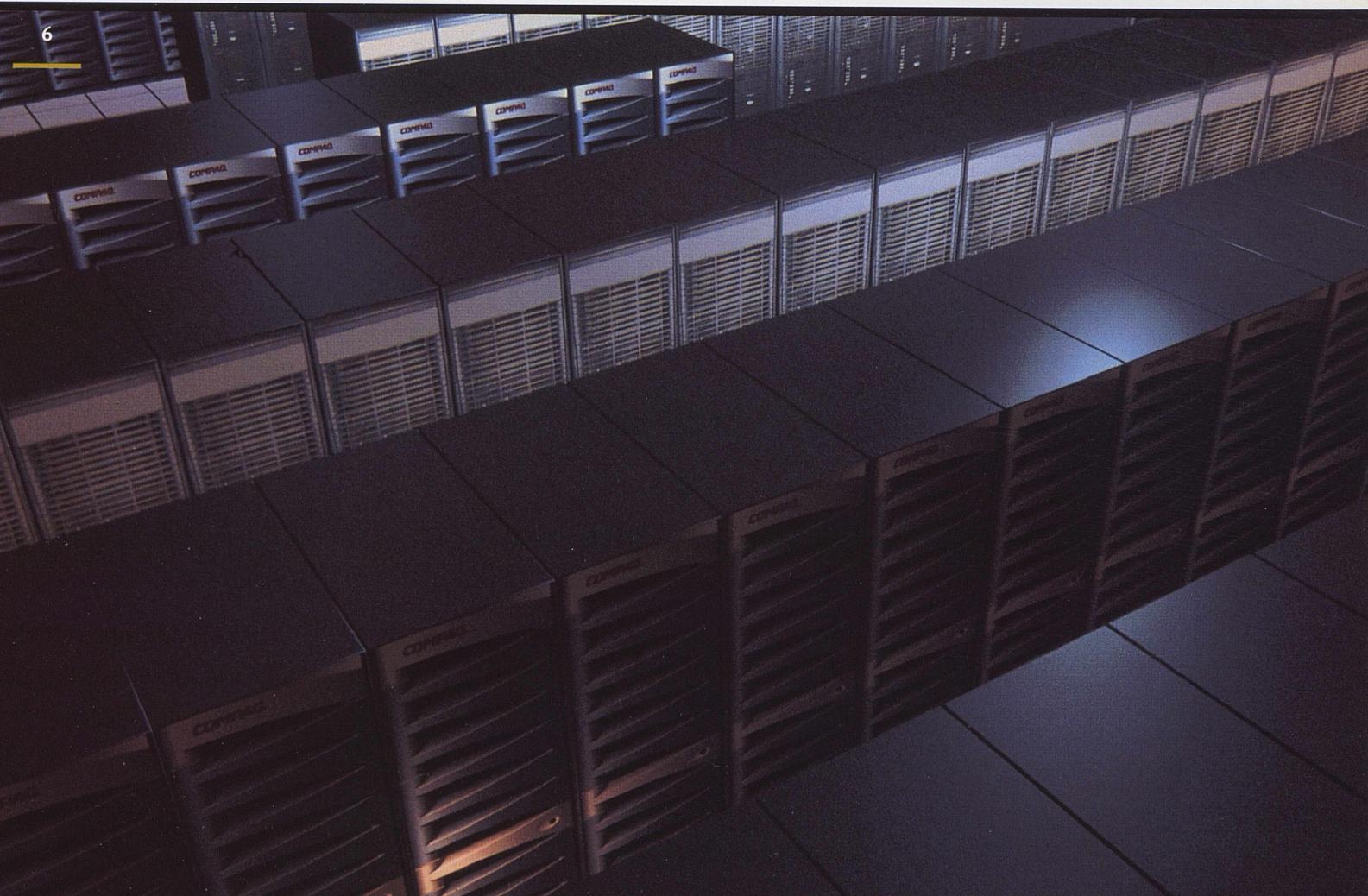
Investing in Ideas

Terascale Computing in the PACI Program

by Robert R. Borchers, Director
Division of Advanced Computational
Infrastructure and Research
National Science Foundation

The Terascale Computing System continues a history of National Science Foundation support for high-performance computing that began with the Supercomputer Centers program established in 1985. The Partnerships for Advanced Computational Infrastructure (PACI) program replaced this program in FY1998. PACI adds emphasis on the coupling of computational and computer science in order to more effectively exploit the emerging capabilities of scalable-parallel systems, high-performance networking and high-bandwidth, large-capacity mass-storage systems.

Within this environment, simulation and modeling for a vast array of scientific and engineering problems have led to truly revolutionary insights over the last decade, and there is every sign that progress is accelerating. Due to increased computer capability, computational science is experiencing a revolution in its ability to solve new research problems. The recent demonstration of computers with speeds of a teraflop (10^{12} floating-point operations per second) or more has directed attention to important fundamental science and engineering problems which are not amenable to solution with current systems, but would be accessible to terascale range computation.



The Planning Process

In early 1998, NSF began planning to upgrade the availability of very high end computing resources for the academic science and engineering research community. As the NSF strategic plan states, part of our mission is investing in *Ideas* that provide a deep and broad fundamental science and engineering knowledge base. A principal means of enabling those *Ideas* is to make available *Tools* for wide access to state-of-the-art science and engineering infrastructure, e.g., support for user facilities in many fields. One of the most widely used research tools that has emerged over the past decade is high-end computational capability; this broad use of computers by many different areas is referred to as "computational science." Consequently, such activities in computer and information science and engineering have received high priority in NSF planning and budget development.

A workshop at NSF in May 1998 (one of a series that examined various questions relating to terascale computing) identified numerous important computational applications that could take advantage of significantly increased computing power. This is true among traditional users of computing in physics, chemistry (the 1998 Nobel Prize was for computational chemistry), geosciences and engineering, as well as in disciplines such as biology, where computing is still emerging as a critical

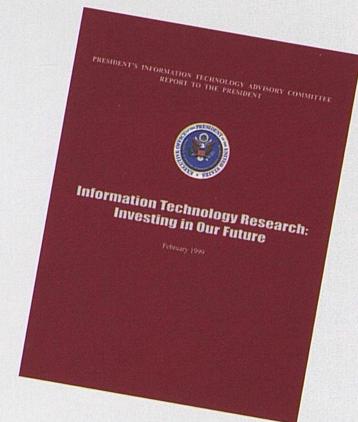
new tool. New application domains such as economics, sociology, and even history are ripe for exploration. These points were reinforced at a joint NSF/DOE National Workshop on Advanced Scientific Computation hosted at the National Academy of Sciences in July 1998.

The PITAC Report

The President's Information Technology Advisory Committee (PITAC) also considered High-End Computing during its deliberations about the current state and future directions of the Nation's Information Technology. PITAC's final report, *Information Technology Research: Investing in Our Future*, released on Feb. 24, 1999, found that:

"high-end computing is essential to science and engineering research"

"both for the sake of fundamental scientific research and to enable applications to benefit from the research, the research community needs access to systems at the leading edge of capability."



See <http://www.ccic.gov/ac/report/>

The report goes on to state that:

"If the United States is to continue as the world leader in basic research, its scientists and engineers must have access to the most powerful computers. Therefore, the Committee recommends that the Federal government continue to provide these computing systems to the research community through major, shared-facility centers. To increase long-term, fundamental research across all science and engineering disciplines, the first priority should be to increase the computing capacity of the centers that can best serve the entire research community."

"These high-end systems will also create an environment for new applications and will expand the critical, domestic high end computing market."

At the National Science Foundation, we expect to see fascinating new science and engineering as a result of the TCS I machine at PSC. The system will be fully integrated as a leading-edge computing resource in the fabric of the PACI program, and we have every confidence that PSC will splendidly fulfill its traditional role of supporting the very high-end users.



[If the United States is to continue as the world leader in basic research, its scientists and engineers must have access to the most powerful computers.]

Research Notes and Highlights

Biomedical Supercomputing

The National Institutes of Health awarded PSC \$8.6 million this year to renew its program in biomedical supercomputing. The grant, which covers five years, supports biomedical research by PSC scientists and PSC collaborations at Scripps Research Institute and Carnegie Mellon University. It also supports PSC research on collaborative tools for biomedical

research and PSC's program to provide computational resources, consulting and training for biomedical researchers around the country.

"For over 12 years, the Pittsburgh Supercomputing Center has provided national leadership in applying advanced computational resources to biomedical research," said Michael Marron, associate director for biomedical technology at

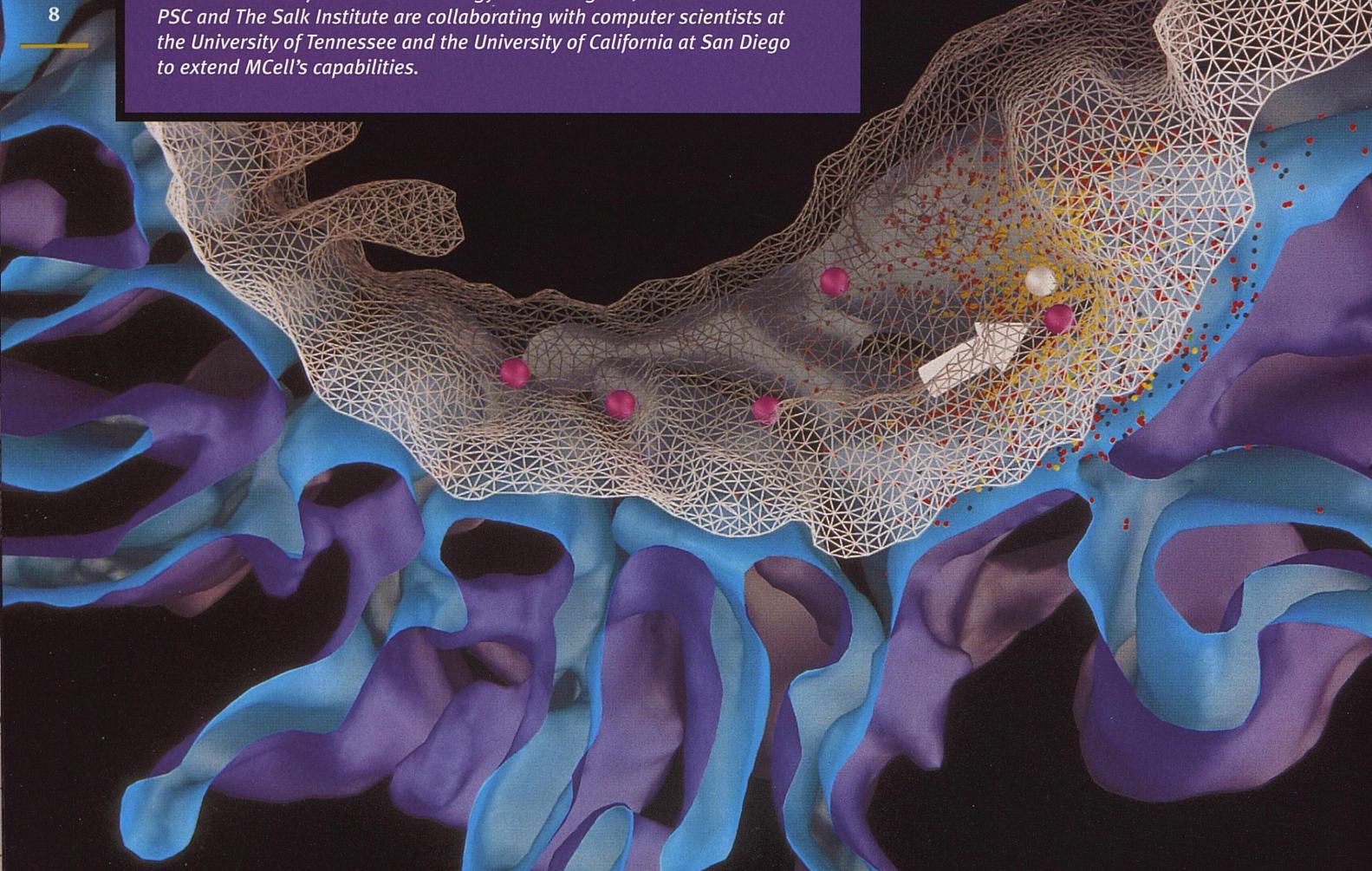
NIH's National Center for Research Resources. "This grant is part of NCRR's ongoing commitment to bring together leading-edge computational resources and experts in computing with experts in biology and medicine to solve some of the important problems in biomedicine facing the nation."

"Our training activities reach hundreds of biomedical researchers each

It Takes Nerve to Flex Your Muscles

Through the PSC biomedical program, researchers are expanding sophisticated software called MCell that simulates chemical reactions and diffusion of molecules within, around and between cells. This MCell visualization depicts the junction, called a synapse, between a nerve-cell membrane (white mesh) and a muscle-cell (folded surface). Neurotransmitter molecules released from sacs (magenta and white) diffuse through this space and activate receptor proteins in the muscle to trigger movement. The arrow indicates the direction of the magnified view of receptors in the muscle.

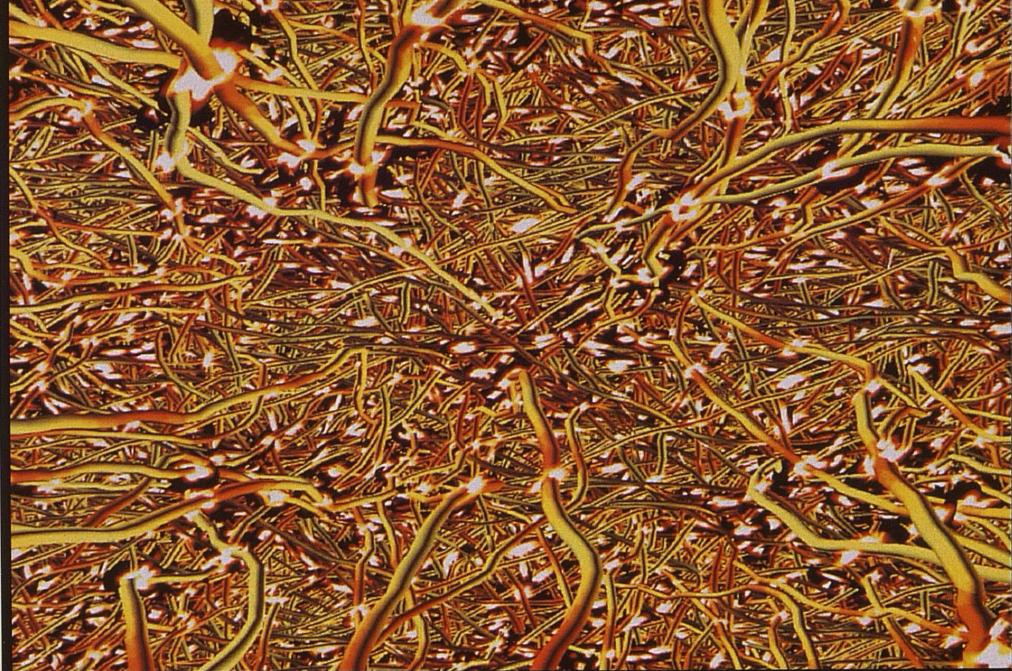
With an NSF Information Technology Research grant, researchers at PSC and The Salk Institute are collaborating with computer scientists at the University of Tennessee and the University of California at San Diego to extend MCell's capabilities.



year," said biochemist David Deerfield, who directs the PSC biomedical program. "Techniques we've developed are helping scientists nationwide cope with the explosion of genome data. The bottom line is that a great deal of important biomedical work over the last decade wouldn't have been done without NIH support for this program."

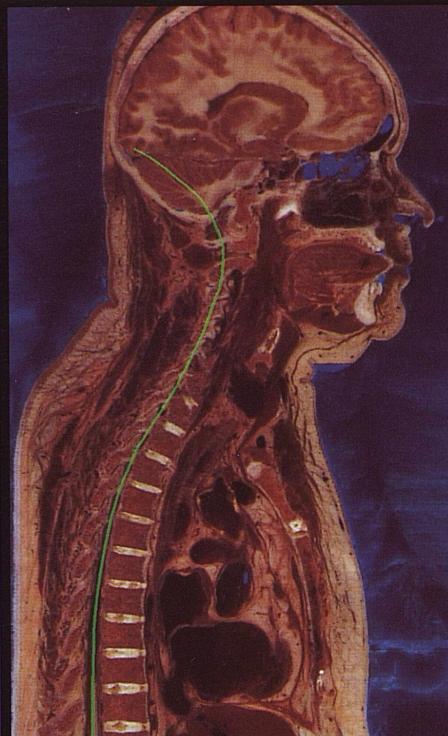
Since its inception in 1987, PSC's biomedical program has provided access to computing resources for more than 800 biomedical research projects involving nearly 1,800 researchers in 43 states and the District of Columbia. The center's workshops on computational biology have trained more than 2,000 researchers in the use of high-performance computing for biomedical research in such areas as sequence analysis in genome research, the structure of proteins and DNA, and biological fluid dynamics.

More information on biomedical supercomputing at PSC:
<http://www.psc.edu/biomed/biomed.html>



Gray Matters: the Museum Show
Through this forest of axons, a volley of spark-like flashes (neural impulses) travel deeper into the cortex where the axons branch to form a dense web. PSC scientists produced this animated view of brain activity for "Gray Matters," a planetarium show that debuted in fall 1999 at the Carnegie Science

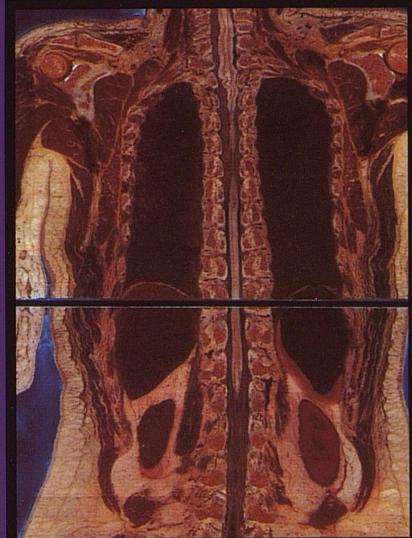
Center in Pittsburgh. With the brain's neural structure projected on the 3D space of the planetarium dome, Gray Matters offers an interactive, multi-media lesson in the science of the brain for children and adults. PSC collaborated on Gray Matters with Carnegie Mellon's Studio for Creative Inquiry and the Center for the Neural Basis of Cognition.



The Visible Human

In collaboration with University of Michigan researchers, PSC is developing technologies for high-speed networking and visualization for the National Library of Medicine's Visible Human data. This project enhances anatomy training through innovative, interactive viewing.

Along with allowing retrieval of anatomical imagery, such as the torso side-view shown here, the PSC team is developing the capability to select views arbitrarily, which are then constructed in real time. In this case, a user selects a curved pathway through the center of the spinal cord (left) and the interface constructs a corresponding warped-plane frontal cross-section (right), providing an extremely useful view that's otherwise unavailable.



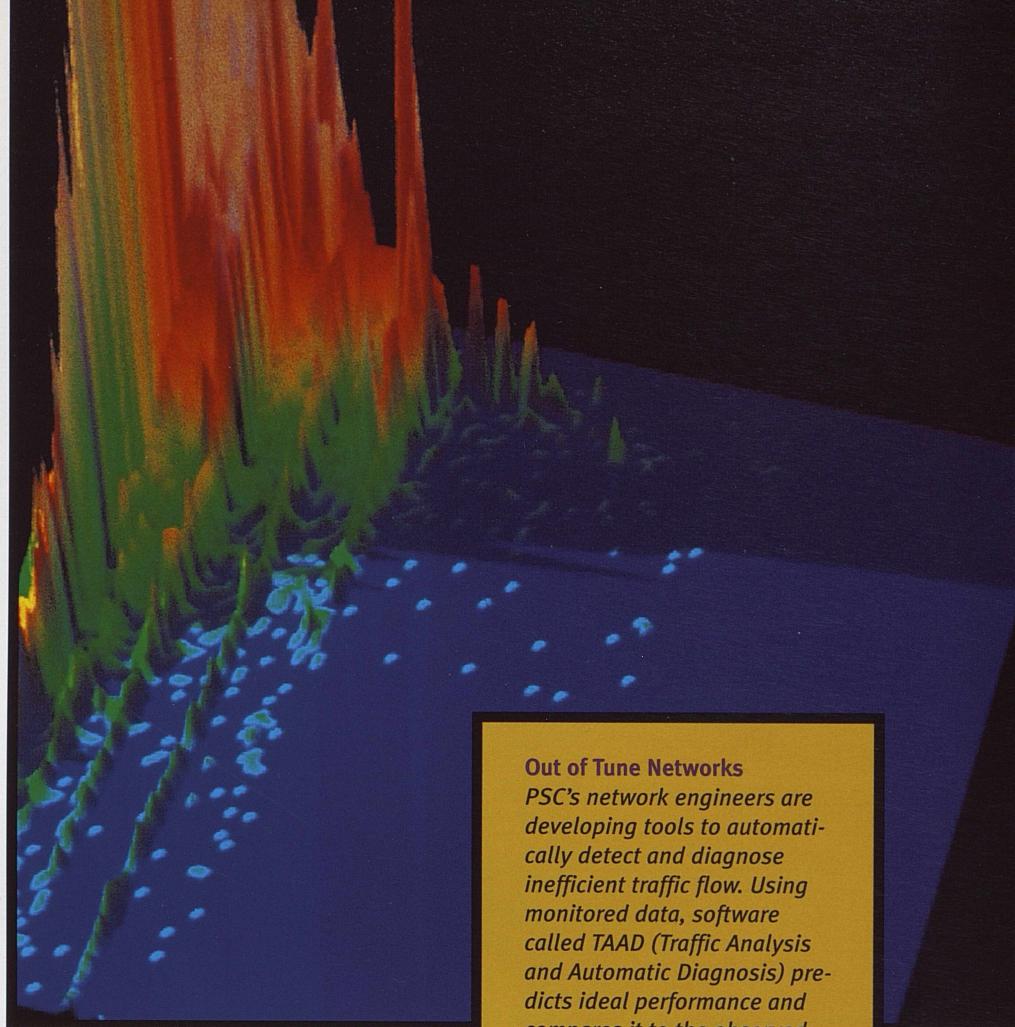
Networking the Future

Through its consulting, training and research initiatives, PSC's network resource group plays a leadership role nationally in network technologies. As the engineering services component of the National Laboratory for Advanced Network Research, PSC staff consult with and provide workshops for network engineers at more than 100 universities and research centers. These institutions are linked to high-performance research and education networks, such as the very-high performance Backbone Network Service (vBNS) and Abilene, which transfer data more than 10 times faster than the current Internet.

This year, through the Web100 project, the PSC network group received a three-year \$2.9 million award from the National Science Foundation. Cisco Systems also provided a \$100,000 gift to PSC as further support for Web100. The objective is to develop software that will transparently "tune" network performance. While high-performance networks have developed in bandwidth and accessibility, it's difficult for users to attain optimal performance without expert help. Web100—a joint effort among PSC, the National Center for Atmospheric Research and the National Center for Supercomputing Applications—is developing software tools to interact with network protocols and user applications and tune them to the network without user intervention.

The PSC group also operates the Pittsburgh gigaPoP, a high-performance network interconnect for western Pennsylvania and surrounding regions. The gigaPoP provides a 100 megabit per second link to the Internet and separate 155 Mbps links to vBNS, Abilene and other advanced networks. Universities connected through the gigaPoP include Carnegie Mellon, the University of Pittsburgh, Penn State and West Virginia University.

More information:
<http://www.ncne.nlanr.net>



Infrastructure for Telemedicine

Through a \$7 million grant from the Defense Advanced Research Projects Agency in April, PSC is collaborating with hospitals in a four-state region (Delaware, Maryland, Pennsylvania, New Jersey) to develop technologies for networking and archiving of medical-image data. The effort is a pilot project to create technological infrastructure for real-time consultation on patient diagnosis and care among specialists at diverse locations, and to promote regional collaborative research and health education.

The grant supports PSC and researchers at the University of Pittsburgh School of Medicine (UPSM) to collaborate with three other hospitals in a "smart region" consortium called HUBS (Hospitals, Universities, Businesses and Schools). The effort includes several overlapping projects aimed at creating enabling technologies for secure, reliable and fast "telemedicine."

Out of Tune Networks

PSC's network engineers are developing tools to automatically detect and diagnose inefficient traffic flow. Using monitored data, software called TAAD (Traffic Analysis and Automatic Diagnosis) predicts ideal performance and compares it to the observed transfer rate. For a well tuned network, this ratio—called the gain ratio—should equal one. This graphic shows a poorly tuned network, with the highest frequency of well tuned transmissions (spikes) clustered at low transfer rates.

PSC's superb resources, including hundreds of terabytes of data-storage and high-performance network connectivity, are a keystone of the HUBS initiative. "Our staff scientists and engineers bring unparalleled expertise in the application of these resources," said PSC scientific directors Michael Levine and Ralph Roskies. "This project will improve quality of life in the four-state region and set the pace for similar efforts across the nation."

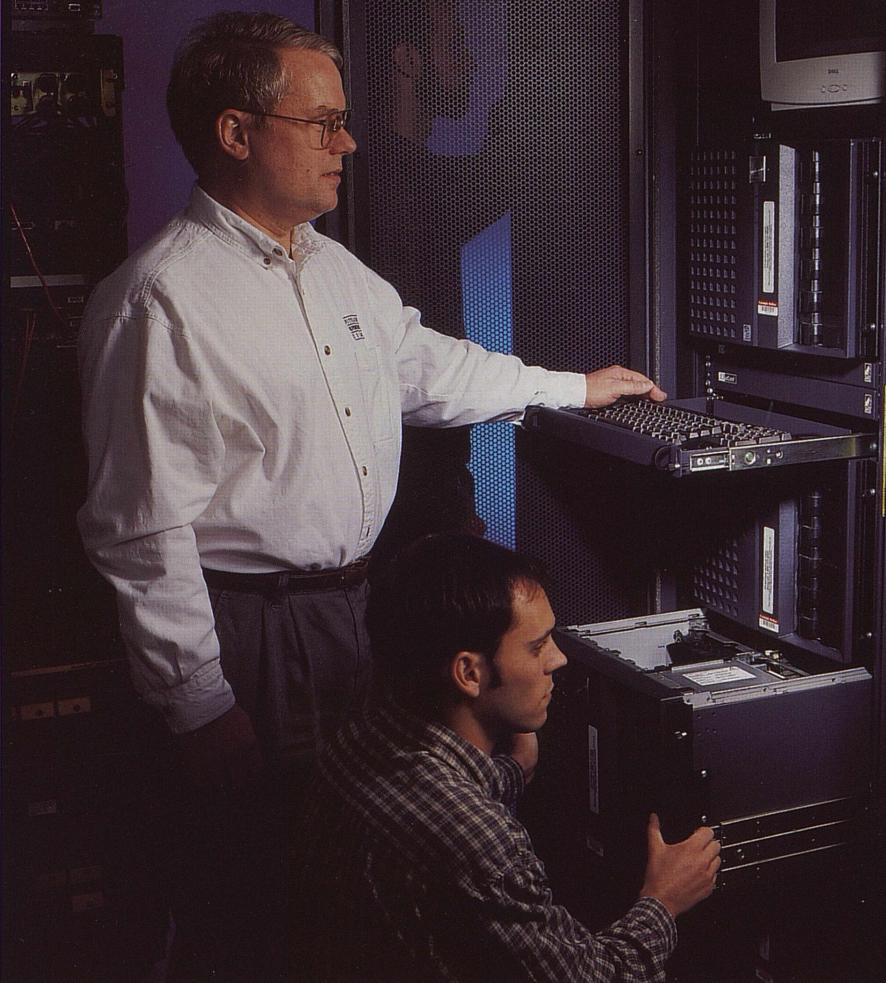
Overall project management is provided by Science Applications International Corporation through their Valley Forge, Pa. facilities. Other



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

partners along with PSC and UPSM are Telcordia Technologies (formerly Bellcore), Johns Hopkins University School of Medicine, the University of Medicine and Dentistry of New Jersey and the Hospital of the University of Pennsylvania.

In collaboration with Telcordia, PSC networking staff is developing an innovative, high-performance implementation of a technology known as a "virtual private network." The objective is to assure integrity and security of patient data at high levels of network performance. In another effort, called "intelligent archiving," PSC is collaborating with UPSM and Johns Hopkins to create network software that facilitates archival and retrieval of x-rays, PET, MRI and other image data. In a third project (see p. 26), PSC and UPSM are developing search techniques for an archive of pathology images.



The PSC Dual Boot Cluster

In August, the PSC Cluster upgraded from 10 to 20 four-processor Intel SMP (symmetric multi-processor) machines, providing an aggregate of 80 Intel PII 400 MHz processors. Uniquely configured as a "dual boot" system, the PSC Cluster runs either the Linux or NT operating system and switches easily between them. The batch scheduler reboots the operating system as necessary.

"The cluster is a production high-performance computing resource," says J. Ray Scott, PSC assistant director of systems and operations, "not merely backup or load-balancing for other PSC systems." To date, researchers have used the PSC Cluster for a range of computations, including quantum studies of magnetism, computational fluid dynamics, and molecular dynamics.

11



A workshop in progress at the PSC Computer Training Center.

Pittsburgh Supercomputing Center Workshops (1999-2000)

- Parallel Processing on Cray Massively Parallel Systems
- Nucleic Acid and Protein Sequence Analysis
- UNIX Clustering: Building Your Own Parallel Computing Facility
- Methods and Applications of Molecular Dynamics of Biopolymers
- Single Particle Reconstruction from Electron Microscope Images
- Building Computing Clusters for Biomedical Research

Supercomputing 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 use high-performance computing to enhance their competitiveness.

Workforce Development

PSC programs help to develop and retain a technology-ready workforce. 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.

Faculty at Pennsylvania universities offer courses in aspects of computing and computational science using PSC facilities. During the past year, nine instructors at seven higher-learning institutions offered such courses to over 300 students.

PSC educational enrichment programs introduce advanced educational technologies in Pennsylvania high-school

classrooms. PSC learning tours and briefings also acquaint students with advanced computational technology. This past July, for instance, a group of 25 students from several Pittsburgh area high schools toured PSC as part of a "Legislator for a Day" program sponsored by Pa. Rep. Jane Clare Orie.

PSC also provides training to Pennsylvania students through undergraduate internships. Since 1986, over 400 students have interned at PSC, and many have gone on to find jobs in Pennsylvania. During the past year, PSC employed 43 students from Carnegie Mellon, the University of Pittsburgh, Indiana University of Pennsylvania, Penn State, St. Vincent College and the Community College of Allegheny County.

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



On August 2, third-place winners of the national High School Science Bowl, from North Hollywood High School in California, visited PSC as part of a week-long Pittsburgh visit hosted by the National Energy Technology Laboratory. David Deerfield, PSC assistant director, described features of the Visible Human project.



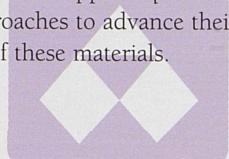
▲ *Beverly Clayton, PSC executive director, coordinates PSC's program to provide advanced training and high-performance computing resources for researchers in Pennsylvania.*

Informatics Inc. This year, for instance, the Pittsburgh office of Deloitte & Touche used the CRAY T3E to assess the potential gain from using a customized software package to calculate tax benefits for their clients.

Research in Pennsylvania

Lubricants in Aluminum Processing

We don't think about the processes that shape aluminum—like rolling, forging and extrusion—when, for instance, we wrap a leftover in aluminum foil and stuff it in the fridge. But an essential ingredient in forming aluminum products is the lubricant chemicals that coat aluminum sheets to protect them from friction and abrasion as they're shaped. Alcoa chemists over many years have built a fund of industrial know-how about these chemicals, but there's little understanding of the molecular details. Using the CRAY T3E, Alcoa researchers applied quantum-mechanical approaches to advance their understanding of these materials.



ALCOA

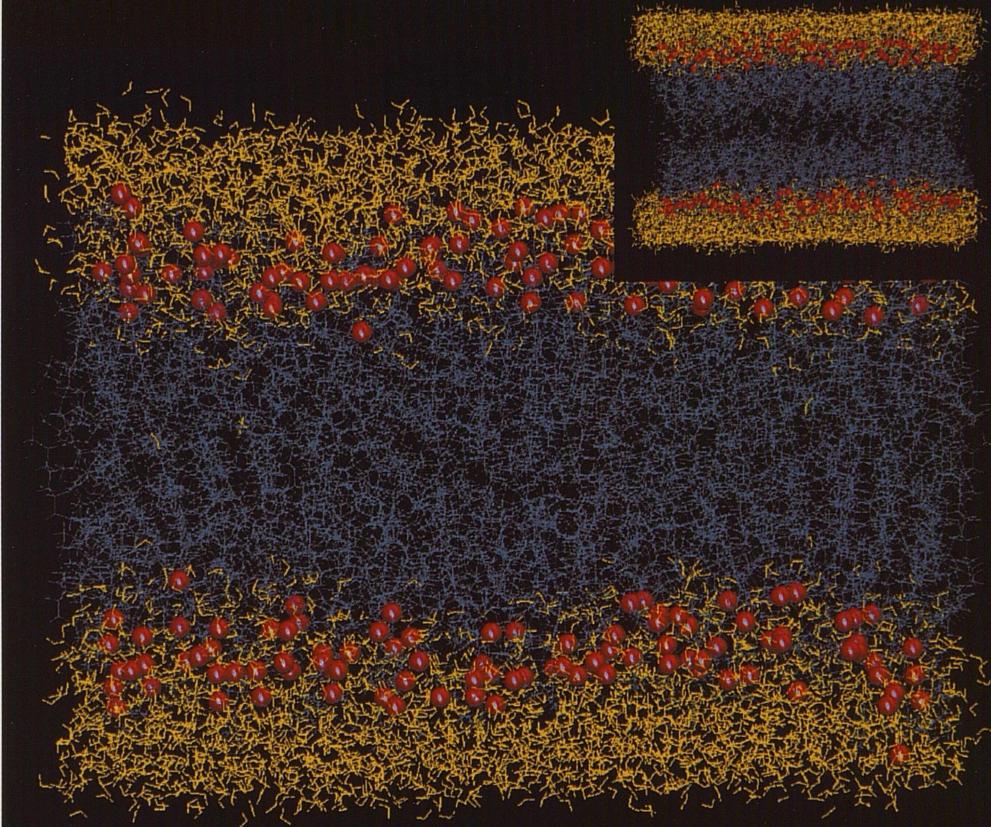
[A visible step in ensuring Pennsylvania's bright, high-technology future.]

Research in Pennsylvania

Biological Membranes

The membranes of living cells are permeable barriers that control the flow of information between cells and their environment. Channels in these membranes, which operate like gates to allow the passage of ions such as calcium and potassium, regulate many biological processes, including heartbeat and nerve impulses. To help with understanding of how drugs interact with these channels, Dr. Pei Tang and colleagues at the University of Pittsburgh School of Medicine collaborated with PSC scientist Marcela Madrid on detailed simulations of a membrane called DMPC (right). Constructed of complex sheetlike assemblies called lipid bilayers, these membranes involve tens of thousands of atoms and can be simulated only with high-performance systems like the CRAY T3E.

Other Pennsylvania research is featured on pages 24-30.



Impact of the Terascale System

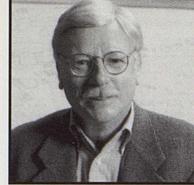
At an August news conference, sponsored by the Pittsburgh Regional Alliance and Greater Pittsburgh Chamber of Commerce, state and regional officials discussed the regional implications of PSC's Terascale Computing System (see pp. 4-7). Speakers included Congressmen Bill Coyne and Mike Doyle, Pennsylvania Secretary Sam McCullough, Allegheny County Executive Jim Roddy and University of Pittsburgh Chancellor Mark Nordenberg.

"This grant, of course, gives a tremendous boost to Western Pennsylvania," said Nordenberg in his remarks. "It brings to Pittsburgh the most powerful unclassified computer in the world. Put another way, the best scientific minds in this country, in a broad range of fields, will be looking to Pittsburgh for help in advancing their important work. On a daily basis, then, PSC will be raising Pittsburgh's image in the minds of the U.S. scientific and technical community. This is one important and very visible step in ensuring that western Pennsylvania more broadly is home to a very bright high-technology future."



▲ Congressman Mike Doyle (left) talks with Pennsylvania Secretary McCullough while University of Pittsburgh Chancellor Nordenberg talks in the background.





▲ Jim Kasdorf, PSC director of special projects. "NETL is a national laboratory in our backyard, and the (SC)² partnership catalyzes high-tech development in this region."

Tools of Discovery

The Super Computing Science Consortium

On April 28, by plugging two cables together in the PSC network room, Undersecretary Ernest J. Moniz of the U.S. Department of Energy inaugurated a high-speed network that will play a vital role in development of technologies for clean, efficient fossil-fuel combustion. "With this link," said Moniz, "we connect DOE's newest national laboratory with new tools of discovery that are transforming science."

Moniz's PSC visit highlighted a new fiber-optic network connecting West Virginia University and the National Energy Technology Laboratory with PSC's Internet hub. As Moniz plugged the cables together, instantaneously a video screen lit up showing Dr. David Hardesty, president of West Virginia University, at the WVU Virtual

Environmental Lab in Morgantown. "This brings greater capability for scientific work," said Hardesty, speaking over the network in real time to an audience at PSC, as he noted that the link will foster collaboration among WVU, NETL and PSC.

In December 1999, NETL—formerly the Federal Energy Technology Center—became DOE's newest national laboratory. Moniz noted that NETL research is directed toward a critical environmental challenge: innovative technologies for fossil-fuel combustion. "Developing technologies to use fossil-fuel more efficiently and cleanly is NETL's key mission," said Moniz. "With the combined brain-power of West Virginia University and Pittsburgh Supercomputing Center, we will press forward to solve these problems."

With offices in Pittsburgh and Morgantown, NETL solves national energy and environmental problems with emphasis on developing cleaner, more efficient technologies for fossil-fuel

combustion. With DOE sponsorship, NETL, PSC, CMU, Pitt, WVU and the West Virginia Governor's Office of Technology formed a regional partnership, the Super Computing Science Consortium, to foster research and economic development in the southwest Pennsylvania-West Virginia region. (SC)² paved the way for research collaboration between PSC and NETL (see p.18) and for the new network link.

More information: <http://www.sc-2.psc.edu>

Bubble Transport in Fluidized-Bed Combustion

One of the technologies in development at NETL is fluidized-bed combustion, an environmentally friendly system to burn biomass in a powdered bed by heating it from the bottom in a cylinder. This combustion process can use virtually any fuel and burn it efficiently.

"New advances in computer simulation technologies, in revolutionary new concepts for coal and natural gas fueled power plants, in carbon sequestration technologies, and in the production of ultra-clean transportation fuels and chemicals now allow us to envision a future in which the economic advantages of fossil fuel can continue to be enjoyed without environmental concerns."

Bill Richardson
U.S. Secretary of Energy
(from his Dec. 10, 1999 statement, designating the National Energy Technology Laboratory as DOE's 15th national laboratory)

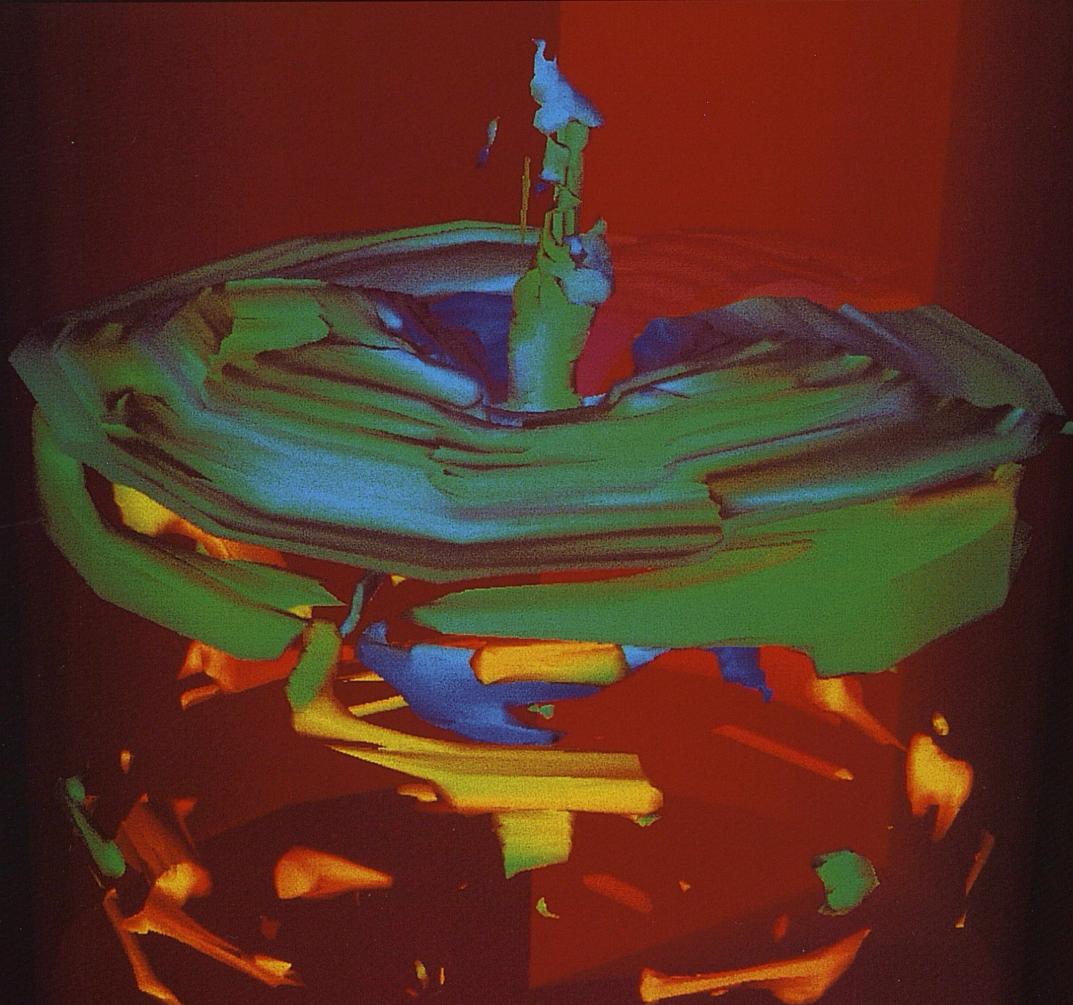
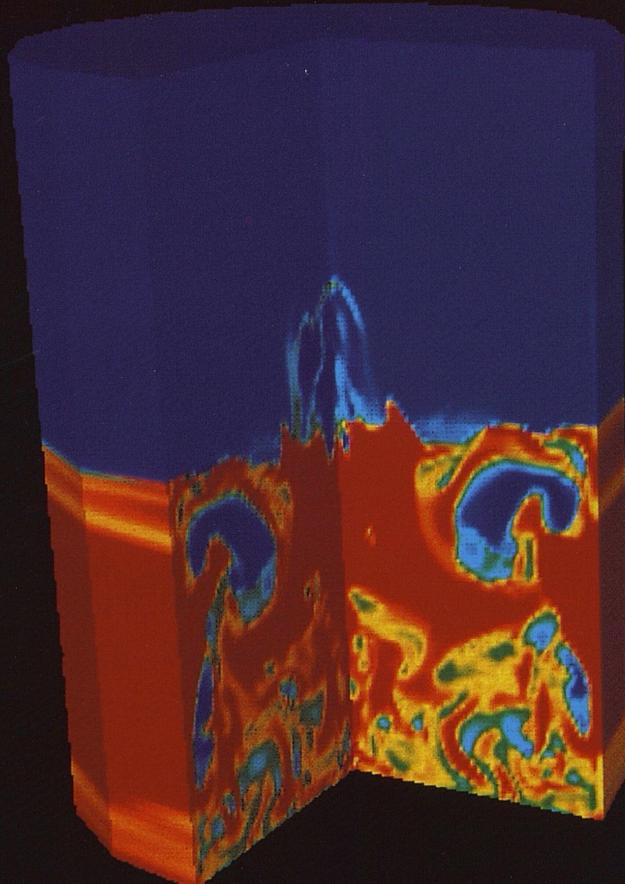
["With this link, we connect DOE's newest national laboratory with new tools of discovery that are transforming science."]



DOE Undersecretary Ernest J. Moniz (left) plugs cables together to officially inaugurate the network linking West Virginia University and the National Energy Technology Laboratory with the Pittsburgh gigaPoP. Looking on are PSC network engineer Steve Cunningham; Lynn Layman, Westinghouse Electric Company; Ralph Roskies, PSC scientific director; and Rita Bajura, director, National Energy Technology Laboratory.

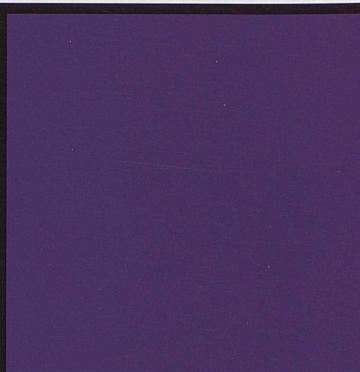
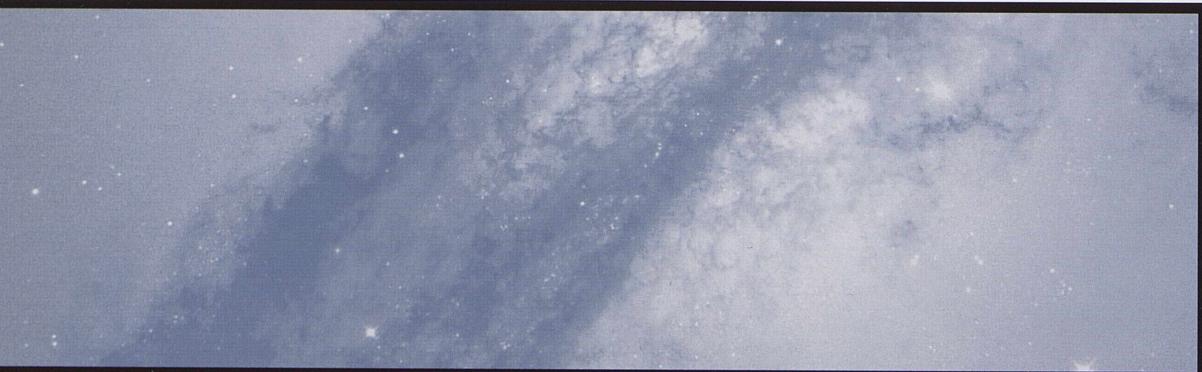
PSC and NETL scientists collaborated to simulate this process using sophisticated software called MFIX. In one animation from these computations (right), color corresponds to the percentage of gas versus solid as the powdered fuel (red) in the bottom of the cylinder heats and forms bubbles that rise into the upper-half (blue) of the cylinder.

An important objective of fluidized-bed combustion is minimizing ozone emissions, and the simulation tracks ozone as it rises—to see if it escapes the cylinder. In another animation (below), color represents relative ozone concentration. This is the first 3D simulation of this process, and it shows asymmetries in the gas plume rising from the bed, which couldn't be detected in earlier 2D simulations.



[2 0 0 0]

Projects



MANUFACTURING PROCESS DESIGN

18 Clean Power

Simulations of Lean, Pre-Mix Combustion in Power-Generating Turbines
Combustion Dynamics Group, National Energy Technology Laboratory



STRUCTURE OF PROTEINS & DNA

22 The Rude Mechanicals

Mechanical Unfolding Intermediates in Titin Modules
Klaus Schulten, University of Illinois, Urbana-Champaign



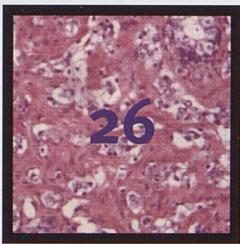
24 Getting a Grip on AIDS

Major Subdomain Rearrangement in HIV-1 Reverse Transcriptase
Marcela Madrid, Pittsburgh Supercomputing Center
Jonathan Lukin, Carnegie Mellon University

PATHOLOGY

26 An Objective View of Cancer

Pathology Image Classification and Evaluation by Content-Based Retrieval
Arthur W. Wetzel, Pittsburgh Supercomputing Center
Michael J. Becich, University of Pittsburgh Medical Center



AT THE FRONTIER OF PHYSICS AND CHEMISTRY

28 Getting the Jump on Superacids

Ab Initio Molecular Dynamics Study of the Superacids
Dongsup Kim and Michael L. Klein, University of Pennsylvania

EVOLUTION AND STRUCTURE OF THE UNIVERSE

30 The Dance of Two Black Holes

Grazing Collisions of Black Holes via the Excision of Singularities
Researchers of the AGAVE Numerical Relativity Collaboration
Penn State, University of Pittsburgh, University of Texas



Clean Power

On the voyage home to Ithaca, Odysseus and his sailors had to navigate between Scylla and Charybdis—dangerous rocks and a whirlpool. Maneuver to avoid one peril and you risk the other. Researchers at the National Energy Technology Laboratory in Morgantown, West Virginia know the feeling. Their job is to steer the course of environmental stewardship in the face of accelerating demands for electrical power around the globe.

“America is running short of electricity,” said a front-page story in the Wall Street Journal a few months ago (May 11, 2000). The information age—temperature controlled machine rooms and offices—and surging appliance purchases have juiced power consumption. Summertime U.S. peak demand is now about 700,000 megawatts, up from 525,000 in 1989, a rise that threatens to outstrip capacity, now about 780,000 megawatts. Complicating things is that deregulation of the electric utility indus-

try has spawned uncertainty about the return on investment in new plants.

Adding fuel to the fire, literally, developing countries are a burgeoning market for energy. One recent projection holds that over the next few years 300 megawatts of new electric generating capacity will be installed somewhere in the world each day!

What about acid rain? What about greenhouse gases? These and other environmental imperatives drive research that will provide clean power options for the world’s energy. At present, 85 percent of U.S. consumption and 90 percent of the world’s comes from fossil fuel, and as the president’s commission of science and technology advisors reported last year, the current best opportunity for environmental progress in power generation is high-efficiency, low-emission combustion.

“The challenge is to convert fuel to energy without creating pollutants,” says George Richards, who leads NETL’s com-

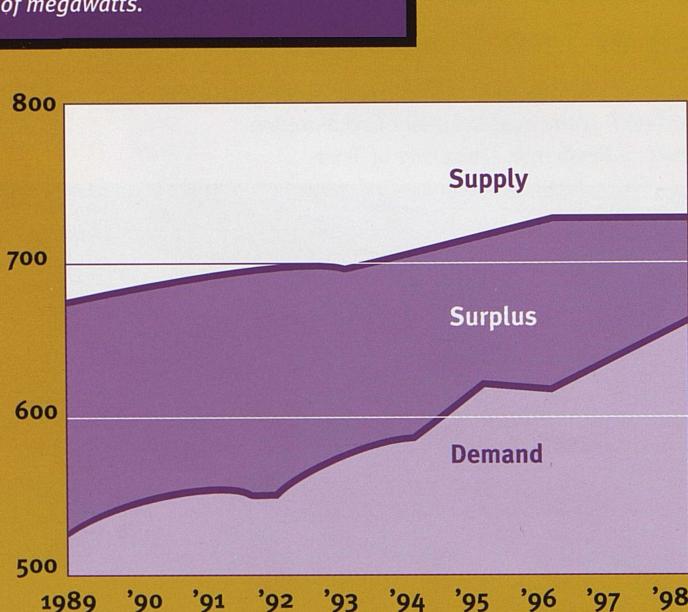
bustion dynamics team. The workhorses of electrical-power generation are the jet-engine-like gas turbines that convert fossil fuel into megawatts of electricity, and the mission of Richards’ team is to help develop the engineering knowledge to make 21st century turbines more efficient, cleaner and cheaper to operate. In a recent series of simulations at the Pittsburgh Supercomputing Center, they’ve made progress toward this goal.

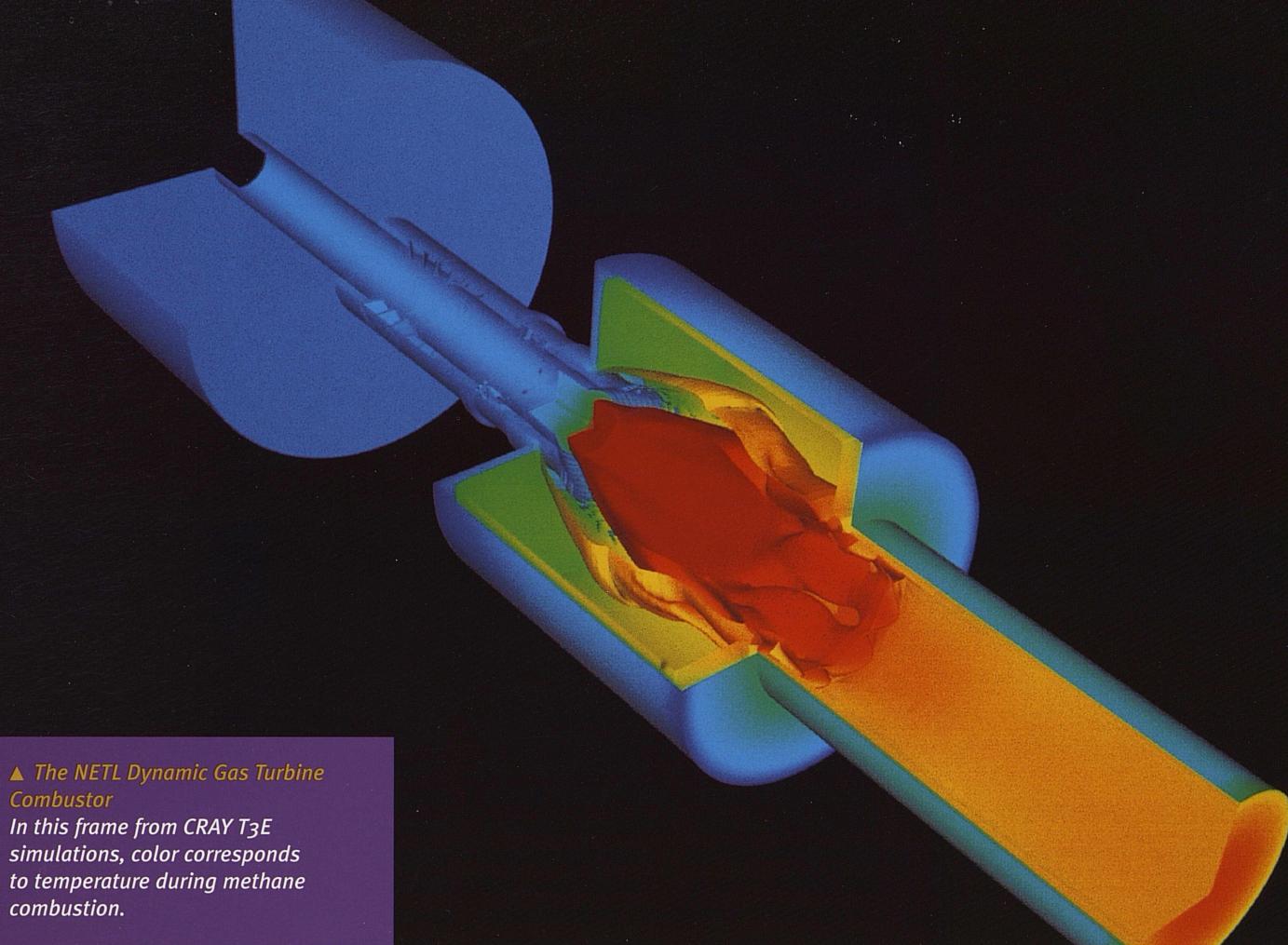
Lean, Pre-Mixed Combustion

The power industry began to shift its new installations toward low-emission technology about 10 years ago, says Richards, and many new power plants employ low-emission turbines. The key to these advanced systems is “lean, pre-mixed combustion”—mixing the fuel, typically natural gas, with a relatively high proportion of air prior to burning. This substantially reduces nitrogen oxide pollutants (known as NO_x) while allowing high-efficiency operation. The high efficiency reduces carbon dioxide, a major greenhouse gas, and lowered NO_x alleviates smog and decreases other byproducts that affect air quality.

But a nasty problem bedevils these systems. With a lean-fuel mix, the combustor flame burns on the thin edge of not having enough fuel to keep burning, and a phenomenon analogous to a flickering candle sets up pressure oscillations—like a series of very rapid small explosions rather than a steadily burning flame. These oscillations can resonate with the vibration modes of the combustion unit and, literally, shake it to pieces.

Summertime Blues
U.S. summer electricity demand, supply and surplus, in thousands of megawatts.





▲ The NETL Dynamic Gas Turbine Combustor

In this frame from CRAY T3E simulations, color corresponds to temperature during methane combustion.

"This instability is a major issue that every turbine developer using pre-mix combustion has to face," says Richards. "It comes up in every conceivable stage—in development, during engine commissioning, in engine-fielding applications. It comes up in permitting these engines and in keeping them operating. It's a very tricky problem. I'm happy to say that there's been a lot of progress, and we can now see fielded engines using these incredibly clean combustors. But we also know that avoiding instability places very tight restrictions on how the engine can operate. Adding desirable features, like fuel flexibility, or a wider operating range, can lead to the same old problem."

To zero-in on the problem, NETL researchers conducted extensive experiments with their Dynamic Gas Turbine Combustor. This state-of-the-art test facility makes it possible to adjust parameters involved in turbine-combustor design—such as location of the fuel injector relative to the flame—and to observe and measure what happens.

The experiments revealed an unexpected result. Changing the location of a nozzle component called the "swirl vane" affected the pressure oscillations. The swirl vane—so-called because it swirls the air flow to create aerodynamics that mix the fuel and air—sits upstream of the fuel injector. In experiments comparing two swirl-vane locations, with other parameters unchanged, when the swirl vane was moved two inches farther upstream the pressure oscillations virtually disappeared. Why?

What to Measure?

The objective, stresses Richards, is to understand the physics behind the observed data, so it can be incorporated rationally into turbine design. Moving the swirl vane gave better performance in one set of conditions, but the data was inconclusive when it came to explaining the results. Prior research suggested that the time lag between when fuel is injected and when it burns is a key factor for the oscillations, but presumably, since the fuel-injector didn't

move, the swirl-vane would have little or no effect on this.

"You can place the swirl vane either closer to the flame or farther away," says Richards, "and it makes a difference. But we didn't know why. We had some conjectures, and we tested those, but we still couldn't prove what was going on. There're subtle effects, like decay of turbulence and swirling flow, that impact the important time scales—multiple, simultaneous processes, and you can't interpret the experimental data without quantifying the contributions from these simultaneous events."

To sort out the details, Richards and his colleagues turned to simulations on PSC's CRAY T3E. In recent years, the NETL team worked with consultants for FLUENT, commercial fluid-dynamics software, to develop 3D modeling that realistically simulates experiments in the experimental combustor. In summer 1999, with help from PSC scientists, they adapted FLUENT to the CRAY T3E and ran a series of simulations replicating the experiments.

Each computation—one for each experiment—required about a week of computing on 20 T3E processors to simulate 30 milliseconds of combustion. Each produced 20 gigabytes of compressed data, an enormous amount of information, which itself created a huge post-processing task.

When the results were in, they told an interesting story: The aerodynamics in the nozzle are such that moving the swirl vane, with no change to the fuel injector, significantly affects the time lag between injection and burning. In the two cases of interest, moving the swirl vane two inches upstream slows this lag time by a millisecond, and that millisecond makes a big difference in combustion stability.

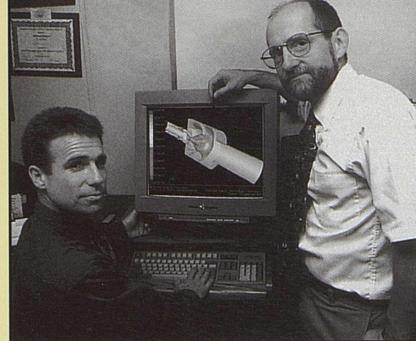
"We looked at the simulations," says Richards, "and said 'ah-ha.' It was obvious. The change in this time lag from

the point of injection is what we need to measure. That's a whole different universe to work in from where we were, a definite conclusion. It helped us set up the next set of experiments in which we've been trying to make a verifiable measurement of those time scales. And we've made some progress on that."

Flame Volume & Reaction Rate

Along with focusing their analysis of the swirl-vane results, the CRAY T3E simulations also provide the NETL team with a way to look deeper yet at the physics of turbine combustion. A key factor in combustor stability is the flame's reaction rate, the speed of burning, which varies with time. The NETL group would like to know what drives this variable. Does the volume of the flame change, such as when a candle-flame flickers, or does the flame volume stay constant as the burning-rate varies?

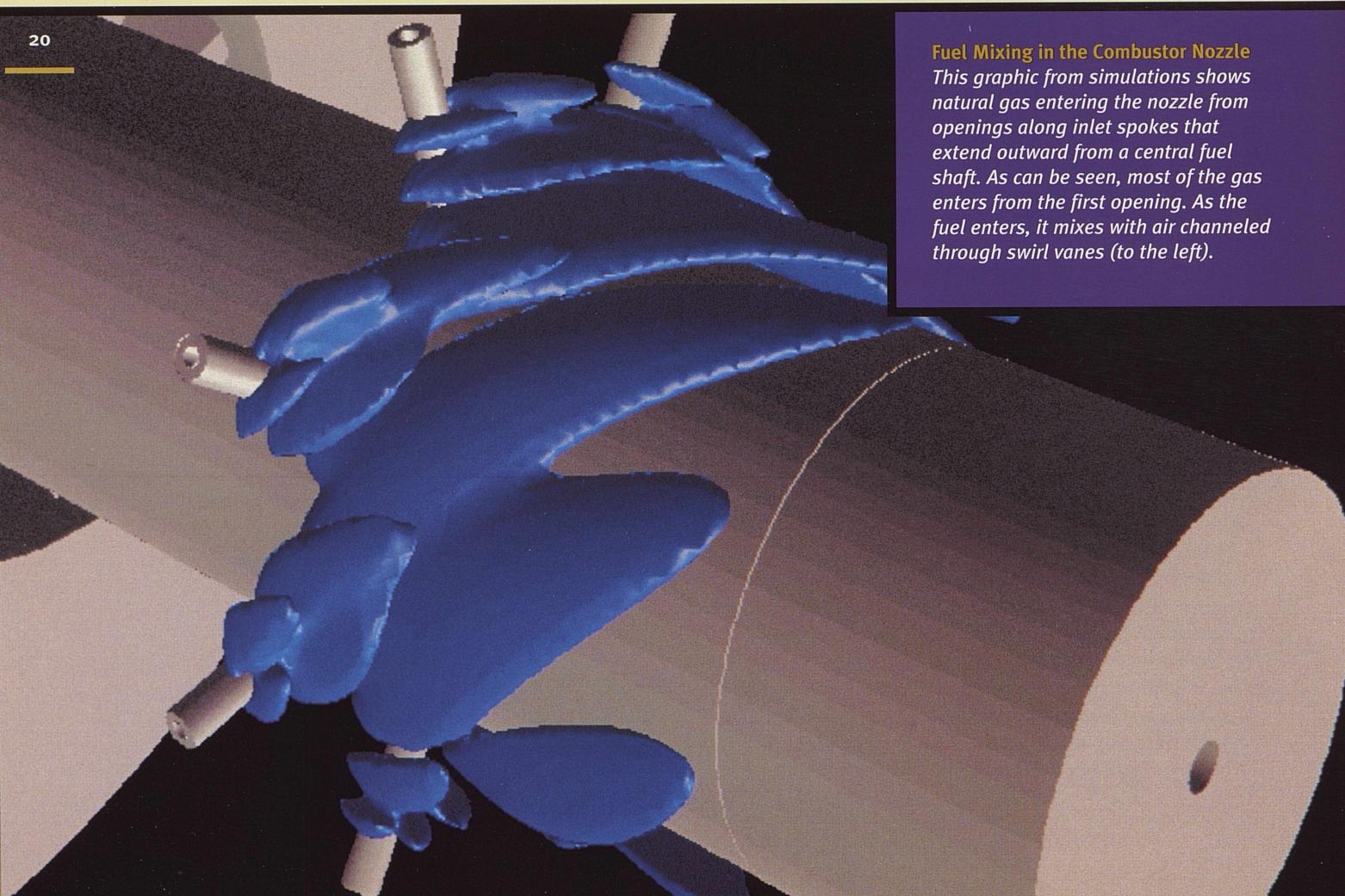
"We don't know which occurs in practical systems," says Richards. "We want to use these simulations and identify



▲ **William Rogers, consulting engineer, Fluent, Inc. (left) and George Richards, team leader for unsteady combustion research, National Energy Technology Laboratory.**

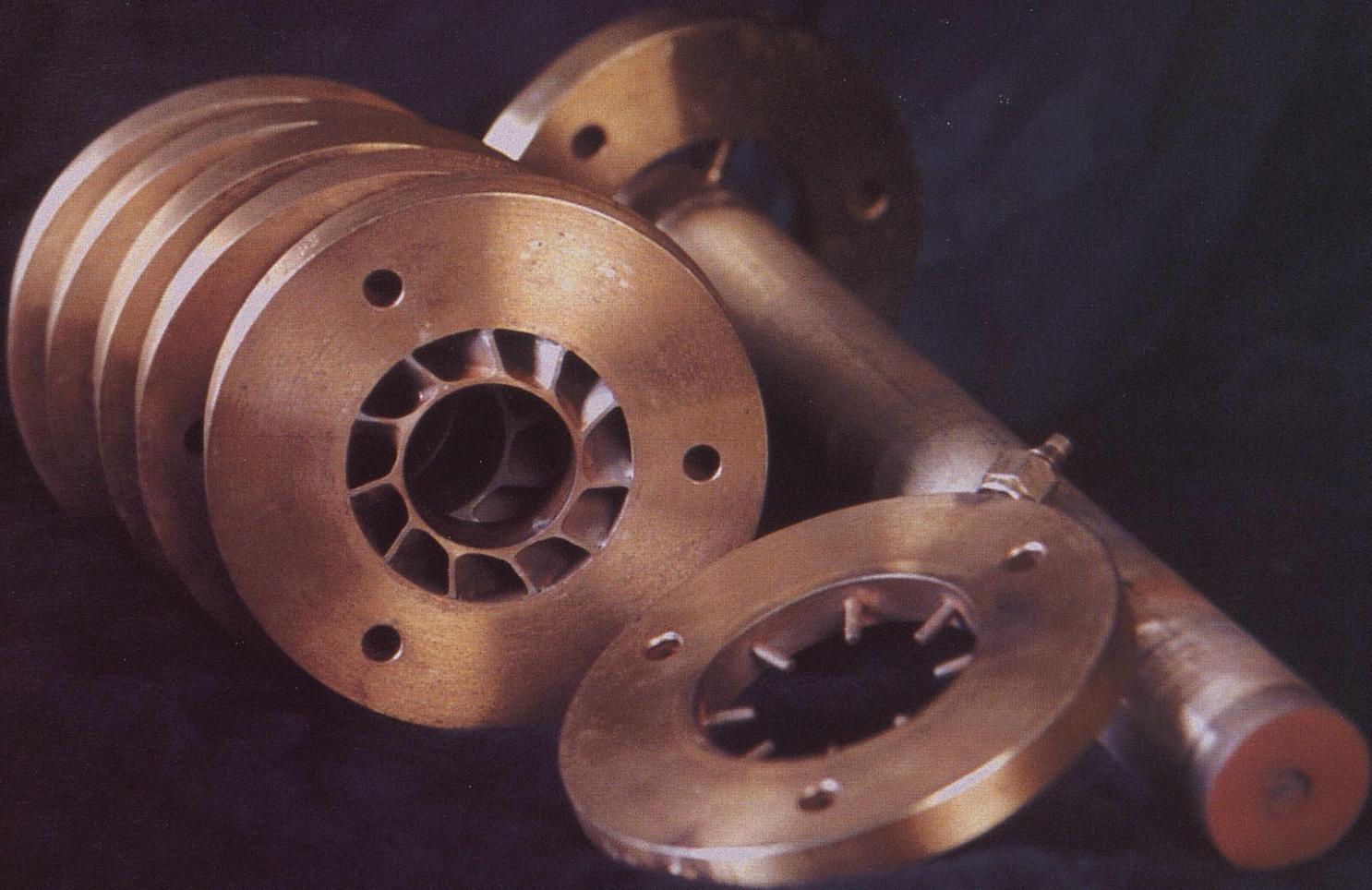
the dominant mechanism. It's probably some of each, but is it 90/10, 50/50 or 20/80? We may find that it's different under different conditions. That's where the simulations really help. If we show that you go from one mechanism to the other in the same combustor, depending on operating conditions, you'd have to do different things to make the system quiet. With simulations, and going back and forth iteratively with the experiments, we're learning a lot about fundamental physics."

Engineering Knowledge for 21st Century Turbines



Fuel Mixing in the Combustor Nozzle

This graphic from simulations shows natural gas entering the nozzle from openings along inlet spokes that extend outward from a central fuel shaft. As can be seen, most of the gas enters from the first opening. As the fuel enters, it mixes with air channelled through swirl vanes (to the left).



▼ Time Lag with Change in Swirl-Vane Location

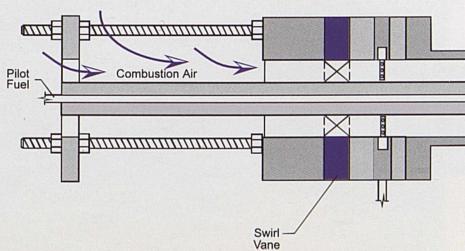
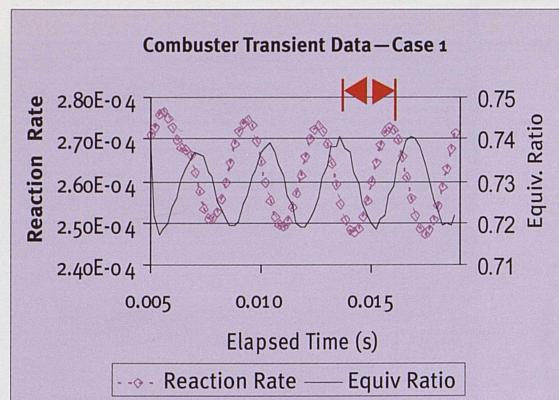
Plots of reaction rate, the speed at which the fuel burns, versus equivalence ratio, which represents the amount of fuel in the air-fuel mix. In the stable case (lower right), the fuel-mix peaks correspond with reaction-rate peaks. Here the time lag between fuel injection and burning is 3.2 milliseconds, versus 2.2 milliseconds in the unstable case.

More information:

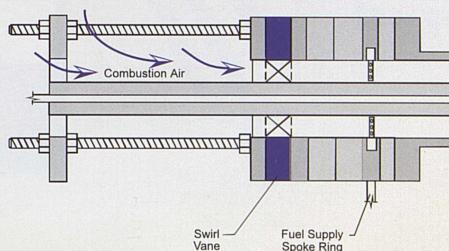
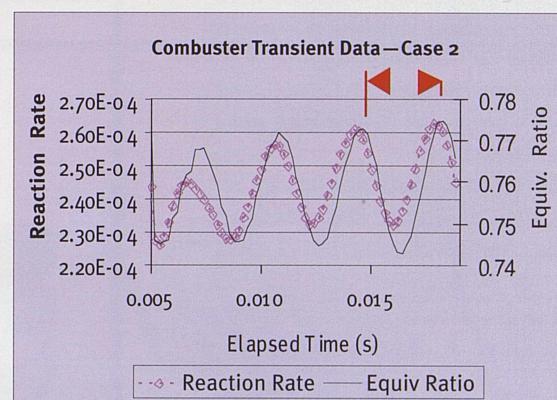
<http://www.psc.edu/science/richards.html>

21

Case 1 Time-lag: 2.2 ms



Case 2 Time-lag: 3.2 ms





The Rude Mechanicals

As with people, some proteins are more well known than others. Among the established stars, for instance, are the enzymes. Less well known are the mechanical proteins—the brawny, hard-hat, working-class heroes. Their job, in some respects, is simple: exert force.

"Enzymes have been studied for many decades," says physicist Klaus Schulten, "but for mechanical proteins there wasn't a good method. Today we have techniques—atomic force microscopy (AFM) and optical tweezers—that allow you to take individual proteins, stretch them and measure their responses."

Schulten directs the Theoretical Biophysics Group at the University of Illinois Beckman Institute for Advanced Science and Technology. In 1999, he and his colleagues carried out a series of molecular dynamics studies of a mechanical protein called titin. Working hand-in-hand with AFM studies, the computer simulations provide a detailed picture of how titin stretches.

Tugging on Titin

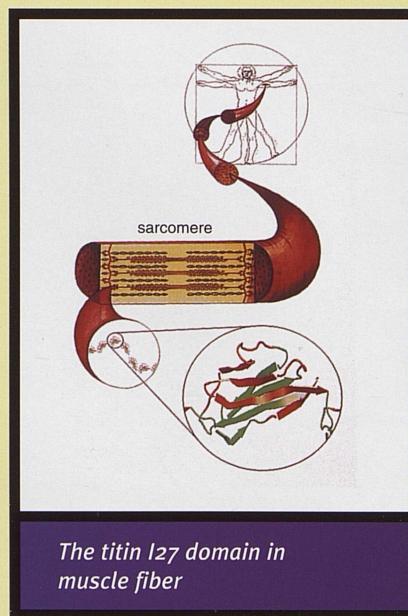
A long filament of roughly 30,000 amino-acids, titin is the largest known protein. In muscle, it helps to form the sarcomere, the integral unit of muscle fiber. When muscle is stretched, titin extends—holding the sarcomere together and providing passive force, like a stretched rubber band, that pulls it back to its unstretched state.

While many large proteins are an aggregate of separate component proteins, titin is a single strand of about

300 tightly-folded domains, each like a molecular coiled spring. Variant types of these spring-like domains are spaced along titin's length, and the Schulten group's recent simulations focus on one called I27, related in structure to the antibody immunoglobulin.

AFM experiments with I27 show a phenomenon called pre-stretch. With a weak force (about 50 picoNewton), the I27 extends slightly (5 Angstrom). With more force (above 100 pN), I27 stretches further (to about 15 Angstrom). Only with further added force (about 200 pN) will I27 let go and unfold all the way (about 300 Angstrom).

This shows, says Schulten, that I27 comes with built-in protection against unfolding. While the AFM experiments provide key information, they don't show how the protein's molecular architecture provides this protection. "Without the simulations," says Schulten, "you have a black box. You know what's happening but not why or how." To fill in the missing atom-by-atom details, Schulten and colleagues used a technique developed by his group called steered molecular dynamics.



▲ Klaus Schulten. "The Pittsburgh Supercomputing Center has helped us enormously in doing our calculations."

Opening a Black Box

If you think of the structure of a protein as a photograph that records the spatial relationships among all the atoms in the molecule, a molecular dynamics simulation is the movie version—it records how the atoms move from one tiny moment to the next. Steered molecular dynamics is a novel approach that can induce molecular changes to occur rapidly that would otherwise take too long to simulate, and it's especially useful for mimicking AFM experiments.

Using SMD, a series of numerical experiments with I27 showed that I27's resistance to unfolding arises from a patch of six hydrogen bonds that bridge between two of the protein's folded strands. With enough force, all six of these bonds rupture simultaneously—the critical event that allows I27 to fully unravel.

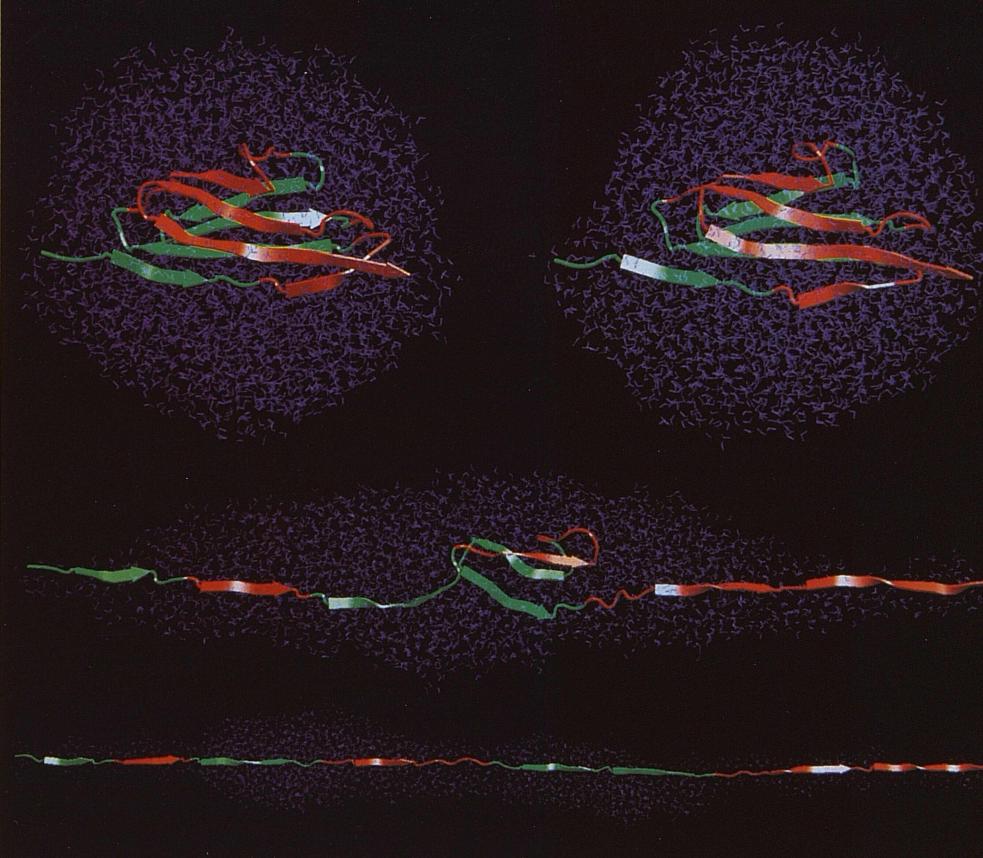
These six hydrogen bonds, explains Schulten, can be thought of as an energy barrier between folded and unfolded I27. With access to Pittsburgh's CRAY T3E, the researchers did further simulations—up to 18 for each numerical experiment—to investigate this barrier and compare the computational results quantitatively with AFM data. The spread in results among individual simulations corresponded within 10 percent to the experimentally observed positions and height of the energy barrier, giving the researchers a high degree of confidence.

The simulations also showed that two other hydrogen bonds, between two different folded strands, break before the six-bond rupture that precipitates unfolding, contributing to a small energy barrier at an extension of about 5 Angstrom. Does this account for the pre-stretch from AFM experiments?

To clinch whether this intermediate I27 state corresponds to the pre-stretch, Schulten's collaborator, J. Fernandez at the Mayo Clinic in Rochester, Minnesota,

Simulations of I27 in Water

The I27 domain is shown as simulated in its unstretched state (upper left), pre-stretch state (upper right) and, at smaller scale, after unfolding and fully extended (bottom). In simulations, the domain is fixed at one end (left) and a stretching force applied at the other end. I27 is a structural motif known as a beta-sandwich: two sheets of ribbon-like strands (orange & green), one on top of the other.



cloned a mutant I27 with a changed amino-acid that eliminated the two hydrogen bonds. AFM experiments showed that this mutant I27 unfolds without an intermediate state.

Compared to enzymes it's a late start, but titin and another mechanical protein, fibronectin, which Schulten and his collaborators have also studied, are

gaining celebrity status. Research shows these proteins play several crucial roles: organizing chromosomes in the nucleus, cell-to-cell communication, and movement of cells relative to each other. SMD simulations have proven themselves as a partner with AFM experiments in learning about these proteins. "Together these research techniques open the door to an

aspect of the cell, protein mechanics, which couldn't be investigated before," says Schulten. "This is an exciting and rich field."

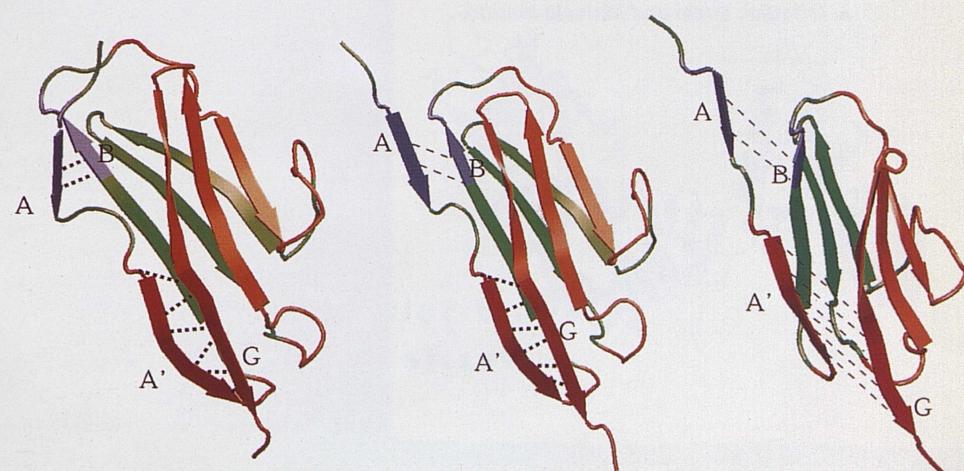
23

More information:

<http://www.psc.edu/science/schulten.html>

Hydrogen Bonds

The I27 domain in its native state (left), pre-stretch at 10 Angstrom, and unfolded to 25 Angstrom (right). The simulations reveal that unfolding occurs when six hydrogen bonds (dotted lines) between two of I27's strands (A' and G) rupture simultaneously. Two hydrogen bonds (between A & B strands) break at pre-stretch.



Getting a Grip on AIDS

AIDS remains one of the most pressing U.S. public-health problems and is virtually out of control in less developed parts of the world. In South Africa, AIDS affects a third of the current populace of 43 million and is the leading cause of death, with no abatement in sight.

Because of their high cost, the drug "cocktails" that control HIV have been of little use in Africa. Even at their most successful, furthermore, these drugs aren't a cure-all, and the quest for AIDS researchers remains, as it has been, to find a cure—not just therapeutic agents that manage the disease, but a knockout punch.

PSC physicist Marcela Madrid and Carnegie Mellon biologist Jonathan Lukin are contributing to this effort. Their computer simulations have revealed new understanding of an enzyme, HIV-1 reverse transcriptase (RT), that is an important target for AIDS drugs.

24



▲ Jonathan Lukin and Marcela Madrid.

A Target for Drugs

Like many viruses, HIV carries its genetic information as RNA, which is single-stranded and must be converted to double-stranded DNA before the virus can reproduce. This is where RT comes into play. The enzyme takes its name from its function: It transcribes RNA to form the DNA version of HIV's genome. This is "reverse" transcription because in most cells transcription goes the other direction—DNA to RNA. While the outlines of this RT copy-and-paste process are understood, the fine-grained molecular details remain a mystery.

What's clear is if you find a foolproof way to stop RT from doing its job you will have cured AIDS. Several existing AIDS drugs work by binding to RT and blocking transcription. These therapies are less than fully effective, however, because HIV transcription is highly prone to error, giving the virus a protean ability to mutate and, thereby, to defeat drugs. For this reason, a detailed atom-by-atom picture of how RT does its job is a key goal of AIDS research.

A major step since the late 80s has been to deduce the structure of RT, called crystal structure since it's obtained by crystallographic techniques. This work shows that RT's active domain, where transcription takes place, is analogous to a hand, with subdomains that roughly represent a thumb, fingers and palm. Interestingly, two versions of this structure—one bound with DNA, one not—show a big difference: With DNA, the thumb is extended and open, making space for the DNA to fit into the palm; without DNA, the thumb is folded over to almost touch the fingers.

This was Madrid's entry to HIV research. Her CRAY T3E simulations at PSC last year showed that removing the DNA from the open-thumb structure caused the thumb to close, in agreement with the closed-thumb crystal structure, indicating this is RT's "native-state." These simulations, furthermore, support thinking that RT's movable thumb may be a key to transcription, allowing the enzyme to slide along the RNA strand as it adds bases one-by-one to form DNA.

Concerted Motion

Along with being involved in transcription, the joint-like flexibility of the RT thumb may be a factor in how one class of AIDS drugs inhibits HIV reproduction. It's believed that these drugs lodge in the palm of RT and lock the thumb in an open position. "We want to understand RT's flexibility," says Madrid, "when it's by itself, when it's bound with DNA and when it's bound with drugs."

With this objective, Madrid and Lukin used the SGI Origin 2000 at NCSA in Illinois to simulate RT structures with and without DNA. By adding 37,000 water molecules to immerse the RT molecule in a water bath, these simulations had a degree of realism missing from the prior study. The plan was to compare the fluctuations of each atom as shown in molecular dynamics movies to the crystal structures, which deduce atomic position according to electron density, and therefore include a measure—called the crystallographic B-factor—that represents uncertainty in the position of an atom, evidence for motion.

[Toward Solving the
Puzzle of HIV Transcription]



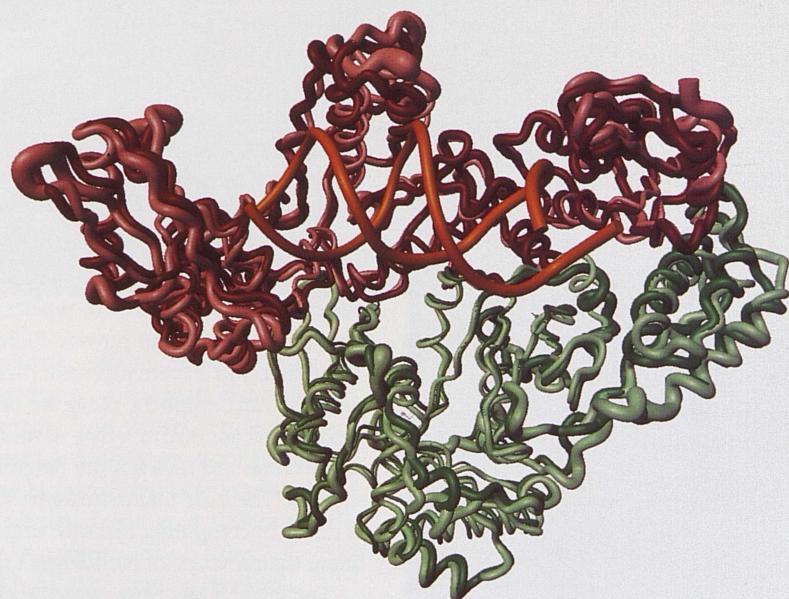
Flexibility of HIV-1 Reverse Transcriptase

The crystal structure of RT compared to the simulated structure without (left) and with DNA (orange). Coil thickness corresponds to the crystallographic B-factors and simulated movements. The entire molecule is represented, with the active-site subunit shown in red (crystal structure) and pink (simulation) and RT's other subunit in corresponding dark and light green. Analysis of these simulations finds a pattern of motions among regions in concert with each other.

For the structure without DNA, the results showed good correlation between the simulations and B-factors, giving confidence in the accuracy of the simulation. For the structure with DNA, the simulation showed greater flexibility in the tips of the thumb and fingers than suggested by the crystal structure, and also more flexibility than the structure without DNA. This simulation showed, furthermore, that the thumb and fingers and a binding pocket in the palm are moving at the same time—information that can't be obtained experimentally.

Though well short of a complete solution to the HIV transcription process, this dynamic picture complements the crystal structures and pulls together much of the available information into a consistent whole, offering a blueprint for future work. In their next project, Madrid and Lukin plan to simulate RT structures that include drug molecules that inhibit transcription. "We'll be looking for clues," says Madrid, "to see how the drugs work."

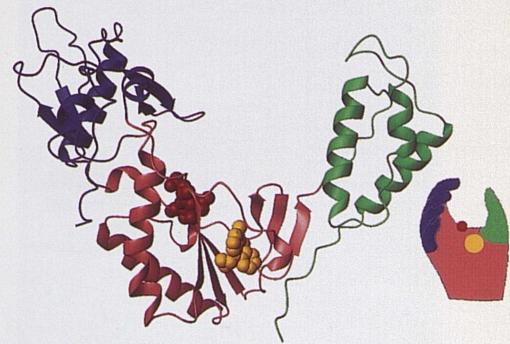
More information: <http://www.psc.edu/science/madrid2000.html>



25

► A Handy Enzyme

This ribbon representation of the RT active domain illustrates its hand-like structure, showing fingers (blue), palm (pink) and thumb (green). The active site (red atoms), where DNA is elongated, is in the palm region. Also shown is an RT-inhibitor drug (yellow) in the pocket where it binds.



An Objective View of Cancer

There's considerable research showing what should come as no surprise: Pathologists are human. Differences between malignant and benign tissue can be subtle. Interpretations are often subjective. A recent study at Johns Hopkins found that 2 to 3 percent of diagnoses are either wrong—misreading malignant tissue as benign or vice-versa—or could have been more accurate, potentially leading to unnecessary or inappropriate treatment.

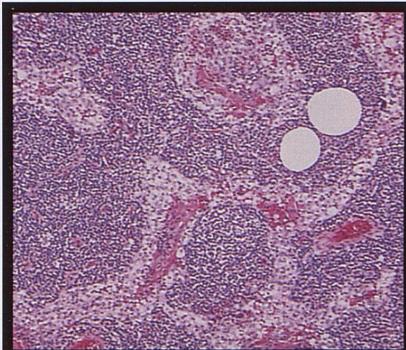
What if you could build a database of accurately diagnosed tissue and do a computerized search with undiagnosed samples? Several years ago, Dr. Michael Becich, a pathologist at the University of Pittsburgh Medical Center took

this question to the Pittsburgh Supercomputing Center. The result is Computer-Based Image Retrieval, a still-evolving software method that does essentially what pathologists do when they flip pages in a reference book, but objectively, more comprehensively and in a matter of seconds.

"We've developed a tool," says Becich, director of genitourinary pathology and informatics at UPMC, "that classifies microscopic images by image content. Instead of using text, where you'd go to a database and say show me your images of prostate cancer, for instance, CBIR uses computerized image-classification to create image signatures. For an unknown sample, the image signature acts as the search key."

The aim, emphasizes PSC scientist Art Wetzel, who developed the signature-matching software, is not to replace the skill of doctors, but to provide more information in a more objective manner: "A book can't cover the range we build into a database. We're trying to save search time and broaden the potential choices. There are many situations where a dollar of computing can save much more than a dollar of pathologist's time."

The PSC-UPMC team, which includes pathologists John Gilbertson and Rebecca Crowley and grad student Lei Zheng, has built an archive of about 120,000 images, representing a range of tissue categories—such as muscular, glandular, and gastrointestinal. Using this data, CBIR has shown impressive ability at matching unknown samples with the correct category. Going beyond classification to diagnosis, the team has focused on prostate tissue, Becich's specialty.



Lymph Node Tissue

An example of glandular tissue, which CBIR can differentiate from other tissue categories, such as muscular and gastrointestinal, on the basis of image features.



▲ Art Wetzel, Pittsburgh Supercomputing Center, and Michael Becich, University of Pittsburgh Medical Center.

Toward Consistent, Objective Diagnosis

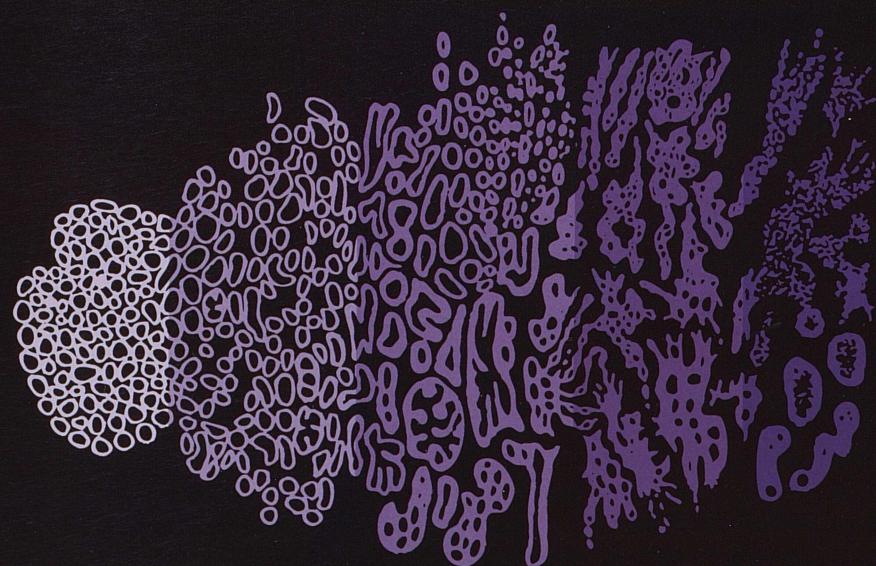
Prostate cancer is the most frequently occurring non-skin cancer in American men. Recent data indicates more than 200,000 new cases each year and more than 40,000 deaths, second only to lung cancer for cancer-related mortality in U.S. males.

The best available method for evaluating long-term prognosis is Gleason grading. Pathologists examine the tissue under a microscope and assign a number—from one through five, one being least severe—according to the structural patterns they observe. By adding the two most predominant Gleason grades in a sample, the tissue gets a score between two and 10. This score is a major factor in treatment planning. The problem, however, is that research shows a 20 to 40 percent variation when different pathologists score the same samples.

To create image signatures, Wetzel implemented a series of image-classification algorithms, some based on coloration, others on mathematical methods, which allow the computer to discern features that often aren't visually apparent. For prostate grading, he turned to a method called "spanning trees" by which, in effect, the software draws lines connecting cell nuclei. The distribution of lengths and angles of these lines correlates well with Gleason grades.

On samples where pathologists have identified a region of interest, CBIR achieves 80 percent agreement with Gleason grading by UPMC pathologists,

[Bringing Consistency, Comprehensiveness and Cost-Saving to Diagnosis]



1

2

3

4

5

The original of this graphic is courtesy of Lippincott, Williams & Wilkins.

and most disagreements are confined to one grade level. It's difficult to gauge, says Becich, what this means with respect to accurate diagnosis: "It's hard to know if we're doing better or worse, until we develop a new, objective scale, but it's clear—since pathologists have this 20 percent variability—that CBIR can do as well as humans."

"Part of what CBIR offers," says Wetzel, "is putting a stake in the ground and getting an objective evaluation of where the boundaries are between grades. It's not so much that there's always a right or wrong answer; it's that we need a consistent answer."

What's Ahead

CBIR's ability at Gleason grading is so far confined to cases in which a human picks out the region of interest. Though relatively easy for pathologists, it's a

challenging software problem: "How do you know which parts of a slide are important," says Wetzel, "and ignore artifacts like torn edges, poor staining? Things human vision can do automatically are hard to program. We need to move into some of these areas."

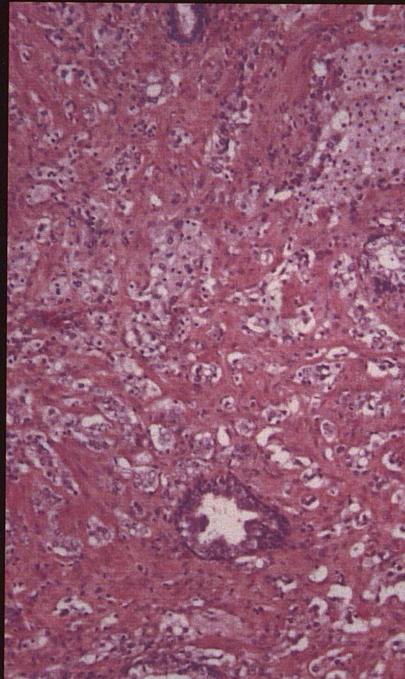
With prostate grading, Becich is tracking the results of cases in which CBIR picks out matches using signatures unrelated to the structural patterns of Gleason grading. "We're currently looking at a large number of cases in which we have clinical follow-up to see if CBIR picks out poor prognosis tumors better than human grading."

The PSC-UPMC team sees CBIR as a tool that will eventually work with high-performance networking to provide "telemicroscopy" for pathologists at remote locations, to save time and improve diagnoses. "When you broaden the range of choice," says Becich, "you improve the ability to arrive at correct diagnosis and appropriate treatment."

More information:
<http://www.psc.edu/science/wetzel.html>

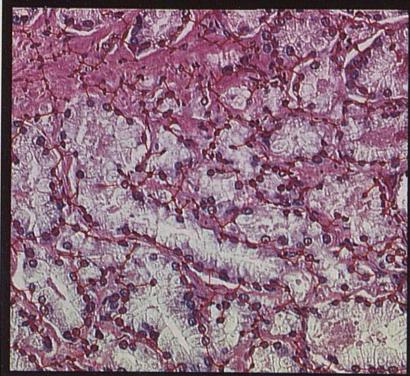
Gleason Grading

At less severe grades of prostate cancer (left), glandular structure is relatively organized and differentiated from the background. At higher grades, structure breaks down. The large image (below) is an example of a Gleason grade 4 prostate.



▼ Spanning Trees

Through digital manipulation, CBIR differentiates cell nuclei, which form roughly circular structural patterns evaluated in Gleason grading, by their dark coloration. The software then connects the nuclei positions, taking account of the characteristics of the space between nuclei, to arrive at a weighted-length tree structure that correlates well with Gleason grading by pathologists.





▲ Michael Klein, Hepburn Professor of Physical Science, University of Pennsylvania, and winner of the 1999 Rahman Prize for Computational Physics.

Getting the Jump on Superacids

They don't come from the planet Krypton or leap tall buildings, but it's not a big stretch to think of superacids as chemical superheroes. Since George Olah's work in the 1960s, which led to the 1994 Nobel Prize in Chemistry, these fascinating compounds have become an essential tool of the chemical industry. Their powerful ability to react with and break down raw petroleum brings us such products as high-strength plastics and lead-free, high-octane gas. Exotic processes like coal gasification are unthinkable without superacids.

"Hundreds of thousands of tons of this material are used in the chemical industry on a routine basis," says Michael Klein, Hepburn Professor of Physical Science at the University of Pennsylvania and director of Penn's Center for Molecular Modeling. "Yet they're mysterious. Why do different

superacids have different strength? How does a superacid actually work at the atomistic level?"

Superacids aren't well understood in large part because they are what they are—extremely strong acids. By the accepted definition, superacid means stronger than 100 percent sulfuric acid; many are a billion or more times stronger than that. "They are very toxic, very volatile," says Klein's collaborator, post-doctoral fellow Dongsup Kim, "and it's difficult to do experiments."

Using PSC's CRAY T3E, Klein and Kim provided new insight into the relation between the molecular structure of superacids and acid strength, and produced the first detailed picture of a fundamental superacid property called proton jumping.

Naked Protons

Acids are defined by their ability to "protonate" bases—that is, to donate protons. The key chemical species is the hydrogen nucleus, naked H^+ , stripped of its single electron. In liquids, the naked proton is always bound with either the acid or solvent. Free protons exist only in the gas phase, one reason why experiments have focused there, yielding information not available experimentally with liquids.

"Chemistry for 50 or 60 years has had a diversion into the gas phase," says Klein. "Real chemistry happens in solutions, but we couldn't do studies

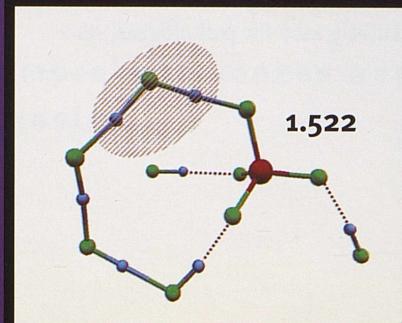
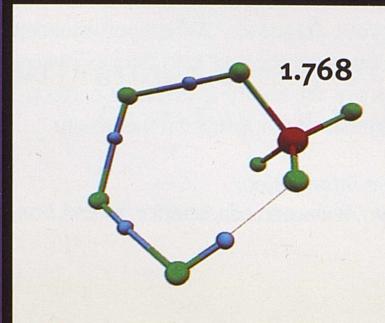
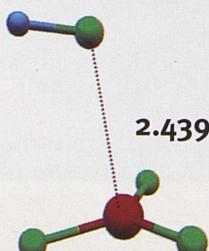
of liquids 25 years ago, because the computing hardware wasn't capable enough, and we didn't have the methods." Today's computational approaches, however, open a window closed to experiment, and Klein and Kim set out to fill in some of the blanks.

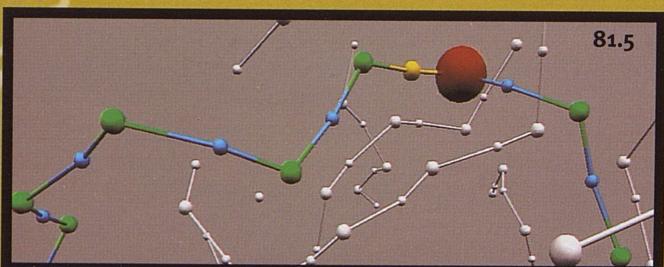
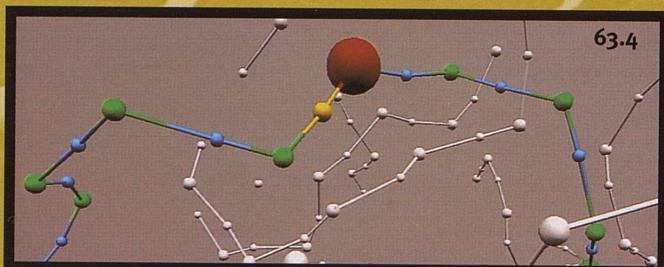
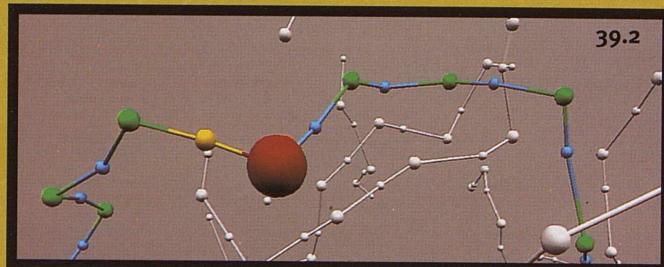
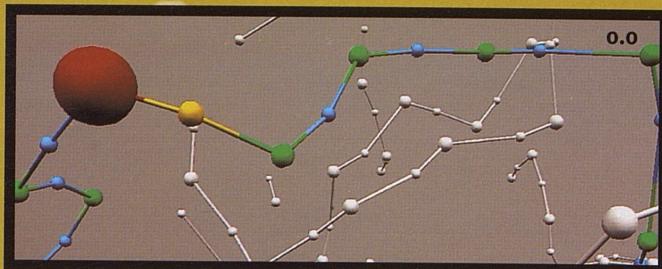
Experiments with one of the superacids, boric trifluoride in hydrogen fluoride (BF_3/HF) indicated a feeble, electrostatic attraction between BF_3 and HF in the gas phase—called a van der Waals attraction. One might expect to see a stronger bond, a chemical bond in which the fluorine atom from HF shares electrons with the boron of BF_3 , yielding a BF_4^- ion and a naked proton. And this posed a question: If these molecules don't react, why do they form a superacid in solution?

Kim and Klein confirmed the experiment and also showed a more complex picture. With a single HF, there's a weak attraction between BF_3 and HF. With more HF molecules, however, the picture changes. With four HFs, the intermolecular distance decreases, and a chemical bond forms, induced by a ring structure of HF molecules, bound together by hydrogen bonds, that stabilizes the BF_4^- ion. With six and seven HFs, the ring architecture is further secured by hydrogen bonds, leading to H_2F^+ like structures, suggesting that this protonated HF is a key to superacids. "You need to have a solvent shell of neighbors," says Klein. "That's why the reaction takes place in the liquid."

▼ Superacid Clusters

In interaction with a single HF, BF_3 is weakly attracted, as indicated by the bond length (in Angstroms). With four HF molecules, however, a ring structure forms with hydrogen bonds between some of the H (blue) and F (green) molecules. This stabilizes and increases interaction strength between B (red) and the added F atom, as indicated by the shortened bond length. With six and seven HFs, an H_2F^+ like structure (shading) further stabilizes the cluster.





◀ Proton Leapfrog

An excess proton (yellow) in solution with 54 HF molecules attaches to a fluorine atom (red) in a pre-existing HF chain (blue and green). These two atoms stay next to each other as they jump down the chain. Time is shown in femtoseconds (a billionth of a millionth of a second). The excess proton and its fluorine partner travel across three bonds in only 81.5 femtoseconds.

[Computation Opens a Window Closed to Experiment]

How Protons Jump

A quantum approach called ab initio molecular dynamics allowed Klein and Kim to delve further into superacids. Ab initio means from first principles or from the beginning, without empirical data. Input to the calculation is solely the atomic numbers of the molecules. It's a computationally intense method, made possible by recent advances in numerical approaches to quantum theory, and it can produce detailed pictures of how molecular structure evolves.

With this approach, the researchers looked at how protons move in superacids. The strongest superacid is antimony pentafluoride in hydrogen fluoride (SbF_5/HF), and experiments have shown that these solutions conduct electricity better than can be accounted for by ionic diffusion, the normal process by which electrons in solution roam from ion to ion.

A plausible scenario, says Kim, arises from the reaction between SbF_5 and HF. When the SbF_5 becomes fluorinated to form an SbF_6^- ion, the free H^+ can move into an HF chain, which becomes a

pathway for the proton to jump from bond to bond like molecular leapfrog. Chemists have postulated that this "proton jump" scenario underlies the abnormal conductivity and other properties of SbF_5/HF . Using 64 CRAY T3E processors, it took a week of computing to simulate four picoseconds (four million-millionths of a second) of this game of proton leapfrog. The results offer the first detailed picture of proton jumping.

More information:

<http://www.psc.edu/science/klein2000.html>

The Dance of Two Black Holes

Once upon a time, black holes were a fascinating theoretical artifact from the mathematics of general relativity—an interesting concept, great stuff for science fiction. We've come a long way since 1915 when Einstein laid out his theory that rocked our world.

The Hubble Space Telescope and NASA's Chandra X-ray Observatory have convincingly lifted black holes from theory into reality. Still, the evidence is circumstantial. Looking for a black hole, says Stephen Hawking, is like trying to find a black cat in a coal cellar. As Penn State astrophysicist Pablo Laguna and post-doctoral fellow Deirdre Shoemaker like to point out, the way to clinch, indisputably, that black holes exist and that Einstein's equations are right is to detect gravity waves from two black holes.

Tuning the Gravity-Wave Radio

Detecting gravity waves is the job cut out for LIGO (Laser Interferometer Gravitational-Wave Observatory), Virgo and GEO600. LIGO is two NSF-funded gravity-wave detectors—in Louisiana and Hanford, Washington—now undergoing testing. Virgo and GEO600 are under construction in Europe. These projects represent a pioneering effort to develop an invaluable new set of eyes—gravity eyes—for seeing the universe.

Einstein's theory predicts that accelerating movements of massive objects in space, such as supernova explosions and black holes, will produce ripples traveling at light-speed through space-time. As with black holes, there's indirect evidence he was right, but compared to

other wave phenomena, like electromagnetism, which brings us radio and TV, gravity waves are very weak. Einstein speculated they might never be detected. If you think of LIGO as the gigantic antenna for a radio receiver, the strongest possible signal might be a faint crackle as you turn the dial.

To improve chances of hearing the first crackle of gravity from the cosmos, researchers like Laguna and Shoemaker are using the most powerful supercomputers they can find to solve Einstein's equations. Their field is called numerical relativity, and with collaborators at the University of Texas and the University of Pittsburgh, the Penn State team used systems at PSC, at NCSA in Illinois and elsewhere, to simulate two black holes merging in what's called a grazing collision—only the second time this has been accomplished. Their numerical approach, called black-hole excision, makes a notable dent in the two-black-hole problem, the major challenge of this challenging field.

"Einstein's equations describe gravity via an elegant but complicated set of non-linear partial differential equations," says Laguna. "Their complexity requires the most powerful supercomputers available. Accurately solving the two-black-hole problem, formulated conceptually by Einstein 80 years ago, will represent an historic moment in the development of general relativity theory, with extremely important implications for astrophysics and cosmology."

A single black hole doesn't make gravity waves, and colliding black holes



▲ The Penn State Numerical Relativity Group. Jorge Pullin, Deirdre Shoemaker, Kenneth Smith, David Garrison, Pablo Laguna, Keith Lockitch, Erik Schnetter, Gioel Calabrese and Bernard Kelly. Not present: Manuel Tiglio.

may be the best shot at detecting them. Theory says it's one of the strongest signals on the gravity-wave dial. To know if a crackle of static is the dance of two black holes or cosmic noise, the detectors need the answers numerical relativists are working to provide.

Black Holes without the Holes

The killer for simulating black holes is the singularity, the point of infinite density and space-time curvature that, mathematically speaking, makes a black hole a black hole. "Simply put," says Shoemaker, "the numbers get too big too fast, and the computation crashes." One approach, employed by researchers at the Albert Einstein Institute near Berlin with some success, is to exploit the relativity of time by slowing down how fast clocks tick near the singularity. This has the drawback that it adds to the already severe computational demands. With software they call AGAVE, the Penn State-Pittsburgh-Texas team took the less-traveled road of surgically removing the singularity from the domain of the calculation.

With 40 processors of NCSA's SGI Origin 2000, their simulation required nearly 100 hours. There's simplifying assumptions, such as two equal mass black holes, but the result is, you might say, a smashing success that pushes beyond prior work. Excision tamed the numerical instabilities long enough for

[Formulated by Einstein, the two-black-hole problem holds extremely important implications for astrophysics and cosmology.]



A jumble of blue star-clusters, glowing gas clouds and dust lanes surround an apparent black hole at the center of galaxy Centaurus A, a mere 10 million light years from Earth, recorded by the Hubble Space Telescope.

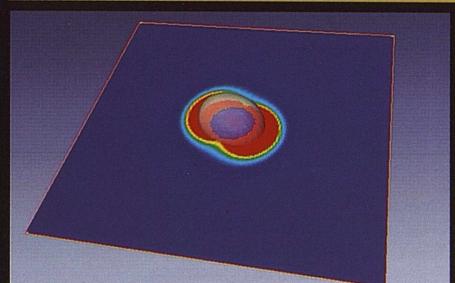
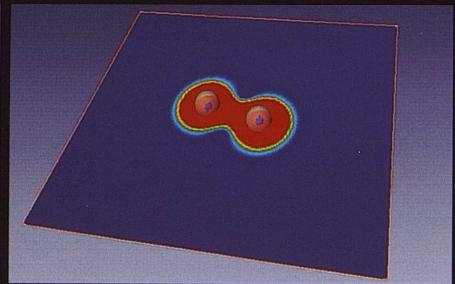
the black holes to merge and evolve for a short period as one large black hole before the simulation crashed. There's not yet accurate gravity-wave predictions to hand over to LIGO, but the next mountain now looks more climbable. That mountain, two black holes that orbit each other before they coalesce, is a few years away say the researchers.

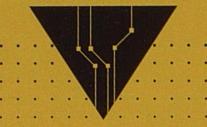
Laguna lights up thinking of PSC's new terascale system, a leap forward that will allow the team to push further. "We believe one of the severe problems we have now is that the merged black hole gets too close to the boundaries of the computational domain. With the new machine, we can shift the outer boundary outward."

More information:
<http://www.psc.edu/science/laguna.html>

► *Grazing Collision of Two Black Holes*

In these two snapshots from the simulation, transparent spheres represent the "apparent" horizon of the black holes. The first snapshot shows two equal-mass black holes caught in each other's gravitational pull; the second shows the large black hole formed as they merge. The bluish area inside the spheres represents the excised region. Color gradations (from red to purple) indicate relative strength of the gravitational field.

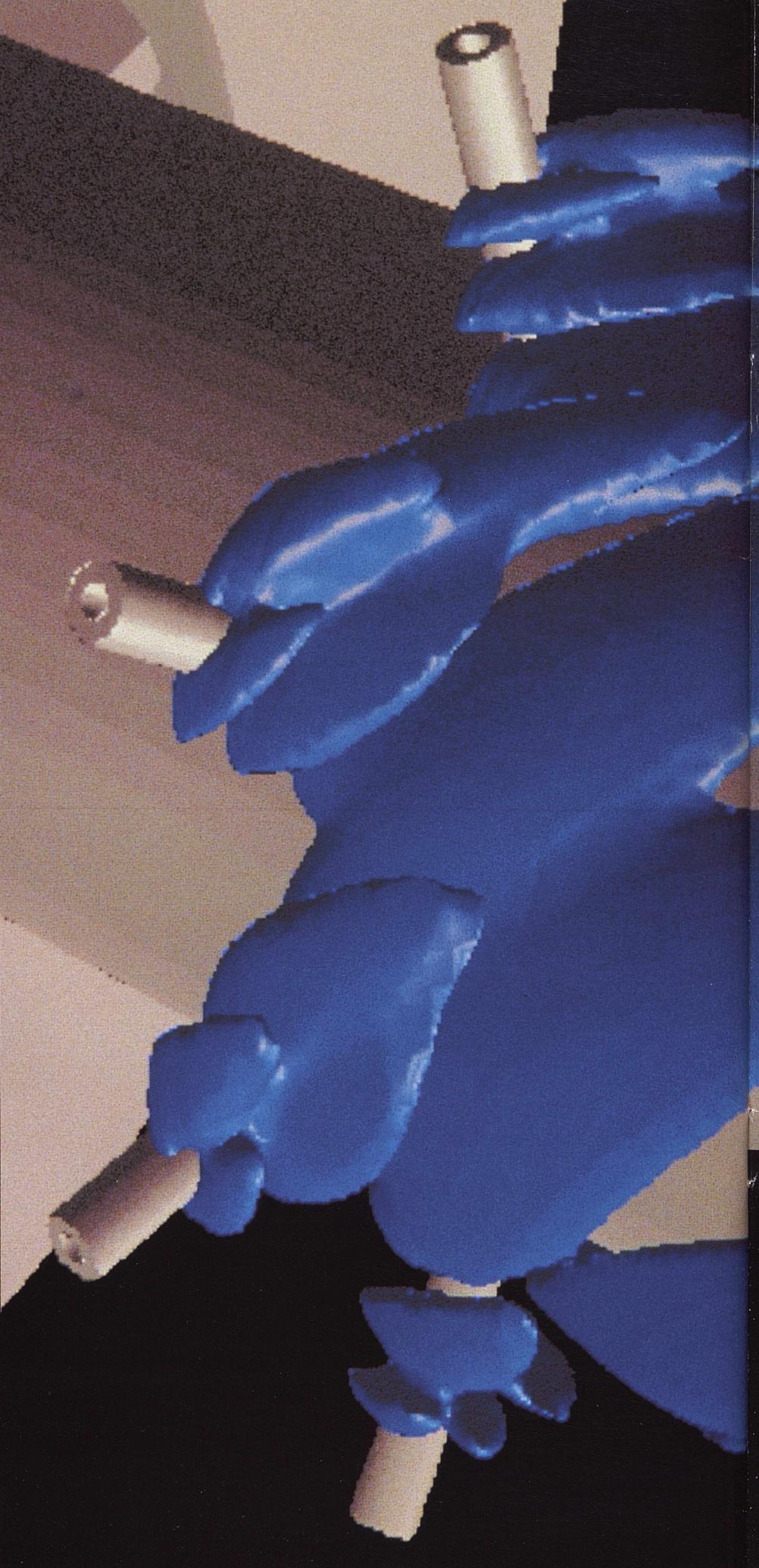


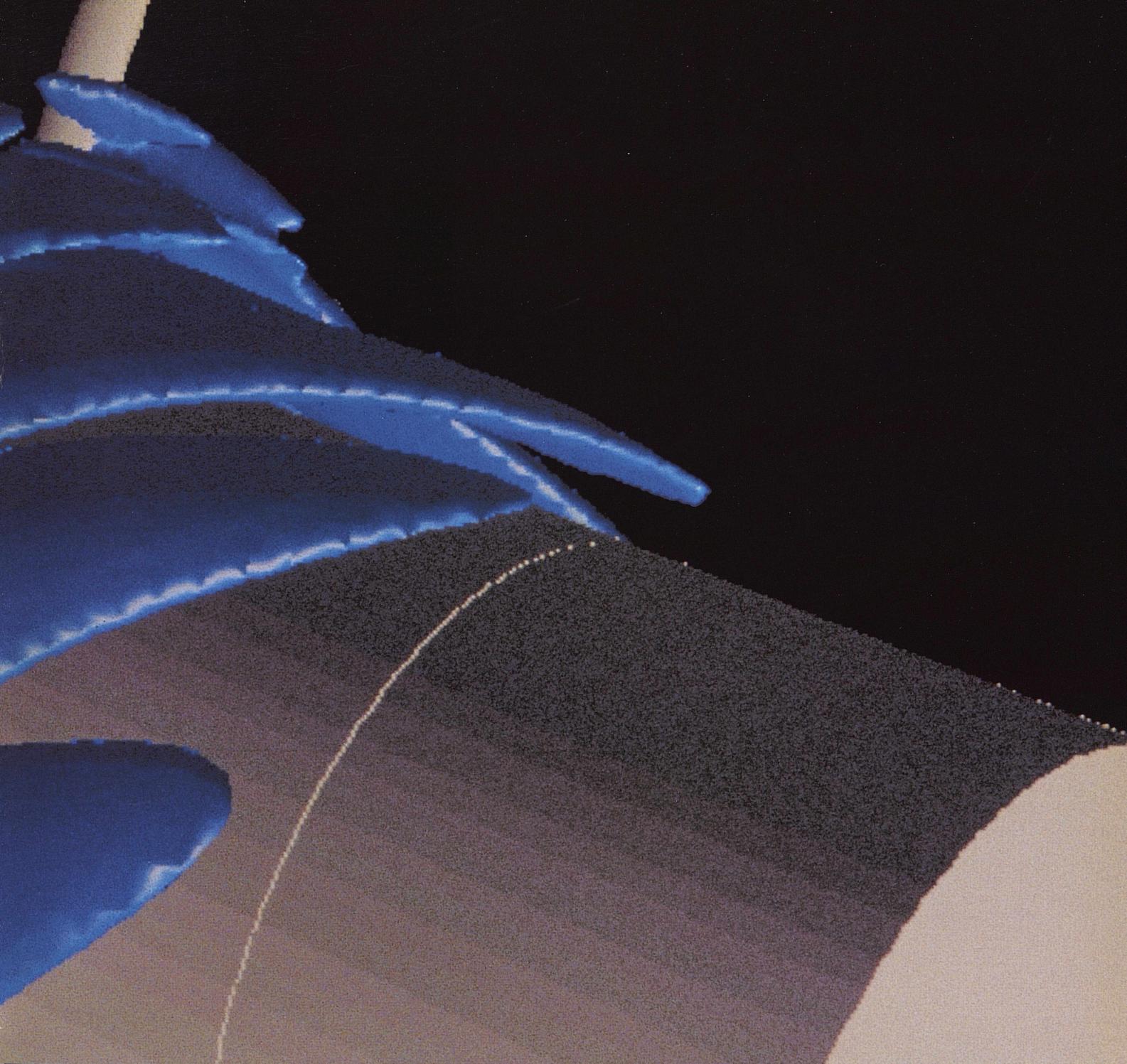


PITTSBURGH
SUPERCOMPUTING
C E N T E R

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
Sandia National Laboratories
The U.S. Department of Defense
The U.S. Department of Energy
The Science Applications International Corporation
The Howard Hughes Medical Institute
Compaq Computer Corporation
Cisco Systems
Cray Research, Incorporated
Intel Corporation
International Business Machines Corporation
Microsoft Corporation
Silicon Graphics Incorporated
Sun Microsystems, Incorporated
Storage Technology Corporation
Dell Marketing L.P.
Gateway Companies Inc.
Hewlett Packard Co.





Credits

Editor/Writer: Michael Schneider, PSC

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

Cover image: Joel Stiles, PSC, and
Tom Bartol, Salk Institute

Teryscale images: Rendered by Greg
Foss, PSC

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

Photography: Graphic Services at Mellon
Institute, Jim Schafer

Printing: Hoechstetter Printing

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